Methylomic and Transcriptomic Insights Linking Air Pollution and Atherosclerosis

Gloria C. Chi

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Joel D. Kaufman, Chair

Daniel A. Enquobahrie

Annette L. Fitzpatrick

Program Authorized to Offer Degree:

Epidemiology

University of Washington

Abstract

Methylomic and Transcriptomic Insights Linking Air Pollution and Atherosclerosis

Gloria C. Chi

Chair of the Supervisory Committee:
Professor Joel D. Kaufman
Departments of Environmental and Occupational Health Sciences,
Epidemiology, and Internal Medicine

Background: Current literature demonstrates a compelling relationship between long-term air pollution exposure and cardiovascular diseases such as atherosclerosis. DNA methylation is a potential mediator of the air pollution-atherosclerosis association because 1) exposure to particulate matter is associated with alterations in DNA methylation, and 2) alterations in DNA methylation are found in atherosclerotic lesions and are associated with ischemic heart disease and blood pressure. However, findings of past studies may be confounded by cell type composition due to the use of mixed leukocytes that combine distinct methylation profiles.

Objectives: In this dissertation, we tested the associations between long-term ambient air pollution exposure and DNA methylation (global, epigenome-wide, and candidate site) in purified monocytes. We also tested the associations between air pollution-associated DNA methylation patterns and gene expression and subclinical atherosclerosis.

Methods: The Multi-Ethnic Study of Atherosclerosis (MESA) and its ancillary studies provided rich information on air pollution, methylomics, transcriptomics, and coronary artery calcium (CAC) from a diverse cohort of adults. A sample of 1,264 participants were randomly selected for DNA methylation and gene expression assessment using monocytes. Of these, 1,207 participants with air pollution exposure predictions were included in the analytic sample. Ambient fine particulate matter ($PM_{2.5}$) and oxides of nitrogen (NO_X) concentrations were

predicted at participants' residences using sophisticated spatiotemporal models that incorporated both monitoring data and geographical variables. Estimates were averaged over one year prior to blood draw. Epigenome-wide DNA methylation and genome-wide gene expression were measured using the Illumina Infinium HumanMethylation450 BeadChip (450k) and the Illumina HumanHT-12 v4 Expression BeadChip, respectively. Candidate site DNA methylation levels were assessed from the 450k array. Mean methylation of probes mapping to long interspersed element 1 (LINE-1) and Alu repetitive elements on the 450k array were used as surrogates for global DNA methylation. CAC was measured using computed tomography following standardized protocols. Linear models were used to investigate the cross-sectional associations between air pollution and global and site-specific DNA methylation, DNA methylation and *cis*-gene expression, and DNA methylation and CAC. An innovative "bump hunting" method was adapted to the methylation array to identify regions differentially methylated with respect to long-term $PM_{2.5}$ and/or NO_X exposure. Effect modification by sex and race/ethnicity were also investigated by including interaction terms with air pollution. The association between DNA methylation and CAC progression was analyzed using linear mixed models with random coefficients and slopes for each participant. The false discovery rate (FDR) was controlled at 0.05 following the Benjamini-Hochberg method to account for multiple comparisons.

Results: The study included a sample of 1,207 participants (mean age=69.6) that was 21.2% Black, 31.6% Hispanic, 47.2% White, and 51.6% female. In this study, global DNA methylation (LINE-1 and Alu) was not significantly associated with long-term ambient air pollution exposure in monocytes ($p>0.05$). We identified nine methylations sites and four methylation regions significantly associated with long-term ambient air pollution exposure in monocytes (FDR<0.05). In general, we did not find strong evidence of effect modification by sex or race/ethnicity. Cg05926640 (near *LOC101927851*) was the CpG site with the smallest p-value for the association with $PM_{2.5}$ (β=0.049 M-value units per 2.5 $\mu g/m^3$; 95% confidence interval [CI]:

0.032, 0.067; p=5.6x10^{-8}). Methylation of cg05926640 was also associated with mRNA expression of genes located within 1 Mb including *ARID4B* and *IRF2BP2*. Cg11756214 (within *ZNF347*) was the only CpG site with methylation significantly associated with NO_X (β=0.078 M-value units per 30 ppb; 95% CI: 0.050, 0.106; p=5.6x10^{-8}), but it was not associated with *cis*-gene expression. The differentially methylated region (DMR) with the smallest p-value for the association with $PM_{2.5}$ was located on chromosome 5, overlapping the 5' untranslated region of *SDHAP3* (average methylation difference=14.6% per 2.5 μg/m^3; p=1.4x10^{-6}), and methylation of this DMR was also associated with expression of *MRPL36*. This DMR and the DMR located on chromosome 6 near *ZFP57* were significantly associated with both $PM_{2.5}$ and NO_X (FDR<0.05). A $PM_{2.5}$-associated DMR located on chromosome 7 overlapping the 5' untranslated region of *HOXA5* was associated with expression of *HOXA5*, *HOXA9*, and *HOXA10*. Methylation of this DMR was also negatively associated with CAC progression (β=-4.6 Agatston units/year per 2-fold higher DNA methylation; 95% CI: -7.5, -1.7; p=0.002).

Conclusions: Our results support the possibility that exposure to long-term ambient air pollution may lead to alterations in DNA methylation with potential functional gene expression consequences. The findings identified novel associations between long-term air pollution exposure and genes such as the *HOXA* cluster genes and *IRF2BP2*, opening up new avenues of research. Genes with DNA methylation associated with both air pollution and gene expression may be more biologically relevant and are involved in various cellular functions including inflammation and hematopoiesis. Our study found methylation of a DMR located on chromosome 7 to be positively associated with $PM_{2.5}$ exposure and negatively associated with CAC progression. However, a prior study in MESA found a positive association between $PM_{2.5}$ exposure and CAC progression; thus the direction of the association between the DMR and CAC does not support this DMR as a potential mediator of the air pollution-CAC association. Future studies are needed to replicate our study results and to explore the biological relevance of these air pollution-associated DNA methylation signals and genes in atherosclerosis.

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my dissertation chair, Joel Kaufman, for his support, encouragement, and thoughtful guidance. His exemplary mentorship was essential for my personal and professional growth. I am thankful for my committee members Daniel Enquobahrie, Annette Fitzpatrick, Katie Kerr, and Marco Carone for generously providing their time and expertise to guide me throughout my dissertation.

I would like to thank the MESA Air team for providing the air pollution exposures used in this dissertation. I am also grateful for Amanda Gassett for her encouragement and biostatistics guidance. I would like to acknowledge MESA study participants who devoted their time, patience, and resources to the study.

I am especially grateful to my parents Bonnie and Phun, my grandmother Siu King, my brother Leo, and my sister-in-law Satori for their love, care, and encouragement. I am deeply indebted to my partner, Orrin, who kept me grounded and whose love and support were crucial in this journey. Finally, I would like to thank my dear friends (Julie, Serena, Ella, Jeanny, Henry, and Nikita) for their friendship and devotion.

Grants

Gloria Chi was supported by the National Institute of Environmental Health Sciences (NIEHS) F31 Individual Predoctoral Fellowship (1 F31 ES025475-01) and the Cardiovascular Epidemiology Training Grant (T32 HL007902) from the National Heart, Lung, and Blood Institute (NHLBI). This research was supported by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, and 2R01 HL071759 from the NHLBI and by grants UL1-TR-000040 and UL1-TR-001079 from NCRR. The author thanks the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

This publication was developed under a STAR research assistance agreement, No. RD831697 (MESA Air), awarded by the U.S Environmental protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the author and the EPA does not endorse any products or commercial services mentioned in this publication. This work was also supported by NIEHS (P30 ES007033, P50 ES015915-01). The MESA Epigenomics & Transcriptomics Study was funded by NHLBI grant 1R01HL101250-01 to Wake Forest University Health Sciences. The contents are solely the responsibility of the author and do not necessarily represent the official views of the NHLBI, USEPA, or NIEHS.

DEDICATION

To my mother Bonnie.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

PUBLIC HEALTH IMPACT OF AMBIENT AIR POLLUTION

Ambient air pollution contributed to over 3.2 million premature deaths and over 76 million years of healthy life lost globally in 2010.[1] It was recently ranked as one of the top ten leading risk factors for global mortality and disease burden.[1] Particulate matter and oxides of nitrogen (NO_X) are among the various pollutants found in the air. Sources of ambient particulate matter include energy generation, combustion of fuel, agricultural and industrial processes, and road and wind-blown dust.[2] Fine particulate matter ($PM_{2.5}$) refers to particles that are less than 2.5 μm in aerodynamic diameter. The predominant anthropogenic sources of NO_X include motor vehicle and industrial fossil fuel combustion, and NO_X levels may be particularly high near heavily traveled roadways.[2] A plethora of studies demonstrated that exposure to air pollution increases the risk of cardiovascular mortality and morbidity, including atherosclerosis-related cardiovascular diseases (CVD) such as ischemic heart disease and stroke.[3] Ischemic heart disease and stroke are the top two leading causes of death globally.[4] Several studies have associated air pollution with subclinical atherosclerosis and/or its progression.[5–12] Given the global impact of air pollution, understanding the mechanism by which it causes diseases has important implications for public health prevention and intervention.

PROPOSED PATHWAYS LINKING AIR POLLUTION TO CARDIOVASCULAR DISEASES

There is increasing evidence suggesting that a systemic inflammatory response plays a role in the pathway between inhalation of air contaminants and CVD.[3] Exposure to ambient air pollution can upregulate cytokine expression in lung cells,[3] which could spill over into the bloodstream and contribute to chronic inflammation and promote progression of

atherosclerosis.[3] Consistent with this hypothesis, studies have linked $PM_{2.5}$ exposure to higher levels of circulating inflammatory cytokines such as C-reactive protein, interleukin-6, and firbingogen.[3,13] Similarly, systemic oxidative stress may result from spillover of localized pulmonary oxidative stress.[3] Moreover, oxidative stress may lead to and be induced by inflammation.[3] Activation of the innate immune response increases generation of reactive oxygen species (ROS), and ROS in the lungs may upregulate a whole host of cytokines and chemokines.[3] In addition, autonomic nervous system imbalance may contribute to systemic oxidative stress and inflammation.[3] Increased systemic inflammation and excess ROS may lead to vasoconstriction by reducing vasodilator bioavailability (e.g. nitric oxide) due to excess ROS formation.[3] These pathways involving inflammation, oxidative stress, and endothelial dysfunction likely promote the development of atherosclerosis, which is a progressive inflammatory disease characterized by build-up of calcium, lipids, inflammatory smooth muscle cells, and white blood cells in the arteries.[14]

DNA METHYLATION

DNA methylation is the most well-studied epigenetic mechanism and occurs when a methyl group is added to DNA without changing the sequence of DNA.[15] DNA methylation can modulate gene expression and can be heritable through cell division.[15] DNA methylation is commonly believed to silence gene expression. While this is generally the case in CpG-rich promoters, DNA methylation may increase or decrease expression in other genomic regions.[16] Alterations in DNA methylation are known to be crucial to mammalian development, but there is increasing recognition that DNA methylation is also sensitive to environmental exposures and may link environmental pollutants to disease. While mammalian DNA is primarily methylated at cytosine guanine dinucleotide (CpG) sites, recent evidence shows that DNA methylation in non-CpG sites are present in mammalian cells and may have important roles in regulating gene expression.[17] DNA methylation assessment techniques in this dissertation only measure CpG

methylation. However, in a study of cultured human monocytes, only 10% of cytosine methylation was attributable to non-CpG methylation.[18] Thus, CpG methylation may be more relevant in this dissertation which focuses on DNA methylation of human monocytes, which are cells central to atherosclerosis.

AIR POLLUTION AND DNA METHYLATION

Several previous studies have linked air pollution exposure to alterations in DNA methylation.[19–22] In a study of 1,097 blood samples from 718 elderly men from the Boston area, long interspersed nucleotide element (LINE)-1 methylation decreased after recent exposure to black carbon and $PM_{2.5}$.[19] In two pollutant models, only black carbon remained significantly associated with LINE-1 methylation with an effect size that was equivalent to a decline in LINE-1 methylation corresponding to aging 3.4 years.[19] Black carbon and $PM_{2.5}$ were not associated with Alu element methylation.[19] In the same cohort, a separate analysis was conducted using intermediate-term exposures (averaged over 28-180 day intervals); the investigators reported that exposure to black carbon was associated with lower methylation of both LINE-1 and Alu methylation, and higher exposure to SO_4 was associated with a decrease in LINE-1 methylation.[20] According to the authors, the effect of this magnitude of hypomethylation would be similar to a 2-year increase in age and a 5% increased risk of developing ischemic heart disease in this cohort.[20] Intermediate-term $PM_{2.5}$ exposure was not associated with either LINE-1 or Alu element methylation.[20] Candidate gene analysis using the same cohort found that acute exposure to both $PM_{2.5}$ and black carbon was associated with hypomethylation of the inducible nitric oxide synthase gene (*iNOS*), which is related to inflammation and the stress response.[21] Occupational studies also provide evidence that exposure to particulate matter led to changes in DNA methylation of LINE-1 and Alu repetitive elements[23] and the *iNOS* gene.[23,24]

More recently, an epigenome-wide analysis of DNA methylation and $PM_{2.5}$ (averaged over 2, 7, and 28 days) was conducted using the Illumina Infinium HumanMethylation450

BeadChip (450k) in three independent cohorts: KORA F3 (n=500) and F4 (n=1,799) from Germany and Normative Aging Study (n=657) from the U.S.[22] The Normative Aging Study only included male participants whereas roughly half of KORA participants were female. Daily fixed monitors were used to measure $PM_{2.5}$ concentrations. Separate epigenome-wide analyses were conducted in each of the three cohorts and results were pooled using meta-analysis. The investigators identified 12 significant CpG sites that reached genome-wide Bonferroni significance. These CpGs were located in genes such as *NSMAF* (Neutral Sphingomyelinase (N-SMase) Activation Associated Factor) and *NXN* (nucleoredoxin), which may be involved in processes such as oxidative stress and cancer.[22]

DNA METHYLATION AND CARDIOVASCULAR DISEASE

Lower global DNA methylation has been demonstrated in atherosclerotic arteries compared to normal arteries.[25] Blood LINE-1 hypomethylation was associated with prevalent ischemic heart disease and higher risk of incident ischemic heart disease, stroke, and total mortality.[26] However, a study using whole-genome shotgun bisulfite sequencing and the Illumina 450k array observed more highly methylated CpGs and differentially methylated regions in atherosclerotic aortas compared to normal aortas.[27] They also used whole-genome shotgun bisulfite sequencing to derive DNA methylation profiles of several repeated sequences and found Alu to be hypermethylated and LINE to be hypomethylated in atherosclerotic tissues compared to control tissues.[27] In addition, the epigenome-wide study identified many locus-specific DNA methylation differences in atherosclerotic lesions, and these loci mapped to genes involved in vascular homeostasis such as *HOXA9*.[27] Another recent study found genes involved in inflammation and immune processes to be differentially methylated between tissue from advanced atherosclerotic plaques and atherosclerotic-resistant internal mammary arteries that were obtained from the same patients with coronary heart disease.[28] Other studies have found associations between candidate gene DNA methylation, such as hypomethylation of the *AHRR*

(aryl hydrocarbon receptor repressor) and hypermethylation of *NR3C1* (glucocorticoid receptor genes), and subclinical atherosclerosis.[29,30] While different results have been found, these findings suggest that DNA methylation might modulate genes and pathways involved in the pathogenesis of atherosclerosis.

THE MULTI-ETHNIC STUDY OF ATHEROSCLEROSIS (MESA)

The Multi-Ethnic Study of Atherosclerosis (MESA) is a population-based prospective cohort study of characteristics and risk factors related to subclinical and clinical CVD.[31] It dedicated substantial resources to quantify subclinical CVD accurately over time in a diverse multi-ethnic cohort from six U.S. regions (Baltimore, Maryland; Chicago, Illinois; Forsyth County, North Carolina; Los Angeles County, California; New York, New York; and St. Paul, Minnesota). The study of risk factors related to disease before clinical manifestation holds promise in identifying targets for potential clinical and public health intervention. Coronary artery calcium (CAC) was measured using computed tomography (CT) and is a marker of subclinical atherosclerosis.[31]

Beginning in July 2000, MESA recruited 6,814 participants between the ages of 45-84. The cohort was 28% Black, 12% Chinese-American, 22% Hispanic, and 38% White with a roughly equal proportion of men and women. Participants were excluded if they had cardiovascular disease at baseline, were undergoing active treatment for cancer, had a serious medical condition that would prevent long-term participation, weighed over 300 pounds, were cognitively impaired, lived in or was on the waitlist for a nursing home, planned to leave the community within five years, or did not speak at least one of the following languages: English, Spanish, Cantonese, or Mandarin.

At baseline, participants underwent a blood draw and an extensive exam that measured demographics, social factors, lifestyle factors, medications, medical history, psychosocial factors, and anthropometrics. Participants also received a CT scan at baseline. There have been four

follow-up visits spanning 10 years since baseline that included both questionnaires and physical examination to ascertain risk factors and outcome measures longitudinally. Half of the cohort were randomly selected to receive a follow-up CT scan at exam 2 while the other half received a scan at exam 3. During exam 4, those without exam 3 scans were preferentially selected for CT scans. At exam 5, participants with scans at exams 3 and 4 were prioritized for CT scans. These CT scans were used to assess CAC.

MESA AIR

MESA Air is an ancillary study that started in 2004 and recruited additional participants and added outcome measurements and air pollution exposure assessments to the original MESA study.[32] Most participants were recruited into MESA Air from the parent MESA study during the end of the third and throughout the fourth examinations, where 93% (n= 5,478) of eligible participants agreed to participate.[32] Others were recruited from another large MESA ancillary study, the MESA Family study, which investigates genetic aspects of subclinical CVD in African-American and Hispanic-American participants; 84% (n= 491) of eligible subjects agreed to participate. Additional participants (n=257) were recruited specifically for MESA Air to exploit exposure heterogeneity upwind and/or downwind of the surrounding areas of two existing MESA sites: New York and Los Angeles. Community-based recruitment emphasized balancing recruitment across census blocks, ethnicity, and gender.

MESA EPIGENOMICS AND TRANSCRIPTIOMICS STUDY

A second ancillary study, the MESA Epigenomics and Transcriptomics Study, obtained CD14+ purified monocytes from a randomly selected subsample of 1,264 MESA participants at the 5th examination (April 2010-February 2012) and measured epigenome-wide DNA

methylation and genome-wide gene expression.[33] Four of the six MESA sites (Baltimore, MD; New York, NY; St. Paul, MN; and Winston-Salem, NC) were included in the study.

AMBIENT AIR POLLUTION ASSESSMENT

Many prior studies of the association between air pollution and DNA methylation are affected by measurement error and potential exposure misclassification of air pollution exposures. Some studies related methylation outcomes to only temporal variability in exposure, which typically rely on exposures aggregated over space such as daily monitors over a metropolitan area. MESA addresses these limitations through its state-of-the-art exposure estimation based on cohort-specific air pollution monitoring and spatiotemporal modeling. Our study used ambient $PM_{2.5}$ and NO_X predictions estimated at participants' residences using spatiotemporal models optimized via maximum likelihood that incorporate monitoring data, cohort-specific modeling, geographical variables such as roadway density, meteorological data, emission data, and land use variables.[34–36] Monitoring data were obtained from regulatory monitors from the U.S. Environmental Protection Agency's (EPA) Air Quality System (AQS) and cohort-specific monitors deployed by MESA Air. MESA Air monitors were positioned at fixed sites, participants' homes, and specific locations that measure roadway concentration gradients. These efforts greatly enhanced characterization of fine-scale spatial differences in concentrations for each participant. Predictions were available as 2-week averages.

EPIGENOME-WIDE PROFILING OF DNA METHYLATION IN PURIFIED MONOCYTES

Tissue type and cell type composition greatly influence DNA methylation patterns and may confound studies of mixed cells and reduce generalizability across studies.[37] Inability to adjust for cell type composition may also mask important findings. Measuring DNA methylation from target cell types that play vital roles in the disease pathway can improve sensitivity of

studies. Past studies of DNA methylation and air pollution are limited by the use of mixed leukocytes. In this dissertation, DNA methylation was measured from monocytes isolated with anti-CD14-coated magnetic beads using AutoMACs automated magnetic separation unit (Miltenyi Biotec, Bergisch Gladbach, Germany).[33] Monocytes were chosen because they are vital to the initiation and progression of atherosclerosis.[14] Epigenome-wide DNA methylation of a subsample of 1,264 randomly selected MESA participants were measured using the Infinium 450K array. The Infinium 450K is a microarray-based platform that provides an epigenome-wide assessment of DNA methylation of over 480,000 CpG sites covering 99% of RefSeq genes.[38] The array also covers 19,755 CpG islands in addition to shore regions, miRNA promoters, and non-CpG sites.[38]

Previous work in the MESA Epigenomics and Transcriptomics study identified 11,203 *cis*-acting expression-associated methylation sites (eMS).[33] DNA methylation of the 2,713 top (smallest p-values) unique CpG sites were associated with expression of 3,093 gene transcripts. We selected these 2,713 eMS as candidate sites for the assessment of differential methylation with respect to air pollution exposure.

GLOBAL DNA METHYLATION ASSESSMENT

Some researchers have concentrated on genome-wide reduction in DNA methylation, or global hypomethylation, as the focus of studies of environmental agents because DNA hypomethylation may lead to reactivation of transposable elements, deregulation of genes, chromosomal instability, and increased mutations.[39] It is estimated that over one-third of DNA methylation occurs in repetitive elements (e.g. LINE-1 and Alu elements), which comprise over half of the human genome.[40] Repetitive element methylation has been suggested to correlate with global genomic DNA methylation and has been used as a surrogate marker for global DNA methylation.[40,41] Prior studies have found associations between exposure to air pollutants and global DNA methylation, as measured by Alu and LINE-1 methylation.[19,20,23]

In this dissertation, we used DNA methylation of Alu and LINE-1 repetitive elements as surrogates of global DNA methylation. We identified over 12,000 and 9,000 probes that mapped to Alu and LINE-1 repetitive elements, respectively, by intersecting probe locations from the Infinium 450K array with RepeatMasker.[42]

BUMP HUNTING METHOD

Studies using the Infinium 450k typically utilize a site-by-site approach similar to genome-wide association studies. However, DNA methylation data are inherently very different from that of single nucleotide polymorphism (SNP) data. SNP data are categorical, but DNA methylation data from microarrays are continuous.[43] While a specific CpG site is either methylated or not, DNA methylation arrays aggregate data from millions of cells (e.g. monocytes) and the resulting measure is continuous and represents the proportion of cells methylated at the CpG site in question.[43] DNA methylation data is also measured with greater uncertainty.[43] In addition, DNA methylation measurements are strongly correlated throughout the epigenome and are also more densely spaced than SNPs.[43] Measurements of DNA methylation of individual CpG sites may be quite noisy, and significant results may be due to noise rather than a true association. Owing to these differences, a "bump hunting" method proposed by Jaffe et al. complements standard site-specific methods.[43] The bump hunting method exploits the correlation between adjacent CpG probes to search for associations at the region level rather than the individual site level.[43]

The bumphunter package in R was used to identify DMRs associated with long-term air pollution exposure. The bump hunter approach has four steps.[43] First, it regresses methylation on the predictor of interest (e.g. air pollution) for each CpG site just like in CpG site-specific analyses to obtain a coefficient for each CpG site. In the second step, the coefficients of the CpG sites are smoothed according to genomic position using loess. The third step uses the smoothed estimate to identify candidate regions. Candidate regions are regions of consecutive

measured locations where the smoothed estimate exceeds a predefined threshold (such as 97.5th percentile of the smoothed estimate). We defined consecutive sites as those that are separated by no more than 500 bp. The last step assigns statistical uncertainty to each candidate region by permuting the exposure (here it will be air pollution) or by bootstrapping. To account for multiple testing, we used an FDR approach with a cutoff at 0.05.

SUMMARY OF CHAPTERS

The analyses described in this dissertation leverage the rich air pollution, methylomics, transcriptomics, and subclinical atherosclerosis data sets available in MESA to investigate the methylomic signature of ambient air pollution exposure and its association with subclinical atherosclerosis and gene expression in purified monocytes.

In Chapter 2, we assessed the association between long-term exposure to air pollution ($PM_{2.5}$ and NO_X) and DNA methylation in monocytes (globally and at candidate CpG sites). DNA methylation of LINE-1 and Alu repetitive elements were used as surrogates of global DNA methylation. Candidate CpG sites were selected because they were previously associated with *cis*-gene expression in the same sample of MESA participants.[33] We also investigated the association between gene expression and air pollution for CpG sites with DNA methylation associated with air pollution.

Chapter 3 describes an epigenome-wide analysis of associations between long-term exposure to ambient air pollution ($PM_{2.5}$ and NO_X) and DNA methylation in monocytes. We used a false discovery rate of 0.05 to account for multiple comparisons.[44] We also investigated the association between significant air pollution-associated methylation signals and *cis*-gene expression of genes within 1 Mb of the methylation signals in question.

In Chapter 4, we analyzed the cross-sectional and longitudinal associations between air pollution-associated methylation signals and CAC.

Finally, in Chapter 5 we present overarching conclusions and limitations of the studies presented in the dissertation. We also discuss how our results fit into the broader air pollution literature and potential future directions.

CHAPTER 2

Long-term Outdoor Air Pollution and Global and Candidate Site DNA Methylation in Circulating Monocytes: The Multi-Ethnic Study of Atherosclerosis (MESA)

ABSTRACT

Background: DNA methylation may mediate effects of air pollution on atherosclerosis and cardiovascular diseases. The association between air pollution and DNA methylation in monocytes, which are central to atherosclerosis, has not been studied.

Objectives: We investigated the association between long-term ambient air pollution exposure and DNA methylation in purified monocytes of adults age 55 and over.

Methods: One-year average ambient fine particulate matter ($PM_{2.5}$) and oxides of nitrogen (NO_X) concentrations were predicted at participants' (n=1,207) addresses using spatiotemporal models. We assessed DNA methylation in circulating monocytes at 1) 2,713 CpG sites associated with mRNA expression of nearby genes and 2) probes mapping to Alu and LINE-1 repetitive elements (surrogates for global DNA methylation) using Illumina's Infinium 450K methylation array. We used linear regression models adjusted for potential confounders such as demographics, smoking, and socioeconomic status. For significant air pollution-associated CpGs, we assessed the association between expression of gene transcripts previously associated with these CpG sites and air pollution.

Results: At a false discovery rate of 0.05, five candidate CpGs had methylation significantly associated with $PM_{2.5}$ and none with NO_X. The nominal p-value for the CpG with the strongest signal, cg20455854, was 2.77×10^{-5}. mRNA expression profiles of genes near three $PM_{2.5}$-associated CpGs (*ANKHD1*, *LGALS2*, and *ANKRD11*) were also significantly associated with $PM_{2.5}$ exposure. These genes may be involved in pathways such as inflammation and

monocyte endothelial adhesion. Alu and LINE-1 methylation were not associated with air pollution.

Conclusions: We observed associations between long-term ambient air pollution exposure and site-specific DNA methylation, but not global DNA methylation, in purified monocytes. DNA methylation may potentially mediate the associations of ambient air pollution with atherosclerosis and CVD.

INTRODUCTION

Ambient air pollution ranks among the top ten leading risk factors for global mortality and disease[1] and has been linked to increased risk of cardiovascular disease (CVD).[3] Although much uncertainty exists regarding the underlying biological mechanisms for this association, DNA methylation has been postulated to mediate the effects of air pollution on CVD.[39]

DNA methylation is a chemical modification of DNA that involves the addition of a methyl group to cytosine and predominantly occurs at cytosine-guanine dinucleotide (CpG) sites.[15] DNA methylation in CpG-rich promoters tends to repress transcription whereas the effect is context-dependent in other genomic regions.[16]

Previous studies show that DNA methylation patterns are associated with atherosclerosis[45] and ischemic heart disease[26] while others have suggested DNA methylation as a biomarker for cardiovascular disease.[46] Moreover, exposure to air pollutants such as black carbon, particulate matter, benzene, and SO_4 led to an epigenome-wide reduction in DNA methylation, or global hypomethylation, and were associated with methylation levels of candidate genes.[19,20,23,47,48] Global hypomethylation may lead to reactivation of transposable elements and increased mutations[39] and has also been linked to atherosclerosis.[25]

Results from past studies, however, are difficult to interpret due to several limitations. First, they were limited by the use of peripheral blood leukocytes, which are composed of a mixture of cells that have unique DNA methylation profiles. Evaluating samples with different

cell mixture proportions may lead to confounding. Since DNA methylation is cell-specific, it is more relevant to study purified samples of a single cell type such as monocytes. Monocytes play an important role in atherogenesis by promoting chronic inflammation and differentiating into macrophages that accumulate in plaques.[14] In addition, most prior studies did not measure gene expression to investigate the potential biological relevance of methylation differences. Some studies were further limited by potential exposure misclassification and measurement error in air pollution exposures.

We evaluated the association between long-term exposure to ambient air pollution and global and candidate CpG site DNA methylation in circulating monocytes of 1,207 participants from the Multi-Ethnic Study of Atherosclerosis (MESA). Candidate sites were those with CpG methylation previously associated with expression of nearby genes in the MESA cohort. For significant air pollution-associated methylation sites, we also assessed the association between air pollution and expression of genes previously associated with the methylation sites.

METHODS

Study Population

MESA is a prospective cohort study of risk factors for cardiovascular disease and atherosclerosis in multi-ethnic adults.[31] We also utilized data from two MESA ancillary studies: MESA Air and the MESA Epigenomics and Transcriptomics Study. Please refer to Chapter 1, pages 5-7, for additional details about the recruitment and design of these studies. The analysis in this chapter included a subsample of MESA participants (n=1,207) who had available methylomic, transcriptomic, and air pollution data from four MESA field centers (Baltimore, MD; New York, NY; St. Paul, MN; and Winston-Salem, NC) at the 5th examination (April 2010-February 2012). Since DNA methylation and gene expression data were not available for participants from the Los Angeles and Chicago MESA sites, there were no Asian participants in our analytic sample. Those included in the study were generally representative of all

participants from the 5th examination at the four MESA field centers represented except for having fewer Black participants, more Hispanic participants, and fewer participants from Winston-Salem, NC (Table 2.1).

Air Pollution Assessment

$PM_{2.5}$ and NO_X predictions were available through MESA Air as likelihood-based 2-week averages at participant residences from 1999-2012. Additional details about air pollution assessment can be found in Chapter 1, page 7. For the present analyses, individual-level outdoor residential concentrations of $PM_{2.5}$ and NO_X were averaged over the 12 months prior to blood draw.

DNA Methylation and Gene Expression Quantification

Blood was drawn at the 5th examination (April 2010-February 2012). Monocytes were purified on-site by trained technicians following standardized protocols with extensive quality control measures using anti-CD14-coated magnetic beads and AutoMACs automated magnetic separation unit (Miltenyi Biotec, Bergisch Gladbach, Germany).[33] Monocytes samples were consistently >90% pure based on flow cytometry of 18 specimens. A NanoDrop spectrophotometer measured optical density measurements for DNA and RNA quality control. The integrity of 18s and 28s ribosomal RNA was also evaluated using the Agilent 2100 Bioanalyzer with RNA 6000 Nano chips (Agilent Technology, Inc., Santa Clara, CA, USA), and RNA with RNA Integrity Numbers (RIN) >9.0 were included in the expression microarrays. DNA and RNA were extracted simultaneously from purified monocytes using AllPrep DNA/RNA Mini Kit (Qiagen, Inc., Hilden, Germany).

Genome-wide methylation profiles of over 485,000 CpG sites were characterized in purified monocytes using the Infinium HumanMethylation450 BeadChip (450k; Illumina, Inc. CA, USA) as previously described.[33] This methylation data has been deposited in the NCBI Gene

Expression Omnibus and is accessible through GEO Series accession number (GSE56046). Bisulfite conversion of DNA fragments was performed with the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA). Bead-level methylation data were summarized using the Illumina *GenomeStudio* software, and then raw methylation calls were normalized and converted to M-values (log ratio of methylated to unmethylated intensities). The M-value was used in statistical analyses due to better statistical performance in differential methylation analyses.[49] The Illumina HumanHT-12 v4 Expression BeadChip and Ilumina Bead Array Reader were used to obtain genome-wide expression profiles of over 48,000 transcripts (Illumina, Inc. CA, USA).[33] A random sampling technique was used to assign samples to chips and positions to mitigate batch effects for both arrays.

Data pre-processing was performed using Bioconductor[50] in R.[51] For expression data, Illumina's GenomeStudio was used to account for local background and the *beadarray* package[52] was used for quality check analyses and bead type summarization. Negative controls on the array were used to compute detection p-values. The *neqc* function of the *limma*[53] package was used to perform normal-exponential convolution model analysis to estimate non-negative signal, conduct quantile normalization using all probes (gene and control, detected and not detected) and samples, add a recommended (small) offset, perform log base 2 transformation, and eliminate control probe data from the normalized expression matrix. Multidimensional scaling plots showed the five common control samples were highly clustered together and identified three outlier samples which were subsequently excluded.

Normalization of the Illumina HumanMethylation450 Beadchip two-channel systems was performed by 1) smooth quantile normalization, 2) subtraction of the median intensity value of negative control probes, and 3) standard quantile normalization applied to bead-type intensities and combined across Infinium I and II assays and both colors using the *lumi* package.[54] Quality control involved checking for sex and race/ethnicity mismatches and outlier identification by multidimensional scaling plots. We also excluded probes designed for sequences on the X or

the Y chromosome, uninformative probes (defined as those probes with detected methylation levels in <90% of participants using a detection p-value cutoff of 0.05), and probes with single nucleotide polymorphisms within 10 bp.

To reduce dimensions of our epigenome-wide methylation data, we focused on CpG sites more likely to be functionally relevant. Previous work in this cohort identified 11,203 methylation sites that were associated with the *cis*-expression of 3,093 gene transcripts in a sample of 1,264 randomly selected MESA participants.[33] There were 2,713 unique CpG sites among the most significant CpG sites (ranked by smallest p-values) associated with each of the 3,093 transcripts. We tested the association of these 2,713 expression-associated methylation sites (eMS) with $PM_{2.5}$ and NO_X. A database of the 11,203 significant methylation-gene expression association results can be downloaded at http://www.wakehealth.edu/Research/Public-Health-Sciences/MesaEpiGenomics/eMS-database---The-MESA-EpiGenomics-Project.htm. In our study, eMS were annotated with respect to the gene with expression most significantly associated (based on smallest p-value) to the eMS.

Alu and LINE-1 repetitive element DNA methylation were used as surrogates for global DNA methylation. By intersecting probe locations from the Infinium 450K array with RepeatMasker,[42] we identified 12,456 and 9,507 probes for Alu and LINE-1 repetitive elements, respectively, that passed quality control exclusion. Median Alu and LINE-1 DNA methylation were calculated for each participant for use in regression analyses.

Other Measures

Information on demographics, socioeconomic status, medication, lifestyle factors, and body mass index were obtained through questionnaires (self- and interviewer-administered) and physical examination. Race/ethnicity, sex, and education were obtained at baseline. Food frequency questionnaire was used to estimate intake of methyl nutrient intake at exam 5.

Nutrients were calculated using the Nutrition Data System for Research (NDS-R database; Nutrition Coordinating Center, Minneapolis, MN, USA) as previously described.[55] A physical activity survey adapted from the Cross-Cultural Activity Participation Study[56] was used to assess the frequency and time spent in various physical activities during a typical week in the previous month. Recent infection was defined as having a fever, cold or flu, urinary infection, bronchitis, sinus infection, pneumonia, tooth infection, or arthritis flare-up within the past two weeks. Enrichment scores for neutrophils, B cells, T cells, and natural killer cells were calculated from a gene set enrichment analysis using the gene expression signature of each cell type obtained from previously defined lists.[33]

Statistical Analysis

The associations between air pollution ($PM_{2.5}$ or NO_X) and methylation M-values at candidate CpGs were assessed with least squares regression and robust empirical Bayes moderated t-statistics using the *limma* package from Bioconductor.[53] To account for multiple testing of the 2,713 eMS, we controlled the false discovery rate (FDR) at 0.05 using the Benjamini-Hochberg method.[44] For eMS that were significantly associated with $PM_{2.5}$ or NO_X, we also tested the association between the transcript previously associated with that eMS and $PM_{2.5}$ or NO_X using linear models with least squares regression and robust empirical Bayes moderated t-statistics using *limma*. The associations between air pollution and median Alu and median LINE-1 methylation were examined using multiple linear regression with robust standard errors.

In all analyses, we adjusted for age, sex, race/ethnicity (Black, Hispanic, White), household income, education, neighborhood socioeconomic status, smoking (smoking status and pack-years), secondhand smoke, body mass index, recent infection (summary index derived from factor analysis),[57] methyl nutrient intake calculated from food frequency questionnaire (continuous folate, vitamin B12, vitamin B6, methionine, zinc), physical activity,

study site, microarray chip, and chip position (DNA methylation analysis only). We also adjusted for residual sample contamination by non-monocytes by adjusting for enrichment scores for neutrophils, B cells, T cells, and natural killer cells. Gene set enrichment analysis of gene expression values for each participant was used to estimate enrichment scores for neutrophils, B cells, T cells, and natural killer cells based on the gene signature of each blood cell type from previously defined lists.[33] The *ComBat* function in the package *SVA*[58] was used to obtain methylation chip-adjusted M-values for use in regression analyses.

No covariate had more than 4% missing data. Multivariate imputation using chained equations was used to impute missing values in Stata 13.[59] Since the proportion of missing data was low, only one iteration was used in association analyses.

Sensitivity analyses were conducted to investigate whether the association between exposure to $PM_{2.5}$ and NO_X and Alu and LINE-1 methylation would be sensitive to a shorter averaging period of 2 weeks for the MESA Air likelihood-based air pollution predictions. In addition to including all covariates in the main analysis, this analysis also adjusted for temperature, relative humidity, month of blood draw, and day of week of blood draw. Secondary analyses were conducted to investigate potential effect modification sex and race/ethnicity by fitting interaction terms for air pollution and sex or race/ethnicity. If an overall interaction p-value was less than 0.05, stratified analyses were conducted.

Functional Annotation Analysis

We investigated the potential functional relevance of air pollution-associated CpGs by using data from publicly available databases. *In silico* functional prediction of chromatin states in monocytes was performed using ChromHMM,[60] which is based on a hidden Markov model, to predict segmentation among six states based on histone modifications in monocyte samples from the BLUEPRINT[61,62] (H3K27ac, H3K4me1, H3K4me3) and ENCODE[63] (H3K36me3) projects. Annotation also included DNase hypersensitive hotspot data in a monocyte sample

(Sample ID RO01746, data generated by the UW ENCODE group) and transcription factor binding sites detected in any cell type available from ENCODE.[63] Data were accessed from the UCSC Genome Browser[64] and the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/).

RESULTS

The analytic sample consisted of 1,207 participants with mean age of 69.6 (Table 2.2). Of these participants, 21.2% were Black, 31.6% were Hispanic, and 47.2% were White. Roughly half were female, and most were former (50.3%) or never (40.3%) smokers. The average body mass index was 29.7 kg/m^2. Over 65% of participants received education beyond high school, and nearly 40% had incomes over $50,000. Median Alu and median LINE-1 methylation were 2.4 and 2.6 M-value units, respectively. Figure 2.1 shows plots of site-specific 12-month average ambient $PM_{2.5}$ and NO_X predictions. The overall average $PM_{2.5}$ and NO_X predictions were 10.7 μg/m^3 (interquartile range [IQR]=2.2 μg/m^3) and 28.7 ppb (IQR=31.9 ppb), respectively. $PM_{2.5}$ and NO_X were positively correlated (correlation coefficient=0.82). Table 2.2 summarizes participant characteristics by study sites.

Figure 2.2 shows a plot of the $-\log_{10}$(p-values) against genomic position from linear regression analysis of 2,713 eMS with $PM_{2.5}$. We identified five eMS with methylation significantly (FDR<0.05) associated with $PM_{2.5}$ (four positively and one negatively), and they included cg20455854 (*ANKHD1*; ankyrin repeat and KH domain containing 1), cg07855639 (*LGALS2*; lectin, galactoside-binding, soluble, 2), cg07598385 (*ANKRD11*; ankyrin repeat domain 11), cg17360854 (*BAZ2B*; bromodomain adjacent to zinc finger domain, 2B), and cg17360854 (*PPIE*; peptidylprolyl isomerase E). The statistically significant eMSs associated with $PM_{2.5}$ are shown in Table 2.3. The eMS cg20455854 had methylation associated with *ANKHD1* expression and was the eMS most significantly associated with $PM_{2.5}$ (β=0.139, 95% confidence interval [CI]: 0.074, 0.203; per 2.5 μg/m^3). Cg20455854 is located within a DNase

hypersensitivity site (monocyte data from ENCODE), a transcription factor binding site (any cell type from UCSC Genome Browser), and a predicted strong enhancer region (Table 2.4). NO_X exposure was not significantly associated with methylation of any eMS (Table 2.5). We did not identify significant interactions between air pollution and race/ethnicity. However, the overall interaction p-value for $PM_{2.5}$ and sex was 0.01 for cg07855639 (*LGALS2*). In sex-stratified analyses, the effect estimates for females and males were 0.104 (95% CI: 0.051, 0.156; p=1.3e-4) and 0.071 (95% CI: 0.014, 0.129; p=0.015), respectively.

Although all candidate CpGs were previously associated with mRNA expression of a nearby gene in this cohort, not all of the mRNA expression profiles associated with $PM_{2.5}$-associated eMS were significantly (FDR<0.05) differentially expressed with respect to air pollution. Three of the five genes with mRNA expression previously associated with the significant $PM_{2.5}$-eMS had mRNA expression also associated with $PM_{2.5}$. $PM_{2.5}$ exposure was negatively associated with *ANKHD1* (β=-0.048; 95% CI: -0.074, -0.022; per 2.5 μg/m^3; p=3.71E-04), *LGASL2* (β=-0.147; 95% CI: -0.240, -0.053; per 2.5 μg/m^3; p=0.002), and *ANKRD11* expression (β=-0.075; 95% CI: -0.142, -0.008; per 2.5 μg/m^3; p=0.028) (Table 2.6).

Median DNA methylation of Alu was not associated with $PM_{2.5}$ (β=-0.003, 95% CI: -0.006, 0.001, p=0.155) or NO_X (β=0.001, 95% CI: -0.006, 0.008, p=0.719) after covariate adjustment (Table 2.7). Median DNA methylation of LINE-1 was also not associated with $PM_{2.5}$ (β=-0.003, 95% CI: -0.007, 0.001, p=0.177) or NO_X (β=-0.0004, 95% CI: -0.009, 0.008, p=0.927) after covariate adjustment (Table 2.7). Results were not sensitive to using a 2-week averaging period for air pollutants (Table 2.7). We did not identify interactions between air pollution and sex or race/ethnicity (p>0.10).

DISCUSSION

We examined the associations between long-term air pollution exposure and DNA methylation in monocytes and found significant site-specific associations. In order to identify

potentially functionally relevant genes involved in the pathogenesis of air pollution-related cardiovascular disease, we focused on a set of CpG sites that were previously associated with expression of nearby genes. We detected five CpGs with methylation associated with long-term $PM_{2.5}$ exposure in monocytes. Expression of three genes related to these methylation sites was associated with air pollution. No site-specific methylation site was significantly associated with long-term NO_X exposure. We did not find evidence of an association between exposure to long-term ambient air pollution and global DNA methylation in monocytes.

Genes with both regulatory DNA methylation site(s) and mRNA expression that are associated with air pollution may be more plausibly involved in the pathogenesis of air pollution-related disease. In this study, we identified three such genes: *ANKHD1*, *LGALS2*, and *ANKRD11*. While we did not formally evaluate whether DNA methylation mediates the association of air pollution with atherosclerotic cardiovascular disease in this chapter and temporal relationships are not clear, our findings suggest that DNA methylation may potentially link air pollution inhalation with changes in monocyte gene expression that could contribute to disease.

ANKHD1 function in monocytes is unknown, but it is highly expressed in acute leukemia, multiple myeloma, and prostate cancer and may support proliferation and cell cycle progression of cancer cells.[65–68] It may modulate yes-associated protein (YAP) expression and activation,[66] a finding potentially relevant to atherosclerosis. During leukocyte transmigration (a critical process in the pathogenesis of atherosclerosis), chemokines, selectins, and adhesion molecules interact to recruit leukocytes to the site of inflammation.[69] In a mouse model, YAP knockdown decreased shear stress- and cytokine-induced production of pro-inflammatory adhesion molecules in monocytes and endothelial cells, limiting monocyte endothelial cell adhesion.[70] Furthermore, overexpression of constitutively active YAP increased endothelial inflammation and monocyte adhesion to endothelial cells.[70] Plausibly, ANKHD1 may modulate monocyte adhesion through YAP.

In our study, $PM_{2.5}$ exposure correlated with higher methylation of *ANKHD1*-eMS cg20455854 and lower mRNA expression of *ANKHD1*. The location of cg20455854 within a DNase hypersensitivity site, transcription factor binding site, and predicted strong enhancer region provides support of a potentially functional $PM_{2.5}$-eMS. Together, our results suggest a possible role of $PM_{2.5}$ in modifying monocyte adhesion during atherosclerosis.

LGALS2 encodes the protein galectin-2, which is part of the galectin family of proteins that typically binds β-galactose-containing glycoconjugates.[71] Galectins affect many cellular functions including the inflammatory response.[71] Galectin-2 can bind inflammatory cytokine lymphotoxin-α and regulate its secretion *in vitro.*[72] In addition, case-control studies reported that the 3279C>T (rs7291467; chromosome 22, position 37576621) polymorphism in *LGALS2* was negatively associated with risk for myocardial infarction[72] and coronary artery disease.[73,74] This SNP is located over 380 kb away from the *LGALS2*-eMS cg07855639. The CC genotype of this polymorphism also increases expression of galectin-2, and higher levels of galectin-2 exist in monocytes and macrophages of coronary artery disease patients with low arteriogenic response.[75] Galectin-2 also induced a pro-inflammatory phenotype in monocytes and macrophages *in vitro* and in mice.[76] However, the TT genotype, which leads to lower levels of galectin-2, was associated with more severe coronary stenosis[77] and higher levels of C-reactive protein.[73] Moreover, addition of galectin-2 to activated T cells downregulated production of the pro-inflammatory cytokines interferon-γ and tumor necrosis factor-α, while increasing secretion of the anti-inflammatory cytokine interleukin-10.[78] This suggests that galectin-2 may modulate both pro-inflammatory and anti-inflammatory cytokines.

In our study, $PM_{2.5}$ exposure was positively associated with methylation of the *LGALS2*-eMS cg07855639 and negatively associated with *LGALS2* mRNA expression in monocytes. The identification of *LGALS2* suggests a potential role of the inflammation pathway in mediating air pollution-related CVD, which is consistent with the current literature.[3]

ANKRD11 regulates chromatin modification and has been found to be associated with autism[79] and KBG syndrome (rare disorder marked by dental, neurobehavioral, craniofacial and skeletal anomalies).[80] In our cohort, $PM_{2.5}$ was positively associated with *ANKRD11* methylation and negatively associated with *ANKRD11* mRNA expression. Although air pollution is a risk factor for cardiovascular disease, air pollution has also been linked to other complex diseases such as autism, where a suggested common mechanism might be inflammation.[81] It is also possible that *ANKRD11* or cg07598385 may exhibit pleiotropic effects on autism, KBG syndrome, and potentially air pollution-related diseases.

BAZ2B and *PPIE* methylation, but not mRNA expression, were significantly associated with $PM_{2.5}$ exposure. BAZ2B is a bromodomain containing chromatin remodeling protein that epigenetically regulates transcription. Although the function of BAZ2B remains unclear, single nucleotide polymorphisms in *BAZ2B* were associated with sudden cardiac death.[82] PPIE belongs to the family of PPIases with proline isomerase activity that stimulates folding and conformational changes in proteins and may be linked to leukemia,[83] colorectal cancer,[84] and body mass index.[85]

Previous studies have associated air pollution exposure with DNA methylation of a number of genes including iNOS,[21,24] tissue factor, intercellular adhesion molecule 1, toll-like receptor 2, interferon-γ, and interleukin-6.[47] In addition, a prior epigenome-wide study of whole blood using the 450k array identified 12 CpG sites associated with short- and mid-term $PM_{2.5}$ exposure at genome-wide significance.[22] Of these genes and CpG sites, only one CpG site associated with intercellular adhesion molecule 1 were among the list of eMS analyzed in this study, and it was not associated with long-term $PM_{2.5}$ or NO_X exposure.

We were unable to replicate prior observational studies that found significant associations between air pollution exposure and global DNA methylation using Alu and LINE-1 methylation as surrogates.[19,23,86] Short-term controlled human experiments (blood collected 1-30 hours post-exposure) showed that exposure to air pollutants changed DNA methylation

levels of Alu and LINE-1 elements.[87,88] However, other studies did not find an association between air pollution and LINE-1 and/or Alu DNA methylation.[19,24,86]

Differences in platforms may partially explain our null associations with global DNA methylation. We assessed global DNA methylation using probes on the Illumina 450k array that map to Alu and LINE-1 repetitive elements, unlike most previous studies that used pyrosequencing. We also adjusted for a wider array of potential confounders not available in prior studies. Furthermore, our study examined DNA methylation in purified monocytes whereas previous studies used a mixture of blood cells. Monocytes constitute about 5-10% of peripheral blood leukocytes,[89] highlighting the possibility that the association between air pollution and global DNA hypomethylation may be specific to other leukocytes. Moreover, our study population differs from those of prior studies. Our study included men and women age 65 and over who reside in six U.S. regions. Prior studies included a variety of study populations including an elderly cohort of men from the Boston area (mean age 73),[19,20] healthy male workers from an electric steel furnace plant in Italy (mean age 44),[23] and a predominantly male group of workers from the Ma Ta Phut industrial estate in Thailand (mean age 31).[86]

We considered the possibility that the difference in air pollution exposure time windows may explain our non-significant findings for global DNA methylation. Previous studies found significant associations for short- and intermediate-term air pollution exposure, rather than long-term exposures that are the primary focus of the MESA Air project. A study of air pollution and percent 5-methyl-2'-deoxycytidine methylation found the strongest associations above 7 days, with most of the effects observed for the 5- to 30-day moving average exposures.[90] However, in our study, the associations between 2-week average air pollutant concentrations and global DNA methylation were not significant in sensitivity analyses.

In our study, we only identified CpGs significantly associated with $PM_{2.5}$ but not NO_X (at FDR of 0.05). Sources of ambient $PM_{2.5}$ include energy generation, combustion of fuel, agricultural and industrial processes, and road and wind-blown dust.[2] While motor vehicle

emissions contribute to ambient $PM_{2.5}$, these emissions are one of the main anthropogenic sources of ambient NO_X. Ambient air is a mixture of pollutants, and this mixture may be responsible for the observed associations between air pollution and health.[91] Single pollutants such as $PM_{2.5}$ or NO_X may serve as proxies for this mixture.[91] A study conducted using data from Quebec, Canada observed that air pollutant mixtures varied substantially within the city, and although none of the single pollutants were perfect proxies for urban ambient air pollution mixtures, NO_X seemed to be the best proxy.[91] However, we found methylation signals to be associated with $PM_{2.5}$ but not NO_X, suggesting that in our study $PM_{2.5}$ may be a better proxy of the mixture of air pollutants that affect DNA methylation levels.

There are limitations to our study. Our analysis was cross-sectional, and we were unable to make inferences about the effect of air pollution on changes in DNA methylation. In addition, our study was comprised of individuals in the United States, which has lower air pollution concentrations than many parts of the world. Higher concentrations of air pollutants may trigger alternate disease mechanisms not captured in our data. Additionally, there may be effect modifiers of the air pollution-DNA methylation association not investigated in this study. However, in our sensitivity analyses we did not find evidence of effect modification by sex or race/ethnicity. There may also be residual confounding by factors not included in the analysis. Our study may be further limited by selection bias; our analytic sample included slightly fewer Black participants and slightly more Hispanic participants compared to all participants from the four participating MESA field centers at exam 5. However, secondary analyses did not provide evidence of effect modification by race/ethnicity of the association between air pollution and DNA methylation. Moreover, our analytic sample was similar to the MESA participants from the four field centers in all other covariates and our analysis adjusted for race/ethnicity. Thus, we expect selection bias to have a minimal influence in our analysis. Finally, the functional assessment is limited to previously described *cis* associations.

Our study is distinct in several ways. Our focus on purified monocytes greatly enhances the interpretability of our results and the ability to tease out monocyte-specific differentially methylated signals that are important in atherosclerosis. Furthermore, the population-based sampling, younger age, and multi-ethnic nature of the MESA cohort improves generalizability of study results to a diverse adult population. The coupling of DNA methylation and gene expression data aids the detection of differentially methylated sites that may be functionally relevant. Finally, the state-of-art air pollution assessment in MESA Air provides individual-level exposures that capture fine-scale spatial variability in air pollution to greatly reduce measurement error.

CONCLUSIONS

Long-term ambient air pollution exposure was associated with site-specific DNA methylation, but not global DNA methylation, in purified blood monocytes obtained from a multi-ethnic adult population. We identified genes that may be functionally relevant to mediating air pollution health effects by focusing on expression-associated methylation sites. These genes may participate in inflammation and possibly monocyte endothelial adhesion. Future research in blood monocytes could address whether air pollution-associated diseases—such as atherosclerosis—may be mediated in part by DNA methylation epigenetic reprogramming.

TABLES AND FIGURES

Table 2.1. Demographics of analytic sample and all participants from the 5th examination from four Multi-Ethnic Study of Atherosclerosis (MESA) study sites.

	Analytic Sample (n=1,207)	Exam 5 Participants[a] (n=3,052)
	n (%) or mean ± SD	n (%) or mean ± SD
Age (y)	69.6 ± 9.4	69.8 ± 9.4
Race/ethnicity, %		
Black	256 (21.2)	943 (30.9)
Chinese-American	0 (0.0)	2 (0.1)
Hispanic	381 (31.6)	699 (22.9)
White	570 (47.2)	1,408 (46.1)
Sex, %		
Female	623 (51.6)	1,660 (54.4)
Male	584 (48.4)	1,392 (45.6)
Smoking status, %[b]		
Never	484 (40.3)	1,249 (41.5)
Former	604 (50.3)	1,493 (49.6)
Current	112 (9.3)	266 (8.8)
Secondhand smoke (hours per week)[b]	3.7 ± 20.5	3.2 ± 16.8
Body mass index (kg^2/m^2)[b]	29.7 ± 5.5	29.3 ± 5.7
Physical activity (MET-min/wk m-su)[b]	5,696.2 ± 7205.6	5,458.2 ± 6566.0
Education, %[b]		
Less than high school	176 (14.6)	402 (13.2)
High school	236 (19.6)	611 (20.1)
Some college but no degree	212 (17.6)	524 (17.2)
Bachelor's/Associate/Technical	373 (31.0)	962 (31.6)
Advanced degree	208 (17.3)	546 (17.9)
Income, %[b]		
< $24,999	297 (25.5)	721 (25.0)
$25,000 - $49,999	371 (31.9)	912 (31.6)
$50,000 - $99,999	333 (28.6)	815 (28.2)
$100,000 or more	163 (14.0)	437 (15.1)
Neighborhood socioeconomic status factor score[b]	-0.41 ± 1.1	-0.34 ± 1.1
Folate (mcg)[b]	317.1 ± 169.2	311.7 ± 214.7
Vitamin B12 (mcg)[b]	3.9 ± 3.8	3.9 ± 5.5
Vitamin B6 (mg)[b]	1.5 ± 0.8	1.5 ± 1.0
Methionine (g)[b]	1.4 ± 0.8	1.4 ± 1.1
Zinc (mg)[b]	8.9 ± 5.2	8.7 ± 6.3
Recent infection, %[b]		
No	922 (77.2)	2,322 (77.3)
Yes	273 (22.8)	682 (22.7)

Study site, %		
Winston-Salem, NC	49 (4.1)	813 (26.6)
New York, NY	398 (33.0)	810 (26.5)
Baltimore, MD	303 (25.1)	658 (21.6)
St. Paul, MN	457 (37.9)	771 (25.3)
$PM_{2.5}$ ($\mu g/m^3$)[b]	10.7 ± 1.7	10.7 ± 1.5
NO_X (ppb)[b]	28.7 ± 18.5	24.8 ± 18.5

[a] All participants at 5th exam from four Multi-Ethnic Study of Atherosclerosis study sites: Baltimore, MD; New York, NY; St. Paul, MN; and Winston-Salem, NC.
[b] Contains missing values.

Table 2.2. Selected characteristics of 1,207 Multi-Ethnic Study of Atherosclerosis (MESA) participants.

	Total	New York	Maryland	Minnesota	North Carolina
	n (%) or mean ± SD	n (%) or mean ± SD	n (%) or mean ± SD	n (%) or mean ± SD	n (%) or mean ± SD
Age (y)	69.6 ± 9.4	69.7 ± 9.7	70.7 ± 9.0	68.5 ± 9.5	71.1 ± 7.0
Race/ethnicity, %					
White	570 (47.2)	82 (20.6)	173 (57.5)	267 (58.4)	48 (94.1)
Black	256 (21.2)	125 (31.4)	128 (42.5)	0 (0)	3 (5.9)
Hispanic	381 (31.6)	191 (48)	0 (0)	190 (41.6)	0 (0)
Sex, %					
Female	623 (51.6)	231 (58)	156 (51.8)	208 (45.5)	28 (54.9)
Male	584 (48.4)	167 (42)	145 (48.2)	249 (54.5)	23 (45.1)
Smoking status, %[a]					
Never	484 (40.3)	181 (45.8)	120 (40.1)	165 (36.3)	18 (35.3)
Former	604 (50.3)	184 (46.6)	149 (49.8)	244 (53.6)	27 (52.9)
Current	112 (9.3)	30 (7.6)	30 (10)	46 (10.1)	6 (11.8)
Secondhand smoke (hours per week)[a]	3.7 ± 20.5	2.3 ± (11.1)	5.9 ± 35.3	3.3 ±12.3	5.5 ±14.2
Body mass index (kg^2/m^2)[a]	29.7 ± 5.5	29.3 ± 5.6	30.1 ± 5.7	29.8 ± 5.3	28.5 ± 5.8
Physical activity (MET-min/wk m-su)[a]	5,696.2 ± 7,205.6	5,364.6 ± 7,139.0	5,978.8 ± 9,405.6	5,942.0 ± 5,770.6	4,405.0 ± 3,314.3
Education, %[a]					
Less than high school	176 (14.6)	94 (23.6)	20 (6.7)	61 (13.3)	1 (2.0)
High school	236 (19.6)	76 (19.1)	59 (19.7)	95 (20.8)	6 (11.8)
Some college but no degree	212 (17.6)	68 (17.1)	59 (19.7)	75 (16.4)	10 (19.6)
Bachelor's/Associate /Technical	373 (31.0)	99 (24.9)	81 (27.1)	171 (37.4)	22 (43.1)
Advanced degree	208 (17.3)	61 (15.3)	80 (26.8)	55 (12.0)	12 (23.5)
Income, %[a]					
< $25,000	297 (25.5)	131 (33.5)	49 (17.3)	110 (24.9)	7 (14.3)
$25,000 - $49,999	371 (31.9)	125 (32.0)	77 (27.2)	150 (34.0)	19 (38.8)
$50,000 - $99,999	333 (28.6)	89 (22.8)	104 (36.7)	125 (28.3)	15 (30.6)
$100,000 or more	163 (14.0)	46 (11.8)	53 (18.7)	56 (12.7)	8 (16.3)
Neighborhood socioeconomic status factor score[a]	-0.41 ± 1.09	-1.06 ± 1.42	-0.27 ± 0.85	0.004 ± 0.62	-0.21 ± 0.91
Recent infection, %[a]					
No	922 (77.2)	293 (73.8)	225 (76.0)	365 (80.8)	39 (78.0)
Yes	273 (22.8)	104 (26.2)	71 (24.0)	87 (19.2)	11 (22.0)
Folate (mcg)[a]	317.1 ± 169.2	308.5 ± 162.7	301.9 ± 165.0	333.6 ± 175.6	324.1 ± 175.9
Vitamin B12 (mcg)[a]	3.9 ± 3.8	4.1 ± 5.1	3.6 ± 3.2	4.0 ± 2.8	3.9 ± 2.8

Vitamin B6 (mg)[a]	1.5 ± 0.8	1.4 ± 0.7	1.5 ± 0.9	1.6 ± 0.8	1.5 ± 0.7
Methionine (g)[a]	1.4 ± 0.8	1.3 ± 0.8	1.3 ± 0.9	1.5 ± 0.8	1.4 ± 0.8
Zinc (mg)[a]	8.9 ± 5.2	8.4 ± 4.9	8.3 ± 5.6	9.6 ± 5.2	8.8 ± 4.3

[a] Contains missing values.

Table 2.3. Expression-associated methylation sites (eMS) significantly associated with $PM_{2.5}$ (per 2.5 μg/m^3).

Gene[a]	Chr	CpG	β (95% CI)[b]	P-value	Adjusted P-value[c]
ANKHD1	5	cg20455854	0.139 (0.074, 0.203)	2.77E-05	0.034
LGALS2	22	cg07855639	0.081 (0.043, 0.120)	3.28E-05	0.034
ANKRD11	16	cg07598385	0.108 (0.056, 0.160)	4.97E-05	0.034
BAZ2B	2	cg17360854	0.081 (0.042, 0.120)	5.08E-05	0.034
PPIE	1	cg23599683	-0.057 (-0.085, -0.029)	7.17E-05	0.039

$PM_{2.5}$, fine particulate matter; chr, chromosome; CI, confidence interval.
[a] Gene corresponding to Illumina transcript associated with CpG methylation.
[b] Models adjusted for age, race/ethnicity, sex, study site, income, education, neighborhood socioeconomic status factor score, cigarette smoking, secondhand smoke, body mass index, physical activity, methyl nutrient intake (folate, vitamin B12, vitamin B6, methionine, zinc), residual cell contamination by non-monocytes, recent infection, methylation chip position (for DNA methylation analysis only), and gene expression chip (for gene expression analysis only). Methylation values were adjusted for methylation chip prior to regression analysis.
[c] Based on false discovery rate cutoff of 0.05 using the Benjamini-Hochberg method.

Table 2.4. Nearest gene and functional annotation of significant expression-associated methylation sites (eMS) related to long-term $PM_{2.5}$ exposure.

CpG	Chr	Position (hg19)	Transcript Gene[a]	Illumina Transcript[b]	Nearest TSS Gene[c]	Distance TSS[d]	CD14+ Chromatin State [e]	DNase HS CD14+[f]	TFBS Any Cell[g]
cg20455854	5	139040849	*ANKHD1*	ILMN_1765091	*CXXC5*	12330	Strong enhancer	Yes	Yes
cg07855639	22	37959465	*LGALS2*	ILMN_1687306	*CDC42EP1*	2993	Heterochromatin	Yes	Yes
cg07598385	16	88566483	*ANKRD11*	ILMN_2108709	*MIR5189*	31156	Heterochromatin	No	Yes
cg17360854	2	161224771	*BAZ2B*	ILMN_1720850	*MIR4785*	39620	Strong enhancer	Yes	Yes
cg23599683	1	40235372	*PPIE*	ILMN_1680341	*OXCT2*	1646	Heterochromatin	No	No

$PM_{2.5}$, fine particulate matter; chr, chromosome; TSS, transcription start site; DNAse HS, DNase hypersensitivity; TFBS, transcription factor binding site; ENCODE, The Encyclopedia of DNA Elements.

[a] All CpG sites have previously been associated with *cis*-gene expression. This annotation is based on the gene that the associated transcript maps to.

[b] Illumina transcript identification number for transcript associated with CpG.

[c] Annotation based on gene with the nearest TSS.

[d] Distance to nearest TSS.

[e] Prediction based on histone modifications in monocyte samples from the BLUEPRINT (H3K27ac, H3K4me1, H3K4me3) and ENCODE (H3K36me3) projects.

[f] DNase hypersensitivity reported in a CD14+ monocyte sample (ENCODE).

[g] TFBS reported in any cell type available from the UCSC Genome Browser.

Table 2.5. Associations between top five[a] expression-associated methylation sites (eMS) and NO_X (per 30 ppb).

CpG	Chr	Position	Transcript Gene[b]	Illumina Transcript[c]	Nearest TSS Gene[d]	Distance TSS[e]	β[f]	95% CI	P-value	Adjusted P-value[g]
cg16859127	12	113897132	*OAS1*	ILMN_1658247	*LHX5*	12743	-0.201	(-0.304, -0.099)	1.27E-04	0.346
cg16734451	8	32430022	*NRG1*	ILMN_1737252	*NRG1*	23322	0.098	(0.044, 0.152)	3.70E-04	0.502
cg22243996	6	26757644	*HIST1H2BD*	ILMN_1758623	*ZNF322*	97663	0.147	(0.058, 0.236)	1.27E-03	0.604
cg03014680	12	10122522	*CLEC12A*	ILMN_2292178	*CLEC12A*	1484	0.103	(0.039, 0.167)	1.55E-03	0.604
cg01092361	6	42185687	*MRPS10*	ILMN_1663664	*MRPS10*	53	0.115	(0.043, 0.186)	1.64E-03	0.604

NO_X, oxides of nitrogen; chr, chromosome; TSS, transcription start site; CI, confidence interval.

[a] These associations did not meet statistical significance at false discovery rate cutoff of 0.05, but have the smallest p-values.

[b] All CpG sites have previously been associated with cis-gene expression. This annotation is based on the gene that the associated transcript maps to.

[c] Illumina transcript identification number for transcript associated with CpG.

[d] Annotation based on gene with the nearest TSS.

[e] Distance to nearest TSS.

[f] Adjusted for age, sex, race/ethnicity, study site, residual cell contamination by non-monocytes (neutrophils, B cells, T cells, and natural killer cells), and expression chip.

[g] Based on false discovery rate cutoff of 0.05 using the Benjamini-Hochberg method.

Table 2.6. Association between $PM_{2.5}$ (per 2.5 μg/m^3) and expression of transcript corresponding to expression-associated methylation sites (eMS) associated with $PM_{2.5}$.

Gene[a]	Chr	Illumina Transcript	eMS	β (95% CI)[b]	P-value	Adjusted P-value[c]
ANKHD1	5	ILMN_1765091	cg20455854	-0.048 (-0.074, -0.022)	3.71E-04	0.002
LGALS2	22	ILMN_1687306	cg07855639	-0.147 (-0.240, -0.053)	0.002	0.005
ANKRD11	16	ILMN_2108709	cg07598385	-0.075 (-0.142, -0.008)	0.028	0.046
BAZ2B	2	ILMN_1720850	cg17360854	-0.016 (-0.060, 0.027)	0.463	0.578
PPIE	1	ILMN_1680341	cg23599683	0.004 (-0.020, 0.028)	0.728	0.728

$PM_{2.5}$, fine particulate matter; eMS, expression-associated methylation sites; chr, chromosome; CI, confidence interval.
[a] Gene corresponding to Illumina transcript associated with CpG methylation.
[b] Models adjusted for age, race/ethnicity, sex, study site, income, education, neighborhood socioeconomic status factor score, cigarette smoking, secondhand smoke, body mass index, physical activity, methyl nutrient intake (folate, vitamin B12, vitamin B6, methionine, zinc), residual cell contamination by non-monocytes, recent infection, methylation chip position (for DNA methylation analysis only), and gene expression chip (for gene expression analysis only). Methylation values were adjusted for methylation chip prior to regression analysis.
[c] Based on false discovery rate cutoff of 0.05 using the Benjamini-Hochberg method.

Table 2.7. Association between global DNA methylation (Alu and LINE-1) and one-year and 2-week average $PM_{2.5}$ (per 2.5 µg/m^3) and NO_X (per 30 ppb).

	Alu		LINE-1	
	β (95% CI)[a]	P-value	β (95% CI)[a]	P-value
One-year average				
$PM_{2.5}$	-0.003 (-0.006, 0.001)	0.155	-0.003 (-0.007, 0.001)	0.177
NO_X	0.001 (-0.006, 0.008)	0.719	-0.0004 (-0.009, 0.008)	0.927
2-week average				
$PM_{2.5}$	0.0004 (-0.002, 0.002)	0.694	0.001 (-0.001, 0.003)	0.415
NO_X	0.002 (-0.004, 0.008)	0.464	0.003 (-0.004, 0.010)	0.391

$PM_{2.5}$, fine particulate matter; NO_X, oxides of nitrogen; CI, confidence interval; LINE-1, long interspersed element 1.
[a] Models adjusted for age, race/ethnicity, sex, study site, income, education, neighborhood socioeconomic status factor score, cigarette smoking, secondhand smoke, body mass index, physical activity, methyl nutrient intake (folate, vitamin B12, vitamin B6, methionine, zinc), residual cell contamination by non-monocytes, recent infection, and methylation chip position. Methylation values were adjusted for methylation chip prior to regression analysis.

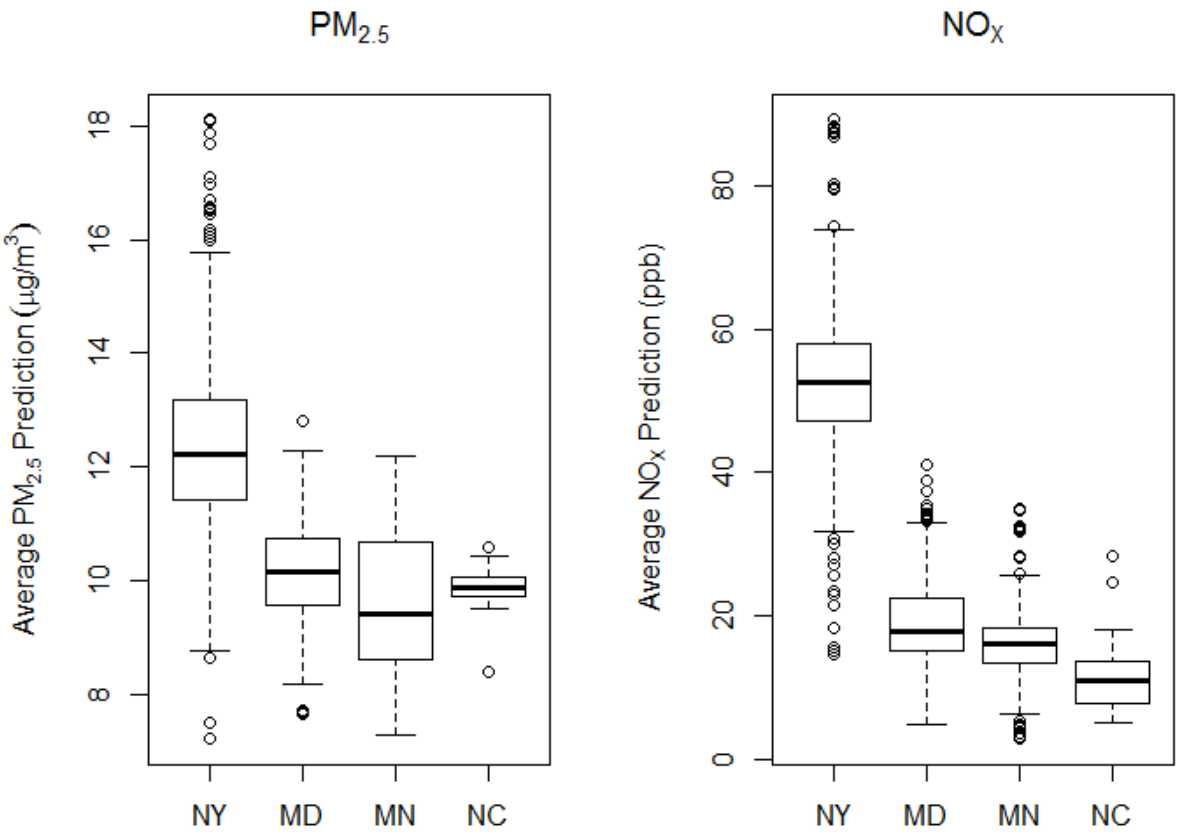

Figure 2.1. Ambient predictions of $PM_{2.5}$ and NO_X by site, averaged over 12 months prior to blood draw. Site abbreviations: NY, New York; MD, Maryland; MN, Minnesota; NC, North Carolina. The lower and upper ends of the box represent the 25th and 75th percentiles, respectively, and the center bar represents the median. The bottom and top whiskers represent 1.5 times the interquartile distance above and below the 25th and 75th percentiles, respectively, and the circles indicate outliers.

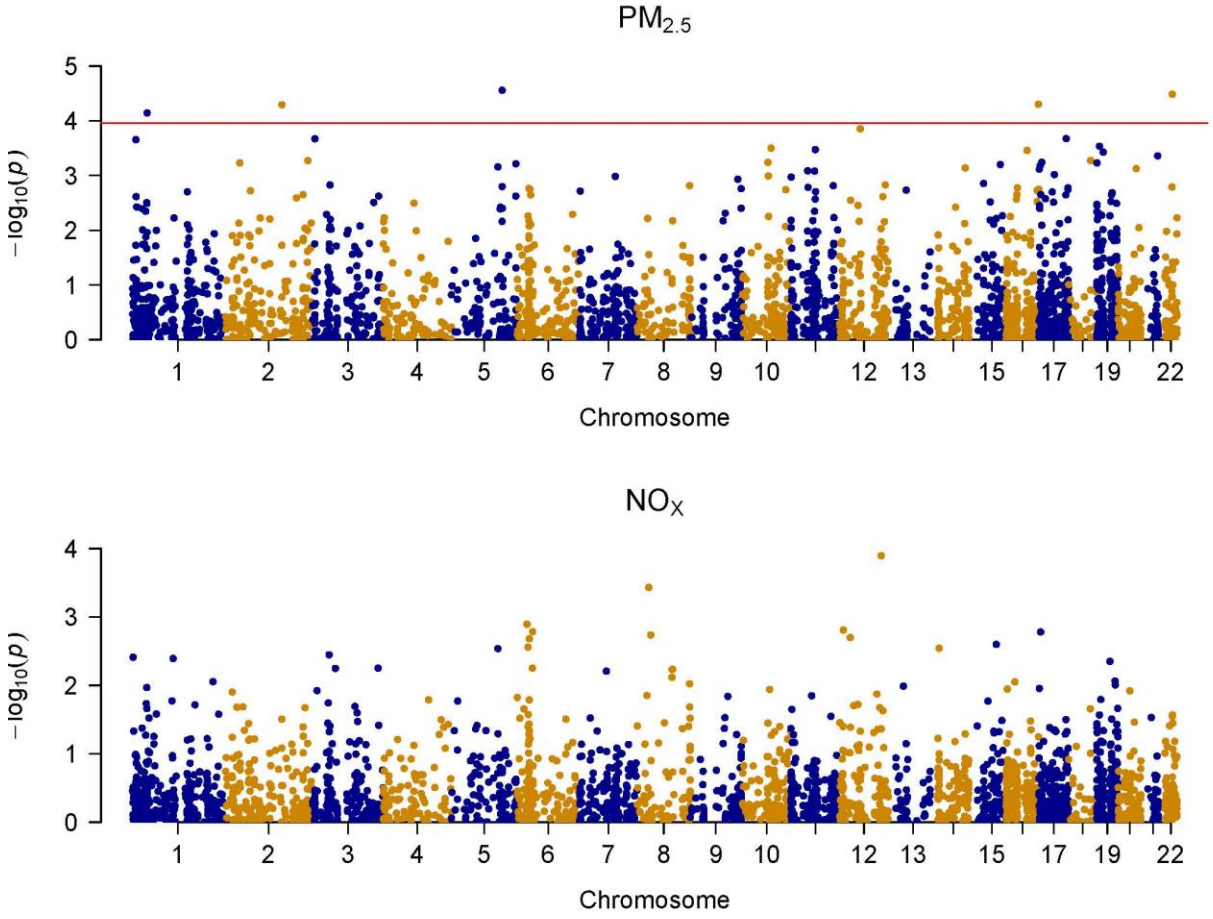

Figure 2.2. Association between $PM_{2.5}$ and NO_X and 2,713 expression-associated methylation sites. Red line demarcates statistical significance at a false discovery rate of 0.05. Plot for NO_X does not have a red line because no sites were statistically significant.

CHAPTER 3

Long-term Air Pollution Exposure and Epigenome-wide DNA Methylation in Monocytes: Results from the Multi-Ethnic Study of Atherosclerosis (MESA)

ABSTRACT

Background: Air pollution exposure may potentially affect atherosclerosis through alterations in DNA methylation. Monocytes are crucial in atherosclerosis pathogenesis, but an epigenome-wide study of DNA methylation in monocytes and long-term air pollution exposure has not been conducted.

Objectives: We investigated epigenome-wide DNA methylation in purified CD14+ monocytes of adults over age 55 in relation to long-term ambient air pollution exposure. We also assessed the association between air pollution-associated methylation signals and *cis*-expression of genes within 1 Mb of the methylation signal in question.

Methods: Using spatiotemporal models, one-year average ambient fine particulate matter ($PM_{2.5}$) and oxides of nitrogen (NO_X) concentrations were estimated at participants' homes in the Multi-Ethnic Study of Atherosclerosis. We assessed epigenome-wide DNA methylation and genome-wide gene expression in circulating monocytes of 1,207 participants using Illumina's Infinium HumanMethylation450 BeadChip and HumanHT-12 v4 Expression BeadChip, respectively. "Bump hunting" and site-specific approaches adjusting for demographics, smoking, body mass index, physical activity, socioeconomic status, methyl-nutrients, recent infection, and technical variables were used to identify differentially methylated genomic regions and CpG sites. We considered findings significant at a false discovery rate of 0.05. Associations between significant DNA methylation signals and gene expression were assessed using linear models.

Results: Four differentially methylated regions (DMRs) located on chromosomes 5, 6, 16, and 7 near *SDHAP3*, *ZFP57*, *PRM1*, and *HOXA5*, respectively, were positively associated with

$PM_{2.5}$. For the DMRs located on chromosomes 5, 6, 16, and 7, the average methylation difference for each 2.5 μg/m^3 higher exposure to $PM_{2.5}$ was 14.6% ($p=1.4x10^{-6}$), 10.9% ($p=2.1x10^{-5}$), 26.5% ($p=3.2x10^{-5}$), and 4.4% ($p=4.6x10^{-5}$), respectively. NO_X was also positively associated with the DMRs located on chromosomes 5 (average methylation difference=25.7% per 30 ppb NO_X; $p=7.7x10^{-6}$) and 6 (average methylation difference =25.0% per 30 ppb NO_X, $p=9.8x10^{-6}$). Three differentially methylated sites were identified for $PM_{2.5}$: cg05926640 (β=0.049; 95% CI: 0.032, 0.067; $p=5.6x10^{-8}$; near *LOC101927851*), cg04310517 (β=0.052; 95% CI: 0.033, 0.072; $p=2.2x10^{-7}$; near *KREMEN2*), and cg09509909 (β=0.079; 95% CI: 0.049, 0.110; $p=4.6x10^{-7}$; near *FGFBP3*). Cg11756214 within *ZNF347* was associated with NO_X (β=0.078; 95% CI: 0.050, 0.106; $p=5.6x10^{-8}$). Methylation of regions or sites near *SDHAP3*, *PRM1*, *HOXA5*, and *LOC101927851* were also associated with *cis*-gene expression.

Conclusions: We identified genomic regions and sites with methylation associated with long-term air pollution exposure. Future studies are needed to investigate whether DNA methylation effects in monocytes mediate air pollution associations with atherosclerosis.

INTRODUCTION

Air pollution exposure is a risk factor for cardiovascular disease (CVD), and several potential pathways have been proposed to explain these associations, including systemic inflammatory responses, oxidative stress,[3] and progression of atherosclerosis.[92] However, mechanisms for pathophysiological effects of air pollution inhalation are still unclear.

Growing literature suggests that environmental pollutants may cause disease through pathways involving epigenetic regulation of gene expression, such as that through DNA methylation.[39] Several studies have shown associations between exposure to air pollutants and global and site-specific DNA methylation in mixed blood leukocytes.[19,48,93–95] In addition, studies have associated DNA methylation patterns with cardiovascular outcomes.[19,45,96] However,

these studies measured DNA methylation in heterogeneous cell types, each with their own epigenetic signature. Shifts in cell type composition may confound results of these studies.

In Chapter 2, we assessed the association between long-term ambient air pollution exposure and global DNA methylation and candidate-site DNA methylation in purified monocytes. In this chapter, we expand on our previous work to conduct an epigenome-wide study of DNA methylation and long-term exposure to ambient fine particulate matter ($PM_{2.5}$) and oxides of nitrogen (NO_X) in monocytes. Monocytes were selected because of their crucial role in atherosclerosis pathogenesis.[14] We utilized data from the Multi-Ethnic Study of Atherosclerosis (MESA), MESA Air, and the MESA Epigenomics and Transcriptomics Study. We also investigated potential biological relevance of significant air pollution-associated methylation signals by analyzing their associations with the expression of nearby genes.

METHODS

Study Population

MESA is a prospective longitudinal study of a diverse population-based cohort consisting of 28% Black, 12% Chinese-American, 22% Hispanic, and 38% White participants from six field centers (Baltimore, MD; Chicago, Illinois; Los Angeles, CA; New York, NY; St. Paul, MN, and Winston-Salem, NC).[31] Data were also obtained from MESA Air and the MESA Epigenomics and Transcriptomics Study. Please see Chapter 1, pages 5-7 for additional details about these studies.

Our analytic sample included 1,207 participants with DNA methylation, gene expression, and air pollution data. Participants included in this study are generally representative of all participants at the four MESA field centers from the 5th examination except for having fewer Black participants, more Hispanic participants, and fewer participants from Winston-Salem, NC (Table 2.1, Chapter 2). In addition, there are no Chinese-American participants in our study

because methylation and gene expression data were not obtained from participants in Los Angeles or Chicago, the main sources of Chinese-American participants in MESA.

Air Pollution Assessment

We used ambient estimates of $PM_{2.5}$ and NO_X concentrations at participants' homes that were averaged over the one year prior to blood draw. These estimates were predicted using spatiotemporal models optimized via maximum likelihood.[34,36] Additional information can be found in Chapter 1, page 7.

DNA Methylation and Gene Expression Quantification

Monocytes were purified from blood drawn at the 5th exam. DNA and RNA were extracted for use in an epigenome-wide DNA methylation array (Infinium HumanMethylation450 BeadChip) and a genome-wide gene expression array (Illumina HumanHT-12 v4 Expression BeadChip), respectively.[33] A detailed description of the measurement, pre-processing, and normalization of microarray data can be found in Chapter 2, pages 15-17.

For quality control, we excluded allosome probes, probes with non-detectable methylation in over 90% of samples (using a detection p-value cutoff of 0.05), and probes located within 10 bp of single nucleotide polymorphisms. Of the 480,000 CpG sites on the Infinium 450k, 437,600 passed our quality control criteria.

Other Measures

A description of covariates used in this analysis can be found in Chapter 2, pages 17-18.

Statistical Analysis

Identification of Differentially Methylated Regions

Epigenome-wide studies using data from the Illumina HumanMethylation450 BeadChip typically conduct site-by-site analyses, following the standard genome-wide association study (GWAS) approach. However, unlike single-nucleotide polymorphism data, DNA methylation measurements from microarrays are continuous when aggregated over a large number of cells (e.g. proportion of cells methylated), are more vulnerable to measurement error, are more densely spaced across the genome, and have more complex correlation structures.[43] The standard GWAS site-by-site analytic approach does not account for these complications. Since DNA methylation of nearby CpG sites are strongly correlated, power can be increased by borrowing information from adjacent CpG sites.[43] Thus a "bump hunting" method to identify differentially methylated regions (DMRs) may improve the sensitivity of epigenome-wide analyses.

We adapted the bump hunting approach previously described[43] to the Infinium 450k array to search for associations at the region level rather than the individual site level. The *bumphunter* package in R was used to identify DMRs associated with long-term air pollution exposure.[43] Consecutive probes on the 450k array were grouped into clusters where the maximum distance between two neighboring probes was 300 bp and each cluster contained at least 5 probes. Of the 437,600 sites that passed our quality control criteria, 152,636 sites (35% of total) were successfully grouped into 18,293 clusters. For each probe within a cluster, we regressed the M-value on air pollution ($PM_{2.5}$ or NO_X) to estimate a coefficient for each CpG site, adjusting for age, sex, race/ethnicity (Black, Hispanic, White), household income, education, neighborhood socioeconomic status, smoking (smoking status and pack years), secondhand smoke, body mass index, recent infection, methyl nutrient intake (continuous folate, vitamin B12, vitamin B6, methionine, zinc), physical activity, study site, and chip position. We also adjusted for residual sample contamination by non-monocytes by adjusting for enrichment scores for neutrophils, B cells, T cells, and natural killer cells. The *ComBat* function in the package *SVA*[58] was used to obtain methylation chip-adjusted M-values for use in regression analyses. These

estimated coefficients were smoothed using loess, and candidate DMRs were identified as regions above or below the 97.5% percentile of the empirical distribution of the smoothed estimate. Statistical uncertainty for each candidate DMR was assessed by bootstrapping. We conducted 1,000 bootstraps, and each bootstrap produced null candidate regions. The empirical p-value for each observed candidate DMR was the percent of null regions that were as extreme (longer and higher average value) as the observed region. A false discovery rate (FDR) cutoff of 0.05 was used to account for multiple testing.[44]

Site-specific Analysis

Since 65% (284,964) of the 437,600 probes that passed quality control were not grouped into clusters for DMR analysis (based on criteria described above), we analyzed them as individual CpG sites to maximize use of the Infinium 450k array. Linear models and robust empirical Bayes moderated t-statistics were used to assess the associations between $PM_{2.5}$ or NO_X and methylation chip-adjusted M-values at individual CpGs using the *limma* package from Bioconductor.[53] A separate linear model was fit for each CpG site, adjusting for the same set of covariates as above. A FDR cutoff of 0.05 was used to account for multiple comparisons.[44]

Gene Expression

For differentially methylated regions and sites, we used linear models to correlate their methylation with expression of genes within a 1 Mb genomic region centered on the CpG region or site in question, adjusting for age, sex, race/ethnicity, study site, residual cell contamination by non-monocytes (neutrophils, B cells, T cells, and natural killer cells), methylation chip position, and expression chip. For DMRs, the average M-value of all CpG sites within the DMR was used. To account for multiple comparisons, we used a FDR threshold of 0.05.[44]

Secondary Analysis

Secondary analyses using an interaction term for air pollution and sex were conducted to assess evidence that sex modifies the effect of air pollution on methylation. We used the same approach to examine evidence of effect modification by race/ethnicity.

Functional Annotation Analysis

We evaluated the sites of significant air pollution-associated signals for overlap with known and predicted potentially functional genomic regions. Details of our approach can be found in Chapter 2, page 19-20.

RESULTS

Demographics of study participants (n=1,207) can be found in Table 2.2 (Chapter 2). The mean age of our analytic sample was 69.6. The sample was diverse, consisting of 31.6% Hispanic, and 21.2% Black, and 47.2% White participants. The participants were well-educated – over 65% of participants received education beyond high school. Over 90% of the participants were non-smokers. Plots of site-specific one-year average ambient $PM_{2.5}$ and NO_X predictions are shown in Figure 2.1 (Chapter 2). Mean $PM_{2.5}$ and NO_X predictions were 10.7 $\mu g/m^3$ (interquartile range [IQR]=2.2 $\mu g/m^3$) and 28.7 ppb (IQR=31.9 ppb), respectively. Participant characteristics by study site can be found in Table 2.2 (Chapter 2).

We identified four DMRs with methylation significantly (FDR<0.05) associated with $PM_{2.5}$ using the bump hunter method (Table 3.1, Figures 3.1-3.4). The DMR with the smallest p-value ($1.4x10^{-6}$) was located on chromosome 5 and overlaps the 5' untranslated region (UTR) of *SDHAP3* (succinate dehydrogenase complex, subunit A, flavoprotein pseudogene 3) and a predicted weak promoter (based on histone modifications in monocyte samples from the BLUEPRINT (H3K27ac, H3K4me1, H3K4me3) and ENCODE (H3K36me3) projects) (Table 3.1). On average, a 2.5 $\mu g/m^3$ higher exposure to $PM_{2.5}$ was associated with a 14.6% higher methylation at this DMR (Table 3.1, Figure 3.1). The second DMR was located 3,448 bases

upstream of *ZFP57* (ZFP57 zinc finger protein) on chromosome 6, and was on average 10.9% more methylated in those with a 2.5 μg/m^3 higher exposure to $PM_{2.5}$ ($p=2.1x10^{-5}$) (Table 3.1, Figure 3.2). This DMR also overlaps a predicted weak promoter. The last two DMRs were found in the exon of *PRM1* (protamine 1) in chromosome 16 and the 5' UTR region of *HOXA5* (homeobox A5) in chromosome 7, and these two DMRs were associated with a 26.5% ($p=3.2x10^{-5}$) and 4.4% ($p=4.6x10^{-5}$) higher average methylation for a 2.5 μg/m^3 higher exposure to $PM_{2.5}$, respectively (Table 3.1, Figures 3.3 and 3.4). The DMR on chromosome 16 is predicted to overlap heterochromatin while the one on chromosome 7 is predicted to overlap both a weak promoter and heterochromatin. The DMR on chromosome 16 only spans one bp, but it belongs to a cluster of 5 CpGs. The identification of this DMR was based on borrowed information from other CpGs within the cluster. However, the DMR seems to be due only to the methylation status of one CpG, which may make this DMR less informative (Figure 3.3).

For NO_X, we identified two significant DMRs (FDR<0.05) that were the same as the top two significant (smallest p-values) DMRs for $PM_{2.5}$, which were located on chromosomes 5 and 6 (Table 3.1). A 30 ppb higher exposure to NO_X was associated with a 25.7% higher average methylation in the DMR located on chromosome 5 ($p=7.7x10^{-6}$) and a 25.0% higher average methylation in the DMR located on chromosome 6 ($p=9.8x10^{-6}$).

Of these six air pollution-associated DMRs, three were also associated with mRNA expression of nearby genes (Table 3.2). The DMR located on chromosome 7 was negatively associated with *HOXA5* (β=-0.288; 95% CI: -0.332, -0.243; $p=1.5x10^{-34}$) and positively associated with *HOXA9* (β=0.168; 95% CI: 0.118, 0.218; $p=8.1x10^{-11}$) and *HOXA10* (β=0.079; 95% CI: 0.047, 0.112; $p=2.1x10^{-6}$) mRNA expression in monocytes. The DMR located in chromosome 16 was positively associated with *DEXI* (dexi homolog (mouse)) mRNA expression (β=0.012; 95% CI: 0.004, 0.019; $p=2.8x10^{-3}$). The DMR located in chromosome 5 was positively associated with mRNA expression of *MRPL36* (mitochondrial ribosomal protein L36) (β=0.014; 95% CI: 0.005, 0.024; $p=3.7x10^{-3}$).

Site-specific analyses identified three CpGs significantly associated with $PM_{2.5}$ and one CpG associated with NO_X at FDR<0.05 (Table 3.3). The $PM_{2.5}$-associated CpG with the smallest p-value was cg05926640 (β=0.049; 95% CI: 0.032, 0.067; $p=5.6x10^{-8}$) and is located 1,903 bases downstream of the *LOC101927851* (uncharacterized LOC101927851) (Table 3.3). Cg05926640 is located within a DNase hypersensitivity site (monocyte data from ENCODE), a transcription factor binding site (any cell type from UCSC Genome Browser), and a predicted strong enhancer region (Table 4). Methylation of cg04310517 (β=0.052; 95% CI: 0.033, 0.072; $p=2.2x10^{-7}$) and cg09509909 (β=0.079; 95% CI: 0.049, 0.110; $p=4.6x10^{-7}$) were also significantly positively associated with $PM_{2.5}$ exposure (Table 4). NO_X exposure was positively associated with methylation of cg11756214 (β=0.078; 95% CI: 0.050, 0.106; $p=5.6x10^{-8}$) (Table 3.3). Additional details such as nearest gene annotation and chromatin state can be found in Table 3.3.

Of these air pollution-associated CpGs, only cg05926640 methylation was associated with *cis*-gene expression of genes within 1 Mb of the CpG site (Table 3.4). Cg05926640 methylation was negatively associated with mRNA expression of *ARID4B* (AT-rich interaction domain 4B) at three transcripts and the coefficient for the transcript with the smallest p-value was -0.323 (95% CI: -0.453, -0.192; $p=1.3x10^{-6}$). It was also associated with *IRF2BP2* (interferon regulatory factor 2 binding protein 2; β=-0.213; 95% CI: -0.318, -0.108; $p=7.4x10^{-5}$) and *TOMM20* (β=0.118; 95% CI: 0.044, 0.191; $p=1.7x10^{-3}$) mRNA expression in monocytes.

In secondary analyses, we did not find evidence of sex or race/ethnicity differences in the association of long-term air pollution exposure and DNA methylation in monocytes (p>0.06).

DISCUSSION

We present results of the first epigenome-wide analysis of DNA methylation and long-term $PM_{2.5}$ and NO_X exposure in monocytes of a multi-ethnic adult population. We identified four DMRs associated with $PM_{2.5}$, two of which were also associated with NO_X. Of these, three

were associated with expression of nearby genes. Site-specific analyses identified three CpGs with methylation associated with $PM_{2.5}$ and one CpG associated with NO_X. DNA methylation of one of the $PM_{2.5}$-associated CpGs was also associated with mRNA expression of nearby genes. Air pollution-associated methylation signals that are also associated with *cis*-gene expression may potentially be more biologically relevant.

Three DMRs, located on chromosomes 7, 16, and 5 were associated with *cis*-expression of nearby genes. Methylation of the DMR located on chromosome 7 was associated with both $PM_{2.5}$ exposure and expression of nearby homeobox cluster A (*HOXA*) genes including *HOXA5*, *HOXA9*, and *HOXA10*. This DMR is predicted to overlap a weak promoter region, providing supportive evidence that this DMR may affect gene expression. *HOXA* genes encode transcription factors crucial for patterning processes during vertebrate development[97] and hematopoietic differentiation.[98] Forced expression of *HOXA10* in CD34+ progenitor cells increases the number of monocytes and reduces the number of B cells and natural killer cells produced.[99] In the carotid artery, disturbed flow, which regulates endothelial inflammation and monocyte-endothelial cell adhesion, induces hypermethylation of *HOXA5* and downregulates its expression in endothelial cells.[100] In addition, lower DNA methylation and higher expression of *HOXA9* was found in atherosclerotic aortas compared to normal aortas.[27] *HOX* genes are differentially methylated in monocytes following lipopolysaccharide activation.[101] Taken together, these suggest that environmental exposures may induce differential methylation of *HOX* genes, which may modulate hematopoietic differentiation and inflammation.

Both $PM_{2.5}$ and NO_X were positively associated with methylation of the DMR located on chromosome 5, which was positively associated with mRNA expression in *MRPL36. MRPL36*, a nuclear gene, encodes a member of the mitrochondrial ribosomal proteins which are involved in mitochondrial protein synthesis.[102] The $PM_{2.5}$-associated DMR located on chromosome 16 was associated with mRNA expression of *DEXI*, which encodes a transcript that is upregulated

in emphysema tissue compared to normal tissue and may play a role in regulating inflammation.[103]

$PM_{2.5}$ exposure is positively associated with methylation at cg05926640, which is located near an uncharacterized locus *LOC101927851* with unknown function. However, this CpG may potentially be involved in modulating gene expression because it is located within several potential functional features including a DNase hypersensitivity site (in monocytes), a transcription-factor binding site (any cell), and a predicted strong enhancer. Cg05926640 methylation was negatively associated with mRNA expression of three nearby *ARID4B* transcripts. *ARID4B* encodes a chromatin remodeling protein that may regulate genomic imprinting,[104] is linked to various cancers,[105,106] and is involved in spermatogenesis.[107] Interestingly, ARID4B has also been suggested to regulate hematopoiesis by controlling expression of *HOX* genes.[106]

DNA methylation at cg05926640 was also positively associated with *TOMM20* mRNA expression. TOMM20 is a subunit of the translocase of the outer mitochondrial membrane, which is the main import gate for most nuclear-encoded mitochondrial proteins.[108] Mitochondrial dysfunction can disrupt macrophage cholesterol homeostasis and contribute to inflammation and atherosclerosis.[109] Methylation of cg05926640 was also negatively associated with mRNA expression of nearby *IRF2BP2*. IRF2BP2 is a transcriptional corepressor that binds interferon regulatory factor 2, a negative regulator of many interferon-responsive genes.[110] In mice, IRF2BP2-deficient macrophages are inflammatory, have impaired cholesterol efflux, and worsened atherosclerosis.[111] Humans who are homozygous for a *IRF2BP2* deletion polymorphism have lower IRF2BP2 protein levels in peripheral blood mononuclear cells and increased risk of coronary artery disease.[111] Taken together, this suggests that $PM_{2.5}$ exposure may affect cholesterol homeostasis to promote inflammation and atherosclerosis.

Prior analysis in this cohort did not find long-term ambient air pollution to be associated with global DNA methylation as measured by methylation in Alu and LINE-1 repetitive elements

(Chapter 2). However, $PM_{2.5}$ was associated with methylation of several CpGs that were previously associated with expression of nearby genes. The mRNA expression of two of these genes, *ANKHD1* (ankyrin repeat and KH domain containing 1) and *LGALS2* (lectin, galactoside-binding, soluble, 2), was also associated with $PM_{2.5}$ exposure. While none of the previously identified air pollution-associated CpGs reached statistical significance (FDR<0.05) in this analysis, some of their linked genes may be involved in pathways that may also be related to genes identified in the current analysis. For instance, *LGALS2* and *ANKHD1* may play roles in regulating inflammation which is similar to the *HOXA* genes and *DEXI* that were identified in this analysis.

Prior research in this cohort also demonstrated that methylation of cg05575921 was associated with smoking, carotid plaque score, and with mRNA expression of *AHRR.*[29] However, we did not find long-term exposure to ambient $PM_{2.5}$ (β=-0.054; 95% CI: -0.116, 0.008) and NO_X (β= -0.097; 95% CI: -0.218, 0.025) to be associated with methylation at cg05575921, suggesting that this methylation signal may be specific to cigarette smoking and not ambient sources of air pollution.

A recent study identified 12 CpGs on the Illumina 450k array that were associated with short- and intermediate-term $PM_{2.5}$ exposure in whole blood at genome-wide significance.[22] Among these, cg04423572, cg11046593, cg13169286, cg13527922, cg19215199, cg20680669, cg23276912, and cg26003785 were among the CpGs analyzed in this study. However, none of these had nominal p-values less than 0.05 (data not shown). We may not have observed these associations due to several fundamental differences between our studies which include the use of long-term (1-year average) versus shorter term air pollution exposures and assessment of DNA methylation in monocytes versus mixed leukocytes. Moreover, differences in study population may also explain these differences. The prior study included individuals from the Normative Aging Study and the KORA F3 and F4 cohorts. The Normative Aging Study included elderly men (mean age 72.4 ± 6.9, n=657) from the greater Boston area while the KORA F3 and

F4 participants included both male and female participants from Germany (KORA F3 mean age 53.1 ± 9.6, n=500; KORA F4 mean age 60.9 ± 8.9, n=1,799). The current study includes female and male participants age 55 and older from four U.S. regions (Baltimore, MD; New York, NY; St. Paul, MN; and Winston-Salem, NC).

The cross-sectional nature of this study limits our ability to infer the causal effect of air pollution on changes in DNA methylation. We note, however, that our exposure estimates are for the one-year period preceding exposure. Since it is unlikely that DNA methylation causes air pollution, there is little potential for reverse causation. Another limitation of our study is the potential for selection bias. Our analytic sample included slightly fewer Black participants and slightly more Hispanic participants than overall participants from the four participating MESA field centers from the 5th examination. However, we did not find evidence that race/ethnicity modifies the effect of air pollution on DNA methylation and all other covariates were similar between our analytic sample the original sample at four MESA centers. In addition, we adjusted for race/ethnicity in our analysis. Thus, we expect the impact of selection bias to be minimal in our analysis. Although we have adjusted for a wide range of potential confounders, there may be residual confounding by unmeasured factors or factors not included in the analysis and misclassification of confounder measurements. The bump hunting and site-specific analyses interrogated mutually exclusive sets of CpG sites. Sites on the Infinium 450k array that could be not be grouped into clusters were not analyzed using the bump hunting approach (see page 43 for details about clustering); instead, non-clustered CpG sites were analyzed using a site-specific approach. Since the bump hunting and site-specific analyses used separate sets of probes on the Infinium 450k array, we could not see if results based on one analysis were replicated by the other. We chose to analyze probes only once, either in the bump hunting or site-specific analysis (based on whether sites could be grouped into clusters), to avoid unnecessary multiplicity. Finally, our analysis is exploratory, and future studies are needed to

replicate our results and further elucidate the biological relevance of air pollution-associated methylation signals.

Our study has several strengths. We have methylomic and transcriptomic data available from purified monocytes, which greatly improves the interpretability of our results especially with respect to signals that may be relevant to atherosclerosis. Moreover, we adapted a novel bump-hunting approach to identify differentially methylated regions to improve the sensitivity of our study. We also explored the functional relevance of significant DNA methylation signals using gene expression data. In addition, we utilized MESA Air's sophisticated air pollution assessment, which provides participant-specific exposures that capture fine-scale spatial variability in air pollution. Finally, our results have greater generalizability than previous studies due to the population-based and multi-ethnic nature of the MESA sample.

CONCLUSIONS

Using purified blood monocytes obtained from a multi-ethnic adult population, we identified differentially methylated regions and sites associated with long-term air pollution exposures—some of which are also associated with expression of nearby genes. These genes are involved in inflammation and cholesterol homeostasis, which are relevant in atherosclerosis pathogenesis. Additional research in circulating monocytes is needed to elucidate the functional relevance of these air pollution-associated methylation signals and whether DNA methylation mediates air pollution-associated diseases such as atherosclerosis.

TABLES AND FIGURES

Table 3.1. Differentially methylated regions for $PM_{2.5}$ and NO_X with FDR<0.05.

Chr	Start	End	Nearest Gene	Location Relative to Gene (hg19)	CD14+ Chromatin State[a]	Average Δ M[b]	P-value
$PM_{2.5}$							
5	1594282	1594863	*SDHAP3*	Overlaps 5' UTR	Weak promoter	14.6%[c]	1.4E-06
6	29648379	29649024	*ZFP57*	Upstream	Weak promoter	10.9%[c]	2.1E-05
16	11374865	11374865	*PRM1*	Inside exon	Heterochromatin/low	26.5%[c]	3.2E-05
7	27183274	27184109	*HOXA5*	Overlaps 5' UTR	Weak promoter and Heterochromatin/low	4.4%[c]	4.6E-05
NO_X							
5	1594282	1595048	*SDHAP3*	Overlaps 5' UTR	Weak promoter	25.7%[d]	7.7E-06
6	29648379	29649024	*ZFP57*	Upstream	Weak promoter	25.0%[d]	9.8E-06

$PM_{2.5}$, fine particulate matter; NO_X, oxides of nitrogen; FDR, false discovery rate; chr, chromosome; M, methylation; UTR, untranslated region.
[a] Prediction based on histone modifications in monocyte samples from the BLUEPRINT (H3K27ac, H3K4me1, H3K4me3) and ENCODE (H3K36me3) projects.
[b] Models adjusted for age, race/ethnicity, sex, study site, income, education, neighborhood socioeconomic status factor score, cigarette smoking, secondhand smoke, body mass index, physical activity, methyl nutrient intake (folate, vitamin B12, vitamin B6, methionine, zinc), residual cell contamination by non-monocytes, recent infection, and methylation chip position. Methylation values were adjusted for methylation chip prior to regression analysis.
[c] Average methylation difference for every 2.5 $\mu g/m^3$ higher exposure to $PM_{2.5}$.
[d] Average methylation difference for every 30 ppb higher exposure to NO_X.

Table 3.2. Significant (FDR<0.05) associations between methylation of $PM_{2.5}$-associated DMRs and mRNA expression of genes within 1 Mb of the DMR in question.

Chr	Start, End	Illumina Transcript	Gene	β (95% CI)[a]	P-value
7	27183274, 27184109	ILMN_1753613	*HOXA5*	-0.288 (-0.332, -0.243)	1.5E-34
7	27183274, 27184109	ILMN_1739582	*HOXA9*	0.168 (0.118, 0.218)	8.1E-11
7	27183274, 27184109	ILMN_1689336	*HOXA10*	0.079 (0.047, 0.112)	2.1E-06
16	11374865, 11374865	ILMN_1738866	*DEXI*	0.012 (0.004, 0.019)	2.8E-03
5	1594282, 1594863	ILMN_1800197	*MRPL36*	0.014 (0.005, 0.024)	3.7E-03

DMR, differentially methylated region; FDR, false discovery rate; CI, confidence interval.
[a] Models adjusted for age, race/ethnicity, sex, study site, income, education, neighborhood socioeconomic status factor score, cigarette smoking, secondhand smoke, body mass index, physical activity, methyl nutrient intake (folate, vitamin B12, vitamin B6, methionine, zinc), residual cell contamination by non-monocytes, recent infection, methylation chip position, and gene expression chip. Methylation values were adjusted for methylation chip prior to regression analysis.

Table 3.3. Differentially methylated sites for $PM_{2.5}$ and NO_X with FDR<0.05.

CpG	Chr	Nearest Gene	Distance	Location Relative to Gene (Hg19)	CD14+ Chromatin State[a]	DNase HS CD14+[b]	TFBS Any Cell[c]	β (95% CI)[d]	P-value
$PM_{2.5}$									
cg05926640	1	*LOC101927851*	1,903	Downstream	Strong enhancer	Yes	Yes	0.049 (0.032, 0.067)	5.6E-08
cg04310517	16	*KREMEN2*	3,954	Upstream	Heterochromatin /low	No	No	0.052 (0.033, 0.072)	2.2E-07
cg09509909	10	*FGFBP3*	3,489	Upstream	Heterochromatin /low	Yes	Yes	0.079 (0.049, 0.110)	4.6E-07
NO_X									
cg11756214	19	*ZNF347*	0	5' UTR	Weak promoter	Yes	Yes	0.078 (0.050, 0.106)	5.6E-08

$PM_{2.5}$, fine particulate matter; NO_X, oxides of nitrogen; FDR, false discovery rate; chr, chromosome; DNAse HS, DNase hypersensitivity; TFBS, transcription factor binding site; ENCODE, The Encyclopedia of DNA Elements; CI, confidence interval.

[a] Prediction based on histone modifications in monocyte samples from the BLUEPRINT (H3K27ac, H3K4me1, H3K4me3) and ENCODE (H3K36me3) projects

[b] DNase hypersensitivity reported in a CD14+ monocyte sample (ENCODE).

[c] TFBS reported in any cell type available from the UCSC Genome Browser.

[d] Models adjusted for age, race/ethnicity, sex, study site, income, education, neighborhood socioeconomic status factor score, cigarette smoking, secondhand smoke, body mass index, physical activity, methyl nutrient intake (folate, vitamin B12, vitamin B6, methionine, zinc), residual cell contamination by non-monocytes, recent infection, and methylation chip position. Methylation values were adjusted for methylation chip prior to regression analysis.

Table 3.4. Significant (FDR<0.05) associations between cg05926640 methylation and mRNA expression of nearby genes.

CpG	Illumina Transcript	Gene	β (95% CI)[a]	P-value
cg05926640	ILMN_2269564	*ARID4B*	-0.323 (-0.453, -0.192)	1.3E-06
cg05926640	ILMN_2362982	*ARID4B*	-0.227 (-0.333, -0.120)	3.2E-05
cg05926640	ILMN_1761334	*ARID4B*	-0.274 (-0.406, -0.143)	4.8E-05
cg05926640	ILMN_1671005	*IRF2BP2*	-0.213 (-0.318, -0.108)	7.4E-05
cg05926640	ILMN_1679796	*TOMM20*	0.118 (0.044, 0.191)	1.7E-03

FDR, false discovery rate; CI, confidence interval.

[a] Models adjusted for age, race/ethnicity, sex, study site, income, education, neighborhood socioeconomic status factor score, cigarette smoking, secondhand smoke, body mass index, physical activity, methyl nutrient intake (folate, vitamin B12, vitamin B6, methionine, zinc), residual cell contamination by non-monocytes, recent infection, methylation chip position, and gene expression chip. Methylation values were adjusted for methylation chip prior to regression analysis.

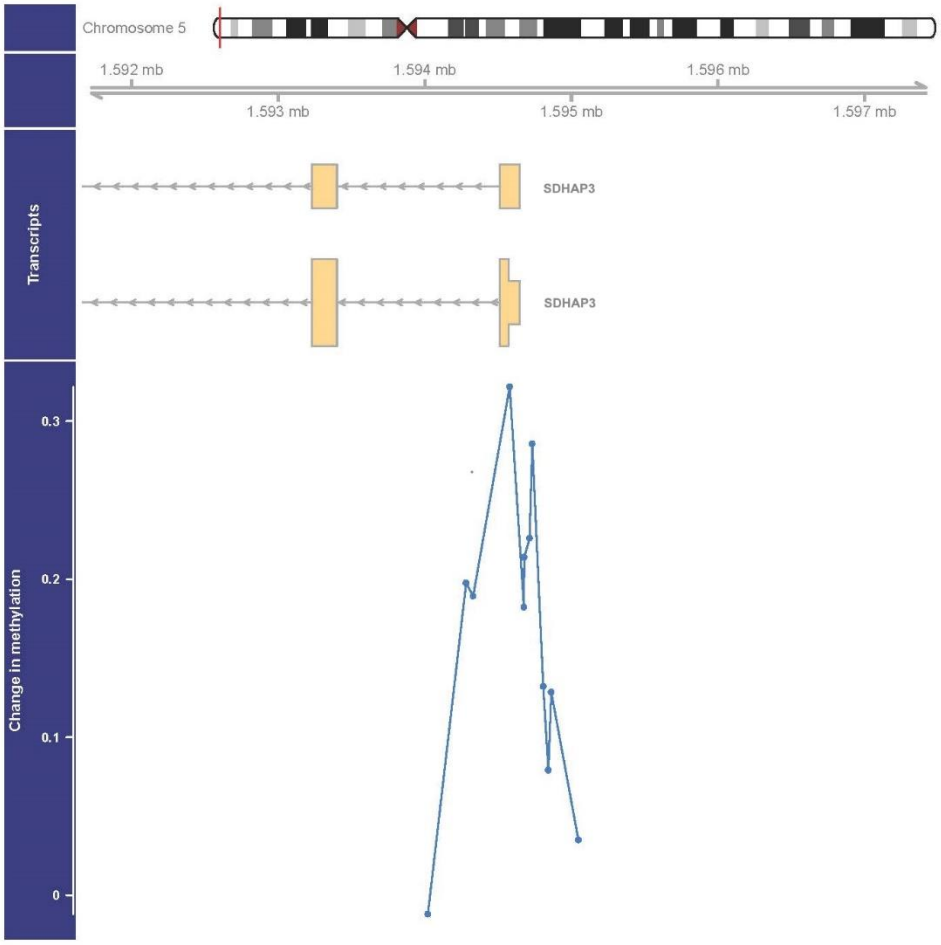

Figure 3.1. Epigenome-wide significant differentially methylated region associated with $PM_{2.5}$ located at chromosome 5 (location: 1,594,282-1,594,863).

The top portion of this plot is a karyogram of chromosome 5, with a red vertical bar indicating the location of interest. Below that is a graphic showing the exact chromosomal location in Mb, followed by a pictogram of the *SDHAP3* gene. Tan boxes represent exons (vertically narrow portions are untranslated regions), and grey lines with arrows represent intronics, as well as the direction of transcription (arrows pointing to the left indicate that this gene is transcribed from right to left, implying it is on the minus strand). The bottom section presents evidence for differential methylation of a DMR (blue line), based on a cluster of 12 CpGs (blue dots). The vertical axis for this section reflects the $PM_{2.5}$ coefficient from the linear model.

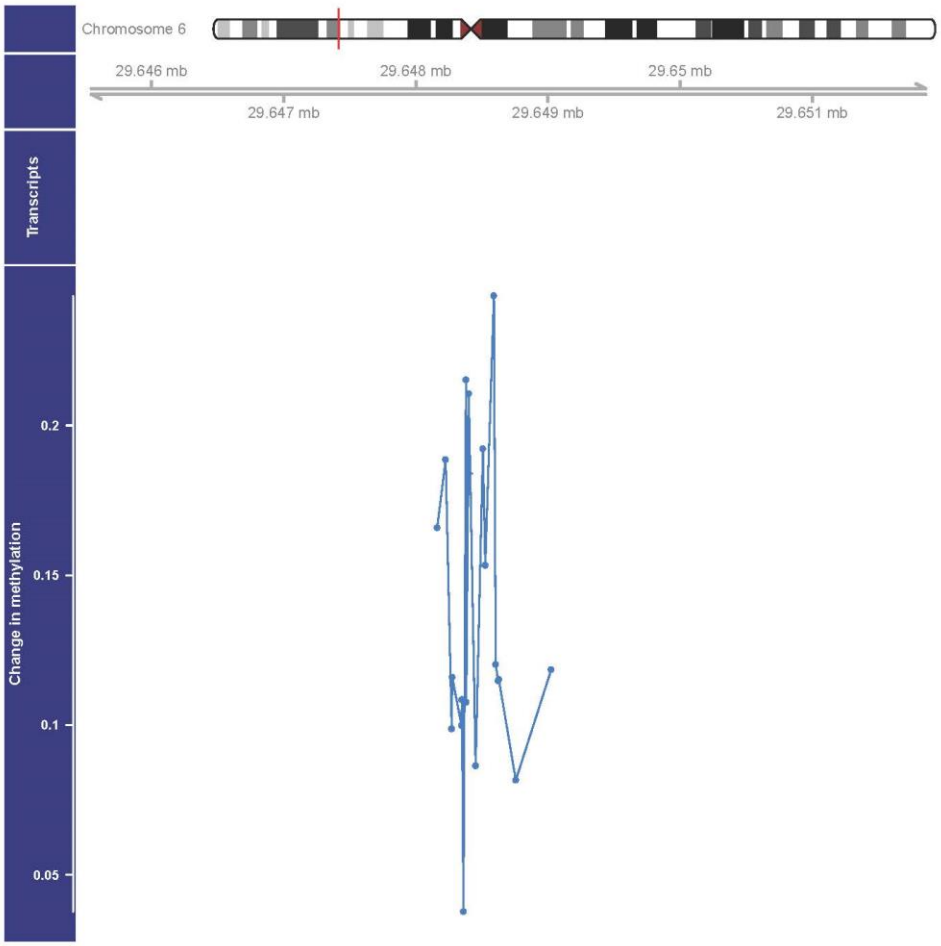

Figure 3.2. Epigenome-wide significant differentially methylated region associated with $PM_{2.5}$ located at chromosome 6 (location: 29,648,379-29,649,024).

The top portion of this plot is a karyogram of chromosome 6, with a red vertical bar indicating the location of interest. Below that is a graphic showing the exact chromosomal location in Mb. The bottom section presents evidence for differential methylation of a DMR (blue line), based on a cluster of 19 CpGs (blue dots). The vertical axis for this section reflects the $PM_{2.5}$ coefficient from the linear model.

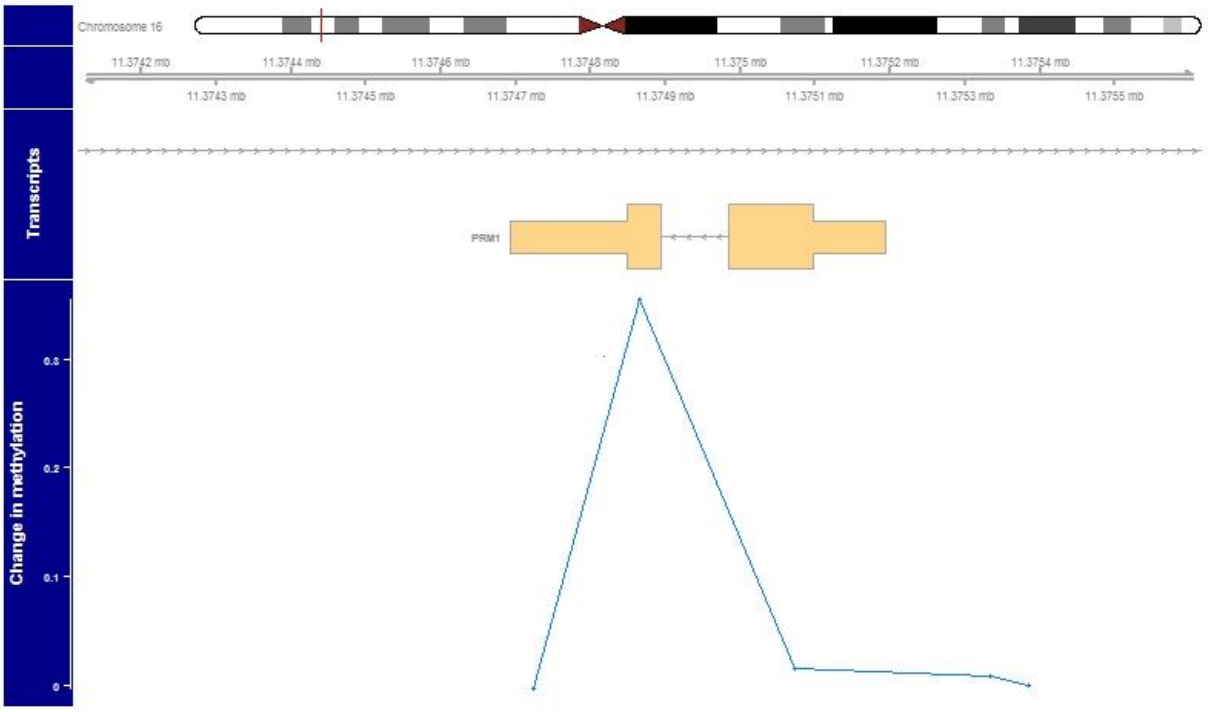

Figure 3.3. Epigenome-wide significant differentially methylated region associated with $PM_{2.5}$ located at chromosome 16 (location: 11,374,865-11,374,865).

The top portion of this plot is a karyogram of chromosome 16, with a red vertical bar indicating the location of interest. Below that is a graphic showing the exact chromosomal location in Mb, followed by a pictogram of the *PRM1* gene. Tan boxes represent exons (vertically narrow portions are untranslated regions), and grey lines with arrows represent intronics, as well as the direction of transcription (arrows pointing to the left indicate that this gene is transcribed from right to left, implying it is on the minus strand). The bottom section presents evidence for differential methylation of a DMR (blue line), based on a cluster of 5 CpGs (blue dots). The vertical axis for this section reflects the $PM_{2.5}$ coefficient from the linear model.

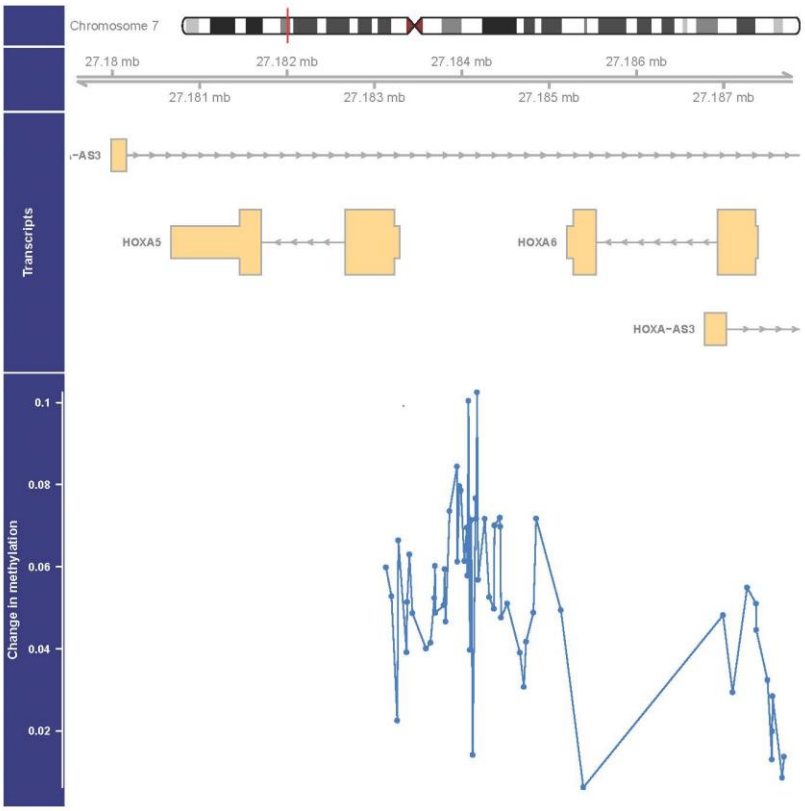

Figure 3.4. Epigenome-wide significant differentially methylated region associated with $PM_{2.5}$ located at chromosome 7 (location: 27,183,274-27,184,109).

The top portion of this plot is a karyogram of chromosome 7, with a red vertical bar indicating the location of interest. Below that is a graphic showing the exact chromosomal location in Mb, followed by a pictograms of the *HOXA5, HOXA6,* and *HOXA-AS3* genes. Tan boxes represent exons (vertically narrow portions are untranslated regions), and grey lines with arrows represent intronics, as well as the direction of transcription (arrows pointing to the left indicate that this gene is transcribed from right to left, implying it is on the minus strand, and vice versa for arrows point to the right). The bottom section presents evidence for differential methylation of a DMR (blue line), based on a cluster of 47 CpGs (blue dots). The vertical axis for this section reflects the $PM_{2.5}$ coefficient from the linear model.

CHAPTER 4

Air Pollution-Associated Methylation Signals and Subclinical Atherosclerosis: The Multi-Ethnic Study of Atherosclerosis (MESA)

ABSTRACT

Background: Atherosclerotic diseases including ischemic heart disease and stroke are among the top leading causes of death globally. Air pollution exposure has been linked to increased risk of atherosclerosis and its progression, and DNA methylation has been proposed as a mechanism underlying this association. Previous results from the Multi-Ethnic Study of Atherosclerosis (MESA) identified nine methylation sites and four methylation regions associated with long-term ambient air pollution exposure in purified monocytes of adults (age 55+).

Objectives: In this study, we investigated the association between methylation of these air pollution-associated signals and coronary artery calcium (CAC) scores (Agatston units) measured by computed tomography.

Methods: For cross-sectional associations, we used multiple linear regression with robust standard errors adjusted for demographics, socioeconomic status, cigarette smoke, secondhand smoke, body mass index, physical activity, methyl nutrient intake (folate, vitamin B12, vitamin B6, methionine, zinc), recent infection, and technical factors (n=1,094). We used linear mixed-effects models to investigate the association between DNA methylation and CAC progression (Agatston units/year) with random intercepts and slopes for each participant (n=1,020). CAC was modeled on both the natural scale (primary) and logarithmic scale (secondary).

Results: In cross-sectional analyses, a differentially methylated region (DMR) on chromosome 7 (position 27,183,274-27,184,109) was negatively associated with log-transformed CAC score (β=-0.13; 95% confidence interval [CI]: -0.25, -0.01; p=0.03). This DMR was also associated

with CAC progression on the natural (β=-4.6 Agatston units/year per 2-fold higher DNA methylation; 95% CI: -7.5, -1.7; p=0.002) and logarithmic scales (β=-0.010 ln(CAC+25) units/year per 2-fold higher DNA methylation; 95% CI: -0.018, -0.003; p=0.01). This DMR is located within the 5' untranslated region of the *HOXA5* (homeobox A5) gene and was previously associated with expression of *HOXA5*, *HOXA9*, and *HOXA10.*

Conclusions: The DMR located on chromosome 7 near the HOXA cluster genes is negatively associated with CAC as measured by the Agatston score. In prior work, this DMR was positively associated with long-term $PM_{2.5}$ exposure, which had been positively linked to CAC progression. Thus, the negative direction of the association between methylation of the DMR and CAC progression would not support a role for this DMR as a mediator in the positive association between $PM_{2.5}$ and CAC progression. The availability of one cross-sectional measurement of DNA methylation limits the interpretability of our results. Future longitudinal studies that measure air pollution, DNA methylation, and CAC prospectively are necessary to understand the relationships between these factors and their implications for atherosclerosis.

INTRODUCTION

Atherosclerosis, a chronic inflammatory disease characterized by deposition of plaques in the artery, is the leading cause of cardiovascular diseases[14] and cost the world $863 billion in 2010.[112] Two major atherosclerotic cardiovascular diseases, ischemic heart disease and stroke, are the top two causes of death worldwide and contributed to 14.6 million deaths globally in 2013.[4] Coronary artery calcium (CAC) reflects coronary atherosclerotic burden and has been shown to predict cardiovascular disease events in a longitudinal cohort.[113,114] Identifying factors associated with CAC, a measure of subclinical atherosclerosis, may help identify new targets for preventive intervention.

Current theoretical models recognize that both genetic and environmental factors influence cardiovascular disease development.[115] Commonly studied environmental risk factors

include diet, physical activity, and other factors modifiable at the individual level. However, a large body of literature suggests that environmental pollutants, such as air pollution, also contribute to cardiovascular mortality and morbidity.[3]

Proposed pathways between air pollution exposure and cardiovascular disease include systemic inflammation from the spill-over of pulmonary inflammation into the bloodstream and autonomic nervous system imbalance.[3] Systemic inflammation may involve cascades of responses involving activated leukocytes and cytokine expression, which could contribute to atherosclerosis progression.[3] Evidence that air pollution exposure is associated with increased inflammation[3,13] and risk of atherosclerosis[3] provides support for this hypothesis. However, the exact mechanisms by which air pollution inhalation leads to physiological changes are still unclear.

DNA methylation changes may link air pollution to atherosclerosis. DNA methylation is a chemical modification of DNA that does not alter DNA sequence but can affect gene expression.[15] DNA methylation has been associated with atherosclerosis,[45] ischemic heart disease,[26] and blood pressure.[96] In addition, short- and intermediate-term air pollution exposure has been linked to global hypomethylation and methylation of candidate genes.[19,20,23,47,48] However, most prior studies used a mixture of blood cells, making the results hard to interpret.[19,20,23,26,47,48,96] Monocytes play a crucial role in atherosclerosis development by promoting chronic inflammation and differentiating into macrophages that accumulate in plaques.[14] By identifying monocyte-specific DNA methylation signals, we may identify genes or pathways more biologically relevant to atherosclerosis pathogenesis.

Our prior work in the Multi-Ethnic Study of Atherosclerosis (MESA) identified differentially methylated regions and loci on the Illumina 450k array that were associated with fine particulate matter ($PM_{2.5}$) and/or oxides of nitrogen (NO_X) exposure in CD14+ purified monocytes (Chapters 2 and 3). In this paper, we investigate the association between these air

pollution-associated DNA methylation signals and CAC in the same prospective multi-ethnic cohort.

METHODS

Study Population

MESA is a longitudinal multi-site study of risk factors that predict progression of subclinical and clinical cardiovascular disease and began in 2000.[31] We also used data from two ancillary studies, MESA Air and the MESA Epigenomics and Transcriptomics Study. Additional detail of these studies can be found in Chapter 1, pages 5-7.

Our study is restricted to individuals with methylomic, transcriptomic, and computed tomography (CT) scan data (n=1,203). Those with missing covariates were further excluded and the analytic sample for the baseline and longitudinal analyses included 1,094 and 1,020 participants, respectively. With the exception of race/ethnicity and site, the analytic samples were generally representative of all participants at exam 5 from the four MESA sites. DNA methylation and gene expression data were not available for participants from Los Angeles and Chicago, and thus there are no Asian participants in the analytic sample.

DNA Methylation Quantification

DNA extraction and DNA methylation quantification has been described in detail in Chapter 2, pages 15-17. Briefly, trained technicians isolated CD14+ monocytes from blood obtained at the 5th examination following strict protocols.[33] Methylomic profiles of monocytes were assessed using the Infinium HumanMethylation450 BeadChip after bisulfite conversion of methylated DNA.[33]

In our prior work in this cohort (Chapters 2 and 3), we identified nine differentially methylated sites (DMSs) associated with $PM_{2.5}$ and/or NO_X exposure. The nine CpG sites include cg20455854, cg07855639, cg07598385, cg17360854, cg23599683, cg05926640,

cg04310517, cg09509909, and cg11756214. We also identified four differentially methylated regions (DMRs) associated with air pollution and they are located on chromosome 5 (position 1,594,282-1,594,863), chromosome 6 (position 29,648,379-29,649,024), chromosome 16 (11,374,865-11,374,865), and chromosome 7 (position 27,183,274-27,184,109). For DMSs, we used the M-value for each CpG site in association analyses. For DMRs, we used the average M-value for all CpGs within the region.

Assessment of Coronary Artery Calcium

Coronary artery calcium was assessed using chest CT as previously described.[116] CAC was measured by certified technologists who conducted CT scans using multi-detector and electron-beam CT scanners following a standardized protocol. All participants were scanned twice on separate breath-holds at full inspiration. The field of view included a phantom of known physical calcium concentration, which was used to adjust the average of the Agatston score of the two scans. The CAC scored was calculated according to the methodology of Agatston et al.[117,118] Different scanners were used across studies and over time, but consistent calibration using a phantom was conducted to calibrate the scans.

All participants received CT scans at baseline. By design, approximately half of the participants were scanned during exam 2 and the other half scanned during exam 3. Those without exam 3 scans were preferentially selected for CT scans during exam 4. For exam 5, participants who had CT scans for exam 3 and 4 were more likely to be selected for CT scans.

Other Measures

A description of covariates used in this analysis are described in Chapter 2, page 17-18.

Statistical Analysis

Cross-sectional Model

The cross-sectional association between DNA methylation and CAC at exam 5 was assessed using multiple linear regression with robust standard errors.[119] A separate model was fit for each DMS and DMR. Given the small number of candidate sites and regions investigated, we report nominal p-values unadjusted for multiple comparisons. Our primary analysis modeled CAC continuously on the natural scale as opposed to the commonly used logarithmic scale. The logarithmic scale may exaggerate differences at the low end of the scale, where the noise is, and where the differences in scanner technologies are amplified. The natural scale performed well in an analysis of CAC progression and cardiovascular risk factors.[120] However, relative differences in CAC scores are also of interest and adding a constant to the score prior to logarithmic transformation may prevent very small CAC scores from causing unreasonably large relative differences.[121] Thus, in secondary analyses, we modeled CAC as ln(CAC+25).

Our models adjusted for age (years), race/ethnicity (Black, Hispanic, White), sex (female, male), study site (NY, MD, MN, NC), income (<$25,000, $25,000-$49,000, $50,000-$99,999, $100,000+), education (less than high school, high school, some college, bachelor's/association/technical, advanced degree), pack-years of cigarettes, smoking status (never, former, current), secondhand smoke (hours per week), body mass index (kg/m^2), physical activity (MET-min/wk m-su), methyl nutrient intake (folate, vitamin B12, vitamin B6, methionine, zinc), residual cell contamination (by enrichment scores non-monocytes: neutrophils, B cells, T cells, and natural killer cells[33]), recent infection (yes/no), and methylation chip and position. We did not adjust for CT scanner type in this analysis because it was perfectly predicted by study site, which was already included as an adjustment variable. Results did not differ by choosing study site or CT scanner type. Participants with missing data on any variables were excluded.

We considered methyl nutrient intake as a potential confounder because it can affect S-adenosylhomocysteine (SAH) levels. SAH is a byproduct of transmethylation reactions that use S-adenosylmethionine (SAM) as a methyl donor and is the precursor to homocysteine.[122]

Apolipoprotiein E-deficient mice that were fed a high methionine diet deficient in folate had significantly increased SAH levels and increased atherosclerotic lesion areas.[123] SAH is considered a sensitive marker of atherosclerosis.[122] SAH levels may affect the pathogenesis of atherosclerosis through alterations in DNA methylation such as through regulation of DNA methyltransferases.[122] However, SAH may also affect atherosclerosis development through pathways besides DNA methylation, such as by modifying histone methylation, which can activate or repress genes involved in inflammation.[122] Moreover, SAH accumulation can induce oxidative stress and apoptosis in endothelial cells.[122] Thus, methyl nutrient intake is a plausible confounder of the association between DNA methylation and atherosclerosis.

Longitudinal Model

The association between DNA methylation and CAC progression was assessed using linear mixed-effects models with a random intercept and random slope for each participant. A separate model was fit for each DMS or DMR. In addition to modeling the longitudinal association between DNA methylation and CAC, the mixed-effects model also simultaneously modeled and controlled for the baseline association between DNA methylation and CAC.[120]

Again, in primary analyses CAC was modeled continuously on the natural scale to avoid emphasizing low CAC scores, and CAC on the natural scale performed well in a study of CAC progression in the MESA cohort.[120] In secondary analyses, we modeled CAC as ln(CAC+25). The mixed-effects model allowed for participants with variable numbers of observations to still be included in the analysis. Any scan that was performed after a coronary revascularization procedure was removed from the analysis.

Age, sex, race/ethnicity, income, education, neighborhood socioeconomic status, smoking, secondhand smoke, body mass index, and physical activity were obtained from baseline. These variables were considered time-invariant and were included in the model as a

main effect and interaction term with time. Variables that may specifically affect exam 5 methylation values were included as main effects only, and these included methylation chip, chip position, residual cell contamination, recent infection, and methyl nutrient intake.

Functional Annotation Analysis

Differentially methylated regions and sites were analyzed for overlap with potentially functional genomic regions. A description of our functional analysis approach can be found in Chapter 2, page 19.

RESULTS

MESA participants with available CAC scores, DNA methylation measurements, and no missing data in covariate were included in the analysis. Participant demographics and other characteristics of participants included in the cross-sectional (n=1,094) and longitudinal analyses (n=1,020) can be found in Table 4.1. In general, participants included in the cross-sectional analysis were very similar to participants in the longitudinal analysis. The mean age of participants at baseline was 60 years old. Approximately half of the sample were female, and the sample consisted of roughly 21% Black, 33% Hispanic, and 46% White participants. At baseline, 55% of participants had CAC scores of 0, and mean CAC was 115 ± 335 Agatston units (median 0, 25th-75th percentile 0-62) for participants in the cross-sectional analysis. At the 5th exam, about 30% of participants had CAC scores of zero, and the mean CAC score increased to 299 ± 626 Agatston units (median 44, 25th-75th percentile 0–293) for participants in the cross-sectional analysis. Mean methylation M-values for each air pollution-associated DMR or DMS were generally consistent across study sites (Table 4.2).

The cross-sectional associations between air pollution-associated methylation signals and CAC scores can be found in Table 4.3. In the ln(CAC+25) analyses, the DMR on chromosome 7 (position 27,183,274-27,184,109) was significantly associated with CAC score at

exam 5 (β=-0.13; 95% CI: -0.25, -0.01; p=0.03). None of the other air pollution-associated regions or sites were significantly associated with CAC in cross-sectional analyses.

In the longitudinal analysis, DNA methylation at the DMR located on chromosome 7 was statistically significantly and inversely associated with CAC progression (β=-4.6 Agatston units/year; 95% CI: -7.5, -1.7; per 2-fold higher methylation; p=0.002) on the natural scale after adjustment for potential confounders (Table 4.4). This association was slightly attenuated when CAC was modeled on the logarithmic scale (p=0.01; Table 4.4). This DMR overlaps the 5' untranslated region of *HOXA5,* a predicted weak promoter region, and heterochromatin (Table 3.1, Chapter 3). DNA methylation of this DMR was also positively associated with long-term exposure to $PM_{2.5}$ (Table 3.1, Chapter 3) and *cis*-gene expression of three *HOXA* cluster genes (*HOXA5*, *HOXA9*, and *HOX10*) (Table 3.2, Chapter 3). Other air pollution-associated sites or regions were not associated with CAC progression.

DISCUSSION

We investigated the association between methylation of air pollution-associated methylation signals and CAC, a measure of subclinical atherosclerosis. DNA methylation of an air pollution-associated DMR, located on chromosome 7 that overlaps the 5' untranslated region of *HOXA5*, was negatively associated with CAC in both cross-sectional and longitudinal analyses. DNA methylation of other air pollution-associated signals were not associated with CAC.

Methylation of the DMR located on chromosome 7 was positively associated with long-term ambient $PM_{2.5}$ exposure in a prior cross-sectional analysis (Chapter 3). A prior study in the MESA cohort found long-term $PM_{2.5}$ exposure to be positively associated with CAC progression.[124] Thus, we would expect methylation at this DMR to be positively associated with CAC as well. Contrary to expectations, methylation in this region was negatively associated with CAC in both the cross-sectional and longitudinal analyses.

The inverse association between methylation at the DMR on chromosome 7 and CAC is also contrary to what we would expect based on current biological knowledge. Disturbed flow in the carotid artery regulates endothelial inflammation and monocyte-endothelial cell adhesion and also induces hypermethylation of *HOXA5* and downregulation of its expression in endothelial cells.[100] Higher methylation of this DMR was associated with reduced *HOXA5* expression (Chapter 3), so we would expect a positive association between this DMR and CAC progression. In addition to being associated with *HOXA5* expression, methylation of this DMR was also positively associated with *HOXA9* expression (Chapter 3). *HOXA9* regulates pro-inflammatory factors in endothelial cells and was found to be more highly expressed in atherosclerotic aortas compared to normal aortas.[27] This further supports our expectation that methylation of this DMR should be positively associated with to CAC progression.

It is possible that methylation of this DMR may be accompanied by compensatory mechanisms that mitigate the effect of air pollution exposures on CAC. It is also possible that an unknown confounder explains the negative association between DNA methylation of the DMR on chromosome 7 and CAC. Likewise, an unknown confounder may also explain the positive association observed between the DMR at chromosome 7 and air pollution exposure from Chapter 3. Moreover, we cannot rule out the possibility that the relationship we observed between methylation of the DMR on chromosome 7 and CAC is a chance finding. Future studies that measure DNA methylation, CAC, air pollution, and potential confounders over time are needed to disentangle the relationship between air pollution-associated DNA methylation signals and CAC.

Our study is limited by the cross-sectional measurement of DNA methylation at exam 5. DNA methylation is pliable and may vary over time and in response to environmental exposures.[39] Given the cross-sectional measurement of DNA methylation, we cannot establish temporality of DNA methylation and CAC. Moreover, for our longitudinal analyses, we modeled the association between exam 5 DNA methylation and CAC progression from baseline to exam

5 (a 10 year period). Thus, we cannot rule out the possibility of reverse causation, where the association observed between the *HOXA5* DMR and CAC is due to the influence of CAC on DNA methylation of the DMR. While the current literature focuses on the potential for DNA methylation to be play a role in disease pathogenesis, it is also possible that disease processes may affect DNA methylation patterns. Future studies can address this gap in the literature to provide insight into whether diseases, especially progressive conditions like atherosclerosis, can alter DNA methylation during disease development.

Batch effects are known to cause spurious associations in microarray studies.[58] In this study, monocyte purification and DNA extraction occurred at each study location before DNA methylation was assessed, and this may be a source of unwanted variability in methylation measurements. However, the procedures were conducted by trained technicians following standardized protocols, and methylation samples were randomized to methylation chips and positions to minimize the potential for confounding by batch effects. In addition, our analysis adjusted for site, methylation chip, and methylation chip position to reduce potential confounding by these technical factors. Finally, there is no evidence that methylation differed across study sites (Table 4.2).

Selection bias may also partially explain our results. Observations with missing data were not included in the longitudinal analysis. However, in minimally adjusted models that included all participants with CAC measurements, the point estimates for the effect of the methylation of the DMR on Chromosome 7 on CAC were not materially different (results not shown). In addition, our study included approximately 15% of the original sample of MESA participants. While our study participants were generally similar to exam 5 participants from the four sites represented, we did not have any participants from Los Angeles and Chicago. The longitudinal study of air pollution exposure and CAC progression included participants from all six MESA sites.

Besides the DMR on chromosome 7, we did not find any significant associations between other air pollution-associated signals and CAC. Our reduced sample size may have decreased our power to detect associations between DNA methylation and CAC. In addition, the CAC score in Agatston units does not provide information on the low end of the atherosclerosis distribution (30% of participants at exam 5 had a CAC score of zero), and this reduced contrast may further diminish our power.

Our study has several strengths. MESA is a well-characterized prospective cohort with unusual dedication of resources for measurement of subclinical atherosclerosis. MESA had standardized protocols for CAC scanning and interpretation, which were coupled with a substantial follow-up time of ten years. In addition, the availability of DNA methylation from purified monocytes allowed us to study monocyte-specific associations, which are more relevant for atherosclerosis. The linear mixed-effects model permitted variable follow-up times and accounted for repeated measures within a participant.

CONCLUSION

We investigated the cross-sectional and longitudinal association between air pollution-associated DNA methylation signals and CAC. We identified one DMR that overlaps the 5' untranslated region of *HOXA5* with DNA methylation that was negatively associated with cross-sectional and longitudinal CAC. However, since DNA methylation was only measured at exam 5, these results should be interpreted with caution. Additional longitudinal studies that measure both DNA methylation and CAC over time are needed to elucidate the association between DNA methylation of air pollution-associated DNA methylation and CAC.

TABLES

Table 4.1. Descriptive characteristics of MESA participants in cross-sectional (n=1,094) and longitudinal analyses (n=1,020).

	Cross-sectional (n=1,094)	Longitudinal (n=1,020)
	n (%) or mean ± SD	n (%) or mean ± SD
Age at baseline (y)	59.9 ± 9.5	60.3 ± 9.6
Race/ethnicity, %		
White	501 (45.8)	477 (46.8)
Black	236 (21.6)	209 (20.5)
Hispanic	357 (32.6)	334 (32.7)
Sex, %		
Female	564 (51.6)	536 (52.5)
Male	530 (48.4)	484 (47.5)
Smoking status at baseline, %[a]		
Never	502 (45.9)	528 (51.8)
Former	440 (40.2)	440 (43.1)
Current	152 (13.9)	52 (5.1)
Secondhand smoke at baseline (hours per week)[a]	4.7 ± 14.3	4.6 ± 13.9
Education, %[a]		
Less than high school	159 (14.5)	143 (14.0)
High school/GED	208 (19.0)	189 (18.5)
Some college but no degree	201 (18.4)	181 (17.7)
Bachelor's/Associate/Technical	333 (30.4)	317 (31.1)
Advanced degree	193 (17.6)	190 (18.6)
Body mass index at baseline (kg/m^2)	29.3 ± 5.2	29.3 ± 5.2
Physical activity at baseline (MET-min/wk m-su)[a]	6,919.70 ± 7,037.9	6,907.50 ± 7,099.8
Income at baseline, %[a]		
< $24,999	259 (24.0)	238 (23.3)
$25,000 - $49,999	375 (34.8)	359 (35.2)
$50,000 - $99,999	330 (30.6)	307 (30.1)
$100,000 or more	115 (10.7)	116 (11.4)
Recent infection, %[a]		
Yes	243 (22.2)	235 (23.0)
Folate (mcg)[a]	315.9 ± 167.5	316.1 ± 164.4
Vitamin B12 (mcg)[a]	3.9 ± 3.8	4.0 ± 3.9
Vitamin B6 (mg)[a]	1.5 ± 0.8	1.5 ± 0.8
Methionine (g)[a]	1.4 ± 0.8	1.4 ± 0.8
Zinc (mg)[a]	8.8 ± 5.0	8.8 ± 5.0
Coronary artery calcium at baseline (Agatston unit)[a]	114.6 ± 334.8	112.9 ± 350.9
Coronary artery calcium at exam 5 (Agatston unit)	299.4 ± 626.4	279.0 ± 573.9

[a] Contains missing values.

Table 4.2. Average methylation for air pollution-associated methylation sites and regions.

	Total		NY		MD		MN		NC	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
cg04310517	3.2	0.2	3.2	0.2	3.2	0.2	3.2	0.2	3.1	0.2
cg05926640	-2.9	0.2	-2.9	0.2	-2.9	0.2	-2.9	0.2	-2.9	0.2
cg07598385	-3.0	0.6	-3.0	0.7	-2.9	0.6	-3.1	0.6	-3.7	0.5
cg07855639	1.3	0.4	1.2	0.4	1.3	0.4	1.3	0.4	1.2	0.3
cg09509909	-2.2	0.3	-2.2	0.3	-2.2	0.4	-2.3	0.3	-2.3	0.2
cg11756214	-2.6	0.1	-2.6	0.2	-2.6	0.2	-2.6	0.1	-2.6	0.2
cg17360854	-4.4	0.5	-4.5	0.5	-4.3	0.4	-4.5	0.4	-4.7	0.4
cg20455854	-3.3	0.8	-3.4	0.8	-3.1	0.8	-3.4	0.8	-4.1	0.5
cg23599683	1.8	0.3	1.8	0.3	1.8	0.3	1.7	0.3	1.8	0.3
Chr5: 1594282-1594863	-2.7	1.2	-2.6	1.3	-2.7	1.2	-2.8	1.2	-2.9	1.1
Chr6: 29648379-29649024	1.5	1.4	1.5	1.3	1.6	1.4	1.5	1.4	1.4	1.4
Chr7: 27183274-27184109	1.6	0.7	1.5	0.7	1.6	0.7	1.6	0.7	1.5	0.8
Chr16: 11374865-11374865	-0.1	2.5	0.2	2.5	-0.6	2.4	0.0	2.4	-1.3	2.3

SD, standard deviation.

Table 4.3. Cross-sectional association between air pollution-associated methylation sites and regions (per 2-fold higher DNA methylation) and coronary artery calcium in MESA.

		Natural Scale		ln(CAC+25)	
Methylation Site or Region	N	β (95% CI)	P-value	β (95% CI)	P-value
cg04310517	1094	-20.3 (-205.4, 164.7)	0.83	-0.02 (-0.41, 0.38)	0.93
cg05926640	1094	-75.7 (-295.4, 144.1)	0.50	0.21 (-0.26, 0.68)	0.38
cg07598385	1094	44.7 (-29.4, 118.7)	0.24	0.14 (-0.02, 0.30)	0.09
cg07855639	1094	20.8 (-79.8, 121.4)	0.69	0.09 (-0.13, 0.30)	0.42
cg09509909	1094	-109.4 (-234.0, 15.2)	0.09	-0.13 (-0.39, 0.14)	0.36
cg11756214	1094	71.2 (-200.9, 343.3)	0.61	-0.35 (-0.93, 0.24)	0.24
cg17360854	1094	83.7 (-15.7, 183.1)	0.10	0.11 (-0.10, 0.33)	0.30
cg20455854	1094	-13.3 (-72.8, 46.3)	0.66	0.01 (-0.11, 0.14)	0.84
cg23599683	1094	41.1 (-97.1, 179.2)	0.56	0.05 (-0.25, 0.34)	0.76
Chr5: 1594282-1594863	1094	4.2 (-26.5, 34.9)	0.79	0.02 (-0.05, 0.08)	0.58
Chr6: 29648379-29649024	1094	-0.5 (-28.1, 27.1)	0.97	-0.01 (-0.07, 0.05)	0.81
Chr7: 27183274-27184109	1094	-30.8 (-85.9, 24.4)	0.27	-0.13 (-0.25, -0.01)	0.03
Chr16: 11374865-11374865	1094	2.9 (-13.0, 18.8)	0.72	0.01 (-0.03, 0.04)	0.74

CI, confidence interval.

[a] Models adjusted for age, race/ethnicity, sex, study site, income, education, cigarette smoking, secondhand smoke, body mass index, physical activity, methyl nutrient intake (folate, vitamin B12, vitamin B6, methionine, zinc), residual cell contamination by non-monocytes, recent infection, and methylation chip and position.

Table 4.4. Association between air pollution-associated methylation sites and regions (per 2-fold higher DNA methylation) and coronary artery calcium progression in MESA.

			Natural Scale		ln(CAC+25)	
Methylation Site or Region	N	N Observations	β (95% CI)	P-value	β (95% CI)	P-value
cg04310517	1020	2204	1.5 (-8.2, 11.1)	0.76	0.014 (-0.012, 0.039)	0.30
cg05926640	1020	2204	0.3 (-9.5, 10.1)	0.95	-0.007 (-0.033, 0.019)	0.61
cg07598385	1020	2204	1.7 (-1.5, 4.9)	0.29	0.000 (-0.008, 0.009)	0.92
cg07855639	1020	2204	1.8 (-3.5, 7.1)	0.50	0.009 (-0.005, 0.023)	0.22
cg09509909	1020	2204	-0.9 (-6.6, 4.8)	0.76	0.007 (-0.008, 0.022)	0.35
cg11756214	1020	2204	12.9 (0.0, 25.9)	0.05	-0.011 (-0.046, 0.024)	0.54
cg17360854	1020	2204	0.9 (-3.5, 5.4)	0.68	-0.010 (-0.022, 0.002)	0.10
cg20455854	1020	2204	-1.3 (-3.9, 1.3)	0.32	-0.004 (-0.011, 0.003)	0.23
cg23599683	1020	2204	3.3 (-3.8, 10.4)	0.36	-0.005 (-0.024, 0.014)	0.63
Chr5: 1594282-1594863	1020	2204	-0.2 (-1.9, 1.4)	0.77	0.000 (-0.005, 0.004)	0.91
Chr6: 29648379-29649024	1020	2204	0.7 (-0.8, 2.1)	0.35	0.000 (-0.004, 0.004)	0.95
Chr7: 27183274-27184109	1020	2204	-4.6 (-7.5, -1.7)	0.002	-0.010 (-0.018, -0.003)	0.01
Chr16: 11374865-11374865	1020	2204	-0.3 (-1.2, 0.5)	0.41	0.000 (-0.002, 0.002)	0.94

CI, confidence interval.

[a] Model adjusted for age, race/ethnicity, sex, study site, income, education, cigarette smoking, secondhand smoke, body mass index, physical activity, methyl nutrient intake (folate, vitamin B12, vitamin B6, methionine, zinc), residual cell contamination by non-monocytes, recent infection, and methylation chip and position.

CHAPTER 5

Conclusions and Implications for Future Research

In this dissertation, we identified differentially methylated DNA regions and loci in monocytes associated with exposure to long-term ambient $PM_{2.5}$ and/or NO_X. We identified novel associations between air pollution and methylation sites/regions near genes that have never been reported to be associated with air pollution exposure before. Several of these air pollution-associated methylation signals were also associated with *cis*-gene expression profiles and overlapped predicted functional regions such as promoters and enhancers. One of the differentially methylated regions was also associated with CAC scores. This study not only identifies new epigenetic markers to be targeted for future functional studies, but also supports the possibility that epigenetic alterations may link air pollution inhalation to pathophysiological changes.

A large body of literature has documented that exposure to air pollution elevates risk of cardiovascular outcomes,[3] inducing policymakers to progressively reduce allowable pollutant concentrations over the years to safeguard public health. The Clean Air Act (CAA) requires the U.S. Environmental Protection Agency (EPA) to develop and enforce air quality regulations to protect the public from hazardous air pollutants. The National Ambient Air Quality Standards (NAAQS) regulate levels of six primary pollutants considered harmful to public health and the environment: sulfur dioxide, particulate matter, carbon monoxide, nitrogen dioxide, lead, and ozone. However, critics of these standards have cited a lack of understanding of mechanisms to be a major limitation.[125]

While a contextual framework for biological pathways linking air pollution to cardiovascular outcomes has been laid out in an update to the American Heart Association scientific statement on “Air Pollution and Cardiovascular Disease,”[3] there are still many gaps in

our understanding of how these pathways are activated by air pollution inhalation. Growing literature suggests that epigenetic mechanisms such as DNA methylation may mediate the association between environmental pollutants and health effects.[39]

This dissertation contributes to this gap in understanding by providing evidence associating long-term air pollution exposure with differential DNA methylation of specific regions and sites. Previous studies have found associations between air pollution exposure and global and site-specific DNA methylation.[19,48,93–95] However, none of these were replicated by our study. Substantial differences in study design and study population could partially explain these results. Past studies focused on short- and intermediate-term air pollution exposure while we concentrated on long-term (one-year average) exposure, as is appropriate for studies of chronic conditions such as atherosclerosis. Chronic and acute exposures may have different epigenetic signatures. Previous studies have also been conducted using mixed blood cells, where the cell type composition of samples may confound results. In contrast, we analyzed DNA methylation from purified monocytes, which improves our ability to identify methylation signals important in atherosclerosis. We also utilized sophisticated individual-level ambient air pollution exposure predictions from MESA Air which were able to characterize fine spatial variations in pollutant concentrations to reduce measurement error. Prior studies of shorter term exposures typically used central monitors to assign air pollution exposures to all participants within a specific region. Moreover, many prior studies used data from the Normative Aging Study, a mostly White cohort of elderly male participants from the Boston area.[19] Our study population includes both female and male participants and is ethnically diverse, consisting of Black, Hispanic, and White participants from four U.S. regions.

Given the multiethnic nature of our participants, our results are more generalizable to a multiethnic population than past studies that included mostly White participants from Boston or Germany.[19–22] However, our study only included only Blacks, Hispanics, and Whites and may not be representative of other racial/ethnic groups. Our study included both females and males,

thereby improving our generalizability compared to results from the Normative Aging Study which only included men.[19] Our study also includes participants age 55 and over, constituting a younger adult population than that of the Normative Aging Study. Since our participants reside within four U.S. regions, our results may not be generalizable to the overall nation or to other countries. Moreover, the air pollution exposures in our study are lower than exposures in many parts of the world, thus our study results may not be representative of populations experiencing higher air pollution exposures.

Few prior studies coupled DNA methylation data with mRNA expression profiles to investigate potential functional relevance. We investigated the association between DNA methylation of air pollution-associated sites and regions with *cis*-gene expression. Three out of six DMRs and six out of nine DMSs were associated with *cis*-expression of gene(s) within 1 Mb of the methylation signal in question. Moreover, we incorporated publicly available information on functional regions (such as histone modifications and DNAse hypersensitive sites) to gain a better understanding of the possible biological relevance of our results.

Consistent with the current hypothesis that inflammation plays a key role in air pollution-related cardiovascular disease,[3] we identified several air pollution-associated methylation signals that were linked with expression of genes involved in inflammation. These genes include *LGALS2*, *HOXA* genes, *DEXI*, and *IRF2BP2*. *LGALS2* encodes the protein galectin-2, which can regulate both pro- and anti-inflammatory proteins and has been found to be differentially expressed in monocytes and macrophages of coronary artery disease patients with low arteriogenic response.[75] *HOXA* genes have been found to be responsive to inflammatory activation,[101] and endothelial inflammation alters *HOXA5* methylation and expression.[100] *DEXI* is upregulated in emphysema tissue and has been suggested to modulate inflammation.[103] *IRF2BP2* binds a negative regulator of interferon-responsive genes, and macrophages that lack IRF2BP2 are inflammatory.[111] In addition, *ANKRD11* has been linked to autism, and

inflammation has been a common pathway suggested to mediate air pollution effects on both cardiovascular disease and autism.[81]

DNA methylation of two air pollution-associated methylation signals, cg05926640 and a DMR located on chromosome 5, were associated with mRNA expression of *TOMM20* and *MRPL36*, respectively. *TOMM20* encodes a subunit of the translocase of a major import channel for nuclear-encoded mitochondrial proteins,[108] and *MRPL36* participates in mitochondrial protein synthesis.[102] Aberrant mitochondrial function can disturb cholesterol balance in macrophages and promote inflammation and atherosclerosis.[109] In addition, cg05926640 methylation was also associated with expression of *IRF2BP2*, which has been linked to impaired cholesterol efflux in macrophages, atherosclerosis, and coronary artery disease.[111] Moreover, emerging evidence suggests that disrupted cholesterol homeostasis in hematopoietic stem and multipotential progenitor cells can lead to increased production of monocytes, which has been linked to increased cardiovascular risk and atherosclerosis severity.[126]

Interestingly, we identified two air pollution-associated signals that may regulate genes involved in hematopoiesis. The DMR located on chromosome 7 was associated with expression of three *HOXA* genes. *HOXA* genes are involved in hematopoiesis, and *HOXA10* overexpression in CD34+ progenitor cells can augment differentiation into monocytes and reduce differentiation into B cells and natural killer cells.[99] In addition, methylation of cg05926640, which was associated with $PM_{2.5}$ exposure, was associated with mRNA expression of three *ARID4B* transcripts. *ARID4B* has been suggested to regulate hematopoiesis by modulating *HOX* gene expression.[106]

In Chapter 4, we assessed whether methylation of air pollution-associated DMSs and DMRs were also associated with CAC in cross-sectional and longitudinal analyses. We observed a statistically significant negative association between methylation at the DMR located on chromosome 7 (near *HOXA5*) and both cross-sectional and longitudinal CAC. However, the

direction of this association does not support this DMR as a potential mediator of the $PM_{2.5}$-CAC association because $PM_{2.5}$ was positively associated with both methylation at this DMR (Chapter 3) and CAC progression.[124] This result must be interpreted very cautiously because DNA methylation was only measured at exam 5 whereas time-varying CAC was measured from exam 1 to exam 5, which spans a 10-year period. We did not detect associations for other DMSs and DMRs.

While we did not find evidence of a methylation target that is likely responsible for the association between $PM_{2.5}$ exposure and CAC progression, we cannot dismiss the possibility that DNA methylation may link air pollution inhalation to CVD. We focused on subclinical atherosclerosis as measured by CAC score in Agatston units. The CAC score is not informative about the lower end of the distribution of atherosclerosis, and 30% of participants had a CAC score of zero at exam 5. In addition, CAC scores are difficult to analyze because of their high proportion of zero values and highly skewed distribution.[127] There are additional measures of subclinical atherosclerosis, such as common carotid intima-thickness, that could be studied. However, we chose to analyze CAC because its progression was linked to long-term $PM_{2.5}$ exposure in a prior MESA analysis.[124] Air pollution exposure has also been linked to a variety of clinically overt outcomes including cardiovascular mortality, ischemic heart disease, heart failure, and cerebrovascular disease which were not analyzed in this study.[3] Investigating the association between air pollution-associated methylation signals and other CVD events and measures would contribute to the broader understanding of mechanisms underlying air pollution-related CVD.

In this dissertation, we demonstrated evidence that long-term ambient air pollution exposure is associated with DNA methylation of specific regions and loci, but it is likely that we missed signals with smaller effect sizes. Our sample size of 1,207 participants represents about one-fifth of the MESA sample and provides limited power. We also adjusted for site, which removed the between-community variation in air pollution exposure and greatly diminished our

statistical power. As we can see from Figure 2.1, there is great contrast in air pollution exposure comparing New York to the rest of the other sites. We adjusted for study site because monocytes were purified on-site and that may introduce site-specific differences in the measurement of DNA methylation and gene expression that could confound our analyses. CVD is a complex disease with multiple potential causal pathways, and the presence of numerous pathways can diminish our ability to detect a component of one pathway. Our analyses were exploratory in nature, and our statistical power was further reduced by accounting for multiple comparisons. To increase study sensitivity to identify true associations, future studies could be hypothesis-driven and focus on genes/pathways that could plausibly link air pollution exposure and CVD, using data form a larger sample of participants exposed to a greater range of air pollution concentrations. Results from this dissertation, in addition to prior research, suggest several genes and pathways that are good candidates for more focused investigation.

Study limitations should be considered when interpreting results of this study. First, DNA methylation was measured at one point in time during the 5th study exam and may not reflect a subject's usual methylation status. Due to the cross-sectional nature of our analysis, we could not establish temporality or study the effect of long-term air pollution exposure and change in DNA methylation. In the cross-sectional analysis of DNA methylation and CAC, we could not rule out the possibility of reverse causation where atherosclerosis might alter DNA methylation. Although the literature thus far has focused on disease as a potential outcome of aberrant DNA methylation, it is possible that disease processes could potentially alter DNA methylation. Future studies are needed to address this gap in the literature. The longitudinal analysis of DNA methylation and CAC progression in Chapter 4 should be interpreted carefully because DNA methylation was only measured at the 5th exam. In Chapter 4, our study may be underpowered to detect associations as our sample included about 15% of the original MESA sample, and 30% of our analytic sample had a CAC score of zero at exam 5.

Our air pollution predictions are only estimates, and do not reflect the true exposures of participants. First, we estimated exposures for outdoor air pollution concentrations at participant homes, but we did not estimate the amount of indoor air pollution that is of ambient origin. Our cohort of older adults likely spent most of their time indoors, and the outdoor air pollution at their residence may not reflect their true exposure. Unmeasured individual level characteristics such as infiltration (e.g. open windows, air conditioning) may affect personal exposure. In addition, we did not take into account individual time-activity patterns. Exposures measured at participant homes may not be a good estimate of individual exposure if participants spend much of their time away from home. Furthermore, not taking into account exposures while participants are commuting or driving may underestimate individual exposures because exposures tend to be high in those environments. Additional participant-level information such as time-location activity and home characteristics can be used to obtain improved exposure predictions in future studies.

In light of these limitations, future studies are needed to confirm our results. Future studies will ideally be prospective longitudinal studies that measure air pollution, DNA methylation, risk factors, and cardiovascular outcomes over time. Air pollution exposure should clearly precede differences in DNA methylation which clearly precede disease or markers of disease. Ideally, more effort should be devoted to reducing exposure measurement error by estimating air pollution concentrations at different microenvironments and incorporating time-location information. Future studies should also include participants who are exposed to a wide range of air pollution concentrations to increase power and to detect associations that are present at varying levels of exposure.

There are still many unanswered questions regarding the relationship of air pollution exposure, DNA methylation, and CVD. This study addressed long-term air pollution exposures because we were focused on chronic effects that relate to atherosclerosis, a progressive inflammatory process. However, it is unclear whether differences in DNA methylation

attributable to short-term exposures accumulate over time and are reflected in DNA methylation differences associated with long-term exposures. On the other hand, it is possible that short-term and long-term air pollution activate different pathways. A study incorporating different exposure windows could be informative about DNA methylation changes in response to different air pollution averaging periods, the magnitudes of these changes, the ability of these changes to persist over time, and the reversibility of these changes. In addition, our study is limited by the investigation of linear relationships between DNA methylation and air pollution exposure. In this cohort, the distribution of median methylation percent across CpG sites is bimodal, but the distribution of median methylation percent of a particular site between individuals is quite uniform.[33] Future studies may elucidate whether non-linear relationships exist between environmental exposures and DNA methylation. The reversibility of methylation changes associated with disease, either by lowering exposures or adopting healthy habits, would be important in the prevention and control of air pollution-related disease. In addition, little is known about the relationships between DNA methylation and distinct sources or components of air pollutants, which may have disparate toxicity and operate through different pathways. The associations between $PM_{2.5}$-associated methylation signals and other CVD measures and diseases (such as cancer) also remain to be elucidated. Many parts of the world experience much higher pollution levels than the United States. It is unknown whether very high concentrations of air pollution lead to alterations in DNA methylation that are different from those elicited by low concentrations.

Our analysis reported associations between $PM_{2.5}$ exposure and methylation of several genes with unclear function, especially in monocytes. The availability of gene expression data and our functional analysis were useful in providing evidence of potential biological relevance, but these results could be supplemented by additional functional experiments that knock down or overexpress these genes of interest in monocytes. Animal and *in vitro* studies can contribute greatly to our understanding of the functions of these genes and their gene products. These

studies would be an important step in identifying potential targets for therapeutic of preventive interventions.

In conclusion, we identified several methylation regions and sites in monocytes that are associated with exposure to long-term ambient air pollution exposure. Several are also related to expression of genes involved in various cellular pathways including inflammation, cholesterol homeostasis, mitochondrial function, and hematopoiesis. The results provide supportive evidence that DNA methylation may plausibly link air pollution inhalation to physiological changes. The analyses described herein are exploratory, and future studies are necessary to confirm these results and further elucidate the biological significance of identified genes and whether they participate in the pathogenesis of atherosclerosis.

REFERENCES

1. Lim SS, Vos T, Flaxman AD, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2013;380(9859):2224-2260.

2. Brook RD, Franklin B, Cascio W, et al. Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association. *Circulation*. 2004;109(21):2655-2671. doi:10.1161/01.CIR.0000128587.30041.C8.

3. Brook RD, Rajagopalan S, Pope CA, et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation*. 2010;121(21):2331-2378. doi:10.1161/CIR.0b013e3181dbece1.

4. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*. 2015;385(9963):117-171. doi:10.1016/S0140-6736(14)61682-2.

5. Diez Roux AV, Auchincloss AH, Franklin TG, et al. Long-term exposure to ambient particulate matter and prevalence of subclinical atherosclerosis in the Multi-Ethnic Study of Atherosclerosis. *Am J Epidemiol*. 2008;167(6):667-675.

6. Bauer M, Moebus S, Möhlenkamp S, et al. Urban particulate matter air pollution is associated with subclinical atherosclerosis: results from the HNR (Heinz Nixdorf Recall) study. *J Am Coll Cardiol*. 2010;56(22):1803-1808.

7. Künzli N, Jerrett M, Garcia-Esteban R, et al. Ambient air pollution and the progression of atherosclerosis in adults. *PLoS One*. 2010;5(2):e9096.

8. Hoffmann B, Moebus S, Mohlenkamp S, et al. Residential exposure to traffic is associated with coronary atherosclerosis. *Circulation*. 2007;116(5):489-496.

9. Rivera M, Basagaña X, Aguilera I, et al. Association between long-term exposure to traffic-related air pollution and subclinical atherosclerosis: the REGICOR study. *Environ Health Perspect.* 2012;121:223-230.

10. Adar SD, Sheppard L, Vedal S, et al. Fine particulate air pollution and the progression of carotid intima-medial thickness: a prospective cohort study from the multi-ethnic study of atherosclerosis and air pollution. *PLoS Med*. 2013;10(4):e1001430.

11. Kaufman JD, Adar S, Barr RG, et al. Air Pollution and Progression of Coronary Artery Calcification: MESA Air. In: *Abstracts of the 2014 Conference of the International Society of Environmental Epidemiology (ISEE).* Vol Abstract Number: 3002 | ID: S-008. Research Triangle Park, NC: Environmental Health Perspectives; 2014. http://ehp.niehs.nih.gov/isee/s-008/. Accessed January 27, 2016.

12. Su T-C, Hwang J-J, Shen Y-C, Chan C-C. Carotid Intima-Media Thickness and Long-Term Exposure to Traffic-Related Air Pollution in Middle-Aged Residents of Taiwan: A Cross-Sectional Study. *Environ Health Perspect*. 2015;123(8):773-778. doi:10.1289/ehp.1408553.

13. Hajat A, Allison M, Diez-Roux AV, et al. Long-term Exposure to Air Pollution and Markers of Inflammation, Coagulation, and Endothelial Activation: A Repeat-measures Analysis in the Multi-Ethnic Study of Atherosclerosis (MESA). *Epidemiology*. 2015;26(3):310-320. doi:10.1097/EDE.0000000000000267.

14. Woollard KJ, Geissmann F. Monocytes in atherosclerosis: subsets and functions. *Nat Rev Cardiol*. 2010;7(2):77-86.

15. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet*. 2013;14(3):204-220.

16. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. 2012;13(7):484-492.

17. Pinney SE. Mammalian Non-CpG Methylation: Stem Cells and Beyond. *Biology*. 2014;3(4):739-751. doi:10.3390/biology3040739.

18. Laurent L, Wong E, Li G, et al. Dynamic changes in the human methylome during differentiation. *Genome Res*. 2010;20(3):320-331. doi:10.1101/gr.101907.109.

19. Baccarelli A, Wright RO, Bollati V, et al. Rapid DNA Methylation Changes after Exposure to Traffic Particles. *Am J Respir Crit Care Med*. 2009;179(7):572-578. doi:10.1164/rccm.200807-1097OC.

20. Madrigano J, Baccarelli A, Mittleman MA, et al. Prolonged Exposure to Particulate Pollution, Genes Associated with Glutathione Pathways, and DNA Methylation in a Cohort of Older Men. *Environ Health Perspect*. 2011;119(7):977-982. doi:10.1289/ehp.1002773.

21. Madrigano J, Baccarelli A, Mittleman MA, et al. Air Pollution and DNA Methylation: Interaction by Psychological Factors in the VA Normative Aging Study. *Am J Epidemiol*. 2012;176(3):224-232. doi:10.1093/aje/kwr523.

22. Panni T, Mehta AJ, Schwartz JD, et al. A Genome-Wide Analysis of DNA Methylation and Fine Particulate Matter Air Pollution in Three Study Populations: KORA F3, KORA F4, and the Normative Aging Study. *Environ Health Perspect*. January 2016. doi:10.1289/ehp.1509966.

23. Tarantini L, Bonzini M, Apostoli P, et al. Effects of Particulate Matter on Genomic DNA Methylation Content and iNOS Promoter Methylation. *Environ Health Perspect*. 2009;117(2):217-222. doi:10.1289/ehp.11898.

24. Kile ML, Fang S, Baccarelli AA, Tarantini L, Cavallari J, Christiani DC. A panel study of occupational exposure to fine particulate matter and changes in DNA methylation over a single workday and years worked in boilermaker welders. *Environ Health*. 2013;12:47. doi:10.1186/1476-069X-12-47.

25. Aavik E, Lumivuori H, Leppänen O, et al. Global DNA methylation analysis of human atherosclerotic plaques reveals extensive genomic hypomethylation and reactivation at imprinted locus 14q32 involving induction of a miRNA cluster. *Eur Heart J*. 2015;36(16):993-1000. doi:10.1093/eurheartj/ehu437.

26. Baccarelli A, Wright R, Bollati V, et al. Ischemic heart disease and stroke in relation to blood DNA methylation. *Epidemiology*. 2010;21(6):819.

27. Zaina S, Heyn H, Carmona FJ, et al. DNA Methylation Map of Human Atherosclerosis. *Circ Cardiovasc Genet*. 2014;7(5):692-700. doi:10.1161/CIRCGENETICS.113.000441.

28. Nazarenko MS, Markov AV, Lebedev IN, et al. A Comparison of Genome-Wide DNA Methylation Patterns between Different Vascular Tissues from Patients with Coronary Heart Disease. *PLoS ONE*. 2015;10(4). doi:10.1371/journal.pone.0122601.

29. Reynolds LM, Wan M, Ding J, et al. DNA Methylation of the Aryl Hydrocarbon Receptor Repressor Associations with Cigarette Smoking and Subclinical Atherosclerosis. *Circ Cardiovasc Genet*. August 2015:CIRCGENETICS.115.001097. doi:10.1161/CIRCGENETICS.115.001097.

30. Zhao J, An Q, Goldberg J, Quyyumi AA, Vaccarino V. Promoter methylation of glucocorticoid receptor gene is associated with subclinical atherosclerosis: A monozygotic twin study. *Atherosclerosis*. 2015;242(1):71-76. doi:10.1016/j.atherosclerosis.2015.07.007.

31. Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156(9):871-881.

32. Kaufman JD, Adar SD, Allen RW, et al. Prospective Study of Particulate Air Pollution Exposures, Subclinical Atherosclerosis, and Clinical Cardiovascular Disease: : The Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Am J Epidemiol*. 2012;176(9):825-837. doi:10.1093/aje/kws169.

33. Liu Y, Ding J, Reynolds LM, et al. Methylomics of gene expression in human monocytes. *Hum Mol Genet*. 2013;22(24):5065-5074.

34. Cohen MA, Adar SD, Allen RW, et al. Approach to Estimating Participant Pollutant Exposures in the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Environ Sci Technol*. 2009;43(13):4687-4693. doi:10.1021/es8030837.

35. Sampson PD, Szpiro AA, Sheppard L, Lindström J, Kaufman JD. Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data. *Atmos Environ*. 2011;45(36):6593-6606.

36. Keller JP, Olives C, Kim S-Y, et al. A Unified Spatiotemporal Modeling Approach for Predicting Concentrations of Multiple Air Pollutants in the Multi-Ethnic Study of Atherosclerosis and Air Pollution. *Environ Health Perspect*. 2015;123(4):301-309. doi:10.1289/ehp.1408145.

37. Houseman EA, Kelsey KT, Wiencke JK, Marsit CJ. Cell-composition effects in the analysis of DNA methylation array data: a mathematical perspective. *BMC Bioinformatics*. 2015;16:95. doi:10.1186/s12859-015-0527-y.

38. Morris TJ, Beck S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods San Diego Calif*. 2015;72:3-8. doi:10.1016/j.ymeth.2014.08.011.

39. Hou L, Zhang X, Wang D, Baccarelli A. Environmental chemical exposures and human epigenetics. *Int J Epidemiol*. 2012;41(1):79-105. doi:10.1093/ije/dyr154.

40. Yang AS, Estécio MRH, Doshi K, Kondo Y, Tajara EH, Issa JPJ. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res*. 2004;32(3):e38-e38.

41. Weisenberger DJ, Campan M, Long TI, et al. Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res*. 2005;33(21):6823-6836.

42. Smit AF, Hubley R, Green P. *RepeatMasker Open-30*. 1996.

43. Jaffe AE, Murakami P, Lee H, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol*. 2012;41(1):200-209.

44. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol*. 1995:289-300.

45. Turunen MP, Aavik E, Ylä-Herttuala S. Epigenetics and atherosclerosis. *Biochim Biophys Acta BBA-Gen Subj*. 2009;1790(9):886-891.

46. Kim M, Long TI, Arakawa K, Wang R, Mimi CY, Laird PW. DNA methylation as a biomarker for cardiovascular disease risk. *PloS One*. 2010;5(3):e9692.

47. Bind M-A, Lepeule J, Zanobetti A, et al. Air pollution and gene-specific methylation in the Normative Aging Study. *Epigenetics*. 2014;9(3):448-458. doi:10.4161/epi.27584.

48. Bollati V, Baccarelli A, Hou L, et al. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer Res*. 2007;67(3):876-880.

49. Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11:587-2105-11-587.

50. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5(10):R80.

51. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2015. https://www.R-project.org/.

52. Dunning MJ, Smith ML, Ritchie ME, Tavaré S. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics*. 2007;23(16):2183-2184. doi:10.1093/bioinformatics/btm311.

53. Smyth GK. Limma: linear models for microarray data. In: Bioinformatics and computational biology solutions using R and Bioconductor. Springer; 2005:397-420.

54. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinforma Oxf Engl*. 2008;24(13):1547-1548.

55. Nettleton JA, Rock CL, Wang Y, Jenny NS, Jacobs DR. Associations between dietary macronutrient intake and plasma lipids demonstrate criterion performance of the Multi-Ethnic Study of Atherosclerosis (MESA) food-frequency questionnaire. *Br J Nutr*. 2009;102(8):1220-1227. doi:10.1017/S0007114509382161.

56. Ainsworth BE, Irwin ML, Addy CL, Whitt MC, Stolarczyk LM. Moderate physical activity patterns of minority women: the Cross-Cultural Activity Participation Study. *J Womens Health Gend Based Med*. 1999;8(6):805-813. doi:10.1089/152460999319129.

57. Hajat A, Diez-Roux AV, Adar SD, et al. Air pollution and individual and neighborhood socioeconomic status: evidence from the Multi-Ethnic Study of Atherosclerosis (MESA). *Environ Health Perspect*. 2013;121(11-12):1325-1333.

58. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinforma Oxf Engl*. 2012;28(6):882-883.

59. StataCorp. *Stata Statistical Software: Release 13.* College Station, TX: StataCorp LP; 2013.

60. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods*. 2012;9(3):215-216. doi:10.1038/nmeth.1906.

61. Adams D, Altucci L, Antonarakis SE, et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol*. 2012;30(3):224-226. doi:10.1038/nbt.2153.

62. Saeed S, Quintin J, Kerstens HHD, et al. Epigenetic programming during monocyte to macrophage differentiation and trained innate immunity. *Science*. 2014;345(6204):1251086. doi:10.1126/science.1251086.

63. Rosenbloom KR, Sloan CA, Malladi VS, et al. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res*. 2013;41(Database issue):D56-63. doi:10.1093/nar/gks1172.

64. Karolchik D, Barber GP, Casper J, et al. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res*. 2014;42(Database issue):D764-D770. doi:10.1093/nar/gkt1168.

65. Dhyani A, Duarte ASS, Machado-Neto JA, Favaro P, Ortega MM, Olalla Saad ST. ANKHD1 regulates cell cycle progression and proliferation in multiple myeloma cells. *FEBS Lett*. 2012;586(24):4311-4318. doi:10.1016/j.febslet.2012.10.037.

66. Machado-Neto JA, Lazarini M, Favaro P, et al. ANKHD1, a novel component of the Hippo signaling pathway, promotes YAP1 activation and cell cycle progression in prostate cancer cells. *Exp Cell Res*. 2014;324(2):137-145. doi:10.1016/j.yexcr.2014.04.004.

67. Machado-Neto JA, Lazarini M, Favaro P, et al. ANKHD1 silencing inhibits Stathmin 1 activity, cell proliferation and migration of leukemia cells. *Biochim Biophys Acta BBA - Mol Cell Res*. 2015;1853(3):583-593. doi:10.1016/j.bbamcr.2014.12.012.

68. Traina F, Favaro PMB, de Souza Medina S, et al. ANKHD1, ankyrin repeat and KH domain containing 1, is overexpressed in acute leukemias and is associated with SHP2 in K562 cells. *Biochim Biophys Acta BBA - Mol Basis Dis*. 2006;1762(9):828-834. doi:10.1016/j.bbadis.2006.07.010.

69. Nourshargh S, Hordijk PL, Sixt M. Breaching multiple barriers: leukocyte motility through venular walls and the interstitium. *Nat Rev Mol Cell Biol*. 2010;11(5):366-378. doi:10.1038/nrm2889.

70. Wang K-C, Nguyen P, Limqueco E, Li Y-S, Chien S. Abstract 427: Flow Regulation of YAP/TAZ in Endothelial Inflammation and Atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2015;35(Suppl 1):A427-A427.

71. Cummings RD, Liu F-T. Galectins. In: Varki A, Cummings RD, Esko JD, et al., eds. *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. http://www.ncbi.nlm.nih.gov/books/NBK1944/. Accessed November 6, 2015.

72. Ozaki K, Inoue K, Sato H, et al. Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-α secretion in vitro. *Nature*. 2004;429(6987):72-75. doi:10.1038/nature02502.

73. Asselbergs FW, Pai JK, Rexrode KM, Hunter DJ, Rimm EB. Effects of lymphotoxin-α gene and galectin-2 gene polymorphisms on inflammatory biomarkers, cellular adhesion molecules and risk of coronary heart disease. *Clin Sci*. 2007;112(5):291-298. doi:10.1042/CS20060200.

74. Lian J, Lian J, Fang P, et al. Association between LGALS2 3279C>T and coronary artery disease: A case-control study and a meta-analysis. *Biomed Rep*. August 2014. doi:10.3892/br.2014.325.

75. van der Laan AM, Schirmer SH, Vries MR de, et al. Galectin-2 expression is dependent on the rs7291467 polymorphism and acts as an inhibitor of arteriogenesis. *Eur Heart J*. 2012;33(9):1076-1084. doi:10.1093/eurheartj/ehr220.

76. Yıldırım C, Vogel DYS, Hollander MR, et al. Galectin-2 Induces a Proinflammatory, Anti-Arteriogenic Phenotype in Monocytes and Macrophages. *PLoS ONE*. 2015;10(4). doi:10.1371/journal.pone.0124347.

77. Ikeda S, Tanaka N, Arai T, Chida K, Muramatsu M, Sawabe M. Polymorphisms of LTA, LGALS2, and PSMA6 genes and coronary atherosclerosis: A pathological study of 1503 consecutive autopsy cases. *Atherosclerosis*. 2012;221(2):458-460. doi:10.1016/j.atherosclerosis.2012.01.003.

78. Sturm A, Lensch M, André S, et al. Human Galectin-2: Novel Inducer of T Cell Apoptosis with Distinct Profile of Caspase Activation. *J Immunol*. 2004;173(6):3825-3837. doi:10.4049/jimmunol.173.6.3825.

79. Gallagher D, Voronova A, Zander MA, et al. Ankrd11 Is a Chromatin Regulator Involved in Autism that Is Essential for Neural Development. *Dev Cell*. 2015;32(1):31-42. doi:10.1016/j.devcel.2014.11.031.

80. Ockeloen CW, Willemsen MH, de Munnik S, et al. Further delineation of the KBG syndrome phenotype caused by ANKRD11 aberrations. *Eur J Hum Genet*. 2015;23(9):1176-1185. doi:10.1038/ejhg.2014.253.

81. Allen JL, Oberdorster G, Morris-Schaffer K, et al. Developmental neurotoxicity of inhaled ambient ultrafine particle air pollution: Parallels with neuropathological and behavioral features of autism and other neurodevelopmental disorders. *NeuroToxicology*. doi:10.1016/j.neuro.2015.12.014.

82. Arking DE, Junttila MJ, Goyette P, et al. Identification of a Sudden Cardiac Death Susceptibility Locus at 2q24.2 through Genome-Wide Association in European Ancestry Individuals. *PLoS Genet*. 2011;7(6). doi:10.1371/journal.pgen.1002158.

83. Wang Z, Song J, Milne TA, et al. Pro Isomerization in MLL1 PHD3-Bromo Cassette Connects H3K4me Readout to CyP33 and HDAC-Mediated Repression. *Cell*. 2010;141(7):1183-1194. doi:10.1016/j.cell.2010.05.016.

84. Morris AR, Bos A, Diosdado B, et al. Alternative Cleavage and Polyadenylation during Colorectal Cancer Development. *Clin Cancer Res*. 2012;18(19):5256-5266. doi:10.1158/1078-0432.CCR-12-0543.

85. Naukkarinen J, Surakka I, Pietiläinen KH, et al. Use of Genome-Wide Expression Data to Mine the “Gray Zone” of GWA Studies Leads to Novel Candidate Obesity Genes. *PLoS Genet*. 2010;6(6). doi:10.1371/journal.pgen.1000976.

86. Peluso M, Bollati V, Munnia A, et al. DNA methylation differences in exposed workers and nearby residents of the Ma Ta Phut industrial estate, Rayong, Thailand. *Int J Epidemiol*. 2012;41(6):1753-1760. doi:10.1093/ije/dys129.

87. Bellavia A, Urch B, Speck M, et al. DNA Hypomethylation, Ambient Particulate Matter, and Increased Blood Pressure: Findings From Controlled Human Exposure Experiments. *J Am Heart Assoc Cardiovasc Cerebrovasc Dis*. 2013;2(3). doi:10.1161/JAHA.113.000212.

88. Jiang R, Jones MJ, Sava F, Kobor MS, Carlsten C. Short-term diesel exhaust inhalation in a controlled human crossover study is associated with changes in DNA methylation of circulating mononuclear cells in asthmatics. *Part Fibre Toxicol*. 2014;11. doi:10.1186/s12989-014-0071-3.

89. Gordon S, Taylor PR. Monocyte and macrophage heterogeneity. *Nat Rev Immunol*. 2005;5(12):953-964. doi:10.1038/nri1733.

90. De Prins S, Koppen G, Jacobs G, et al. Influence of ambient air pollution on global DNA methylation in healthy adults: A seasonal follow-up. *Environ Int*. 2013;59:418-424. doi:10.1016/j.envint.2013.07.007.

91. Levy I, Mihele C, Lu G, Narayan J, Brook JR. Evaluating Multipollutant Exposure and Urban Air Quality: Pollutant Interrelationships, Neighborhood Variability, and Nitrogen Dioxide as a Proxy Pollutant. *Environ Health Perspect*. November 2013. doi:10.1289/ehp.1306518.

92. Kaufman JD. Does Air Pollution Accelerate Progression of Atherosclerosis? *J Am Coll Cardiol*. 2010;56(22):1809-1811. doi:10.1016/j.jacc.2010.05.054.

93. Bind M-A, Lepeule J, Zanobetti A, et al. Air pollution and gene-specific methylation in the Normative Aging Study: Association, effect modification, and mediation analysis. *Epigenetics*. 2014;9(3):0–1.

94. Madrigano J, Baccarelli A, Mittleman MA, et al. Prolonged exposure to particulate pollution, genes associated with glutathione pathways, and DNA methylation in a cohort of older men. *Environ Health Perspect*. 2011;119(7):977.

95. Tarantini L, Bonzini M, Apostoli P, et al. Effects of particulate matter on genomic DNA methylation content and iNOS promoter methylation. *Environ Health Perspect*. 2009;117(2):217.

96. Alexeeff SE, Baccarelli AA, Halonen J, et al. Association between blood pressure and DNA methylation of retrotransposons and pro-inflammatory genes. *Int J Epidemiol*. 2013;42(1):270-280. doi:10.1093/ije/dys220.

97. Krumlauf R. Hox genes in vertebrate development. *Cell*. 1994;78(2):191-201. doi:10.1016/0092-8674(94)90290-9.

98. Abramovich C, Humphries RK. Hox regulation of normal and leukemic hematopoietic stem cells. *Curr Opin Hematol*. 2005;12(3):210-216.

99. Taghon T, Stolz F, Smedt MD, et al. HOX-A10 regulates hematopoietic lineage commitment: evidence for a monocyte-specific transcription factor. *Blood*. 2002;99(4):1197-1204. doi:10.1182/blood.V99.4.1197.

100. Dunn J, Simmons R, Thabet S, Jo H. The role of epigenetics in the endothelial cell shear stress response and atherosclerosis. *Int J Biochem Cell Biol*. 2015;67:167-176. doi:10.1016/j.biocel.2015.05.001.

101. Morgan R, Whiting K. Differential expression of HOX genes upon activation of leukocyte sub-populations. *Int J Hematol*. 2008;87(3):246-249. doi:10.1007/s12185-008-0057-8.

102. Kenmochi N, Suzuki T, Uechi T, et al. The Human Mitochondrial Ribosomal Protein Genes: Mapping of 54 Genes to the Chromosomes and Implications for Human Disorders. *Genomics*. 2001;77(1-2):65-70. doi:10.1006/geno.2001.6622.

103. Edgar AJ, Birks EJ, Yacoub MH, Polak JM. Cloning of Dexamethasone-Induced Transcript. *Am J Respir Cell Mol Biol*. 2001;25(1):119-124. doi:10.1165/ajrcmb.25.1.4417.

104. Wu M-Y, Tsai T-F, Beaudet AL. Deficiency of Rbbp1/Arid4a and Rbbp1l1/Arid4b alters epigenetic modifications and suppresses an imprinting defect in the PWS/AS domain. *Genes Dev*. 2006;20(20):2859-2870. doi:10.1101/gad.1452206.

105. Goldberger N, Walker RC, Kim CH, Winter S, Hunter KW. Inherited variation in miR-290 expression suppresses breast cancer progression by targeting the metastasis susceptibility gene Arid4b. *Cancer Res*. 2013;73(8):2671-2681. doi:10.1158/0008-5472.CAN-12-3513.

106. Wu M-Y, Eldin KW, Beaudet AL. Identification of Chromatin Remodeling Genes Arid4a and Arid4b as Leukemia Suppressor Genes. *JNCI J Natl Cancer Inst*. 2008;100(17):1247-1259. doi:10.1093/jnci/djn253.

107. Wu R-C, Zeng Y, Pan I-W, Wu M-Y. Androgen Receptor Coactivator ARID4B Is Required for the Function of Sertoli Cells in Spermatogenesis. *Mol Endocrinol*. 2015;29(9):1334-1346. doi:10.1210/me.2015-1089.

108. Wurm CA, Neumann D, Lauterbach MA, et al. Nanoscale distribution of mitochondrial import receptor Tom20 is adjusted to cellular conditions and exhibits an inner-cellular gradient. *Proc Natl Acad Sci U S A*. 2011;108(33):13546-13551. doi:10.1073/pnas.1107553108.

109. Graham A. Mitochondrial regulation of macrophage cholesterol homeostasis. *Free Radic Biol Med*. 2015;89:982-992. doi:10.1016/j.freeradbiomed.2015.08.010.

110. Childs KS, Goodbourn S. Identification of novel co-repressor molecules for Interferon Regulatory Factor-2. *Nucleic Acids Res*. 2003;31(12):3016-3026. doi:10.1093/nar/gkg431.

111. Chen H-H, Keyhanian K, Zhou X, et al. IRF2BP2 Reduces Macrophage Inflammation and Susceptibility to Atherosclerosis. *Circ Res*. 2015;117(8):671-683. doi:10.1161/CIRCRESAHA.114.305777.

112. Bloom DE, Cafiero ET, Jané-Llopis E, et al. *The Global Economic Burden of Non-Communicable Disease*. Geneva: World Economic Forum; 2011. http://apps.who.int/medicinedocs/en/d/Js18806en/. Accessed January 16, 2016.

113. Budoff MJ, Hokanson JE, Nasir K, et al. Progression of Coronary Artery Calcium Predicts All-Cause Mortality. *JACC Cardiovasc Imaging*. 2010;3(12):1229-1236. doi:10.1016/j.jcmg.2010.08.018.

114. Budoff MJ, Young R, Lopez VA, et al. Progression of Coronary Calcium and Incident Coronary Heart Disease Events: The Multi-Ethnic Study of Atherosclerosis. *J Am Coll Cardiol*. 2013;61(12):1231-1239. doi:10.1016/j.jacc.2012.12.035.

115. Flowers E, Froelicher ES, Aouizerat BE. Gene-Environment Interactions in Cardiovascular Disease. *Eur J Cardiovasc Nurs J Work Group Cardiovasc Nurs Eur Soc Cardiol*. 2012;11(4):472-478. doi:10.1016/j.ejcnurse.2011.06.001.

116. Carr JJ, Nelson JC, Wong ND, et al. Calcified Coronary Artery Plaque Measurement with Cardiac CT in Population-based Studies: Standardized Protocol of Multi-Ethnic Study of Atherosclerosis (MESA) and Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Radiology*. 2005;234(1):35-43. doi:10.1148/radiol.2341040439.

117. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte Jr. M, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol*. 1990;15(4):827-832. doi:10.1016/0735-1097(90)90282-T.

118. Budoff MJ, Nasir K, McClelland RL, et al. Coronary calcium predicts events better with absolute calcium scores than age-gender-race percentiles – The Multi-Ethnic Study of

Atherosclerosis (MESA). *J Am Coll Cardiol*. 2009;53(4):345-352. doi:10.1016/j.jacc.2008.07.072.

119. Zeileis A. Econometric computing with HC and HAC covariance matrix estimators. 2004. http://epub.wu.ac.at/520/. Accessed January 26, 2016.

120. Gassett AJ, Sheppard L, McClelland RL, et al. Risk Factors for Long-Term Coronary Artery Calcium Progression in the Multi-Ethnic Study of Atherosclerosis. *J Am Heart Assoc Cardiovasc Cerebrovasc Dis*. 2015;4(8). doi:10.1161/JAHA.114.001726.

121. Kronmal RA, McClelland RL, Detrano R, et al. Risk Factors for the Progression of Coronary Artery Calcification in Asymptomatic Subjects Results From the Multi-Ethnic Study of Atherosclerosis (MESA). *Circulation*. 2007;115(21):2722-2730. doi:10.1161/CIRCULATIONAHA.106.674143.

122. Xiao Y, Su X, Huang W, et al. Role of S-adenosylhomocysteine in cardiovascular disease and its potential epigenetic mechanism. *Int J Biochem Cell Biol*. 2015;67:158-166. doi:10.1016/j.biocel.2015.06.015.

123. Liu C, Wang Q, Guo H, et al. Plasma S-Adenosylhomocysteine Is a Better Biomarker of Atherosclerosis Than Homocysteine in Apolipoprotein E-Deficient Mice Fed High Dietary Methionine. *J Nutr*. 2008;138(2):311-315.

124. Kaufman JD, Adar SD, Barr RG, et al. Association between air pollution and coronary artery calcification in a longitudinal cohort study within six metropolitan areas in the USA (The Multi-Ethnic Study of Atherosclerosis and Air Pollution). *The Lancet*. 2016;In press.

125. Moolgavkar SH. A review and critique of the EPA's rationale for a fine particle standard. *Regul Toxicol Pharmacol*. 2005;42(1):123-144. doi:10.1016/j.yrtph.2005.02.003.

126. Murphy AJ, Dragoljevic D, Tall AR. Cholesterol Efflux Pathways Regulate Myelopoiesis: A Potential Link to Altered Macrophage Function in Atherosclerosis. *Front Immunol*. 2014;5. doi:10.3389/fimmu.2014.00490.

127. Gudmundsson EF, Gudnason V, Sigurdsson S, Launer LJ, Harris TB, Aspelund T. Coronary artery calcium distributions in older persons in the AGES-Reykjavik study. *Eur J Epidemiol*. 2012;27(9):673-687. doi:10.1007/s10654-012-9730-6.