On Aging: Analyses of Long-Term Fine Particulate Air Pollution Exposure, Genetic Variants, and Blood DNA Methylation Age in the Elderly

On Aging: Analyses of Long-term Fine Particulate Air Pollution Exposure, Genetic Variants, and Blood DNA Methylation Age in the Elderly

A dissertation presented

by

Jamaji Chilaka Nwanaji-Enwerem

to

The Committee on Higher Degrees in Biological Sciences in Public Health

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological Sciences in Public Health

Harvard University

Cambridge, Massachusetts

November, 2017

Dissertation Advisors: Andrea A. Baccarelli, MD, PhD & Marc G. Weisskopf, PhD, ScD

Jamaji Chilaka Nwanaji-Enwerem

On Aging: Analyses of Long-term Fine Particulate Air Pollution Exposure, Genetic Variants, and Blood DNA Methylation Age in the Elderly

Abstract

Human aging is often accompanied by the development of chronic disease. Research has identified molecular processes that are shared by aging-related diseases, and it is widely believed that pre-clinical changes in these aging-related molecular processes (*i.e.* measures of "biological age") may be more informative of morbidity and mortality risks than simple chronological age. DNA methylation age (DNAm-age) is a DNA methylation based predictor of chronological age and a novel measure of biological age. Studies have demonstrated associations of DNAm-age with a host of aging-related health outcomes including all-cause mortality, frailty, cancer, and Parkinson's disease. However, very few studies have examined DNAm-age relationships with aging risk factors.

Fine particulate air pollution ($PM_{2.5}$) is a well-documented aging risk factor and is considered the world's largest singular environmental health risk. This body of work utilized multivariate linear mixed effects models and a well-established aging cohort, the United States Veterans Affairs Normative Aging Study (NAS), to examine the relationship of long-term $PM_{2.5}$ exposure levels with DNAm-age. After determining the direct relationship of $PM_{2.5}$ with DNAm-age in the NAS, we determined which of five major $PM_{2.5}$ component species (ammonium, elemental carbon, organic carbon, sulfate, and nitrate) were most associated with DNAm-age. Finally, we examined if normal genetic variation in aging-related physiological processes (endothelial function, metal processing, oxidative stress, mitochondrial genome physiology, and microRNA processing) impacted the relationships of $PM_{2.5}$ and its component species with DNAm-age.

We found that $PM_{2.5}$ was significantly, positively associated with DNAm-age and that sulfate and ammonium were the component species most associated with DNAm-age. Moreover, endothelial function, mitochondrial genome, and microRNA processing variants significantly modified the association of $PM_{2.5}$ with DNAm-age. DNAm-age was also significantly associated with a number of

serum measures related to these effect modifiers including mitochondrial DNA copy number, intercellular adhesion molecule (ICAM), and vascular cell adhesion molecule (VCAM).

In all, our studies demonstrate a novel association of $PM_{2.5}$ with DNAm-age. Our studies also suggest that DNAm-age has robust relationships with endothelial function, mitochondrial physiology, and miRNA processing – all of which are processes known to play a role in aging-related diseases. Still, future studies will be necessary to further understand what DNAm-age represents and how it can best be used as a biomarker.

Table of Contents

Dedication

This dissertation is dedicated to God, through whom all things are possible. This dissertation is also dedicated to my parents (Pamaji and Chioma Nwanaji-Enwerem), and grandparents (Ezinna Vincent Nwanaji-Enwerem [late], Ezinne Elfreda Nwanaji-Enwerem [late], Chief Engineer Bath Achilike, and Lolo Elfreda Achilike). You have been my foundation in life and the completion of this degree in part honors the sacrifices that you have made for your children. This work is also dedicated to the people of Umuozu, Nigeria and Concord, North Carolina. As I continue my journey, I remain aware of my roots and the environments that shaped my early years.

Acknowledgements

As I reflected on what to write in these acknowledgements, my conscious was immediately drawn to the Bible parable of the talents (Matthew 25:14 – 29):

> *"It will be as when a man who was going on a journey called in his servants and entrusted his possessions to them. To one he gave five talents; to another, two; to a third, one—to each according to his ability. Then he went away. Immediately the one who received five talents went and traded with them, and made another five. Likewise, the one who received two made another two. But the man who received one went off and dug a hole in the ground and buried his master's money. After a long time the master of those servants came back and settled accounts with them. The one who had received five talents came forward bringing the additional five. He said, 'Master, you gave me five talents. See, I have made five more.' His master said to him, 'Well done, my good and faithful servant. Since you were faithful in small matters, I will give you great responsibilities. Come, share your master's joy.' [Then] the one who had received two talents also came forward and said, 'Master, you gave me two talents. See, I have made two more.' His master said to him, 'Well done, my good and faithful servant. Since you were faithful in small matters, I will give you great responsibilities. Come, share your master's joy.' Then the one who had received the one talent came forward and said, 'Master, I knew you were a demanding person, harvesting where you did not plant and gathering where you did not scatter; so out of fear I went off and buried your talent in the ground. Here it is back.' His master said to him in reply, 'You wicked, lazy servant! So you knew that I harvest where I did not plant and gather where I did not scatter? Should you not then have put my money in the bank so that I could have got it back with interest on my return? Now then! Take the talent from him and give it to the one with ten. For to everyone who has, more will be given and he will grow rich; but from the one who has not, even what he has will be taken away."*

The belief that we are not simply given gifts in our lives to be buried away, but rather that we are obligated to share our gifts with the world has resonated with me for much of my life. Throughout my educational career and doctoral training, there have been countless instances when my family, mentors, colleagues, and friends have gifted me with their time, attention, and expertise. Unbeknownst to them, simple conversations, brainstorming sessions, holiday dinners, vacations, and causal outings have meant all the difference to me on this journey. For their contributions, I say thank you and strive to make myself available and willing to help those that I encounter throughout life. My sincerest gratitude goes to my advisors Dr. Andrea Baccarelli and Dr. Marc Weisskopf. I will be ever-appreciative of your allowing me to pursue my research ideas and your guidance in shaping my research questions. A very special thank you also goes to my dissertation advisory and defense committees: Dr. Edward Boyer, Dr. Immaculata De Vivo, Dr. Bernardo Lemos, Dr.

Quan Lu, Dr. Frank Sacks, and Dr. Marianne Wessling-Resnick. Each of you have offered me critical advice at different junctures of my doctoral training and have been stellar examples of scholarly leadership. A warm thank you also goes to the many other professors, faculty, and staff members at Harvard Medical School and Harvard T.H. Chan School of Public Health that have had a valuable impact on my career development.

I have also been privileged to have the support of a loving family. To my parents Pamaji and Chioma Nwanaji-Enwerem, thank you for raising me to be a moral, compassionate, and hardworking individual. To my siblings Onyemaechi, Uzoji, Ugoji, and Ezeji, your support has and will always be invaluable to me. Lastly, but surely not least, thank you to God for all that I have been given. I hope that I do these opportunities justice and can positively impact the world to your pleasing. Again, thank you all.

Chapter 1:

Introduction

1.1. A Synopsis of Fine Particle ($PM_{2.5}$) Exposure and Human Health

In 1993, a prospective study involving 8111 residents of six United States cities reported important associations of air pollutant levels with premature mortality. Even after controlling for smoking and other risk factors, the study suggested that individuals living in highly polluted cities were at a greater risk of premature mortality and that these individuals were dying from cardiopulmonary disease and lung cancer[1]. Since this landmark 1993 paper, reanalysis of the original study data has confirmed the quality of the initial findings. Moreover, it has become widely accepted that fine particulate air pollution ($PM_{2.5}$) is a major global health risk[2]. Continued follow-up of the six city study participants has demonstrated that reductions in the concentrations of ambient $PM_{2.5}$ – due to legislation like the Clean Air Act of 1970 and amendments of 1977 and 1990 – are significantly associated with declines in cardiopulmonary mortality[3].

Researchers have continued to extensively study the relationships of ambient $PM_{2.5}$ exposure with human health. Much of this work has been in observational studies which have reported consistent relationships of $PM_{2.5}$ exposure with adverse cardiopulmonary and cardiometabolic health outcomes[4,5]. Still, emerging data from observational studies has also revealed novel associations of $PM_{2.5}$ exposure with previously unconsidered health outcomes like dementia, Parkinson's disease, and chronic kidney disease[6-8]. Experimental toxicological studies in human cells and animal models have provided supportive evidence that $PM_{2.5}$ exposure can result in health endpoints comparable to those described in human observational studies[9,10]. From this experimental work, we have come to understand that fine particles are particularly dangerous because they are readily inhalable and can penetrate into the lung's alveolar gas exchange regions. In addition to respiratory-related consequences, $PM_{2.5}$ can traverse the respiratory barrier, enter the circulatory system and cause systemic sequelae[11]. Experimental studies have also demonstrated that $PM_{2.5}$ exposure impacts biological process (*e.g.* inflammation and oxidative stress) and that disturbances in these processes are likely to be among the molecular mechanisms that manifest themselves in $PM_{2.5}$-related disease[12,13].

Despite these compelling findings, there are studies that still report no relationships of $PM_{2.5}$ mass with health endpoints including relationships that have been previously well-described. A leading theory to

explain these seemingly contradictory results is based on the premise that the levels and the composition of $PM_{2.5}$ mass are known to vary both spatially and temporally. In fact, the chemical composition of $PM_{2.5}$ is often dependent on particle sources. For instance, primary particles emitted from coal combustion are enriched with arsenic and selenium while those from oil combustion are enriched with nickel and vanadium[14]. Particles from soil sources are enriched for crustal elements like aluminum and silicon while sulfates, nitrates, and organic compounds are usually secondary $PM_{2.5}$ component species from atmospheric photochemical reactions[15]. A study of 45 school children living Southern California with persistent asthma, demonstrated that associations of $PM_{2.5}$ with airway inflammation in asthmatics is missed if only relationships with total $PM_{2.5}$ mass are examined. However, when associations with particular component species (*e.g.* elemental carbon and nitric oxide) were examined, these associations were robust in magnitude and statistically significant[16]. Likewise, another more recently study, based on a population of individuals under the age of 20 living in the Shalu district of Taiwan, found the risk of asthma outpatient visits to be associated with carbon and nitrate $PM_{2.5}$ component species[17]. Differential $PM_{2.5}$ component species toxicity has also been reported in adult populations with respect to many health outcomes including risk of myocardial infarction and pre-term birth[18,19]. Ultimately, these data demonstrate the importance of component species analyses alongside total $PM_{2.5}$ mass analyses especially when attempting to identify causal toxic pollutants[20].

1.2. The Importance of Continued $PM_{2.5}$ Research

It is important to re-emphasize that $PM_{2.5}$ research has informed the implementation of new air quality standards that have helped to save many lives and reduced health risks across the world[21,22]. A critical component of defining meaningful current and future air quality standards will be work not only examining relationships of $PM_{2.5}$ exposure with disease, but work that examines relationships of $PM_{2.5}$ exposure with molecular processes that may precede the manifestation of clinical disease. Understanding relationships with molecular processes like biological aging may also inform additional interventions or therapeutics for individuals currently living in areas that do not adhere to $PM_{2.5}$ air quality standards.

1.3. $PM_{2.5}$ Exposure and Aging

Human aging is often accompanied by the development of multiple chronic conditions including dementia, metabolic syndrome, and cardiovascular disease. Since many of these chronic conditions have been independently associated with $PM_{2.5}$ exposure, exploring relationships of $PM_{2.5}$ with aging remains one promising strategy to further understand the adverse impact of $PM_{2.5}$ on human health. One large study of approximately 28 million adults across the United States found that every interquartile range (4.19 $\mu g/m^3$) increase in $PM_{2.5}$ exposure was significantly associated with a lower probability of exceptional aging, which was defined as living to the age range of 85-94, and a lower probability of becoming a centenarian, living to the age range of at least 100 years[23]. Still, the mechanisms that explain how $PM_{2.5}$ exposure impacts the aging process are not well understood. One major hurdle in addressing this research gap is the variability in how aging is defined in current research. As demonstrated by the previously mentioned exceptional aging study, many studies define aging as the passing of time or chronological age. However, research has identified a number of molecular markers that outperform chronological age in representing morbidity and mortality risk. These molecular markers include telomere length, mitochondrial genome abundance, measures of cellular senescence, stem cell exhaustion, and epigenetic alterations. Furthermore, human observational studies and animal experimental studies have reported relationships of $PM_{2.5}$ exposure with a number of these markers including telomere length, cellular senescence, and cell atrophy[24-26]. Some of the more novel associations of $PM_{2.5}$ exposure with aging markers have involved epigenetic alterations, particularly DNA methylation.

1.4. $PM_{2.5}$ Exposure and DNA Methylation

Epigenetic modifications are alterations to DNA, RNA, or proteins that result in changes in the regulation or function of these molecules. Although these molecules may be modified, their respective nucleic acid or protein sequences remain unchanged. Epigenetic changes are one major way that an organism's internal and external environments can influence their cellular and phenotypic traits. DNA

methylation is one of the three major epigenetic modifications – histone modifications and non-coding RNAs are the other two – and it involves the addition of methyl residues to DNA nucleotides. DNA methylation is catalyzed by enzymes called DNA methyltransferases (DNMTs). Mammals have three major DNMTs: DNMT1, DNMT3A, and DNMT3B. DNMT3A and DNMT3B are involved in *de novo* methylation and are most active during embryogenesis and early life[27]. DNMT1 is the most abundant methyltransferase and it remains active throughout adulthood. DNMT1 is active in *maintenance methylation* which involves maintaining methylation patterns throughout an organism's life. When DNA is replicated, if the template strand is methylated, DNMT1 methylates the newly synthesized strand accordingly. In most organisms, S-adenosylmethionine (SAM) serves as a methyl donor in the methylation processes[28-30].

In mammals, DNA methylation almost exclusively occurs on cytosine residues that are followed by guanine residues (CpG sites). In fact, approximately 80% of CpGs in mammalian genomes are methylated[31]. DNMTs transfer the methyl group from SAM to the 5 position of cytosine residues to form 5-methylcytosine (5mC). Many CpG sites exist in clusters near transcriptionally integral regions of the genome like promoters and enhancers. Clusters of CpGs near transcriptional start sites are called CpG islands. Methylation of CpG islands usually leads to condensed chromatin, delayed replication, and inhibition of transcription initiation[32,33]. By covalently altering the structure of cytosine residues with methyl groups, the interactions of chromatin proteins and transcription factors with these areas of DNA are now altered. Since these interactions are critical for transcription, altering them is how methylation affects transcription and regulates a number of biological processes including development, genomic imprinting and inactivation of X chromosomes. It is also important to note that DNA can be demethylated. This process can occur passively or via enzymes known as Ten-eleven translocation (TET) enzymes. DNA demethylation may also involve the formation of additional DNA modifications like 5-hydroxymethylcyotsine (5hmC)[34].

Both experimental and observational studies have revealed relationships of $PM_{2.5}$ and its component species with DNA methylation. Moreover, the results of these studies have offered much insight

into the pathological effects $PM_{2.5}$[35]. For instance, one epigenome wide study, utilizing the peripheral blood leukocytes of individuals from three independent United States or German cohorts, identified 12 CpG sites with methylation levels associated with various time windows of short-term $PM_{2.5}$ exposure. Methylation at 9 of the sites was positively associated with $PM_{2.5}$ levels while methylation at the remaining 6 sites was negatively associated with $PM_{2.5}$[36]. In an epigenome-wide study made up of samples from 1207 individuals living in Los Angeles and Chicago, methylation at 5 CpG sites was associated with long-term $PM_{2.5}$ exposure levels[37]. Not one of these 5 sites was among the previously identified 12 sites. The difference in these findings could be due a number of differences between the two studies including the length of $PM_{2.5}$ exposure or even the type of tissue that methylation was measured in. The latter study used methylation from CD14+ purified monocytes while the first used a mixture of blood leukocytes. Studies have also examined relationships of $PM_{2.5}$ component species with DNA methylation. In peripheral blood monocytes, Dai *et. al* (2017) examined the relationships of long-term, one year $PM_{2.5}$ component exposure (Al, Ca, Cu, Fe, K, Na, Ni, S, Si, V, and Zn) with methylation in a cohort of community dwelling older men. These authors found 20 CpG sites that were significantly associated with Fe, 8 that were associated with Ni, and 1 that was associated with V[38]. Again, none of these component species-associated CpGs overlapped with the previously reported long-term or short-term total $PM_{2.5}$-associated CpG sites. Together, these and other existing studies highlight that differences in research methods or particle composition can influence the results of $PM_{2.5}$ epigenome-wide studies. Moreover, the biological applicability of findings from any epigenome-wide study should always be carefully considered.

1.5. DNA Methylation and Aging

In addition to its relationships with $PM_{2.5}$ exposure levels, DNA methylation also has relationships with disease and physiological processes. One of the most well-studied of these relationships is the relationship of DNA methylation with aging. Although some site-specific hypermethylation is observed, mammalian aging is generally associated with DNA CpG hypomethylation. More specifically, sites like promoter-associated CpG islands, which normally have low baseline DNA methylation, tend to become

hypermethylated with age. On the other hand, sites with high baseline methylation, like intergenic non-island CpGs, tend to become hypomethylated with age. Since most CpGs in the genome are located outside of CpG islands –and are thus highly methylated at baseline – this translates to an overall loss of DNA methylation with aging[39]. This aging-related change in DNA methylation is called epigenetic drift and to some extent is thought to be due to a decline in DNMT1 that occurs with aging[40]. Even though DNA methylation patterns are highly divergent in different tissues, this phenomenon of epigenetic drift has been replicated in many tissues and is believed to be a general total-organism phenomenon[41]. In addition to the general epigenetic drift patterns of global hypomethylation and local hypermethylation, research has identified specific sites in the genome that are so highly associated with aging that they can be used to predict chronological age[42]. A number of these "epigenetic clock" sites and measures have been described, but one particular measure developed by Steve Horvath, PhD, has demonstrated utility across individuals and tissue/cell types. Hereby, we refer to this Horvath measure as DNA methylation age (DNAm-age)[43].

1.6. DNA Methylation Age (DNAm-Age)

DNA methylation age (DNAm-age) is a measure that arose out of the hypothesis that particular sites in the genome experienced aging-related changes in DNA methylation that were progressive and common across tissues and individuals. Horvath (2013) developed the measure as a predictor of age using data from 82 Illumina DNA methylation array datasets that consisted of 7844 non-cancer human samples from 51 healthy tissues or cell types. 39 of the datasets were used to train the age predictor, 31 were used to validate the measure, and the others were used for additional analyses. Beginning with 21,369 CpGs shared between the Illumina 27K and 450K platforms, a transformed version chronological age was regressed on the CpGs using a penalized regression elastic net model. From these 21,369 CpGs, 353 were selected by the elastic net. 193 CpGs were hypermethylated with age and 160 were hypomethylated. The hypermethylated CpGs were more likely to be in poised promoters and were over-represented near Polycomb-group target genes, which are known to play a critical role during embryonic development. The 160 hypomethylated CpGs were more likely to be in weak promoters or strong enhancers and were over-

represented in CpG shores. Pathway analysis of the genes that co-locate with the 353 CpGs revealed enrichment for the biological processes of organism tissue development; cellular growth and proliferation; cell death and survival; and cancer[43].

The weighted average derived from the regression coefficients of each of the 353 CpG sites was then used to calculate one measure of age prediction, DNAm-age. The measure performed well in the test datasets (age correlation = 0.96, error = 3.6 years) regardless of if the dataset was from mixed tissues (*e.g.* whole blood) or from an individual cell type (*e.g.* CD14+ monocytes). Moreover, early analyses demonstrated that DNAm-age possessed a number of other properties beyond its ability to accurately predict age in a multitude of human tissues including blood, brain, saliva, skin, and bone. First, DNAm-age appeared to be reflecting some intrinsic measurement of the methylome because it able was able to track chronological age in non-proliferative tissues (*e.g.* neurons) while also assigning similar ages to more short-lived tissues (*e.g.* blood cells). Yet, there were some tissues where DNAm-age consistently performed poorly as an age predictor: breast tissue, dermal fibroblasts, uterine endometrium, skeletal muscle, and heart tissue. Heart tissue tended to have a lower DNAm-age than expected while the other mentioned tissues had higher DNAm-ages than expected. However, it is thought that this poor performance is due to some unique property of the methylome in these tissues rather than an error in the metric itself. Second, the DNAm-age of induced pluripotent stem cells and embryonic stem cells was found to be near zero but increased as these cells were passaged following cell culture. Third, DNAm-age could be calculated and accurately perform in chimpanzees. This suggested that the measure was somewhat evolutionarily conserved. Fourth, the number of somatic mutations in a cancer sample tended to be inversely correlated with the sample's DNAm-age even though DNAm-age had very weak relationships with tumor grade and stage.

Since the initial publication that described DNAm-age and its intrinsic properties, researchers have published findings that further suggest that DNAm-age is not simply a predictor of chronological age. Most of the emerging evidence suggests that DNAm-age captures risks associated with the molecular aging process and represents a novel measure of biological aging. Some of the most compelling evidence for this theory has come from studies demonstrating associations of DNAm-age with all-cause mortality. The

largest of such studies is a meta-analysis of 13 population-based cohorts amounting to 13,089 individuals. This study found that increases in DNAm-age were predictive of mortality even when accounting for chronological age and additional risk factors like cancer, coronary artery disease, hypertension, type 2 diabetes, race, sex, physical activity, and body mass index[44]. Additional evidence comes from studies that report that centenarians (long-lived individuals) have DNAm-ages that are lower than their chronological ages[45,46]. Following this trend, aging-related lifestyle factors and health conditions have also been associated with DNAm-age. Negative lifestyle factors like exposure violence and personal life stressors have been associated with increased DNAm-age while more positive lifestyle factors like dietary fish intake or blood carotenoid levels (an indicator of fruit and vegetable intake) are correlated with decreases in DNAm-age[47-49]. Menopause, Huntington's disease, frailty, and Alzheimer's disease-related cognitive decline have all been associated with increased DNAm-age[50-53].

Beyond these associations from observational studies, researchers are also beginning to explore molecular mechanisms related to DNAm-age. In his initial publication, Horvath reported the results from a number of simple studies that ultimately resulted in his epigenetic maintenance system (EMS) hypothesis. The EMS hypothesis states that DNAm-age may represent the cumulative work performed by a yet to be defined epigenetic maintenance system, which plays a role in maintaining epigenetic homeostasis. Any event or exposure that disrupts the epigenome will result in more work being done by the EMS to return the epigenome to homeostasis. The output of this additional work is a higher DNAm-age. In line with his hypothesis, decreases in DNAm-age could be interpreted as epigenome stability or a disruption of the EMS' ability to do work. Although Horvath hypothesized that methyltransferases would be a component of the EMS, no studies have explicitly looked to identify DNAm-age EMS components. One of the first studies to even explore the mechanistic underpinnings of DNAm-age used human cell lines to examine the relationships of three major forms of cellular senescence (DNA damage, oncogene-induced, and replicative) with DNAm-age. These researchers found that replicative senescence and oncogene-induced senescence were associated with increased DNAm-age, but DNA damage senescence was not. As part of their studies, the researchers also demonstrated that DNAm-age was independent of telomere length[54]. A

second study, that begins to explore the mechanistic relationships of DNAm-age, builds upon the widely accepted phenomenon in aging research that caloric restriction extends lifespan in model organisms. It was unknown whether this phenomenon and the biology associated with caloric restriction was at all related to DNAm-age, until a published report demonstrated that 30% caloric restriction since the age of 7-14 years in 22-30 year-old rhesus monkeys can impact DNAm-age. Specifically, caloric restricted monkeys had a blood DNAm-age that was on average 7 years younger than their chronological age when compared to ad libitum-fed controls[55].

1.7. Studying $PM_{2.5}$ and DNAm-age Relationships

The emerging research involving DNAm-age has inspired the pursuit of studies that examine the relationships of DNAm-age with $PM_{2.5}$, an aging and environmental health and risk factor. The work presented in this dissertation is intended to contribute the growing body of research aimed at providing a better understanding of how long-term $PM_{2.5}$ exposure can impact human health.

1.8. References

1 Dockery, D. W. *et al.* An association between air pollution and mortality in six U.S. cities. *N Engl J Med* **329**, 1753-1759, doi:10.1056/NEJM199312093292401 (1993).

2 World Health, O. Burden of disease from household air pollution for 2012. *WHO, Geneva* (2014).

3 Laden, F., Schwartz, J., Speizer, F. E. & Dockery, D. W. Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities study. *Am J Respir Crit Care Med* **173**, 667-672, doi:10.1164/rccm.200503-443OC (2006).

4 Lippmann, M. Toxicological and epidemiological studies of cardiovascular effects of ambient air fine particulate matter (PM2.5) and its chemical components: coherence and public health implications. *Crit Rev Toxicol* **44**, 299-347, doi:10.3109/10408444.2013.861796 (2014).

5 Brook, R. D., Newby, D. E. & Rajagopalan, S. Air Pollution and Cardiometabolic Disease: An Update and Call for Clinical Trials. *Am J Hypertens*, doi:10.1093/ajh/hpx109 (2017).

6 Bowe, B. *et al.* Particulate Matter Air Pollution and the Risk of Incident CKD and Progression to ESRD. *J Am Soc Nephrol*, doi:10.1681/ASN.2017030253 (2017).

7 Chen, H. *et al.* Exposure to ambient air pollution and the incidence of dementia: A population-based cohort study. *Environ Int* **108**, 271-277, doi:10.1016/j.envint.2017.08.020 (2017).

8 Palacios, N. Air pollution and Parkinson's disease - evidence and future directions. *Rev Environ Health*, doi:10.1515/reveh-2017-0009 (2017).

9 Haberzettl, P., O'Toole, T. E., Bhatnagar, A. & Conklin, D. J. Exposure to Fine Particulate Air Pollution Causes Vascular Insulin Resistance by Inducing Pulmonary Oxidative Stress. *Environ Health Perspect* **124**, 1830-1839, doi:10.1289/EHP212 (2016).

10 Weldy, C. S. *et al.* In utero and early life exposure to diesel exhaust air pollution increases adult susceptibility to heart failure in mice. *Part Fibre Toxicol* **10**, 59, doi:10.1186/1743-8977-10-59 (2013).

11 Feng, S., Gao, D., Liao, F., Zhou, F. & Wang, X. The health effects of ambient PM2.5 and potential mechanisms. *Ecotoxicol Environ Saf* **128**, 67-74, doi:10.1016/j.ecoenv.2016.01.030 (2016).

12 Liu, Q. *et al.* Particulate matter 2.5 regulates lipid synthesis and inflammatory cytokine production in human SZ95 sebocytes. *Int J Mol Med* **40**, 1029-1036, doi:10.3892/ijmm.2017.3109 (2017).

13 He, M. *et al.* PM2.5-induced lung inflammation in mice: Differences of inflammatory response in macrophages and type II alveolar cells. *J Appl Toxicol* **37**, 1203-1218, doi:10.1002/jat.3482 (2017).

14 Kim, E., Hopke, P. K., Kenski, D. M. & Koerber, M. Sources of fine particles in a rural midwestern U.S. area. *Environ Sci Technol* **39**, 4953-4960 (2005).

15 Adar, S. D., Filigrana, P. A., Clements, N. & Peel, J. L. Ambient Coarse Particulate Matter and Human Health: A Systematic Review and Meta-Analysis. *Curr Environ Health Rep* **1**, 258-274, doi:10.1007/s40572-014-0022-z (2014).

16 Delfino, R. J. *et al.* Personal and ambient air pollution is associated with increased exhaled nitric oxide in children with asthma. *Environ Health Perspect* **114**, 1736-1743 (2006).

17 Jung, C. R. *et al.* PM2.5 components and outpatient visits for asthma: A time-stratified case-crossover study in a suburban area. *Environ Pollut* **231**, 1085-1092, doi:10.1016/j.envpol.2017.08.102 (2017).

18 Rich, D. Q. *et al.* The triggering of myocardial infarction by fine particles is enhanced when particles are enriched in secondary species. *Environ Sci Technol* **47**, 9414-9423, doi:10.1021/es4027248 (2013).

19 Rappazzo, K. M., Daniels, J. L., Messer, L. C., Poole, C. & Lobdell, D. T. Exposure to Elemental Carbon, Organic Carbon, Nitrate, and Sulfate Fractions of Fine Particulate Matter and Risk of Preterm Birth in New Jersey, Ohio, and Pennsylvania (2000-2005). *Environ Health Perspect* **123**, 1059-1065, doi:10.1289/ehp.1408953 (2015).

20 Grahame, T. J. & Schlesinger, R. B. Health effects of airborne particulate matter: do we know enough to consider regulating specific particle types or sources? *Inhal Toxicol* **19**, 457-481, doi:10.1080/08958370701382220 (2007).

21 Apte, J. S., Marshall, J. D., Cohen, A. J. & Brauer, M. Addressing Global Mortality from Ambient PM2.5. *Environ. Sci. Technol.* **49**, 8057-8066, doi:10.1021/acs.est.5b01236 (2015).

22 Fann, N., Kim, S. Y., Olives, C. & Sheppard, L. Estimated Changes in Life Expectancy and Adult Mortality Resulting from Declining PM2.5 Exposures in the Contiguous United States: 1980-2010. *Environ Health Perspect* **125**, 097003, doi:10.1289/EHP507 (2017).

23 Baccarelli, A. A. *et al.* Particulate Air Pollution, Exceptional Aging, and Rates of Centenarians: A Nationwide Analysis of the United States, 1980-2010. *Environ Health Perspect* **124**, 1744-1750, doi:10.1289/EHP197 (2016).

24 Shan, M. *et al.* A feasibility study of the association of exposure to biomass smoke with vascular function, inflammation, and cellular aging. *Environ Res* **135**, 165-172, doi:10.1016/j.envres.2014.09.006 (2014).

25 Gao, Z. X. *et al.* Assessment of DNA Damage and Cell Senescence in Corneal Epithelial Cells Exposed to Airborne Particulate Matter (PM2.5) Collected in Guangzhou, China. *Invest Ophthalmol Vis Sci* **57**, 3093-3102, doi:10.1167/iovs.15-18839 (2016).

26 Woodward, N. C. *et al.* Traffic-related air pollution impact on mouse brain accelerates myelin and neuritic aging changes with specificity for CA1 neurons. *Neurobiol Aging* **53**, 48-58, doi:10.1016/j.neurobiolaging.2017.01.007 (2017).

27 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).

28 <J. Biol. Chem.-1951-Cantoni-745-54.pdf>.

29 <J. Biol. Chem.-1953-Cantoni-403-16.pdf>.

30 Santamaria, E. *et al.* Functional proteomics of nonalcoholic steatohepatitis: mitochondrial proteins as targets of S-adenosylmethionine. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 3065-3070, doi:10.1073/pnas.0536625100 (2003).

31 Jabbari, K. & Bernardi, G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* **333**, 143-149, doi:10.1016/j.gene.2004.02.043 (2004).

32 Baylin, S. B., Herman, J. G., Graff, J. R., Vertino, P. M. & Issa, J. P. Alterations in DNA methylation: a fundamental aspect of neoplasia. *Advances in cancer research* **72**, 141-196 (1998).

33 Jones, P. A. & Laird, P. W. Cancer epigenetics comes of age. *Nature genetics* **21**, 163-167, doi:10.1038/5947 (1999).

34 Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet*, doi:10.1038/nrg.2017.33 (2017).

35 Leclercq, B. *et al.* Genetic and epigenetic alterations in normal and sensitive COPD-diseased human bronchial epithelial cells repeatedly exposed to air pollution-derived PM2.5. *Environ Pollut* **230**, 163-177, doi:10.1016/j.envpol.2017.06.028 (2017).

36 Panni, T. *et al.* 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* **124**, 983-990, doi:10.1289/ehp.1509966 (2016).

37 Chi, G. C. *et al.* Long-term outdoor air pollution and DNA methylation in circulating monocytes: results from the Multi-Ethnic Study of Atherosclerosis (MESA). *Environ Health* **15**, 119, doi:10.1186/s12940-016-0202-4 (2016).

38 Dai, L. *et al.* Differential DNA methylation and PM2.5 species in a 450 K epigenome-wide association study. *Epigenetics*, 1-10, doi:10.1080/15592294.2016.1271853 (2016).

39 Jones, M. J., Goodman, S. J. & Kobor, M. S. DNA methylation and healthy human aging. *Aging Cell* **14**, 924-932, doi:10.1111/acel.12349 (2015).

40 Pal, S. & Tyler, J. K. Epigenetics and aging. *Sci Adv* **2**, e1600584, doi:10.1126/sciadv.1600584 (2016).

41 Jung, M. & Pfeifer, G. P. Aging and DNA methylation. *BMC Biol* **13**, 7, doi:10.1186/s12915-015-0118-4 (2015).

42 Jylhava, J., Pedersen, N. L. & Hagg, S. Biological Age Predictors. *EBioMedicine* **21**, 29-36, doi:10.1016/j.ebiom.2017.03.046 (2017).

43 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).

44 Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* **8**, 1844-1865, doi:10.18632/aging.101020 (2016).

45 Horvath, S. *et al.* Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. *Aging (Albany NY)* **7**, 1159-1170, doi:10.18632/aging.100861 (2015).

46 Armstrong, N. J. *et al.* Aging, exceptional longevity and comparisons of the Hannum and Horvath epigenetic clocks. *Epigenomics* **9**, 689-700, doi:10.2217/epi-2016-0179 (2017).

47 Jovanovic, T. *et al.* Exposure to Violence Accelerates Epigenetic Aging in Children. *Sci Rep* **7**, 8962, doi:10.1038/s41598-017-09235-9 (2017).

48 Zannas, A. S. *et al.* Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome Biol* **16**, 266, doi:10.1186/s13059-015-0828-5 (2015).

49 Quach, A. *et al.* Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (Albany NY)* **9**, 419-446, doi:10.18632/aging.101168 (2017).

50 Levine, M. E. *et al.* Menopause accelerates biological aging. *Proc Natl Acad Sci U S A* **113**, 9327-9332, doi:10.1073/pnas.1604558113 (2016).

51 Horvath, S. *et al.* Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging (Albany NY)* **8**, 1485-1512, doi:10.18632/aging.101005 (2016).

52 Breitling, L. P. *et al.* Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. *Clin Epigenetics* **8**, 21, doi:10.1186/s13148-016-0186-5 (2016).

53 Levine, M. E., Lu, A. T., Bennett, D. A. & Horvath, S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging (Albany NY)* **7**, 1198-1211 (2015).

54 Lowe, D., Horvath, S. & Raj, K. Epigenetic clock analyses of cellular senescence and ageing. *Oncotarget* **7**, 8524-8531, doi:10.18632/oncotarget.7383 (2016).

55 Maegawa, S. *et al.* Caloric restriction delays age-related methylation drift. *Nat Commun* **8**, 539, doi:10.1038/s41467-017-00607-3 (2017).

Chapter 2:

Long-term Ambient Particle Exposures and Blood DNA Methylation Age: Findings from the VA Normative Aging Study

Jamaji C. Nwanaji-Enwerem[1], BS; Elena Colicino[1], PhD; Letizia Trevisi[1], PhD; Itai Kloog[2], PhD; Allan C. Just[3], PhD; Jincheng Shen[4], PhD; Kasey Brennan[1], BS; Alexandra Dereix[1], BS; Lifang Hou[5], MD, PhD, MS; Pantel Vokonas[6], MD; Joel Schwartz[1], PhD; and Andrea A. Baccarelli[1], MD, PhD†

[1]*Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

[2]*Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Beer Sheva, Israel*

[3]*Department of Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

[4]*Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

[5]*Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA*

[6]*VA Normative Aging Study, Veterans Affairs Boston Healthcare System and the Department of Medicine, Boston University School of Medicine, Boston, MA, USA*

This is a pre-copyedited, author-produced version of an article accepted for publication in *Environmental Epigenetics* following peer review. This version was reproduced with permission from [Nwanaji-Enwerem, Jamaji C., et al. "Long-term ambient particle exposures and blood DNA methylation age: findings from the VA normative aging study." *Environmental epigenetics* 2.2 (2016).] The version of record is available online at: https://academic.oup.com/eep/article/2/2/dvw006/2841034. DOI: 10.1093/eep/dvw006.

2.1. Abstract

Background: Ambient particles have been shown to exacerbate measures of biological aging; yet, no studies have examined their relationships with DNA methylation age (DNAm-age), an epigenome-wide DNA methylation based predictor of chronological age.

Objective: We examined the relationship of DNAm-age with fine particulate matter ($PM_{2.5}$), a measure of total inhalable particle mass, and black carbon (BC), a measure of particles from vehicular traffic.

Methods: We used validated spatiotemporal models to generate 1-year $PM_{2.5}$ and BC exposure levels at the addresses of 589 older men participating in the VA Normative Aging Study with 1 to 3 visits between 2000 and 2011 (n=1032 observations). Blood DNAm-age was calculated using 353 CpG sites from the Illumina HumanMethylation450 BeadChip. We estimated associations of $PM_{2.5}$ and BC with DNAm-age using linear mixed effects models adjusted for age, lifestyle/environmental factors, and aging-related diseases.

Results: After adjusting for covariates, a 1-$\mu g/m^3$ increase in $PM_{2.5}$ (95%CI: 0.30, 0.75, P<0.0001) was significantly associated with a 0.52-year increase in DNAm-age. Adjusted BC models showed similar patterns of association (β=3.02, 95%CI: 0.48, 5.57, P=0.02). Only $PM_{2.5}$ (β=0.54, 95%CI: 0.24, 0.84, P=0.0004) remained significantly associated with DNAm-age in two-particle models. Methylation levels from 20 of the 353 CpGs contributing to DNAm-age were significantly associated with $PM_{2.5}$ levels in our two-particle models. Several of these CpGs mapped to genes implicated in lung pathologies including *LZTFL1*, *PDLIM5*, and *ATPAF1*.

Conclusion: Our results support an association of long-term ambient particle levels with DNAm-age and suggest that DNAm-age is a biomarker of particle-related physiological processes.

2.2. Introduction

Annually, air pollution – including ambient particle exposures – contributes to 3.7 million deaths worldwide and is one of the world's largest single environmental health risks[1]. Emerging evidence has also suggested that ambient particles may have aging-related effects: particulate matter with aerodynamic diameter ≤ 2.5 μm ($PM_{2.5}$) exposures have been associated with age-related outcomes including brain atrophy[2], declines in cognitive performance[3], ischemic heart disease[4], and stroke[5], as well as increases in systolic blood pressure by as much as 4.6 mmHg[6,7]. Moreover, traffic related particle exposures have been associated with hastened lung function decline by 6-7% over a five year period[8], accelerated pigment spot formation, and other clinical hallmarks of premature skin aging[9]. Previous research has used telomere length (TL), a common biomarker of biological aging[10], to characterize the relationship between particle exposures and aging. Nevertheless, data on the associations between ambient particles and TL have been conflicting and thus reflect a need for alternative biological aging markers[6,11,12].

Recent developments in the epigenetics of aging have provided new opportunities to address the relationship between particle exposures and aging biology. DNA methylation is an epigenetic mark involved in regulating genomic structure and transcription[13]. Reproducible changes in DNA methylation have long been associated with chronological aging[14-16] and recent studies report persisting associations even after accounting for age-related cellular heterogeneity, a previously neglected confounder[17-19]. DNA methylation age (DNAm-age) is a novel tissue-independent predictor of chronological age and is calculated by an algorithm that uses methylation values from 353 chronological age-correlated CpG dinucleotides in Illumina's HumanMethylation450 BeadChip[20,21]. Since DNA methylation in blood has been empirically shown to be sensitive to a number of biological processes[22-28], the DNAm-age of blood cells may help in further understanding epigenetic aging relationships with ambient particles. In this study, we investigated the relationship of DNAm-age with ambient particle exposures – $PM_{2.5}$ and Black Carbon (BC) – in a cohort of elderly men. We also examined the relationship of $PM_{2.5}$ and BC with leukocyte TL.

2.3. Materials and Methods

2.3.a. Study population

The Normative Aging Study (NAS) is an ongoing longitudinal cohort study of male volunteers within the Eastern Massachusetts community established in 1963 by the U.S. Department of Veterans Affairs (VA). Participants free of any chronic medical conditions were enrolled in the study and returned for onsite, detailed medical examinations every 3-5 years, during which data on stress levels, diet, physical activity, smoking status, and additional risk factors that may impact health were collected. Participants provided written informed consent to the VA Institutional Review Board (IRB). The Harvard T.H. Chan School of Public Health and the VA IRBs granted human subjects approval.

Eligibility for our study sample required continued participation as of 2000, when $PM_{2.5}$ air pollution levels became available. We excluded NAS participants with a diagnosis of leukemia (11 participants) because of a possible influence on the DNA methylation of blood cells. The remaining 589 participants were used in the analysis (Fig. S1). Study staff measured DNA methylation on blood DNA collected at up to three different visits for the participants. Using all available visits for each participant resulted in 1032 total observations.

2.3.b. Assessment of environmental factors: ambient particles and temperature

We selected $PM_{2.5}$ and BC as our ambient particle exposures because of their global pervasiveness [1] in addition to their status as the leading ambient particles with well-documented relationships with both DNA methylation[53-55] and adverse health outcomes[56-59].

To generate daily $PM_{2.5}$ exposure levels (in $\mu g/m^3$) at each participant's address, we employed a well-validated satellite based hybrid spatiotemporal prediction model with a multi-step approach[60,61]. The hybrid model combined satellite-derived aerosol optical depth (AOD) measurements and local land use regression model variables (*e.g.* traffic density, population density, and elevation) alongside temporal variables (*e.g.* temperature, wind speed, etc.). We fit the models to data from each year separately and generated daily predictions at the 1 x 1 km area resolution. Each participant's residence was geocoded and linked to an area level grid-point. To create a metric of long-term exposure, we averaged daily $PM_{2.5}$ level

predictions at each participant's address over the 365 days prior to the day of visit. The prediction model had an out of sample R^2 of 0.88 for daily samples.

We generated daily black carbon (BC) exposure estimates (in $\mu g/m^3$) based on participants' residences using a validated spatiotemporal land-use regression model[62]. Daily average BC estimates from 83 monitoring sites throughout the Greater Boston area were used to develop a prediction model. The final model included predictors based on information from meteorological conditions (*e.g.* wind speed), land use (*e.g.* traffic density), daily BC concentrations at a central monitor, and additional descriptors (*e.g.* day of the week). The prediction model had a high R^2 of 0.83 based on the training data set and a moderate correlation between predicted values and observed BC levels in four out-of-sample validation samples (R^2 = 0.59). To generate a 1-year BC exposure, we averaged daily BC exposure levels for the 365 days prior to the day of NAS visit.

To generate ambient temperature (in Celsius) for each participant we used a spatiotemporal prediction multi-step approach[51]. We obtained daily physical surface temperature (T_s) data from AOD measurements with 1 x 1 km resolution and daily near surface air temperature (T_a) data from the National Climatic Data Center, Environmental Protection Agency, and Weather Underground Inc. Mixed model regression was first used to calibrate T_s to T_a in 1 x 1 km grid cells where both were available. The model was validated with mean out of sample R^2 for days with available T_s and days without T_s equal to 0.95 and 0.94 respectively. Daily temperature measurements were averaged over the 365 days prior to the visit to generate 1-year temperature exposure estimates to complement the 1-year $PM_{2.5}$ and BC measurements. We selected the 1-year average because it correlates well with averages of $PM_{2.5}$, BC, and temperature over longer time windows and was available for a higher number of participants (Table S1). Moreover, existing studies examining relationships between particle exposures and other biological markers of aging, like telomere length, report more consistent and biologically significant results when a 1-year particle exposure is utilized[6,11,45-48].

2.3.c. DNA methylation and calculation of DNA methylation age (DNAm-age)

Laboratory staff extracted DNA from buffy coat of 7 mL whole blood using the QIAamp DNA Blood Kit (QIAGEN, Valencia, CA, USA). 500 ng DNA samples were then treated for bisulfite conversion using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA). Following bisulfite conversion, DNA samples were hybridized to the 12 sample Illumina HumanMethylation450 BeadChips as per Infinium HD Methylation protocol (Illumina, San Diego, CA, USA). Study staff then used a two-stage age-stratified algorithm to randomize samples to avoid confounding with chip and plate effects while ensuring similar age distribution across chips and plates. For quality control, we removed samples where >5% of probes had beadcount < 3 and >1% of probes had a detection *P*-value >0.05. The Bioconductor minfi package Illumina-type background correction without normalization was used to preprocess the remaining samples and generate methylation beta values to compute DNAm-age[63]. 450k arrays were run in the Genomics Core Facility at Northwestern University.

We calculated DNAm-age through Horvath's publically available online calculator (http://labs.genetics.ucla.edu/horvath/dnamage/)[20]. In short, an elastic net model (penalized regression) was used to regress a calibrated version of chronological age on 21,369 CpG probes shared by Illumina HumanMethylation27 and HumanMethylation450 BeadChip platforms. The elastic net platform selected 353 CpGs that correlate with age (193 positively and 160 negatively). The calculator predicts the age of each DNA sample (DNAm-age) using regression coefficients of the 353 CpGs resulting from the elastic net regression model trained from a number of training data sets. The calculator maintains predictive accuracy (age correlation 0.97, error = 3.6 years) across body tissues including blood[20].

2.3.d. Assessment of leukocyte telomere length (TL)

Laboratory staff performed quantitative real time polymerase chain reaction (qRT-PCR) on DNA extracted from buffy coat of whole blood using the QIAamp DNA Blood Kits[64]. Relative TL was measured on a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) as the qRT-PCR factor by which a sample differs from a reference DNA sample in its ratio of telomere repeat copy number (T) to single *36B4* gene copy number (S)[11,64]. The *36B4* gene is located on chromosome 12 and encodes

acidic ribosomal phosphoprotein PO. Laboratory staff ran all samples in triplicate and derived the average T:S ratio by dividing the average of the three T measurements by the average of the three S measurements. TL was then reported in relative units (qRT-PCR factor) of T:S ratio in the test sample to T:S ratio in the reference DNA pool. Batches for participant qRT-PCR telomere measurements were also recorded.

2.3.e. Assessment of smoking status

Smoking histories were collected on all study participants at NAS entry and standardized smoking interviews were administered at each subsequent NAS visit. Smoking status was characterized into three groups: **1) never smokers** were individuals who reported at entry and consistently thereafter that their lifetime cigarette consumption was <100 cigarettes; **2) former smokers** reported that they had smoked in the past but quit prior to study entry or they were smokers at entry and quit at some point during the follow up period and remained quit at the present study visit; **3) current smokers** were those who reported smoking regularly at each the follow up visit or those who quit, but reported inability to maintain abstinence at the present study visit. All participants also reported their average number of cigarettes per day at each assessment.

2.3.f. Statistical analysis

We used generalized linear mixed effects models to evaluate the relationship of DNAm-age with 1-year $PM_{2.5}$ and 1-year BC exposure levels, singularly and in two-particle models. To account for within participant correlation between the repeated measurements, the mixed effects models included a random intercept for each participant. DNAm-age, 1-year $PM_{2.5}$, and 1-year BC were all considered as continuous variables in all analyses.

The aforementioned models were adjusted for known confounders and covariates with *a priori* biological/clinical relevance using a tiered approach. Given that results from previous DNA methylation studies have been confounded by blood cell heterogeneity, we obtained cell type estimates for six blood cell types (*i.e.* plasma, CD4T, CD8T, NK, monocytes and granulocytes) using Houseman and Horvath

methods[20,65]. We first constructed chronological age and blood cell type adjusted mixed effects models for the relationships of $PM_{2.5}$ and BC with DNAm-age (Model 1). Next, we built models (Model 2) accounting for environmental/lifestyle factors by adjusting for average 1-year temperature (continuous), cumulative cigarette pack years (continuous), smoking status (current, former, or never), and season of visit (Spring [March-May], Summer [June-August], Fall [September-November], and Winter [December-February]), body mass index (lean [<25], overweight [25-30], obese [>30]), alcohol intake (yes/no ≥ 2 drinks daily), and maximum years of education (continuous) in addition to the Model 1 covariates. We constructed a third (Model 3) and fourth set of models (Model 4) which accounted for aging-related diseases and disease-related medications respectively. Model 3 adjusted for cancer (yes/no history of lifetime cancer diagnosis), coronary heart disease (yes/no based on electrocardiogram, validated medical records, or physical exam), diabetes (physician diagnosis or a fasting blood glucose > 126 mg/dL), and hypertension (yes/no antihypertensive medication use or systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg) in addition to the Model 2 covariates. Model 4 adjusted for subjects taking statins and/or any diabetes and hypertension medications in addition to the Model 2 covariates. Last, we constructed two-particle mixed effects models with both $PM_{2.5}$ and BC as predictors of DNAm-age using the covariates from the Model 1- 4 framework.

To exclude sensitivity of our models to outliers, we repeated all analyses using robust regression. By iteratively reweighting data points such that points far from model predictions in the previous iteration are given smaller weights, robust regression is able to minimize the sensitivity of a model to outlying values. Iterations continue until the values of coefficient estimates meet a specified tolerance and weighted least squares regression is then used to compute model coefficients. We performed a set of additional sensitivity analyses: **1)** we added a random intercept for 450k plate to account for potential batch effects, **2)** we explored our particle DNAm-age associations in participants with only one NAS visit to see how our results compared to the primary analysis on the full study sample and **3)** we stratified our study sample by season of NAS visit to further explore the contribution of season to the relationship between particle exposures

and DNAm-age. We also looked at the Pearson correlation between change in particle exposure and change in DNAm-age between study visits using participants with at least two NAS visits.

Additionally, we evaluated the relationships of DNA methylation values at each of the 353 DNAm-age CpG probes with 1-year $PM_{2.5}$ and 1-year BC exposure levels using the aforementioned Model 2 covariates and technical covariates (450k plate, chip, row, and column). FDR correction was performed to account for multiple hypotheses testing for all CpG methylation analyses. Gene ontology analyses were performed on significant CpG results using the publically available DAVID bioinformatics platform[66,67].

As a means of comparison with the DNAm-age results, we explored the relationships of a standard marker of aging, telomere length, with $PM_{2.5}$ and BC exposure levels. We constructed mixed effects multivariable linear regression models adjusting for chronological age, blood cell type, average 1-year temperature, cumulative cigarette pack years, smoking status, season of visit, telomere batch (categorical with four batches), BMI, alcohol intake, and maximum years of education. Similar to our DNAm-age analyses, we constructed two additional sets of models adjusting for age-related diseases and disease-related medications respectively. There was one relative TL observation of 12.7, while the remaining 856 TL observations were < 4. We kept the outlying observation in the TL mixed effects models, but re-ran the models using robust regression and without the outlying value as sensitivity analyses.

We performed all statistical analyses using R Version 3.1.1 (R Core Team, Vienna, Austria) and considered a *P*-value <0.05 to be statistically significant.

2.4. Results

2.4.a. Baseline characteristics and descriptive statistics

All participants were Caucasian males with a mean age of 74.8 years (SD = 7.06) and a mean DNAm-age of 74.1 years (SD = 7.90, Table 1). Participants with coronary heart disease, hypertension, and a lifetime cancer diagnosis had a significantly higher mean DNAm-age than their respective counterparts (Table S2).

Table 1. Descriptive Statistics of Study Participants

Characteristic	
Number of Observations (participants)	1032 (589)
Total Number of Visits, N (%)	
One	589 (57%)
Two	352 (34%)
Three	91 (9%)
Chronological Age, Mean (SD)	74.8 (7.06)
DNAm-age, Mean (SD)	74.1 (7.90)
1-year $PM_{2.5}$ ($\mu g/m^3$), Mean (SD)	10.7 (1.40)
1-year BC ($\mu g/m^3$), Mean (SD)	0.51 (0.18)
Year Average Temperature (°C), Mean (SD)	11.5 (1.19)
Cigarette Pack Years, Mean (SD)	20.5 (24.4)
Relative Telomere Length, Mean (SD)	1.25 (0.64)
Max Years Education, N (%)	
≤ 12 years	264 (25%)
12 – 16 years	493 (48%)
> 16 years	275 (27%)
Body Mass Index, N (%)	
Underweight	2 (0%)
Healthy/Lean	234 (23%)
Overweight	549 (53%)
Obese	247 (24%)
Alcohol Consumption, N (%)	
< 2 drinks/day	831 (81%)
≥ 2 drinks/day	201 (19%)
Lifetime Cancer Diagnosis, N (%)	
Yes	574 (56%)
No	458 (44%)
Coronary Heart Disease, N (%)	
Yes	355 (33%)
No	677 (67%)
Diabetes, N (%)	
Yes	193 (19%)
No	839 (81%)
Hypertension, N (%)	
Yes	753 (73%)
No	279 (27%)
Smoking Status, N (%)	
Never	294 (29%)
Former	701 (67%)
Current	37 (4%)
Season, N (%)	
Spring	249 (24%)
Summer	245 (24%)
Fall	350 (34%)
Winter	188 (18%)

Furthermore, never smokers had a significantly higher mean DNAm-age when compared to former smokers (Table S2). No significant associations were found when comparing never smokers to current smokers or current smokers to former smokers. Mean 1-year $PM_{2.5}$ and 1-year BC levels were 10.7 $\mu g/m^3$ (SD = 1.40) and 0.51 $\mu g/m^3$ (SD = 0.18) respectively (Table 1). Moreover, 1-year $PM_{2.5}$ and BC levels were significantly correlated (r = 0.41, P <0.0001) in our study sample (Table S3).

Table 2. 1-Year Particulate Matter 2.5 ($PM_{2.5}$) and Black Carbon (BC) as Predictors of DNA Methylation (DNAm) Age

Particle (1 $\mu g/m^3$)	Difference in DNAm-age (95% CI)	*P*	N	AIC
$PM_{2.5}$				
Model 1	0.55 (0.33, 0.77)	**<0.0001**	1032	6346.85
Model 2	0.52 (0.30, 0.75)	**<0.0001**	1032	6360.86
Model 3	0.52 (0.29, 0.74)	**<0.0001**	1032	6361.47
Model 4	0.50 (0.27, 0.72)	**<0.0001**	1032	6362.88
BC				
Model 1	2.49 (0.11, 4.88)	**0.04**	898	5571.94
Model 2	3.02 (0.48, 5.57)	**0.02**	898	5583.16
Model 3	2.92 (0.36, 5.48)	**0.03**	898	5583.51
Model 4	2.83 (0.28, 5.39)	**0.03**	898	5582.92
Two-Particle Model 1			898	5560.38
$PM_{2.5}$	0.56 (0.28, 0.84)	**0.0001**		
BC	0.52 (-2.03, 3.08)	0.69		
Two-Particle Model 2			898	5574.56
$PM_{2.5}$	0.54 (0.24, 0.84)	**0.0004**		
BC	0.62 (-2.24, 3.47)	0.67		
Two-Particle Model 3			898	5575.71
$PM_{2.5}$	0.52 (0.22,0.83)	**0.0007**		
BC	0.61 (-2.25, 3.47)	0.67		
Two-Particle Model 4			898	5575.70
$PM_{2.5}$	0.51 (0.21,0.82)	**0.0009**		
BC	0.60 (-2.25, 3.46)	0.68		

Model 1: adjusted for chronological age and blood cell type.

Model 2: Model 1 but additionally adjusted for temperature, pack years, smoking status, season, BMI, alcohol consumption, and education.

Model 3: Model 2 but additionally adjusted for history of cancer, hypertension, chd, and diabetes.

Model 4: Model 2 but additionally adjusted for statins and medications for diabetes and hypertension.

2.4.b. $PM_{2.5}$ and BC as independent and joint predictors of DNAm-age

Residuals from all models appeared normally distributed. In a model solely adjusted for chronological age and blood cell type, 1 $\mu g/m^3$ increases in 1-year $PM_{2.5}$ exposures were significantly associated with 0.55 year increases in DNAm-age (P <0.0001). Following adjustments in Model 2, $PM_{2.5}$ remained associated with increases in DNAm-age ($\beta = 0.52$, P <0.0001) (Table 2). These results remained consistent in Model 3 ($\beta = 0.52$, P <0.0001) and Model 4 ($\beta = 0.50$, P <0.0001), which were adjusted for aging-related disease covariates and disease medications respectively (Table 2). These $PM_{2.5}$ associations persisted in sensitivity analyses with robust regression (data not shown) and in models adjusting for 450k plate, though the effect estimates were slightly attenuated (Table S4). In a model adjusted for chronological age and blood cell type, BC was a significant predictor of DNAm-age ($\beta = 2.49$, $P = 0.04$), and remained a significant predictor of DNAm-age in subsequent models adjusting for additional covariates (Table 2). Nonetheless, after adjusting for 450k plate, the BC associations with DNAm-age remained marginally significant at best (Tables S4). $PM_{2.5}$ levels remained significantly associated with increases in DNAm-age of 0.51 years or greater (P <0.0001) in two-particle models with BC (Table 2) though the magnitude of the effect estimates were also attenuated following adjustments for 450k plate (Table S4). BC levels were not significantly associated with DNAm-age in any of the two-particle models (Tables 2 & S4).

A sensitivity analysis exploring particle associations with DNAm-age in participants with only one NAS visit, revealed similar, but non-significant trends as the primary analysis (Table S5). A subsequent sensitivity analysis that stratified the study sample by season of NAS visit also revealed similar trends as the primary analysis, but results were only significant for $PM_{2.5}$ associations in the summer and fall NAS visit groups (Table S6). Finally, an analysis using participants with at least two NAS visits and exploring the correlation between the change in particle exposure between visits and the change in DNAm-age between visits, revealed weak and non-significant correlations (Table S7).

2.4.c. Associations between $PM_{2.5}$ levels and methylation values at individual DNAm-age CpG sites

We explored associations between $PM_{2.5}$ levels and the methylation values for the 353 CpG sites that are used to calculate DNAm-age. Methylation of 20 out of 353 CpGs was significantly associated with $PM_{2.5}$ levels in two-particle mixed effects Model 2 (adjusting for BC, age, blood cell type, and lifestyle/environmental characteristics) following FDR correction (Fig. 1). $PM_{2.5}$ levels were positively or negatively associated with CpG methylation depending on the CpG site (Table 3).

Table 3. 1-Year Particulate Matter ($PM_{2.5}$) as a Predictor of CpG Probe Methylation in a Two-Particle Model

CpG*	Gene	Process	Difference in Methylation	*P*	FDR
Negative Association					
cg14163776	ACAP2	GTPase activator activity	-0.0049	**<0.0001**	**0.003**
cg06044899	TMSL3	actin cytoskeleton organization	-0.0048	**<0.0001**	**0.001**
cg01570885	FAM50B	protein binding	-0.0041	**0.001**	**0.041**
cg18139769	SGCE (PEG10)	calcium ion binding	-0.0040	**0.001**	**0.037**
cg22736354	NHLRC1	ubiquitin-protein transferase activity	-0.0032	**0.002**	**0.042**
cg15661409	C14orf105	uncharacterized	-0.0012	**0.002**	**0.041**
Positive Association					
cg02047577	UCKL1	uridine kinase activity	0.0002	**0.001**	**0.041**
cg10940099	CD164	cellular adhesion	0.0002	**0.002**	**0.041**
cg22006386	CATSPERG	ion channel activity	0.0003	**0.002**	**0.044**
cg08186124	LZTFL1	protein binding: cytoplasm	0.0004	**<0.0001**	**0.015**
cg04094160	ZBTB5	transcriptional regulation	0.0005	**<0.0001**	**0.014**
cg16408394	RXRA	DNA binding	0.0005	**0.002**	**0.042**
cg23786576	ATPAF1	ATP synthase complex assembly	0.0006	**0.001**	**0.040**
cg15341340	DNASE2	endodeoxyribonuclease activity	0.0007	**0.002**	**0.041**
cg21395782	NDUFA13	NADH dehydrogenase activity	0.0008	**0.001**	**0.041**
cg26043391	FBXO28	protein binding	0.0009	**0.001**	**0.041**
cg06557358	TMEM132E	integral component of membrane	0.0010	**0.003**	**0.050**
cg14409958	ENPP2	nucleic acid binding	0.0011	**0.002**	**0.041**
cg20305610	PDLIM5	actin binding	0.0013	**<0.0001**	**0.014**
cg05675373	KCNC4	potassium channel activity	0.0032	**0.002**	**0.042**

*adjusted for chronological age, blood cell type, BC, temperature, pack years, smoking status, season, BMI, alcohol consumption, education, and 450k technical covariates.

The 20 CpGs mapped to 20 known genes; nevertheless, gene ontology analysis did not return significant pathway enrichment (data not shown). No CpGs were significantly associated with BC levels in the two-particle mixed effects model.

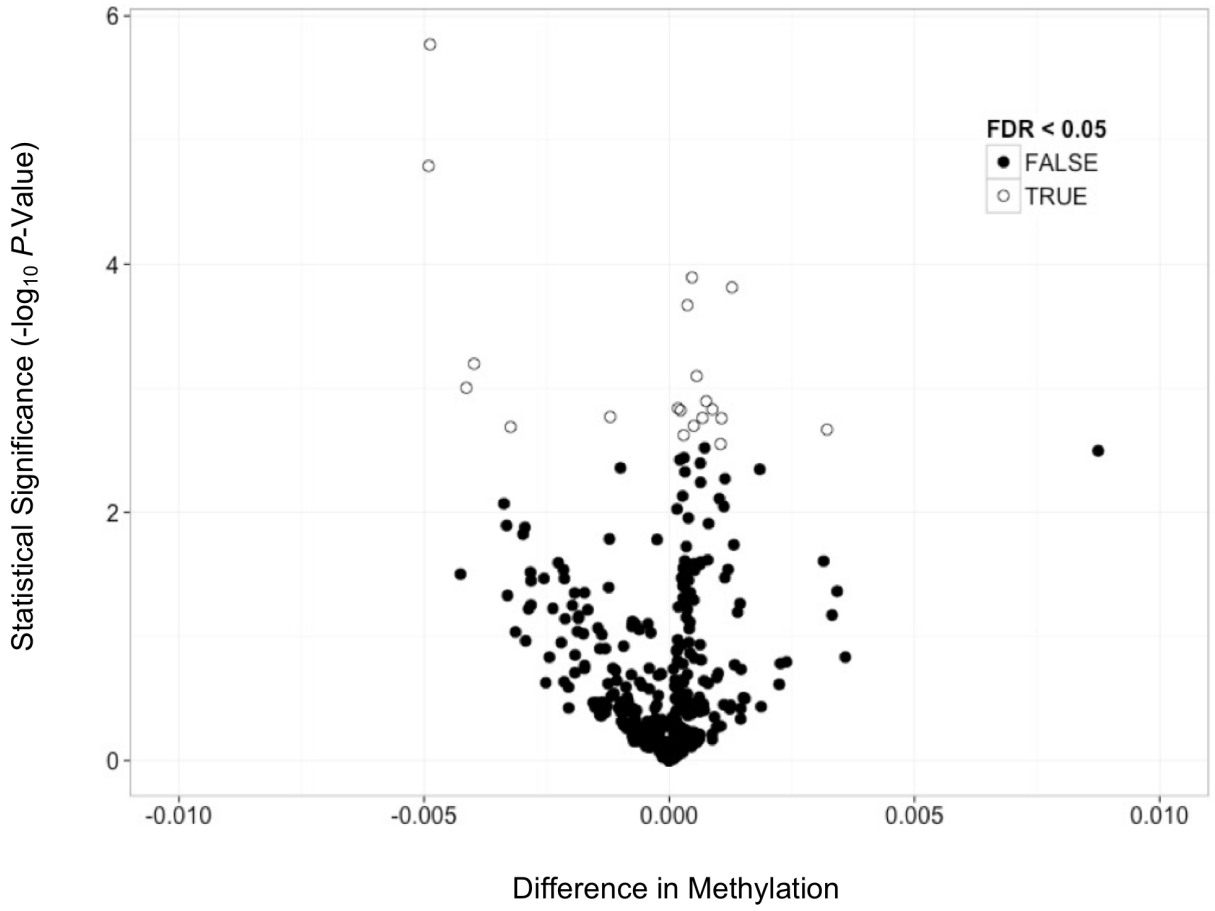

Figure 1 | Volcano plot of Regression Coefficients for Difference in DNA Methylation Beta Values from 353 DNAm-age CpGs Analyzed for Association with 1-Year $PM_{2.5}$ Levels in a Two-Particle Model. Linear mixed effects models were used to explore the associations between 1-Year $PM_{2.5}$ exposure levels and DNA methylation values for the 353 CpG sites used to calculate DNAm-age. The regression coefficient for the difference in DNA methylation beta values given by a 1$\mu g/m^3$ increase in 1-Year $PM_{2.5}$ exposure level is plotted on the x-axis, and the corresponding significance is plotted on the y-axis. CpG probes meeting statistical significance following FDR adjustment are depicted as hollow circles. DNA methylation beta values range from 0 (completely unmethylated) to 1 (completely methylated).

2.4.d. DNAm-age, $PM_{2.5}$, and BC as predictors of relative TL

TL showed a weak and non-significant correlation (r = -0.06, P = 0.08) with DNAm-age in participants' NAS observations (Table S3). Moreover, DNAm-age was not a significant predictor of TL in mixed effects models adjusting for chronological age, blood cell type, and telomere batch (Table 4). TL also showed no significant associations with 1-year $PM_{2.5}$ or 1-year BC levels in any of the single-particle or two-particle models (Table 5).

Table 4. DNAm-age as a Predictor of Relative Telomere Length (TL)

	Change in TL (95% CI)	***P***	**N**	**AIC**
DNAm-age				
Model 1	-0.006 (-0.01, 0.002)	0.14	857	1687.65
Model 2	-0.004 (-0.01, 0.002)	0.23	856	1233.57

Model 1: adjusted for chronological age, blood cell type, and telomere batch

Model 2: Model 1 but excluding one participant with an outlying telomere value.

Table 5. 1-Year Particulate Matter 2.5 ($PM_{2.5}$) and Black Carbon (BC) as Predictors of Telomere Length (TL)

Particle (1 μg/m^3)	Difference in TL (95% CI)	*P*	N	AIC
$PM_{2.5}$				
Model 1	0.02 (-0.01, 0.06)	0.24	857	1750.26
Model 2	0.02 (-0.01, 0.06)	0.20	857	1754.13
Model 3	0.02 (-0.01, 0.06)	0.23	857	1776.27
Model 4	0.02 (-0.01, 0.06)	0.21	857	1771.64
BC				
Model 1	0.11 (-0.16, 0.38)	0.42	770	1616.52
Model 2	0.13 (-0.16, 0.42)	0.37	770	1637.87
Model 3	0.12 (-0.16, 0.41)	0.40	770	1658.35
Model 4	0.13 (-0.16, 0.42)	0.38	770	1654.71
Two-Particle Model 1			770	1623.11
$PM_{2.5}$	0.02 (-0.02, 0.07)	0.32		
BC	0.05 (-0.24, 0.35)	0.71		
Two-Particle Model 2			770	1644.69
$PM_{2.5}$	0.02 (-0.03, 0.07)	0.40		
BC	0.07 (-0.25, 0.38)	0.66		
Two-Particle Model 3			770	1665.33
$PM_{2.5}$	0.02 (-0.03, 0.07)	0.47		
BC	0.07 (-0.24, 0.39)	0.65		
Two-Particle Model 4			770	1661.52
$PM_{2.5}$	0.02 (-0.03, 0.07)	0.39		
BC	0.07 (-0.25, 0.39)	0.67		

Model 1: adjusted for chronological age and blood cell type.

Model 2: Model 1 but additionally adjusted for temperature, pack years, smoking status, season, telomere batch, BMI, alcohol consumption, and education.

Model 3: Model 2 but additionally adjusted for history of cancer, hypertension, chd, and diabetes.

Model 4: Model 2 but additionally adjusted for statins and medications for diabetes and hypertension.

2.5. Discussion

The present study showed a novel positive association between 1-year $PM_{2.5}$ exposure levels and DNAm-age. To the best of our knowledge, this is the first study showing relationships between any environmental pollutant and an epigenetic biomarker of aging. $PM_{2.5}$ remained a statistically significant positive predictor of DNAm-age after adjusting for chronological age and other covariates. The study also

revealed a significant positive association between BC and DNAm-age after adjusting for age and other covariates, but not after adjusting for 450k plate. Moreover, we identified 20 age-related CpG sites whose methylation was significantly associated with $PM_{2.5}$ exposure levels in two-particle models adjusting for BC, age, and other covariates.

Operating under the premise that adverse exposures accelerate aging, we expected 1-year $PM_{2.5}$ and BC exposure levels to be associated with increases in DNAm-age. In alignment with our expectations, both $PM_{2.5}$ and BC exposure levels were positively associated with DNAm-age. Pearson correlations of between visit changes in particle exposures and between visit changes in DNAm-age in participants with multiple visits were not significant potentially due to the smaller number of observations. Nonetheless, compared to the primary analysis, we observed similar trends in the association of our particles with DNAm-age in sensitivity analyses using participants with a single NAS visit. These trends suggest that having a single or multiple visits was not driving the results from the adjusted mixed effects models. Likewise, trends similar to the primary analysis were also observed in our seasonal analysis and were significant for the summer and fall seasons, which had the highest average particle exposures across all observations.

Although DNAm-age is primarily viewed as a predictor of chronological age, emerging research suggests that it reflects underlying physiological processes including metabolic dysregulation, immune dysfunction, and genomic instability[29-32]. To date, two studies have described significant associations between DNAm-age and all-cause mortality[29,33]. Moreover, studies have also demonstrated that DNAm-age may predict or be reflective of various disease processes[31,32,34-37]. It is hypothesized that DNAm-age may measure "the cumulative work done by a particular kind of epigenetic maintenance system [EMS], which helps maintain epigenetic stability"[20]. Under the EMS hypothesis, an increase in DNAm-age suggests that an event or process has occurred and the EMS has completed more work to repair or return the epigenome to homeostasis. Alternatively, a reduction in DNAm-age can be interpreted as epigenetic stability or disrupted activation of the EMS, both of which would result in less maintenance work. Given the known toxicity of ambient particles, our data supports the theory that some particles may disrupt the

epigenome thus requiring more maintenance work. Nevertheless, mechanistic studies are warranted to explicitly identify the components of this system.

Given our interpretation of the relationship between adverse exposures and DNAm-age, it was interesting to find that cigarette pack years was negatively correlated with DNAm-age and that former smokers had a lower mean DNAm-age than never smokers. Though cigarette smoking can be considered a personal form of air pollution, it is also a complex mixture with a composition that differs from that of $PM_{2.5}$. Differences in particle composition can account for differences in the toxicological pathways of these exposures and may be one reason why we observe differences in their DNAm-age relationships. Moreover, individuals who are sick are often urged to quit smoking so there may still be some confounding when observing the unadjusted correlations of pack years and cigarette smoking status with DNAm-age. A number of physiological factors can also affect the epigenome and should be considered when comparing smoking to air pollution exposures. For instance, it is widely known that smoking can account for substantial weight loss and it has been demonstrated that obesity accelerates the DNAm-age of liver cells[31]. Finally, a study sample with 37 current smoker observations may be underpowered to detect differences in mean DNAm-age between current smokers and other groups.

In our two-particle models, BC exposure levels were not significantly associated with DNAm-age while $PM_{2.5}$ remained a significant predictor of DNAm-age. BC is considered a specific marker of traffic-related air pollution, while $PM_{2.5}$ is a heterogeneous mixture of fine particles with component species often including carbonaceous fractions (*e.g.* black carbon), inorganic compounds (*e.g.* sulfate, nitrate, ammonium), and trace metals (*e.g.* nickel, lead, copper)[38]. Research on total $PM_{2.5}$ is more extensive than any work singularly exploring BC or other components. Many studies suggest that BC may be more toxic than $PM_{2.5}$[39], but data also exists where $PM_{2.5}$ associations are stronger than that of BC[40]. The finding that $PM_{2.5}$ was driving the association with DNAm-age in the two particle models may possibly be because other components apart from BC are responsible for the DNAm-age relationship. Another theory is that the mixture of the $PM_{2.5}$ components is more harmful, with regards to DNAm-age, than any of the components

singularly. Ultimately, further work involving a detailed compositional analysis of $PM_{2.5}$ will aid in further understanding what components are driving the associations with DNAm-age.

Although we attributed the observed positive association of $PM_{2.5}$ with DNAm-age to greater cumulative work by the epigenetic maintenance system, we also conducted additional analyses to identify which of the 353 CpG sites contributing to the DNAm-age metric had methylation values that were significantly associated with $PM_{2.5}$ levels. We identified 20 such CpGs through a mixed effects model adjusting for chronological age, blood cell type, and lifestyle/environmental factors. These CpGs mapped to 20 known genes. A gene ontology analysis of these 20 genes did not return any significant enrichment for specific biological pathways. Nevertheless, a literature review revealed relationships between the genes. For instance, *LZTFL1*, *PDLIM5*, and *ATPAF1*, can all be generally characterized as being involved in protein binding. *LZTFL1* (Leucine Zipper Transcription Factor-like 1) is a nuclear gene that encodes a cytoplasmic protein that interacts with other cytosolic proteins to regulate ciliary trafficking and control β-catenin nuclear signaling. *LZTFL1* downregulation has been implicated in non-small cell lung cancer and poor survival. In contrast, *LZTFL1* re-expression in lung tumor cells inhibits tumor growth and lung tissue colonization by circulating tumor cells[41]. *ATPAF1* (ATP Synthase Mitochondrial F1 Complex Assembly Factor 1) encodes a soluble mitochondrial protein that helps prevent abnormal aggregation of F1-ATP synthase subunits, and, like *LZTFL1,* is expressed in many tissues including the lung. *ATPAF1* is highly expressed in bronchial biopsies of individuals with severe asthma and has been found to predispose children of different ancestries to asthma[42]. Unlike *LZTFL1* and *ATPAF1*, *PDLIM5* (PDZ and LIM domain 5) primarily is involved in cardiomyocyte function. Nonetheless, *PDLIM5* still has implications for lung physiology as its downregulation has been linked to hypoxia-induced pulmonary hypertension[43]. Collectively, our data suggests putative relationships between ambient particle levels and genes involved in various elements of lung physiology. Nonetheless, additional methylation and mechanistic studies will be necessary to first confirm these changes in gene methylation and next ascertain if these changes actually manifest themselves as differences in gene expression and protein levels/activity.

Finally, to help interpret our DNAm-age results, we explored the relationship of $PM_{2.5}$ and BC exposure levels with relative telomere length. Telomeres are nucleoprotein structures, at the ends of eukaryotic chromosomes, involved in maintaining genomic fidelity. Telomere shortening has been associated with aging and aging related diseases[44]. Contrary to our expectations, we observed no association of $PM_{2.5}$ and BC with relative TL. As mentioned, the literature examining the relationship between particles and TL has been conflicting. Significant associations between annual $PM_{2.5}$ exposures and decreased TL have been reported[45], but in the NAS the relationship between annual BC exposures and decreased TL was only observed in never smokers[11]. The literature concerning short-term particle exposures is even more obscure. In some cases, short-term particle exposures have been associated with increased TL[46], decreased TL[47], and in other instances no significant association was observed[6,48]. Our findings add to the body of literature that suggests: **1)** that exposure duration and study population characteristics are particularly critical in understanding and interpreting the results of TL studies; and **2)** other measures, like DNAm-age, may offer more advantages for understanding the relationship between particle exposures and biological aging. Moreover, DNAm-age was not associated or correlated with TL in our study sample. Similar non-significant relationships between DNAm-age and TL have also been independently reported in a study conducted in the Lothian Birth Cohorts[49]. The known relationships of DNAm-age and TL with *in vitro* cell passaging also highlight the differences between these markers. As cells are passaged, they divide and in most cases their telomeres shorten[50]. However, DNAm-age increases as cells are passaged and divide *in vitro*[20]. In all, our findings and existing evidence suggests that though DNAm-age and TL are both measures of "aging," the two are not one in the same and may capture different elements of biological processes.

Though we present a study with a number of objective, validated measures and rigorous statistical methods, our study does have a few limitations. First, our $PM_{2.5}$ and BC measurements were generated using spatiotemporal prediction models. Though the models were validated[51], we cannot completely eliminate residual measurement error or discrepancies in calibration coefficients[52]. Also, ambient levels of air pollution at a participant's address may differ from personal exposures, which also depend on time spent at home, rates of penetration of ambient particles into the house, and the presence of indoor sources of

particles. However, we note that the demographics of the Normative Aging Study, which is composed primarily of retired older men, make it more likely that participants spend a large part of their day at home. Our findings are also based on a cohort of older Caucasian males residing in a lightly-polluted urban environment; thus, studies including younger individuals, females, non-Caucasians, and in different environments are warranted to confirm our findings more broadly. Lastly, we attempted to adjust for potential confounders, but cannot rule out the possibility of unknown or residual confounding in our analyses.

2.6. Conclusion

Our data suggests that DNAm-age and TL capture different elements of biological aging; describes novel associations between ambient particles and DNAm-age; and highlights existing limits in interpretations of biological/molecular aging. Further analyses utilizing DNAm-age with $PM_{2.5}$, BC, and other particles may provide much needed insight into fully understanding the biologically adverse nature of ambient particles.

2.7. Contributions and Support

Contributors: JNE, EC, and AAB conceived and designed the study. LT, IK, ACJ, JiS, KB, AD, LH, and PV gathered data. JNE performed the data analyses and drafted the manuscript. JS, EC, LT, JiS and ACJ contributed to the analyses. All authors revised and approved the manuscript.

Conflict of interest statement: None declared

Ethics approval: Boston VA Medical Center, Harvard T.H. Chan School of Public Health (protocol 14027-102).

Acknowledgements: This study is supported by a grant from the National Institute of Environmental Health Sciences (NIEHS) (R01ES021733). Other support comes from NIEHS grants ES015172, ES014663 and ES020010, and Environmental Protection Agency (EPA) grant RD832416. The US Department of Veterans Affairs (VA) Normative Aging Study (NAS) is supported by the Cooperative Studies

Program/ERIC, US Department of Veterans Affairs, and is a research component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). The views expressed in this paper are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs. Additional support was provided by the US Department of Agriculture, Agricultural Research Service (contract 53-K06-510).

Data Availability: Data are from the Normative Aging Study, from which restricted data are available for researchers who meet the criteria. A subset of the methylation data is deposited at NCBI dbGaP (study accession number: phs000853.v1.p1).

2.8. References

1 World Health, O. Burden of disease from household air pollution for 2012. *WHO, Geneva* (2014).

2 Chen, J.-C. & Schwartz, J. Neurobehavioral effects of ambient air pollution on cognitive performance in US adults. *Neurotoxicology* **30**, 231-239 (2009).

3 Weuve, J. *et al.* Exposure to Particulate Air Pollution and Cognitive Decline in Older Women. *Arch Intern Med* **172**, 219-227, doi:10.1001/archinternmed.2011.683 (2012).

4 Brown, D. M. *et al.* Occupational Exposure to PM2.5 and Incidence of Ischemic Heart Disease: Longitudinal Targeted Minimum Loss-based Estimation. *Epidemiology* **26**, 806-814, doi:10.1097/EDE.0000000000000329 (2015).

5 Scheers, H., Jacobs, L., Casas, L., Nemery, B. & Nawrot, T. S. Long-Term Exposure to Particulate Matter Air Pollution Is a Risk Factor for Stroke: Meta-Analytical Evidence. *Stroke* **46**, 3058-3066, doi:10.1161/STROKEAHA.115.009913 (2015).

6 Shan, M. *et al.* A feasibility study of the association of exposure to biomass smoke with vascular function, inflammation, and cellular aging. *Environ. Res.* **135**, 165-172, doi:10.1016/j.envres.2014.09.006 (2014).

7 Wilker, E. H. *et al.* Long-Term Exposure to Fine Particulate Matter, Residential Proximity to Major Roads and Measures of Brain Structure. *Stroke* **46**, 1161-1166 (2015).

8 Lepeule, J. *et al.* Long-term effects of traffic particles on lung function decline in the elderly. *Am J Respir Crit Care Med* **190**, 542-548, doi:10.1164/rccm.201402-0350OC (2014).

9 Vierkotter, A. *et al.* Airborne particle exposure and extrinsic skin aging. *J Invest Dermatol* **130**, 2719-2726, doi:10.1038/jid.2010.204 (2010).

10 Cawthon, R. M., Smith, K. R., O'Brien, E., Sivatchenko, A. & Kerber, R. A. Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* **361**, 393-395, doi:10.1016/S0140-6736(03)12384-7 (2003).

11 McCracken, J. *et al.* Annual ambient black carbon associated with shorter telomeres in elderly men: Veterans Affairs Normative Aging Study. *Environ. Health Perspect.* **118**, 1564-1570 (2010).

12 Zhang, X., Lin, S., Funk, W. E. & Hou, L. Environmental and occupational exposure to chemicals and telomere length in human studies. *Postgrad Med J* **89**, 722-728, doi:10.1136/postgradmedj-2012-101350rep (2013).

13 Robertson, K. D. DNA methylation and human disease. *Nat Rev Genet* **6**, 597-610, doi:10.1038/nrg1655 (2005).

14 Teschendorff, A. E., West, J. & Beck, S. Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Hum Mol Genet* **22**, R7-R15, doi:10.1093/hmg/ddt375 (2013).

15 Alisch, R. S. *et al.* Age-associated DNA methylation in pediatric populations. *Genome Res* **22**, 623-632, doi:10.1101/gr.125187.111 (2012).

16 Bell, J. T. *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* **8**, e1002629, doi:10.1371/journal.pgen.1002629 (2012).

17 Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* **15**, R31, doi:10.1186/gb-2014-15-2-r31 (2014).

18 Lin, Q. & Wagner, W. Epigenetic Aging Signatures Are Coherently Modified in Cancer. *PLoS Genet* **11**, e1005334, doi:10.1371/journal.pgen.1005334 (2015).

19 Bacalini, M. G. *et al.* A meta-analysis on age-associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450k data. *Aging (Albany NY)* **7**, 97-109, doi:10.18632/aging.100718 (2015).

20 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).

21 Horvath, S. Erratum to: DNA methylation age of human tissues and cell types. *Genome Biol* **16**, 96, doi:10.1186/s13059-015-0649-6 (2015).

22 Sanchez-Guerra, M. *et al.* Effects of particulate matter exposure on blood 5-hydroxymethylation: results from the Beijing truck driver air pollution study. *Epigenetics* **10**, 633-642, doi:10.1080/15592294.2015.1050174 (2015).

23 De Prins, S. *et al.* Influence of ambient air pollution on global DNA methylation in healthy adults: a seasonal follow-up. *Environ Int* **59**, 418-424, doi:10.1016/j.envint.2013.07.007 (2013).

24 Unternaehrer, E. *et al.* Dynamic changes in DNA methylation of stress-associated genes (OXTR, BDNF) after acute psychosocial stress. *Transl Psychiatry* **2**, e150, doi:10.1038/tp.2012.77 (2012).

25 Alasaari, J. S. *et al.* Environmental stress affects DNA methylation of a CpG rich promoter region of serotonin transporter gene in a nurse cohort. *PLoS One* **7**, e45813, doi:10.1371/journal.pone.0045813 (2012).

26 van der Knaap, L. J. *et al.* Glucocorticoid receptor gene (NR3C1) methylation following stressful events between birth and adolescence. The TRAILS study. *Transl Psychiatry* **4**, e381, doi:10.1038/tp.2014.22 (2014).

27 Aslibekyan, S. *et al.* Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. *Obesity (Silver Spring)* **23**, 1493-1501, doi:10.1002/oby.21111 (2015).

28 Nicoletti, C. F. *et al.* DNA Methylation and Hydroxymethylation Levels in Relation to Two Weight Loss Strategies: Energy-Restricted Diet or Bariatric Surgery. *Obes Surg* **26**, 603-611, doi:10.1007/s11695-015-1802-8 (2016).

29 Marioni, R. E. *et al.* DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* **16**, 25, doi:10.1186/s13059-015-0584-6 (2015).

30 Nazarenko, M. S. *et al.* A comparison of genome-wide DNA methylation patterns between different vascular tissues from patients with coronary heart disease. *PLoS One* **10**, e0122601, doi:10.1371/journal.pone.0122601 (2015).

31 Horvath, S. *et al.* Obesity accelerates epigenetic aging of human liver. *Proceedings of the National Academy of Sciences* **111**, 15538-15543, doi:10.1073/pnas.1412759111 (2014).

32 Horvath, S. & Levine, A. J. HIV-1 Infection Accelerates Age According to the Epigenetic Clock. *J Infect Dis* **212**, 1563-1573, doi:10.1093/infdis/jiv277 (2015).

33 Christiansen, L. *et al.* DNA methylation age is associated with mortality in a longitudinal Danish twin study. *Aging Cell* **15**, 149-154, doi:10.1111/acel.12421 (2016).

34 Wolf, E. J. *et al.* Accelerated DNA methylation age: Associations with PTSD and neural integrity. *Psychoneuroendocrinology* **63**, 155-162, doi:10.1016/j.psyneuen.2015.09.020 (2016).

35 Levine, M. E. *et al.* DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY)* **7**, 690-700 (2015).

36 Levine, A. J. *et al.* Accelerated epigenetic aging in brain is associated with pre-mortem HIV-associated neurocognitive disorders. *J. Neurovirol.*, doi:10.1007/s13365-015-0406-3 (2015).

37 Marioni, R. E. *et al.* The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol* **44**, 1388-1396, doi:10.1093/ije/dyu277 (2015).

38 Turpin, B. J. & Lim, H.-J. Species contributions to PM2. 5 mass concentrations: Revisiting common assumptions for estimating organic mass. *Aerosol Science & Technology* **35**, 602-610 (2001).

39 Lei, X., Xiu, G., Li, B., Zhang, K. & Zhao, M. Individual exposure of graduate students to PM2.5 and black carbon in Shanghai, China. *Environ Sci Pollut Res Int* **23**, 12120-12127, doi:10.1007/s11356-016-6422-x (2016).

40 Mehta, A. J. *et al.* Associations between air pollution and perceived stress: the Veterans Administration Normative Aging Study. *Environ Health* **14**, 10, doi:10.1186/1476-069X-14-10 (2015).

41 Wei, Q. *et al.* LZTFL1 suppresses lung tumorigenesis by maintaining differentiation of lung epithelial cells. *Oncogene* **35**, 2655-2663, doi:10.1038/onc.2015.328 (2016).

42 Schauberger, E. M. *et al.* Identification of ATPAF1 as a novel candidate gene for asthma in children. *J Allergy Clin Immunol* **128**, 753-760 e711, doi:10.1016/j.jaci.2011.04.058 (2011).

43 Cheng, H. *et al.* A High-Throughput Screening Platform Targeting PDLIM5 for Pulmonary Hypertension. *J Biomol Screen* **21**, 333-341, doi:10.1177/1087057115625924 (2016).

44 Kupiec, M. Biology of telomeres: lessons from budding yeast. *FEMS Microbiol Rev* **38**, 144-171 (2014).

45 Pieters, N. *et al.* Biomolecular Markers within the Core Axis of Aging and Particulate Air Pollution Exposure in the Elderly: A Cross-Sectional Study. *Environ Health Perspect* **124**, 943-950, doi:10.1289/ehp.1509728 (2016).

46 Hou, L. *et al.* Air pollution exposure and telomere length in highly exposed subjects in Beijing, China: a repeated-measure study. *Environ Int* **48**, 71-77, doi:10.1016/j.envint.2012.06.020 (2012).

47 Hoxha, M. *et al.* Association between leukocyte telomere shortening and exposure to traffic pollution: a cross-sectional study on traffic officers and indoor office workers. *Environ Health* **8**, 41, doi:10.1186/1476-069X-8-41 (2009).

48 Xia, Y. *et al.* Ambient air pollution, blood mitochondrial DNA copy number and telomere length in a panel of diabetes patients. *Inhal Toxicol* **27**, 481-487, doi:10.3109/08958378.2015.1075090 (2015).

49 Marioni, R. E. *et al.* The epigenetic clock and telomere length are independently associated with chronological age and mortality. *Int J Epidemiol*, doi:10.1093/ije/dyw041 (2016).

50 Greider, C. W. Telomere length regulation. *Annu Rev Biochem* **65**, 337-365, doi:10.1146/annurev.bi.65.070196.002005 (1996).

51 Kloog, I., Nordio, F., Coull, B. A. & Schwartz, J. Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA. *Remote Sensing of Environment* **150**, 132-139 (2014).

52 Kioumourtzoglou, M. A. *et al.* Exposure measurement error in PM2.5 health effects studies: a pooled analysis of eight personal exposure validation studies. *Environ Health* **13**, 2, doi:10.1186/1476-069X-13-2 (2014).

53 Baccarelli, A. *et al.* Rapid DNA methylation changes after exposure to traffic particles. *Am. J. Respir. Crit. Care Med.* **179**, 572-578, doi:10.1164/rccm.200807-1097OC (2009).

54 Ding, R. *et al.* Characteristics of DNA methylation changes induced by traffic-related air pollution. *Mutat Res Genet Toxicol Environ Mutagen* **796**, 46-53, doi:10.1016/j.mrgentox.2015.12.002 (2016).

55 Panni, T. *et al.* 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* **124**, 983-990, doi:10.1289/ehp.1509966 (2016).

56 Zhong, J. *et al.* Cardiac Autonomic Dysfunction: Particulate Air Pollution Effects Are Modulated by Epigenetic Immunoregulation of Toll-like Receptor 2 and Dietary Flavonoid Intake. *J Am Heart Assoc* **4**, e001423, doi:10.1161/JAHA.114.001423 (2015).

57 Hart, J. E. *et al.* The association of long-term exposure to PM2.5 on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. *Environ Health* **14**, 38, doi:10.1186/s12940-015-0027-6 (2015).

58 Zanobetti, A. *et al.* Associations between arrhythmia episodes and temporally and spatially resolved black carbon and particulate matter in elderly patients. *Occup Environ Med* **71**, 201-207, doi:10.1136/oemed-2013-101526 (2014).

59 Colicino, E. *et al.* Long-term exposure to black carbon, cognition and single nucleotide polymorphisms in microRNA processing genes in older men. *Environ Int* **88**, 86-93, doi:10.1016/j.envint.2015.12.014 (2016).

60 Kloog, I. *et al.* A new hybrid spatio-temporal model for estimating daily multi-year PM 2.5 concentrations across northeastern USA using high resolution aerosol optical depth data. *Atmospheric Environment* **95**, 581-590 (2014).

61 Kloog, I. *et al.* Effects of airborne fine particles (PM2.5) on deep vein thrombosis admissions in the northeastern United States. *Journal of Thrombosis and Haemostasis* **13**, 768-774, doi:10.1111/jth.12873 (2015).

62 Gryparis, A., Coull, B. A., Schwartz, J. & Suh, H. H. Semiparametric latent variable regression models for spatiotemporal modelling of mobile source particles in the greater Boston area. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **56**, 183-209 (2007).

63 Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-1369, doi:10.1093/bioinformatics/btu049 (2014).

64 Cawthon, R. M. Telomere measurement by quantitative PCR. *Nucleic Acids Res* **30**, e47 (2002).

65 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86, doi:10.1186/1471-2105-13-86 (2012).

66 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

67 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).

Chapter 3:

Associations between Long-term Exposure to $PM_{2.5}$ Component Species and Blood DNA Methylation Age in the Elderly: The VA Normative Aging Study

Jamaji C. Nwanaji-Enwerem[a†]; Lingzhen Dai[a], Elena Colicino[b]; Youssef Oulhote[a]; Qian Di[a]; Itai Kloog[c]; Allan C. Just[d]; Lifang Hou[e]; Pantel Vokonas[f]; Andrea A. Baccarelli[b]; Marc G. Weisskopf[a]; and Joel D. Schwartz[a]

[a]*Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

[b]*Department of Environmental Health Sciences, Columbia Mailman School of Public Health, New York, NY, USA*

[c]*Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Beer Sheva, Israel*

[d]*Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

[e]*Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA*

[f]*VA Normative Aging Study, Veterans Affairs Boston Healthcare System and the Department of Medicine, Boston University School of Medicine, Boston, MA, USA*

This is a pre-copyedited, author-produced version of an article accepted for publication in *Environmental Epigenetics* following peer review. This version was reproduced with permission from [Nwanaji-Enwerem, Jamaji C., et al. "Associations between long-term exposure to $PM_{2.5}$ component species and blood DNA methylation age in the elderly: The VA normative aging study." *Environment International* 102 (2017): 57-65.] Copyright 2017 Elsevier. The version of record is available online at: http://www.sciencedirect.com/science/article/pii/S0160412016307012. DOI: 10.1016/j.envint.2016.12.024

3.1. Abstract

Background: Long-term $PM_{2.5}$ exposure and aging have been implicated in multiple shared diseases; studying their relationship is a promising strategy to further understand the adverse impact of $PM_{2.5}$ on human health.

Objective: We assessed the relationship of major $PM_{2.5}$ component species (ammonium, elemental carbon, organic carbon, nitrate, and sulfate) with Horvath and Hannum DNA methylation (DNAm) age, two DNA methylation-based predictors of chronological age.

Methods: This analysis included 552 participants from the Normative Aging Study with multiple visits between 2000 and 2011 (n=940 visits). We estimated 1-year $PM_{2.5}$ species levels at participants' addresses using the GEOS-chem transport model. Blood DNAm-age was calculated using CpG sites on the Illumina HumanMethylation450 BeadChip. We fit linear mixed-effects models, controlling for $PM_{2.5}$ mass and lifestyle/environmental factors as fixed effects, with the adaptive LASSO penalty to identify $PM_{2.5}$ species associated with DNAm-age.

Results: Sulfate and ammonium were selected by the LASSO in the Horvath DNAm-age models. In a fully-adjusted multiple-species model, interquartile range increases in both 1-year sulfate (95%CI: 0.28, 0.74, P<0.0001) and ammonium (95%CI: 0.02, 0.70, P=0.04) levels were associated with at least a 0.36-year increase in Horvath DNAm-age. No $PM_{2.5}$ species were selected by the LASSO in the Hannum DNAm-age models. Our findings persisted in sensitivity analyses including only visits with 1-year $PM_{2.5}$ levels within US EPA national ambient air quality standards.

Conclusion: Our results demonstrate that sulfate and ammonium were most associated with Horvath DNAm-age and suggest that DNAm-age measures differ in their sensitivity to ambient particle exposures and potentially disease.

3.2. Introduction

Fine particulate matter ($PM_{2.5}$) remains an inescapable environmental exposure and an enormous global public health concern[1]. It is estimated that at least 2.1 million lives could be saved annually if $PM_{2.5}$ guidelines were adhered to worldwide[2]. For the millions of people exposed to $PM_{2.5}$ daily, understanding the impact of $PM_{2.5}$ on human health is critical for developing interventions aimed at reducing $PM_{2.5}$-related morbidity and mortality globally. Researchers have consistently demonstrated that long-term $PM_{2.5}$ exposure is a major contributor to cardiopulmonary disease[3-8], and emerging evidence suggests that $PM_{2.5}$ is a risk factor for previously unconsidered disease outcomes like cognitive decline[9-11]. Nevertheless, much remains to be understood about how $PM_{2.5}$ contributes to even its most well-documented disease outcomes. One promising strategy to better understand the adverse impact of $PM_{2.5}$ on human health, is to study the relationship of $PM_{2.5}$ with aging. Many studies have implicated $PM_{2.5}$ as a contributor to accelerated aging[12-17]. Moreover, independent of $PM_{2.5}$ exposures, aging is associated with cardiopulmonary disease, cognitive decline, and many other $PM_{2.5}$-related disease outcomes[18-21]. Thus, understanding how $PM_{2.5}$ can contribute to aging, may provide additional insight into other adverse $PM_{2.5}$-related health effects.

DNA methylation-based biomarkers of age have proved to be promising tools in understanding the relationship of $PM_{2.5}$ with aging. These biomarkers have surpassed their initial utility of simply predicting chronological age, and have demonstrated remarkable usefulness in assessing individuals' risk of mortality, malignancy, neurocognitive disease, and other biologically-relevant health endpoints[22-28]. Evidence also suggests that these biomarkers of age are reflective of individuals' past environmental exposures[29]. One such study by our group demonstrated robust associations between $PM_{2.5}$ exposure levels and Horvath DNA methylation (DNAm) age. Horvath DNAm-age is a tissue-independent predictor of chronological age that is calculated from DNA methylation values at 353 chronological age-correlated CpG dinucleotides in Illumina's HumanMethylation450 BeadChip[30]. Specifically, in an elderly cohort and with fully-adjusted models, we observed that a 1 $\mu g/m^3$ increase in 1-year $PM_{2.5}$ exposure was associated with a 0.52-year increase in Horvath DNAm-age[31].

Still, $PM_{2.5}$ is a heterogeneous mixture of carbonaceous fractions, inorganics, and metals; and it is

widely appreciated that $PM_{2.5}$ component species often differ in their health effects[32-35]. The present study builds upon our previous research and examines the relationships of $PM_{2.5}$ component species with both Horvath and Hannum DNAm-age in elderly men. Hannum DNAm-age is also a DNA methylation-based predictor of chronological age, but it is based on measurements from 71 CpG dinucleotides[36]. Only 6 CpG dinculeotides are shared between the Horvath and Hannum metrics. By investigating the relationships of $PM_{2.5}$ component species with these two forms of DNAm-age, we aim to (1) better understand how specific $PM_{2.5}$ species are related to aging, and (2) demonstrate differences in the biological utility of different DNAm-age measures.

3.3. Materials and Methods

3.3.a. Study population

The participants in this analysis were part of the U.S. Veterans Affairs Normative Aging Study (NAS), a longitudinal investigation of aging men established in Eastern Massachusetts in 1963 [37]. The men were free of known chronic medical conditions at enrollment, and returned for onsite, follow-up visits every 3-5 years. During these visits, detailed physical examinations were performed, bio-specimens including blood were obtained, and questionnaire data pertaining to diet, smoking status, and additional lifestyle factors that may impact health were collected. All participants provided written informed consent to the VA Institutional Review Board (IRB), and both the Harvard T.H. Chan School of Public Health and VA IRBs granted human subjects approval.

All NAS men with continued study participation as of the year 2000, when $PM_{2.5}$ component levels became available, were eligible for our study sample. After excluding participants with a diagnosis of leukemia (n=11), due to its potential influence on the DNA methylation of blood cells[30], and those incomplete for the covariates of interest (n=16), we had a total of 552 participants with 940 observations between the years 2000 and 2011. Of the 552 participants, 249 (45%) had one visit, 218 (40%) had two visits, and 85 (15%) had three or more visits.

3.3.b. DNA Methylation and calculation of DNAm-age

Laboratory staff extracted DNA from the buffy coat of whole blood collected from each participant at each NAS follow-up visit (QIAamp DNA Blood Kit, QIAGEN, Valencia, CA, USA). DNA samples were then treated with bisulfite conversion (EZ-96 DNA Methylation Kit, Zymo Research, Orange, CA, USA) and hybridized to the 12 sample Illumina HumanMethylation450 BeadChips (Infinium HD Methylation protocol, Illumina, San Diego, CA, USA). To ensure a similar age distribution and avoid confounding across chips and plates, study staff employed a two-stage age-stratified algorithm to randomize samples. For quality control, study staff removed samples where >5% of probes had a beadcount < 3 or > 1% of probes had a detection P-value >0.05. The Bioconductor minfi package Illumina-type background correction without normalization was used to preprocess the remaining samples and generate methylation beta values [38]. The beta values represent the percentage of methylation for each of the ~480,000 CpG sites in the BeadChip array. The 450k arrays were run in the Genomics Core Facility at Northwestern University.

To explore potential differences in the relationship of $PM_{2.5}$ and $PM_{2.5}$ species with different forms of DNAm-age, we calculated both Horvath DNAm-age and Hannum DNAm-age using the 450k beta values and Horvath's publically available online calculator (http://labs.genetics.ucla.edu/horvath/dnamage/). Horvath DNAm-age was derived from an elastic net (penalized regression) using multiple data sets of varying tissue and cell types. 21,369 CpG probes, shared by the Illumina HumanMethylation27 and HumanMethylation450 BeadChip platforms were regressed on a calibrated version of chronological age. The elastic net selected 353 CpGs that correlate with age, and the resulting model coefficients are used by the calculator to predict the age of each DNA sample (DNAm-age) [30]. Hannum DNAm-age was also derived using an elastic net. However, Hannum DNAm-age was based on a single cohort where DNA methylation values were calculated from whole blood. This elastic net selected 71 CpG probes in the Illumina HumanMethylation450 BeadChip that are predictive of chronological age. Hannum DNAm-age was calculated as the sum of the beta values multiplied by the reported effect sizes for the Hannum predictor[36]. The Hannum and Horvath DNAm-ages only share 6 CpG probes (cg04474832, cg05442902, cg06493994, cg09809672, cg19722847, and cg22736354).

3.3.c. Assessment of Environmental Factors: Ambient Particles and Temperature

We employed the widely used GEOS-chem chemical transport model (http://www.geoschem.org) to generate 1-year exposure estimates for $PM_{2.5}$ and the following major $PM_{2.5}$ component species: organic carbon (OC), elemental carbon (EC), sulfate, nitrate, and ammonium [39]. These 5 component species were selected because they make up a large fraction of total $PM_{2.5}$ mass (~ 88.6%) and were best predicted by the model. GEOS-chem incorporates nonlinear chemistry, meteorology, and detailed emissions inventories to simulate the formation and transportation of atmospheric components to give raw estimates of $PM_{2.5}$ and its major chemical components. Ten-fold cross-validation demonstrated that the model performs well for $PM_{2.5}$ mass and its component species with R^2s ranging from 0.70 to 0.88 [40]. We generated daily estimates at the 1x1 km area resolution. Each participant's residence was geocoded and linked to an area level grid-point. Time spent away from home (>7 days) and address changes were also accounted for as particle estimates were assigned to each participants' address. Given that >90% of NAS participants are retired, home address exposures are expected to be a good proxy for their individual ambient exposures. We then generated 1-year total $PM_{2.5}$ and $PM_{2.5}$ component species exposure windows by averaging daily exposures for the 365 days prior to the day of each participants' NAS visit. The 1-year $PM_{2.5}$ exposure window was utilized because it has been previously reported to be robustly associated with DNAm-age[31].

We used a spatiotemporal prediction multi-step approach to generate temperature (in Celsius) for each participant [41]. First, we obtained 1x1 km resolution daily physical surface temperature (Ts) data from NASA satellite measurements and daily near surface air (Ta) data from the Environmental Protection Agency, National Climatic Data Center, and Weather Underground Inc. We then used mixed model regression to calibrate Ts to Ta. The model was validated with a mean out of sample R^2 of 0.95. To generate 1-year temperature measurements to complement 1-year particle exposures, we averaged daily temperature measurements over the 365 days prior to participants' NAS visits.

3.3.d. Statistical Analysis

We first used generalized linear mixed effects models to determine the relationship of DNAm-age (Horvath and Hannum independently) with 1-year $PM_{2.5}$ exposure levels and 1-year $PM_{2.5}$ component species exposure levels. All linear mixed effects models included a random participant-specific intercept to account for correlation between repeated measures (*i.e.* multiple visits for a participant).

We adjusted for confounders and covariates that have *a priori* biological/clinical relevance and/or are reported in the existing literature. Specifically, our previous publication was the first study examining associations of ambient particles and DNAm-age[31]. There, we used a tiered approach of adding confounders and covariates based on known relationships of ambient particles with DNA methylation and known relationships of ambient particles with older markers of aging[30,42-45]. Tier one adjusted for chronological age and blood cell types. Tier two made additional adjustments for lifestyle and environmental factors. Tier three expanded on tier two by additionally adjusting for age-related diseases, and tier four expanded on tier two by additionally adjusting for medications of age-related diseases. After considering model fit (assessed via AIC) and considering biological factors that are known to be important, the tier two covariates were deemed to be most appropriate. Thus, in line with the previously published tier two framework[31], the models for this analysis were adjusted for chronological age (continuous), six blood cell type estimates [*i.e.* plasma cells, CD4+ lymphocytes, CD8+ lymphocytes, natural killer (NK) cells, monocytes, and granulocytes] (continuous) determined via Houseman and Horvath methods[30,46], average 1-year temperature (continuous), cumulative cigarette pack years (continuous), smoking status (current, former, or never), season of visit (spring [March-May], Summer [June-August], Fall [September-November], and Winter [December-February]), body mass index (BMI) (lean [<25], overweight [25-30], obese [>30]), alcohol intake (yes/no ≥ 2 drinks daily), and maximum years of education (continuous). All $PM_{2.5}$ component species models were additionally adjusted for $PM_{2.5}$ mass[47].

To more rigorously identify the $PM_{2.5}$ component species that may be associated with DNAm-age, we applied the adaptive LASSO (least absolute shrinkage and selection operator)[48]. Given that $PM_{2.5}$ component species are correlated, placing them together within the same standard linear regression model can result in unaccounted for stochastic errors. The LASSO is a regression shrinkage and selection approach

that helps overcome such limitations. The LASSO applies an l_1 penalty on the component regression coefficients, which minimizes the sum of squared errors subject to the sum of the absolute values of the coefficients being less than a given value[49]. The adaptive LASSO improves upon this procedure by utilizing weights for penalizing different coefficients in the l_1 penalty to identify a subset of model predictors to achieve asymptotic normality[50]. Furthermore, the adaptive LASSO has been successfully applied in air pollution and health research[51,52].

To identify and select $PM_{2.5}$ component species associated with DNAm-age, we applied a penalty to all $PM_{2.5}$ component species, but not to $PM_{2.5}$ mass and the other covariates in the model. λ, the penalty parameter, determines how strongly the magnitude of the $PM_{2.5}$ species regression coefficients are constrained. When λ is small, the regression coefficients are weakly penalized and mirror those that would be given from a standard linear mixed effects model. When λ is large, the coefficients are strongly penalized, shrinkage is maximized, and all coefficients tend towards zero such that the resulting model includes fixed covariates only. When λ takes a value in between the extremes, the result is a penalized model where some $PM_{2.5}$ component species will have coefficients of zero and others will be non-zero. $PM_{2.5}$ component species with non-zero coefficients are considered as "selected" by the adaptive LASSO. We ran the model across a range of λs, beginning with a λ of 0, and selected the λ resulting in the model with the smallest Bayesian Information Criterion (BIC)[53]. Following LASSO selection, we fit a final multiple-species linear mixed effects model using the selected $PM_{2.5}$ component species and our fixed covariates. From this final model, we were able to estimate component species effect sizes and their corresponding 95% confidence intervals.

Additionally, we considered that the LASSO may not select the $PM_{2.5}$ species that are most correlated with total $PM_{2.5}$ mass. Thus, we conducted a sensitivity analysis where we performed LASSO selection without adjusting for $PM_{2.5}$ mass. From this sensitivity analysis model, we fit a multiple-species linear mixed effects model using the selected $PM_{2.5}$ component species and estimated component species effect sizes and their corresponding 95% confidence intervals.

After finding that Horvath DNAm-age alone was significantly associated with $PM_{2.5}$ component species, we evaluated the relationships of the DNA methylation values of each of the 353 Horvath CpG probes with the particles in the aforementioned LASSO-selected multiple-species linear mixed effects model. In addition to the already described covariates, we included technical covariates (450k plate, chip, row, and column) to this analysis. To account for multiple hypothesis testing, we also performed FDR correction in this analysis. We then performed gene ontology analysis on the list of significant CpGs (FDR P-value < 0.05) using the publically available GoTermFinder tool (http://go.princeton.edu/cgi-bin/GOTermFinder).

In an additional sensitivity analyses, we re-ran our models excluding participant visits with $PM_{2.5}$ exposures greater than 12 $\mu g/m^3$. This allowed us to assess if our findings persisted even at the $PM_{2.5}$ levels currently deemed acceptable by the U.S. Environmental Protection Agency (EPA) National Ambient Air Quality Standards (NAAQS)[54].

All statistical analyses were performed using R Version 3.1.1 (R Core Team, Vienna, Austria) and we considered a P-value < 0.05 to be statistically significant.

3.4. Results

3.4.a. Descriptive Results

Table 1 summarizes the characteristics of the study population. All study participants were Caucasian males with a mean (± SD) age of 74.7 ± 6.99 years across all study visits. Average Horvath DNAm-age and Hannum DNAm-age were 74.0 ± 7.92 years and 75.1 ± 8.95 years respectively. Horvath DNAm-age ($r = 0.59$, $p<0.0001$) and Hannum DNAm-age ($r = 0.77$, $p<0.0001$) were both strongly correlated with chronological age in the study population. Both measures of DNAm-age were also strongly correlated to each other ($r = 0.69$, $p<0.0001$).

Table 1. Characteristics of Study Subjects (2000 – 2011)

Variable		First Visit (N = 552)	All Visits (N = 940)
Age (years), mean (SD)		73.3 (6.82)	74.7 (6.99)
Horvath DNAm-age (years), mean (SD)		73.7 (7.77)	74.0 (7.92)
Hannum DNAm-age (years), mean (SD)		73.8 (8.80)	75.1 (8.95)
Temperature (°C), mean (SD)		11.5 (1.12)	11.3 (1.00)
Pack years, mean (SD)		20.7 (24.7)	20.5 (24.4)
Smoking Status, N (%)			
	Current	25 (4)	40 (4)
	Former	355 (64)	614 (65)
	Never	172 (32)	286 (31)
Season, N (%)			
	Spring	145 (26)	241 (26)
	Summer	115 (21)	199 (21)
	Fall	177 (32)	313 (33)
	Winter	115 (21)	187 (20)
BMI, N (%)			
	Healthy/Lean	119 (21)	216 (23)
	Overweight	302 (55)	493 (52)
	Obese	131 (24)	231 (25)
Alcohol Consumption, N (%)			
	< 2 drinks/day	441 (80)	761 (81)
	≥ 2 drinks/day	111 (20)	179 (19)
Education, N (%)			
	≤ 12 years	146 (27)	242 (26)
	12 - 16 years	262 (47)	434 (46)
	> 16 years	144 (26)	264 (28)

Table 2 reports 1-year $PM_{2.5}$ and $PM_{2.5}$ component species exposure levels across all study visits. The participants had a mean 1-year $PM_{2.5}$ exposure level of 10.3 ± 1.60 μg/m^3, with an interquartile range (IQR) of 2.16 μg/m^3. Of the measured $PM_{2.5}$ component species, sulfate accounted for the largest proportion of total $PM_{2.5}$ mass (33%), followed by organic carbon (28.6%), nitrate (11.5%), ammonium (10.1%), and elemental carbon (5.4%). OC was the $PM_{2.5}$ species most correlated with total $PM_{2.5}$ mass (r = 0.67). 1-year $PM_{2.5}$ and $PM_{2.5}$ species Pearson correlations across all visits are reported in Table S1. Moreover, 1-year $PM_{2.5}$ and $PM_{2.5}$ species exposure distributions across first visits are reported in Table S2.

Table 2. Mean 1-Year Particulate Matter 2.5 ($PM_{2.5}$) and Component Species Concentrations Across All Study Visits

Particle ($\mu g/m^3$)	Mean (SD)	IQR	Proportion of $PM_{2.5}$ (%)	Pearson Correlation with $PM_{2.5}$	N
$PM_{2.5}$	10.3 (1.60)	2.16	-	-	940
$PM_{2.5}$ Component Species					
EC	0.56 (0.17)	0.23	5.4	0.62	940
OC	2.94 (0.91)	1.28	28.6	0.67	940
Sulfate	3.40 (1.23)	0.82	33.0	0.30	940
Nitrate	1.18 (0.32)	0.42	11.5	0.46	940
Ammonium	1.04 (0.31)	0.3	10.1	0.53	940

3.4.b. 1-Year $PM_{2.5}$ and $PM_{2.5}$ Component Species as Predictors of DNAm-age

Table 3 summarizes the results of three model frameworks where $PM_{2.5}$ and its component species were modeled as predictors of both Horvath and Hannum DNAm-age. Residuals from all models appeared normally distributed. In the model framework 1, $PM_{2.5}$ was modeled as a predictor of Horvath and Hannum DNAm-age independently. In the fully adjusted model, an IQR increase in 1-year $PM_{2.5}$ exposure was significantly associated with a 0.64-year increase in Horvath DNAm-age (p=0.005). However, an IQR increase in 1-year $PM_{2.5}$ exposure was not significantly associated with Hannum DNAm-age (β=0.06, p=0.74). Under the model framework 2, each $PM_{2.5}$ component species was modeled as an independent predictor of Horvath and Hannum DNAm-age adjusting for all covariates and total $PM_{2.5}$ mass. 1-year IQR increases in OC (β=0.93, p=0.001), sulfate (β=0.59, p<0.0001), nitrate (β=0.58, p=0.01), and ammonium (β=0.59, p=0.0004) were all significantly associated with increases in Horvath DNAm-age of at least 0.58 years. No $PM_{2.5}$ component species were significantly associated with Hannum DNAm-age (Table 3).

The model 3 framework reflects the results of the multiple-species fully-adjusted linear mixed effects models with the $PM_{2.5}$ component species selected by the adaptive LASSO. The adaptive LASSO selected sulfate and ammonium as important predictors of Horvath DNAm-age. Figure 1A depicts the relationship between BIC, the model selection criterion, and λ, the adaptive LASSO penalty parameter. The

model with the smallest BIC had $\lambda = 11$. Figure 1B shows the LASSO coefficient paths for the $PM_{2.5}$ component species. Each component species coefficient is expressed as the difference in mean Horvath DNAm-age per an IQR increase in the 1-year component species exposure level. Each curve depicts the rate at which the component species coefficient shrinks towards zero as λ increases. At $\lambda=0$, all components species have a non-zero coefficient.

Table 3. 1-Year Particulate Matter 2.5 ($PM_{2.5}$) and Component Species as Predictors of DNA Methylation (DNAm) Age

Particle	Difference in Horvath DNAm-age for IQR (95% CI)	*P*	Difference in Hannum DNAm-age for IQR (95% CI)	*P*	N
Model Framework 1					
$PM_{2.5}$	0.64 (0.20, 1.09)	**0.005**	0.06 (-0.28, 0.40)	0.74	940
Model Framework 2					
EC	0.27 (-0.25, 0.80)	0.30	-0.09 (-0.48, 0.29)	0.64	940
OC	0.93 (0.37, 1.50)	**0.001**	0.35 (-0.05, 0.77)	0.09	940
Sulfate	0.59 (0.37, 0.81)	**<0.0001**	0.08 (-0.09, 0.25)	0.36	940
Nitrate	0.58 (0.11, 1.04)	**0.01**	0.30 (-0.04, 0.65)	0.08	940
Ammonium	0.59 (0.26, 0.92)	**0.0004**	0.06 (-0.18, 0.30)	0.63	940
Model Framework 3					
$PM_{2.5}$	0.18 (-0.30, 0.66)	0.45	-	-	940
Sulfate	0.51 (0.28, 0.74)	**<0.0001**	-	**-**	940
Ammonium	0.36 (0.02, 0.70)	**0.04**	-	**-**	940

Model Framework 1: adjusted for chronological age, blood cell types, temperature, pack years, smoking status, season, BMI, alcohol consumption, and education. *Model Framework 2*: $PM_{2.5}$ component species as independent predictors of DNAm-age adjusted for $PM_{2.5}$ in addition to model 1 covariates. *Model Framework 3*: $PM_{2.5}$, sulfate, and ammonium as joint predictors of DNAm-age (given selection of sulfate and ammonium by the adaptive LASSO) adjusted for model 1 covariates. No species were selected as predictors of Hannum DNAm-age.

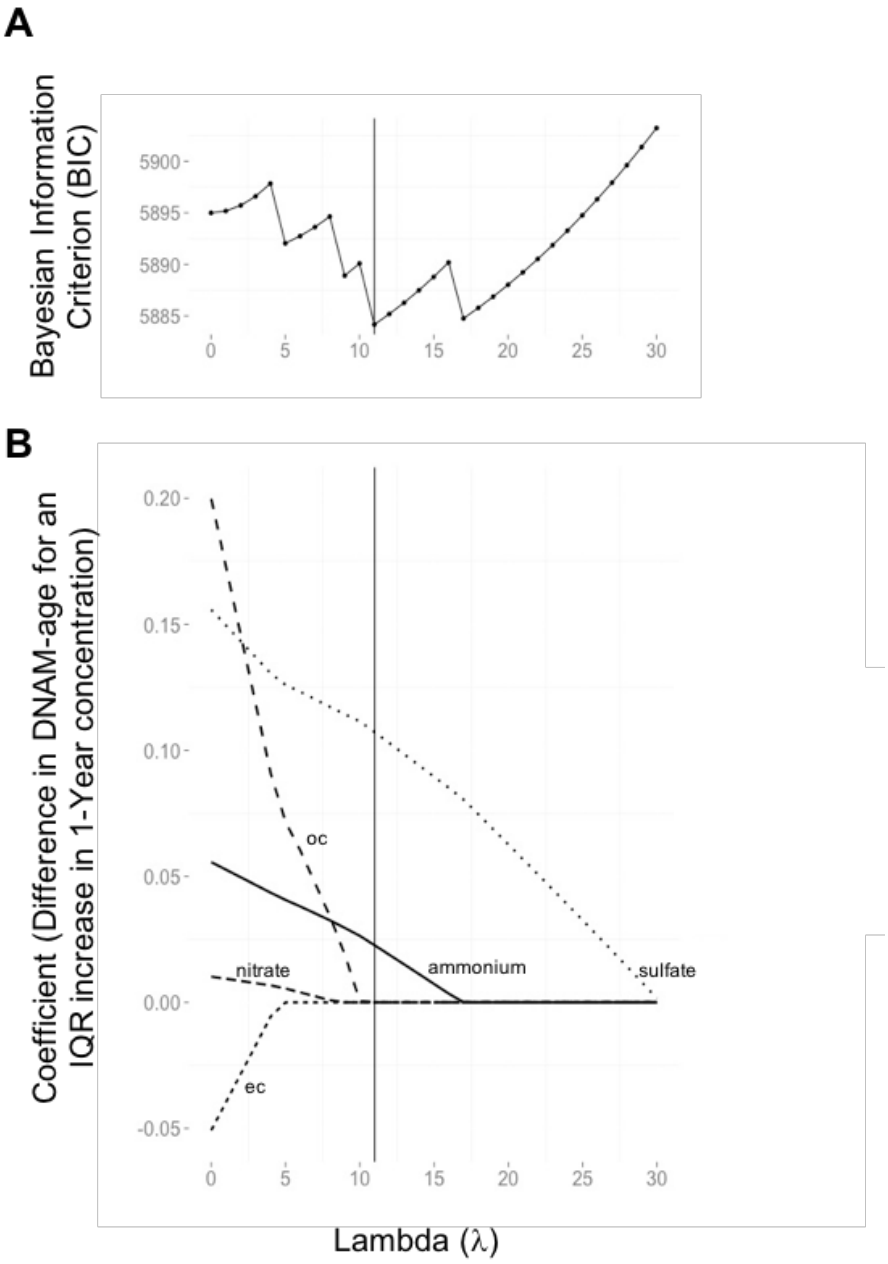

Figure 1 | A) The relationship between BIC, a criterion for model selection and λ (lambda), the adaptive LASSO penalty parameter, for DNAm-age. The vertical line at λ = 11 denotes the penalty parameter with the lowest BIC. B) LASSO coefficient paths: plot of coefficient profiles for $PM_{2.5}$ components as a function of λ. At λ = 11, sulfate and ammonium are the only $PM_{2.5}$ components with a non-zero coefficient.

In the multiple-species fully-adjusted linear mixed effects model, both sulfate (β=0.51, p<0.0001) and ammonium (β=0.36, p=0.04) remain significant positive predictors of Horvath DNAm-age. The adaptive LASSO did not select any $PM_{2.5}$ component species as important predictors of Hannum DNAm-age.

In our sensitivity analysis – where LASSO selection was performed without adjusting for total $PM_{2.5}$ mass – sulfate, ammonium, and OC were selected as important predictors of DNAm-age (Figure S1). Nonetheless, in a multiple-species fully-adjusted linear mixed effects model, both sulfate (β=0.45, p=0.0003) and ammonium (β=0.34, p<0.05) remained significant positive predictors of Horvath DNAm-age, but OC (β=0.42, p=0.16) was not a significant predictor of Horvath DNAm-age (Table S3). Again, the sensitivity analysis adaptive LASSO did not select any $PM_{2.5}$ component species as important predictors of Hannum DNAm-age.

Significant findings from the main analysis multiple-species fully-adjusted linear mixed effects model persisted in the second sensitivity analyses excluding participant visits with $PM_{2.5}$ exposures greater than 12 $\mu g/m^3$, the annual $PM_{2.5}$ exposure level currently deemed acceptable by the U.S. Environmental Protection Agency (EPA) National Ambient Air Quality Standards (NAAQS) (Table S4).

3.4.c. Associations between 1-Year $PM_{2.5}$ and $PM_{2.5}$ Component Species Levels and Methylation Values at Horvath DNAm-age CpG Sites

After FDR correction, 47 out of 353 Horvath DNAm-age CpG sites had methylation values that were significantly associated with total $PM_{2.5}$ levels in the fully-adjusted multiple-species linear mixed effects model. $PM_{2.5}$ levels were positively or negatively associated with CpG methylation values depending on the CpG site (Table 4). 46 of the 47 CpG sites mapped to known

Table 4. 1-Year Particle Exposures as Predictors of Horvath CpG Probe

CpG	Gene	Process/Function	Difference in Methylation per SD (%)	Direction of Association	FDR Adjusted *P*
		$PM_{2.5}$			
cg15262928	TIMM17A	Mitochondrial protein import	24.18	+	0.001
cg14409958	ENPP2*	Nucleic acid binding	19.85	+	0.001
cg01570885	FAM50B*	Protein binding	19.64	-	0.004
cg08186124	LZTFL1*	Protein binding: cytoplasm	19.59	+	0.004
cg18139769	SGCE*	Calcium binding	19.25	-	0.004
cg15547534	PPP1R35	Phosphatase binding	18.79	+	0.004
cg26456957	PPP1R12C	Protein kinase binding	18.74	+	0.001
cg05847778	BBS5	Transcription initiation	18.01	+	0.006
cg15661409	C14orf105*	Uncharacterized	17.52	-	<0.001
cg02335441	NEK11	DNA replication	17.32	+	0.008
cg17285325	TYMP	Phosphorylase activity	17.23	+	0.007
cg04094160	ZBTB5*	Transcriptional regulation	16.92	+	0.003
cg03682823	SGCE	Calcium binding	16.80	-	0.008
cg07663789	NPR3	Hormone binding/blood volume	16.13	+	0.003
cg15703512	PDZD9	Uncharacterized	15.80	+	0.015
cg22190114	NLRP8	ATP binding	15.60	+	0.015
cg19008809	SFMBT1	Transcription corepressor activity	15.48	+	0.013
cg00374717	ARSG	Sulfatase enzyme activity	15.29	-	0.004
cg12985418	MIB1	Protein binding	15.14	+	0.018
cg03588357	GPR68	G-protein coupled receptor activity	15.09	+	0.020
cg14424579	AGBL5	Metallocarboxypeptidase	15.07	+	0.007
cg14597908	GNAS	G-protein binding	15.04	-	0.015
cg19044674	LEPRE1	Oxidoreductase activity	14.88	+	0.023
cg09441152	PQLC1	Membrane component	14.87	+	0.027
cg07849904	MN1	Transcriptional activator	14.85	+	0.015
cg19273182	PAPOLG	Polynucleotide adenylyltransferase activity	14.81	+	0.025
cg17063929	NOX4	Nucleotide binding	14.56	-	0.015
cg24116886	DEFB127	Immunologic response	14.40	-	0.015
cg09191327	PRDM12	Methyltransferase activity	14.30	+	0.027
cg23662675	ZMYND8	Transcription cofactor activity	14.16	+	0.014
cg14992253	EIF3I	Translation initiation	13.12	-	0.018

Table 4. 1-Year Particle Exposures as Predictors of Horvath CpG Probe (Continued)

cg05442902	P2RX6	Channel activity	12.82	-	0.026
cg06557358	TMEM132E*	Integral component of membrane	12.60	+	0.031
cg11932564	TNFRSF13C	Immunologic response	12.51	+	0.037
cg18031008	MRPS21	Mitochondrial ribosome	12.10	+	0.030
cg19945840	SDF4	Calcium binding	12.08	-	0.023
cg19167673	PDGFB	Protein homodimerization activity	12.03	+	0.031
cg25159610	PLK2	Cell division	11.96	+	0.038
cg22006386	CATSPERG*	Ion channel activity	11.89	+	0.026
cg27377450	*unknown*	*unknown*	11.85	-	0.026
cg20100381	NAE1	Protein heterodimerization activity	11.84	+	0.046
cg04268405	CHST3	Sulfotransferase	11.47	-	0.026
cg07595943	ADAD2	RNA binding	11.36	-	0.023
cg25505610	EIF3M	Translation initiation	10.67	+	0.042
cg16744741	PRKG2	Protein kinase activity	10.17	-	0.038
cg21395782	NDUFA13*	NADH dehydrogenase activity	8.27	+	0.027
cg01459453	SELP	Oligosaccharide binding	8.05	-	0.023
		Ammonium			
cg02275294	SOAT1	Fatty-acyl-CoA binding	10.81	+	0.036

All models are fully adjusted. * = CpGs associated with $PM_{2.5}$ levels in a prior publication.

genes. 9 of these 46 genes (*ENPP2*, *FAM50B*, *LZTFL1*, *SGCE*, *C14orf105*, *ZBTB5*, *TMEM132E*, *CATSPERG*, and *NDUFA13)* were previously reported in a similar, previously published $PM_{2.5}$ Horvath CpG analysis[31]. Gene ontology of our 46 genes combined with the genes in the previously reported study returned the GO term "regulation of translational initiation" (Table S5).

Only 1 out 353 CpG sites (cg02275294) had methylation values that were significantly associated with ammonium levels in the fully-adjusted multiple-species linear mixed effects model. No individual CpG sites had methylation values that were significantly associated with sulfate levels after FDR correction.

3.5. Discussion

In this study, we report positive associations of 1-year $PM_{2.5}$ exposure levels with Horvath DNAm-age in a population of community-dwelling, elderly men. Additionally, we utilized the adaptive LASSO to identify 1-year sulfate and ammonium levels as the $PM_{2.5}$ components most robustly associated with Horvath DNAm-age. To our knowledge, this is the first report of associations of multiple $PM_{2.5}$ component species with DNAm-age and the second time that satellite-derived $PM_{2.5}$ exposure levels have been found to be associated with Horvath DNAm-age. In addition to being consistent with the existing literature[31], our findings also demonstrate important public health relevance as they persist in sensitivity analyses including only participant visits with 1-year $PM_{2.5}$ levels within current US EPA national ambient air quality standards[54]. Our study also extends the literature by exploring $PM_{2.5}$ relationships with Hannum DNAm-age although these relationships were found to be null. Furthermore, we identified 47 CpG sites, 9 of which were previously reported, whose methylation values were significantly associated with $PM_{2.5}$ levels in fully-adjusted linear mixed effects models. Only 1 CpG was associated with ammonium levels and 0 were associated with sulfate levels.

Given our prior report of robust associations of $PM_{2.5}$ levels from satellite-based spatiotemporal models with Horvath DNAm-age, we expected to observe a similar positive relationship using $PM_{2.5}$ levels from the GEOS-chem chemical transport model. As expected, we observed that an IQR increase in 1-year $PM_{2.5}$ exposure was associated with a 0.64-year increase in Horvath DNAm-age. Since $PM_{2.5}$ component species are highly related to total $PM_{2.5}$, we also expected that $PM_{2.5}$ component species would be associated with Horvath DNAm-age, even when adjusting for $PM_{2.5}$ mass. Given the existing literature concerning the differential health effects of $PM_{2.5}$ component species, we speculated that some component species may be more robustly associated with Horvath DNAm-age than others. In particular, we expected the carbonaceous fractions to be among the species most robustly associated with DNAm-age due to the extensive literature (including work from our group) on the adverse nature of carbonaceous fraction exposures on health[31,43,55-57]. In our fully adjusted one-species linear mixed effects models, we observed strong positive associations of 4 out of the 5 component species examined with Horvath DNAm-age. IQR range increases in organic

carbon, sulfate, nitrate, and ammonium were all significantly associated with at least a 0.58-year increase in Horvath DNAm-age.

Despite the results from our fully adjusted one-species linear mixed effects models, we desired a method to more comprehensively identify the component species most associated with DNAm-age. Nevertheless, we were aware that simply modeling highly-correlated $PM_{2.5}$ species together would result in unaccounted for stochastic errors. Thus, we employed the adaptive LASSO as a penalized regression method to help overcome this difficulty. The literature has shown that carbonaceous fractions are robustly associated with age-related health outcomes[31,43,55-57]; however, neither elemental or organic carbon were selected in our models. Rather sulfate and ammonium were selected. This difference may be explained by the fact that a majority of the aforementioned studies did not consider other $PM_{2.5}$ component species in addition to the carbonaceous fractions. Even in our single-species linear mixed effects models, we note that organic carbon was among the four species significantly associated with Horvath DNAm-age (Table 3). However, when all five component species are considered together in the adaptive LASSO, only sulfate and ammonium were selected. It is also possible that the LASSO did not select the carbonaceous fractions because the selection was performed under $PM_{2.5}$ adjustment and $PM_{2.5}$ may be capturing most of the variability of organic and elemental carbon. Thus, we performed LASSO selection not adjusting for total $PM_{2.5}$ mass as a sensitivity analysis. This time LASSO did select organic carbon along with sulfate and ammonium. However, when these three component species were modeled with $PM_{2.5}$ in a multiple-species fully-adjusted linear mixed effects model, organic carbon was the only species that was not a significant predictor of DNAm-age. This suggests that organic carbon was selected in the sensitivity analysis because of its strong correlation with $PM_{2.5}$ mass and not because organic carbon itself is a good predictor of DNAm-age. This finding also reiterates the notion that adjustment for $PM_{2.5}$ mass in component species models is very important as $PM_{2.5}$ mass often confounds the relationship between the outcome and species[47]. Failing to include $PM_{2.5}$ mass may lead to misleading findings about species. In all, our data suggests that of the considered species, sulfate and ammonium have the most important relationships with DNAm-age. Furthermore, existing studies that do consider a range of $PM_{2.5}$ components demonstrate that other non-

carbonaceous components are important to age-related outcomes[52,58,59]. These data, together with our findings, also suggests the important need to consider a range of $PM_{2.5}$ components, rather than one or two species, in air pollution and health studies.

Both sulfate and ammonium are classified in the inorganic fraction of $PM_{2.5}$. Sulfates are often produced from oxidation or photochemical reactions involving primary gases derived from sources like coal-burning power plants[60]. Additionally, ammonia from organic sources including animal feeds and fertilizers can contribute to the existence of sulfates in the form of atmospheric ammonium sulfate[61]. As far as direct ambient sulfate and ammonium toxicity to human health is concerned, existing studies are limited. Yet, there has been extensive evidence describing the ability of acidic sulfates, like ammonium sulfate, to increase the number and toxicity of biologically harmful secondary particles[47,62-66]. For instance, ammonium sulfate aerosols have been shown to influence the photo-chemical reactions of nitrogen oxides and toluene hastening the production of secondary organic aerosols[67]. Moreover, sulfur concentrations have been found to be directly proportional to the ability of soluble particle extracts to generate biologically damaging oxidants[68]. Furthermore, a prior study in the NAS has reported a 27% decrease in long interspersed nucleotide element-1 methylation per every IQR increase in 90-day sulfate exposure. This study provides evidence for the influence of sulfates on DNA methylation, which may be a potential pathway for sulfate toxicity[42]. It is still unclear what the molecular relevance of Horvath DNAm-age is, but our findings along with the existing literature will be helpful in providing additional insight for future work.

Following the selection of sulfate and ammonium by the adaptive LASSO, we constructed a final multiple-species linear mixed effects model adjusted for $PM_{2.5}$ mass and all covariates. Even in this model, sulfate and ammonium remained significant positive predictors of Horvath DNAm-age. We then looked to see if there were specific Horvath DNAm-age component CpG sites with methylation values that were associated with $PM_{2.5}$, sulfate, and/or ammonium in our fully-adjusted multiple-species linear model. From this analysis, we identified 47 significant CpG sites after FDR adjustment. These sites mapped to 46 genes, and 9 of them were reported in a previous CpG-level analysis of the same 353 sites in the Horvath algorithm that we conducted using $PM_{2.5}$ levels from a satellite-based spatiotemporal model. To better grasp the

impact of $PM_{2.5}$ levels on methylation, we divided the coefficients for each significant CpG site (*i.e.* difference in methylation per IQR increase in particle level) by the standard deviation of the respective particle level. We were pleased to see that 5 of the 9 CpGs that were shared between both $PM_{2.5}$ prediction models were in the top 20% of our gene list. We then combined the gene lists from both $PM_{2.5}$ prediction models (removing any duplicates) and performed a gene ontology (GO) analysis. The GO analysis returned the term "regulation of translational initiation" with the following genes from our list falling into this category: RXRA, EIF3M, EIF31. Though the GO term itself is not highly specific, combining this pathway with what is known about the toxicity of $PM_{2.5}$ will be useful in further understanding how $PM_{2.5}$ may contribute to aging and disease. Only 1 CpG was associated with ammonium levels and it mapped to the gene SOAT1, which is involved in fatty-acyl-CoA binding. SOAT1 has been implicated in a number of diseases including familial hypercholesterolemia[69]. No CpG sites were specifically associated with sulfate levels. The finding that almost no CpGs sites were associated with ammonium and sulfate further demonstrates that Horvath DNAm-age is simply not a reflection of its 353 component CpGs, and reiterates the need for work focused on defining the molecular relevance of DNAm-age.

Finally, our study demonstrates that all DNAm-age measures are not the same. In the literature, there is evidence of both Horvath and Hannum DNAm-age reflecting the same disease outcome and evidence where they differ in their reporting ability. For instance, both Horvath and Hannum DNAm-age appear to be useful in predicting mortality[70,71]. However, in a study of male and female veterans, Hannum DNAm-age was associated with post-traumatic stress disorder and neural integrity, but Horvath DNAm-age was not[72]. The differences in these two DNAm-age measures may stem from the fact that they are derived from almost entirely different CpG sites or from the fact that Horvath DNAm-age was constructed using many datasets of multiple tissue types and the Hannum DNAm-age was based only on blood from one dataset[30,36]. Our results suggest that Hannum DNAm-age is not sensitive to exposure levels of $PM_{2.5}$ and its component species. Additional studies in different populations will be necessary to confirm these findings more broadly. Nonetheless, continued research exploring the specific sensitivity of DNAm-age measures will be a crucial next step in the growth of this field of research. Once more is known about the

profiles of these markers, we can begin to use them more effectively in answering questions concerning human health.

Strengths of our study include rigorous statistical methods and access to a large cohort with extensive and repeated information regarding pollutant exposures, potential confounders, and DNA methylation data from multiple study visits. However, our study does have several limitations. First, although we used a validated chemical transport model to estimate the levels of ambient $PM_{2.5}$ and its component species at participants' addresses, we recognize that these estimates may differ from personal exposures. Nonetheless, we know that a majority of NAS participants are retired and spend most of their time at home. Moreover, our approach is expected to result in non-differential misclassification that is likely to underestimate the observed associations rather than bias them away from the null[73]. Secondly, it is known that LASSO regression is limited to linear relationships. Given the linear relationship of our particle exposures with DNAm-age and the scope of this paper, the adaptive LASSO was a good tool for identifying $PM_{2.5}$ components that are independently important to DNAm-age. However, for future studies potentially interested in the interactions between $PM_{2.5}$ components, another technique may be necessary as $PM_{2.5}$ species interactions that are important for the prediction of DNAm-age may be more complex (*i.e.* not linear). Third, we note that our findings are based on an elderly cohort of Caucasian males that reside in a lightly-polluted environment. Hence, additional studies involving other demographic groups and in different environments will be necessary to confirm our findings more broadly. Finally, we used the existing literature and *a priori* knowledge of biological/clinical relevance to adjust for potential confounders. Nonetheless, we cannot rule out the possibility of unknown or residual confounding in our analyses.

3.6. Conclusion

Our study utilizes the GEOS-chem chemical transport model to validate novel positive associations between long-term $PM_{2.5}$ exposure levels and Horvath DNAm-age. For the first time, we demonstrate that sulfate and ammonium are among the $PM_{2.5}$ component species most associated with Horvath DNAm-age in this population of elderly men. In contrast, we observed no relationships of long-term $PM_{2.5}$ and $PM_{2.5}$

component species exposure levels with Hannum DNAm-age. These results suggest that DNA methylation-based biomarkers of age differ in their sensitivity to ambient particle exposures and potentially disease outcomes. Future studies in other populations will be critical for defining the environmental and disease sensitivity profiles of DNAm-age measures.

3.7. Contributions and Support

Contributors: JCN and JDS conceived and designed the study. EC, QD, IK, ACJ, LH, and PV gathered data. JCN performed the data analyses and drafted the manuscript. LD, JDS, EC, YO, MGW, and AAB contributed to the analyses. All authors revised and approved the manuscript.

Conflict of interest statement: None declared

Ethics approval: Boston VA Medical Center, Harvard T.H. Chan School of Public Health (protocol 14027-102).

Acknowledgements: This study is supported by grants from the National Institute of Environmental Health Sciences (NIEHS) (R01ES021733 and R01ES025225-01A1). Other support comes from NIEHS grants ES015172, ES014663 and ES020010, and Environmental Protection Agency (EPA) grants RD832416 and RD83587201. The US Department of Veterans Affairs (VA) Normative Aging Study (NAS) is supported by the Cooperative Studies Program/ERIC, US Department of Veterans Affairs, and is a research component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). The views expressed in this paper are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs. Additional support was provided by the US Department of Agriculture, Agricultural Research Service (contract 53-K06-510).

Data Availability: Data are from the Normative Aging Study, from which restricted data are available for researchers who meet the criteria. A subset of the methylation data is deposited at NCBI dbGaP (study accession number: phs000853.v1.p1).

3.8. References

1 World Health, O. Burden of disease from household air pollution for 2012. *WHO, Geneva* (2014).

2 Apte, J. S., Marshall, J. D., Cohen, A. J. & Brauer, M. Addressing Global Mortality from Ambient PM2.5. *Environ. Sci. Technol.* **49**, 8057-8066, doi:10.1021/acs.est.5b01236 (2015).

3 Künzli, N. *et al.* Ambient air pollution and atherosclerosis in Los Angeles. *Environ. Health Perspect.* **113**, 201-206 (2005).

4 Giorgini, P. *et al.* Air Pollution Exposure and Blood Pressure: an Updated Review of the Literature. *Curr. Pharm. Des.* (2015).

5 Martinelli, N., Olivieri, O. & Girelli, D. Air particulate matter and cardiovascular disease: a narrative review. *Eur. J. Intern. Med.* **24**, 295-302, doi:10.1016/j.ejim.2013.04.001 (2013).

6 Zhong, J. *et al.* Cardiac Autonomic Dysfunction: Particulate Air Pollution Effects Are Modulated by Epigenetic Immunoregulation of Toll-like Receptor 2 and Dietary Flavonoid Intake. *J Am Heart Assoc* **4**, e001423, doi:10.1161/JAHA.114.001423 (2015).

7 Kloog, I. *et al.* Effects of airborne fine particles (PM2.5) on deep vein thrombosis admissions in the northeastern United States. *Journal of Thrombosis and Haemostasis* **13**, 768-774, doi:10.1111/jth.12873 (2015).

8 Raaschou-Nielsen, O. *et al.* Particulate matter air pollution components and risk for lung cancer. *Environ Int* **87**, 66-73, doi:10.1016/j.envint.2015.11.007 (2016).

9 Terzano, C., Di Stefano, F., Conti, V., Graziani, E. & Petroianni, A. Air pollution ultrafine particles: toxicity beyond the lung. *Eur Rev Med Pharmacol Sci* **14**, 809-821 (2010).

10 Schikowski, T. *et al.* Association of air pollution with cognitive functions and its modification by APOE gene variants in elderly women. *Environ. Res.* **142**, 10-16, doi:10.1016/j.envres.2015.06.009 (2015).

11 Power, M. C. *et al.* Traffic-related air pollution and cognitive function in a cohort of older men. *Environ. Health Perspect.* **119**, 682-687, doi:10.1289/ehp.1002767 (2011).

12 Chen, J.-C. & Schwartz, J. Neurobehavioral effects of ambient air pollution on cognitive performance in US adults. *Neurotoxicology* **30**, 231-239 (2009).

13 Weuve, J. *et al.* Exposure to Particulate Air Pollution and Cognitive Decline in Older Women. *Arch Intern Med* **172**, 219-227, doi:10.1001/archinternmed.2011.683 (2012).

14 Brown, D. M. *et al.* Occupational Exposure to PM2.5 and Incidence of Ischemic Heart Disease: Longitudinal Targeted Minimum Loss-based Estimation. *Epidemiology* **26**, 806-814, doi:10.1097/EDE.0000000000000329 (2015).

15 Scheers, H., Jacobs, L., Casas, L., Nemery, B. & Nawrot, T. S. Long-Term Exposure to Particulate Matter Air Pollution Is a Risk Factor for Stroke: Meta-Analytical Evidence. *Stroke* **46**, 3058-3066, doi:10.1161/STROKEAHA.115.009913 (2015).

16 Shan, M. *et al.* A feasibility study of the association of exposure to biomass smoke with vascular function, inflammation, and cellular aging. *Environ. Res.* **135**, 165-172, doi:10.1016/j.envres.2014.09.006 (2014).

17 Wilker, E. H. *et al.* Long-Term Exposure to Fine Particulate Matter, Residential Proximity to Major Roads and Measures of Brain Structure. *Stroke* **46**, 1161-1166 (2015).

18 Rowe, J. W. & Kahn, R. L. Successful Aging 2.0: Conceptual Expansions for the 21st Century. *J Gerontol B Psychol Sci Soc Sci* **70**, 593-596, doi:10.1093/geronb/gbv025 (2015).

19 Rowe, J. W. & Kahn, R. L. Successful aging and disease prevention. *Adv Ren Replace Ther* **7**, 70-77 (2000).

20 Rowe, J. W. & Kahn, R. L. Human aging: usual and successful. *Science* **237**, 143-149 (1987).

21 Chung, H. Y. *et al.* Molecular inflammation: underpinnings of aging and age-related diseases. *Ageing Res Rev* **8**, 18-30, doi:10.1016/j.arr.2008.07.002 (2009).

22 Marioni, R. E. *et al.* The epigenetic clock and telomere length are independently associated with chronological age and mortality. *Int J Epidemiol*, doi:10.1093/ije/dyw041 (2016).

23 Marioni, R. E. *et al.* The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol* **44**, 1388-1396, doi:10.1093/ije/dyu277 (2015).

24 Horvath, S. *et al.* The cerebellum ages slowly according to the epigenetic clock. *Aging (Albany NY)* **7**, 294-306 (2015).

25 Horvath, S. & Ritz, B. R. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY)* **7**, 1130-1142 (2015).

26 Levine, A. J. *et al.* Accelerated epigenetic aging in brain is associated with pre-mortem HIV-associated neurocognitive disorders. *J. Neurovirol.*, doi:10.1007/s13365-015-0406-3 (2015).

27 Levine, M. E. *et al.* DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY)* **7**, 690-700 (2015).

28 Levine, M. E., Lu, A. T., Bennett, D. A. & Horvath, S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging (Albany NY)* **7**, 1198-1211 (2015).

29 Horvath, S. *et al.* Obesity accelerates epigenetic aging of human liver. *Proceedings of the National Academy of Sciences* **111**, 15538-15543, doi:10.1073/pnas.1412759111 (2014).

30 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).

31 Nwanaji-Enwerem, J. C. *et al.* Long-term ambient particle exposures and blood DNA methylation age: findings from the VA normative aging study. *Environ Epigenet* **2**, doi:10.1093/eep/dvw006 (2016).

32 Zanobetti, A., Franklin, M., Koutrakis, P. & Schwartz, J. Fine particulate air pollution and its components in association with cause-specific emergency admissions. *Environ Health* **8**, 58, doi:10.1186/1476-069X-8-58 (2009).

33 Ito, K. *et al.* Fine particulate matter constituents associated with cardiovascular hospitalizations and mortality in New York City. *Environ. Health Perspect.* **119**, 467-473, doi:10.1289.ehp.1002667 (2011).

34 Liu, S. *et al.* Fine particulate matter components and emergency department visits among a privately insured population in Greater Houston. *Sci. Total Environ.* **566-567**, 521-527, doi:10.1016/j.scitotenv.2016.05.022 (2016).

35 Laurent, O. *et al.* Low birth weight and air pollution in California: Which sources and components drive the risk? *Environ Int* **92-93**, 471-477, doi:10.1016/j.envint.2016.04.034 (2016).

36 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* **49**, 359-367, doi:10.1016/j.molcel.2012.10.016 (2013).

37 Bell, B., Rose, C. L. & Damon, A. The Veterans Administration Longitudinal Study of Healthy Aging. *The Gerontologist* **6**, 179-184, doi:10.1093/geront/6.4.179 (1966).

38 Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-1369, doi:10.1093/bioinformatics/btu049 (2014).

39 van Donkelaar, A. *et al.* Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environ Health Perspect* **118**, 847-855, doi:10.1289/ehp.0901623 (2010).

40 Di, Q., Koutrakis, P. & Schwartz, J. A hybrid prediction model for PM 2.5 mass and components using a chemical transport model and land use regression. *Atmospheric Environment* **131**, 390-399 (2016).

41 Kloog, I., Nordio, F., Coull, B. A. & Schwartz, J. Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA. *Remote Sensing of Environment* **150**, 132-139 (2014).

42 Madrigano, J. *et al.* Prolonged exposure to particulate pollution, genes associated with glutathione pathways, and DNA methylation in a cohort of older men. *Environ Health Perspect* **119**, 977-982, doi:10.1289/ehp.1002773 (2011).

43 Baccarelli, A. *et al.* Rapid DNA methylation changes after exposure to traffic particles. *Am. J. Respir. Crit. Care Med.* **179**, 572-578, doi:10.1164/rccm.200807-1097OC (2009).

44 Bind, M. A. *et al.* Beyond the Mean: Quantile Regression to Explore the Association of Air Pollution with Gene-Specific Methylation in the Normative Aging Study. *Environ Health Perspect* **123**, 759-765, doi:10.1289/ehp.1307824 (2015).

45 Peng, C. *et al.* Particulate Air Pollution and Fasting Blood Glucose in Nondiabetic Individuals: Associations and Epigenetic Mediation in the Normative Aging Study, 2000-2011. *Environ Health Perspect* **124**, 1715-1721, doi:10.1289/EHP183 (2016).

46 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86, doi:10.1186/1471-2105-13-86 (2012).

47 Mostofsky, E. *et al.* Modeling the association between particle constituents of air pollution and health outcomes. *Am J Epidemiol* **176**, 317-326, doi:10.1093/aje/kws018 (2012).

48 Schelldorfer, J., Bühlmann, P., De, G. & Van, S. Estimation for High-Dimensional Linear Mixed-Effects Models Using ℓ1-Penalization. *Scandinavian Journal of Statistics* **38**, 197-214 (2011).

49 Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267-288 (1996).

50 Zou, H. The adaptive lasso and its oracle properties. *Journal of the American statistical association* **101**, 1418-1429 (2006).

51 Dai, L. *et al.* Fine particles, genetic pathways, and markers of inflammation and endothelial dysfunction: Analysis on particulate species and sources. *J Expo Sci Environ Epidemiol* **26**, 415-421, doi:10.1038/jes.2015.83 (2016).

52 Dai, L. *et al.* Use of the Adaptive LASSO Method to Identify PM2.5 Components Associated with Blood Pressure in Elderly Men: The Veterans Affairs Normative Aging Study. *Environ Health Perspect* **124**, 120-125, doi:10.1289/ehp.1409021 (2016).

53 Schwarz, G. & others. Estimating the dimension of a model. *The annals of statistics* **6**, 461-464 (1978).

54 Us Epa, O. A. R. NAAQS Table.

55 Colicino, E. *et al.* Mitochondrial haplogroups modify the effect of black carbon on age-related cognitive impairment. *Environ Health* **13**, 42, doi:10.1186/1476-069X-13-42 (2014).

56 Zanobetti, A. *et al.* Associations between arrhythmia episodes and temporally and spatially resolved black carbon and particulate matter in elderly patients. *Occup Environ Med* **71**, 201-207, doi:10.1136/oemed-2013-101526 (2014).

57 McCracken, J. *et al.* Annual ambient black carbon associated with shorter telomeres in elderly men: Veterans Affairs Normative Aging Study. *Environ. Health Perspect.* **118**, 1564-1570 (2010).

58 Wu, S. *et al.* Blood pressure changes and chemical constituents of particulate air pollution: results from the healthy volunteer natural relocation (HVNR) study. *Environ Health Perspect* **121**, 66-72, doi:10.1289/ehp.1104812 (2013).

59 Wu, S. *et al.* Association of chemical constituents and pollution sources of ambient fine particulate air pollution and biomarkers of oxidative stress associated with atherosclerosis: A panel study among young adults in Beijing, China. *Chemosphere* **135**, 347-353, doi:10.1016/j.chemosphere.2015.04.096 (2015).

60 Huang, X. H. H., Bian, Q. J., Ng, W. M., Louie, P. K. K. & Yu, J. Z. Characterization of PM2.5 Major Components and Source Investigation in Suburban Hong Kong: A One Year Monitoring Study. *Aerosol Air Qual. Res.* **14**, 237-250, doi:10.4209/aaqr.2013.01.0020 (2014).

61 Frank, N. The Chemical Composition of PM2.5 to support PM Implementation.

62 Popovicheva, O. B., Persiantseva, N. M., Kireeva, E. D., Khokhlova, T. D. & Shonija, N. K. Quantification of the hygroscopic effect of soot aging in the atmosphere: laboratory simulations. *J Phys Chem A* **115**, 298-306, doi:10.1021/jp109238x (2011).

63 Rubasinghege, G., Lentz, R. W., Scherer, M. M. & Grassian, V. H. Simulated atmospheric processing of iron oxyhydroxide minerals at low pH: roles of particle size and acid anion in iron dissolution. *Proc Natl Acad Sci U S A* **107**, 6628-6633, doi:10.1073/pnas.0910809107 (2010).

64 Li, W. *et al.* Internally mixed sea salt, soot, and sulfates at Macao, a coastal city in South China. *J Air Waste Manag Assoc* **61**, 1166-1173 (2011).

65 Lepeule, J., Laden, F., Dockery, D. & Schwartz, J. Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009. *Environ Health Perspect* **120**, 965-970, doi:10.1289/ehp.1104660 (2012).

66 Schwartz, J. & Lepeule, J. Is Ambient PM2.5 Sulfate Harmful? Schwartz and Lepeule Respond. *Environ. Health Perspect.* **120**, a454-a455, doi:10.1289/ehp.1205873R (2012).

67 Wu, S., Hao, J. M., Lu, Z. F., Zhao, Z. & Li, J. H. [Effect of ammonium sulfate aerosol on the photochemical reaction of toluene/ NO(x)/air mixture]. *Huan Jing Ke Xue* **28**, 1183-1187 (2007).

68 Ghio, A. J., Stoneheurner, J., McGee, J. K. & Kinsey, J. S. Sulfate content correlates with iron concentrations in ambient air pollution particles. *Inhal Toxicol* **11**, 293-307, doi:10.1080/089583799197104 (1999).

69 Peters, B. J. *et al.* Genetic variability within the cholesterol lowering pathway and the effectiveness of statins in reducing the risk of MI. *Atherosclerosis* **217**, 458-464, doi:10.1016/j.atherosclerosis.2011.06.023 (2011).

70 Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* **8**, 1844-1865, doi:10.18632/aging.101020 (2016).

71 Wolf, E. J. *et al.* Accelerated DNA methylation age: Associations with PTSD and neural integrity. *Psychoneuroendocrinology* **63**, 155-162, doi:10.1016/j.psyneuen.2015.09.020 (2016).

72 Perna, L. *et al.* Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics* **8**, 64, doi:10.1186/s13148-016-0228-z (2016).

73 Kioumourtzoglou, M. A. *et al.* Exposure measurement error in PM2.5 health effects studies: a pooled analysis of eight personal exposure validation studies. *Environ Health* **13**, 2, doi:10.1186/1476-069X-13-2 (2014).

Chapter 4:

Modifying Role of Endothelial Function Gene Variants on the Association of Long-term $PM_{2.5}$ Exposure with Blood DNA Methylation Age: the VA Normative Aging Study.

Jamaji C. Nwanaji-Enwerem[a†]; Marie-Abele Bind[b]; Lingzhen Dai[a]; Youssef Oulhote[a]; Elena Colicino[c]; Qian Di[a]; Allan C. Just[d]; Lifang Hou[e]; Pantel Vokonas[f]; Brent A. Coull[g]; Marc G. Weisskopf[a]; Andrea A. Baccarelli[c]; and Joel D. Schwartz[a]

[a]*Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

[b]*Department of Statistics, Faculty of Arts and Sciences, Harvard University, Cambridge, MA, USA*

[c]*Department of Environmental Health Sciences, Columbia Mailman School of Public Health, New York, NY, USA*

[d]*Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

[e]*Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA*

[f]*VA Normative Aging Study, Veterans Affairs Boston Healthcare System and the Department of Medicine, Boston University School of Medicine, Boston, MA, USA*

[g]*Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

This is a pre-copyedited, author-produced version of an article accepted for publication in *Toxicological Sciences* following peer review. This version was reproduced with permission from [Nwanaji-Enwerem, Jamaji C., et al. "Modifying Role of Endothelial Function Gene Variants on the Association of Long-Term $PM_{2.5}$ Exposure with Blood DNA Methylation Age: The VA Normative Aging Study." *Toxicological Sciences* (2017).] The version of record is available online at: https://academic.oup.com/toxsci/article-abstract/158/1/116/3806698?redirectedFrom=fulltext. DOI: 10.1093/toxsci/kfx077.

4.1. Abstract

Background: Recent studies have reported robust associations of long-term $PM_{2.5}$ exposure with DNA methylation-based measures of aging; yet, the molecular implications of these relationships remain poorly understood.

Objective: We evaluated if genetic variation in three biological pathways implicated in $PM_{2.5}$-related disease – oxidative stress, endothelial function, and metal processing – could modify the effect of $PM_{2.5}$ on DNAm-age, one prominent DNA methylation-based measure of biological age.

Methods: This analysis was based on 552 individuals from the Normative Aging Study with at least one visit between 2000 and 2011 (n=940 visits). A genetic-score approach was used to calculate aging-risk variant scores for endothelial function, oxidative stress, and metal processing pathways. One-year $PM_{2.5}$ and $PM_{2.5}$ component (sulfate and ammonium) levels at participants' addresses were estimated using the GEOS-chem transport model. Blood DNAm-age was calculated using CpG sites on the Illumina HumanMethylation450 BeadChip.

Results: In fully-adjusted linear mixed-effects models, the effects of sulfate on DNAm-age (in years) were greater in individuals with high aging-risk endothelial function variant scores when compared to individuals with low aging-risk endothelial function variant scores ($P_{interaction}$=0.0007; β_{High}=1.09, 95%CI_{High}: 0.70, 1.48; β_{Low}=0.40, 95%CI_{Low}: 0.14, 0.67). Similar trends were observed in fully-adjusted models of ammonium and total $PM_{2.5}$ alone. No effect modification was observed by oxidative stress and metal processing variant scores. Secondary analyses revealed significant associations of serum endothelial markers, ICAM1 (β=0.01, 95%CI: 0.002, 0.012) and VCAM1 (β=0.002, 95%CI: 0.0005, 0.0026), with DNAm-age.

Conclusion: Our results add novel evidence that endothelial physiology may be important to DNAm-age relationships, but further research is required to establish their generalizability.

4.2. Introduction

Approximately 92% of the world's population lives in areas with ambient fine particle ($PM_{2.5}$) levels higher than currently accepted global standards[1]. In fact, a substantial portion of global mortality and morbidity can be attributed to $PM_{2.5}$ exposure[2]. Research continues to demonstrate that long-term $PM_{2.5}$ exposure is a major risk factor for cardiovascular disease[3,4], and respiratory impairment[5]. $PM_{2.5}$ has also been associated with cognitive decline[6,7], and cancer[8,9]. Still, exactly how $PM_{2.5}$ contributes to these and other important health outcomes is still not fully understood. Addressing this knowledge gap is a critical step for developing interventions to alleviate the disease burden of individuals already exposed to high $PM_{2.5}$ levels. Aging also independently contributes to many $PM_{2.5}$-related health endpoints[10,11]. Thus, studying how $PM_{2.5}$ is related to aging may facilitate a greater understanding of the complex pathophysiology surrounding $PM_{2.5}$-related diseases.

Some of the most recent studies of $PM_{2.5}$ and aging have involved DNA methylation age (DNAm-age), a novel tissue-independent measure of biological age that is calculated using DNA methylation values from 353 age-correlated CpG dinucleotides[12]. Our research group was the first to report significant positive associations of long-term $PM_{2.5}$ exposure levels with DNAm-age[13]. Recently, we have identified sulfate and ammonium as important component species in the $PM_{2.5}$-DNAm-age relationship[14]. In collaboration with another research group, we have also demonstrated associations between other air pollutants (*e.g.* black carbon, PM_{10}, and NO_x) and epigenetic aging measures[15]. Despite this work, very little is known about the molecular implications of the $PM_{2.5}$ and DNAm-age relationship. At the moment, the algorithm used to calculate DNAm-age relies on assays that have only been optimized for humans and chimpanzees[12]. Hence, this field of research is limited by a lack of traditional animal models to study DNAm-age's relationships. It is possible that DNAm-age is simply reflecting a well-studied biological process (*e.g.* oxidative stress), but it may alternatively reflect a completely novel process. Existing epidemiologic studies have demonstrated robust associations of DNAm-age with mortality, cognitive decline, cancer, and other $PM_{2.5}$-related outcomes thereby suggesting that DNAm-age processes are common to many disease pathologies[16-18]. There is also epidemiologic evidence that normal genetic variation may influence DNAm-

age, but this has yet to be explored in the context of the $PM_{2.5}$-DNAm-age association[19]. To our knowledge, only a couple of studies have peripherally examined the molecular implications of DNAm-age, and these studies only conclude that DNAm-age represents a form of biological aging that differs from cellular senescence[12,20].

Even with current limitations, existing technologies can be creatively utilized to begin understanding the molecular implications of DNAm-age. Here, we employ components of a previously developed genetic score approach that categorizes normal genetic variation into three biological pathways – oxidative stress, endothelial function, and metal processing[21]. Oxidative stress, endothelial function, and metal processing are all biological pathways that have been implicated in numerous $PM_{2.5}$-related diseases; thus, they may also be involved in the $PM_{2.5}$-DNAm-age relationship[22]. The existing studies that have used this Bind *et al.* method have suggested potential modification of the associations of $PM_{2.5}$ exposure with inflammatory markers and cardiac autonomic function, but the results have not been statistically significant[21-24]. Given these previous findings, our aim is to use the candidate pathway-specific genetic variants employed by the Bind method to 1) identify variants specifically important to DNAm-age and 2) assess if the normal genetic variation captured by these variants modifies the $PM_{2.5}$-DNAm-age relationship in a population of community-dwelling elderly Caucasian men.

4.3. Materials and Methods

4.3.a. Study Population

Participants included in this analysis were part of the Normative Aging Study (NAS), an ongoing longitudinal cohort study of healthy male volunteers from the Eastern Massachusetts area[25]. The NAS is a closed cohort and participants are now elderly. The NAS was established by the U.S. Department of Veterans Affairs (VA) in 1963, and enrolled men who were free of any chronic disease. Every 3-5 years, NAS participants reported for onsite, detailed medical examinations during which bio-specimens were collected and assessments of lifestyle factors that may affect health were made. All participants provided

written informed consent to the VA Institutional Review Board (IRB), and human subjects approval was granted by the VA and Harvard T.H. Chan School of Public Health IRBs.

All NAS men with continued study participation as of the year 2000, when address-specific $PM_{2.5}$ component species levels became available, were eligible for our study sample. After excluding participants with a diagnosis of leukemia (n=11), due to its potential influence on the DNA methylation of blood cells[12], and those incomplete for the covariates of interest (n=16), we had a total of 552 participants with 940 observations (*i.e.* study visits) between the years 2000 and 2011. This was the study sample that was used in reporting the significant associations between $PM_{2.5}$ component species and DNAm-age in our previous publication[14]. Of these 552 participants, 249 (45%) had one visit, 218 (40%) had two visits, and 85 (15%) had three or more visits. From this sample, we then excluded participants missing pathway specific polymorphism data. This resulted in three distinct, but not mutually exclusive, groups of participants: 1) Oxidative stress subset (n=410, obs=702); 2) Endothelial function subset (n=450, obs=779); and 3) Metal processing subset (n=426, obs=744).

4.3.b. DNA Methylation Assay and Calculation of DNAm-age

Laboratory staff extracted DNA from the buffy coat of whole blood collected from each participant during each NAS follow-up visit (QIAamp DNA Blood Kit, QIAGEN, Valencia, CA, USA). DNA samples underwent bisulfite conversion (EZ-96 DNA Methylation Kit, Zymo Research, Orange, CA, USA) and were hybridized to the 12 sample Illumina HumanMethylation450 BeadChips (Infinium HD Methylation protocol, Illumina, San Diego, CA, USA). A two-stage age-stratified algorithm was used to randomize samples avoiding confounding and ensuring a similar age distribution across chips and plates. For quality control purposes, study staff removed samples where >5% of probes had a beadcount < 3 or > 1% of probes had a detection P-value >0.05. The Bioconductor minfi package Illumina-type background correction without normalization was used to preprocess the remaining samples and generate methylation beta values[26]. Beta values represent the percentage of methylation for each of the ~480,000 CpG sites in the BeadChip array. The 450k arrays were run in the Genomics Core Facility at Northwestern University.

DNAm age was determined using the publically available online calculator (http://labs.genetics.ucla.edu/horvath/dnamage/). DNAm-age was derived from penalized regression (an elastic net) using multiple data sets of varying cell and tissue types. 21,369 CpG probes, shared by the Illumina HumanMethylation27 and HumanMethylation450 BeadChip platforms were regressed on a calibrated version of chronological age. The elastic net selected 353 CpGs that correlated with age (193 positively and 160 negatively), and the resulting model coefficients were used by the calculator to predict the age of each DNA sample (DNAm-age)[12]. Empirical data demonstrated that the calculator maintains predictive accuracy (age correlation 0.97, error = 3.6 years) across almost all body tissues including blood, brain, and bone.

4.3.c. Ambient Particle (Exposure) Assessment

We utilized the GEOS-chem chemical transport model (http://www.geoschem.org) to generate 1-year exposure estimates for $PM_{2.5}$, sulfate, and ammonium. Sulfate and ammonium are the major $PM_{2.5}$ component species previously demonstrated to be most important in predicting DNAm-age[14]. By incorporating meteorology variables, non-linear chemistry, and detailed emissions inventories, GEOS-chem simulated the formation and transportation of atmospheric components and provided raw estimates of $PM_{2.5}$ and its major component species. Ten-fold cross-validation demonstrated that the model performed well for $PM_{2.5}$ mass and its component species with R^2s ranging from 0.70 to 0.88[27]. Each participant's residence was geocoded and linked to an area level grid-point. After accounting for address changes, we assigned particle estimates to each participant's address. Greater than 90% of NAS participants are retired; thus, home address exposures are expected to be a good proxy for their individual ambient exposures. We generated daily estimates at the 1x1 km area resolution and 1-year total $PM_{2.5}$ and $PM_{2.5}$ component species exposure windows by averaging daily exposures for the 365 days prior to the day of each participants' NAS visit. The 1-year $PM_{2.5}$ exposure window was utilized because it has been previously reported to be robustly associated with DNAm-age[13].

4.3.d. Serum Endothelial Function Marker Assays

We used three common plasma endothelial function markers (vascular cell adhesion molecule-1 [VCAM], intercellular adhesion molecule-1 [ICAM], and vascular endothelial growth factor [VEGF]) that were measured in blood collected from NAS participants during their study visits. These markers were selected because of all serum physiologic markers available in the NAS, they are most specific and directly related to the endothelium. Other markers like CRP are non-specific to the endothelium and have more nuanced relationships with other biological processes like general inflammation. These markers have also been consistently associated with $PM_{2.5}$ levels in numerous (NAS and non-NAS) epidemiologic studies and have been extensively used to assess endothelial function[22,28]. VCAM and ICAM are two important cellular adhesion molecules that mediate leukocyte-endothelial cell adhesion and transendothelial migration[29]. Laboratory staff measured VCAM (ng/mL) and ICAM (ng/mL) in serum using the enzyme-linked immunosorbent assay method (R&D Systems, Minneapolis, MN). Sensitivity of the assay was 0.35 ng/mL for ICAM with day-to-day assay variabilities of 10.1, 7.4, 6.0 and 6.1% at concentrations of 64.2, 117, 290 and 453 ng/mL, respectively. Sensitivity of the assay was 2.0 ng/mL for VCAM with day-to-day assay variabilities of 10.2, 8.5 and 8.9% at concentrations of 9.8, 24.9 and 49.6 ng/mL, respectively. VEGF is a signaling protein that stimulates the production of endothelial cells and the formation of blood vessels[30]. VEGF (pg/mL) was quantified using multiplexing technology (MILLIPLEXTM MAP) with commercially available MILLIPLEXTM MAP kits (EMD Millipore, Billerica, MA, USA). The VEGF assay has intra-assay and inter-assay precision of 13% and 19% respectively.

4.3.e. Statistical Analysis

4.3.e.1. Primary Analysis and Variant Scores:

We first used linear mixed effects models to determine if we could observe previously published positive associations of 1-year $PM_{2.5}$, sulfate, and ammonium levels with DNAm-age in each of our three pathway subsets.

In our study reporting a relationship between $PM_{2.5}$ and DNAm-age, we used a tiered framework adjusting for confounders and covariates with: 1) *a priori* biological/clinical relevance and/or 2) reported in the existing literature[13]. Tier one adjusted for chronological age and blood cell types. Tier two made additional adjustments for lifestyle and environmental factors. Tiers three and four expanded on tier two by additionally adjusting for age-related diseases and medications for age-related diseases respectively. After examining model fit (assessed via AIC) and considering the implications of genetic polymorphisms on disease independent of $PM_{2.5}$ and DNAm-age relationships, we employed the tier three covariates for this analysis. In all, these models were adjusted for chronological age (continuous), six blood cell type estimates [*i.e.* plasma cells, CD4+ lymphocytes, CD8+ lymphocytes, natural killer (NK) cells, monocytes, and granulocytes] (continuous) determined via Houseman and Horvath methods[12,31], average 1-year temperature (continuous address-specific satellite measurements[14]), cumulative cigarette pack years (continuous), smoking status (current, former, or never), season of visit (spring [March-May], Summer [June-August], Fall [September-November], and Winter [December-February]), body mass index (BMI) (lean [<25], overweight [25-30], obese [>30]), alcohol intake (yes/no ≥ 2 drinks daily), maximum years of education (continuous), cancer (yes/no history of lifetime cancer diagnosis), coronary heart disease (yes/no based on electrocardiogram, validated medical records, or physical exam), diabetes (physician diagnosis or a fasting blood glucose > 126 mg/dL), and hypertension (yes/no antihypertensive medication use or systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg). The sulfate and ammonium models were additionally adjusted for $PM_{2.5}$ mass. All linear mixed effects models were run using the lme function from the nlme R package[32], and included a random participant-specific intercept to account for correlation between repeated outcome measures (*i.e.* multiple visits for a participant).

We next used fully adjusted mixed effects models to determine: 1) if oxidative stress, endothelial function, or metal processing genetic variants were associated with DNAm-age and 2) the impact of air pollution-genetic pathway interactions on DNAm-age. To accomplish this, we utilized genotyping data from the NAS dataset and components of a novel genetic score developed by Bind *et al.* (2014)[21]. Briefly, genotyping assays were performed using the Sequenom MassArray MALDI-TOF Mass Spectrometer with

semi-automated primer design and implementation of the short extension method (San Diego, CA). The MassArray system has the capacity to analyze multiple classes of genetic markers with high sensitivity.

Bind *et al.* developed a novel approach to investigate interactions between environmental exposures and the biological pathways of oxidative stress, endothelial function, and metal processing[21]. The authors first related genes to one of these three pathways based on the biological functionality provided by GeneCards[33]. Then considering independent outcomes representative of each pathway (8-hydroxy-2'-deoxyguanosine for oxidative stress, augmentation index for endothelial function, and patella lead concentration for metal processing), they used the least absolute shrinkage and selection operator (LASSO) method to select the most important gene variants for each of the outcomes.

Although the Bind *et al.* method would allow us to broadly identify pathways that may be related to the $PM_{2,5}$-DNAm-age relationship, it does not allow us to identify variants that are specifically important to this relationship. Identifying specific variants allows for a more comprehensive understanding of why these pathways are important. In an effort to identify pathway score component variants that were specifically sensitive to DNAm-age relationships, we made one alteration to the Bind *et al*. method. Two major limitations of LASSO selection are that 1) the number of selected variables are bounded by the number of observations and 2) that the LASSO tends to select one variable from a highly related group while ignoring the others[34]. Given our desire to maximize the identification of specific genetic variants important to DNAm-age from three individual groups of pathway-related variants, the latter of these two limitations was a concern to us. To overcome this limitation, we employed an elastic net penalized regression, which allows for the selection of highly-related variables[35]. Thus, starting with the reported Bind *et al*. list of candidate pathway-specific gene variants, we then employed an elastic net (penalized regression) via the glmnet function in the R glmnet package to determine which of these pathway-specific gene variants were also important for DNAm-age[36]. Our method was similar to that described by Lenters and colleagues[37] and the full documentation for running all aspects of the elastic net via glmnet is publically available (https://cran.r-project.org/web/packages/glmnet/index.html). In short, all aforementioned covariates were included in the elastic net regression models as unpenalized variables. The fully-adjusted

elastic net regression linear models utilized a hybrid of ridge and LASSO penalty functions to determine which genetic variants, within each respective pathway, were important predictors of DNAm-age. With ridge, the square of the regression coefficients for predictors are penalized. All predictors are retained but coefficients from highly related predictors are proportionally shrunk towards zero. With LASSO, the absolute value of predictor coefficients is penalized and coefficients are shrunk by a constant factor. Coefficients for the least predictive variants are shrunk to zero and only one predictor from a highly correlated group tends to be selected. By combining both of these penalty functions, the elastic net performed selection while allowing for the inclusion of highly-related genetic variants bin[35,36]. Cross-validation was also performed to determine the optimal degree of penalization. The proportion of ridge and LASSO functions and the corresponding penalty that yielded the minimum mean-squared error (MSE) of prediction from repeated 10-fold cross-validation was used in the final elastic net selection model. Gene variants with non-zero coefficients are considered as "selected" by the elastic net. Following this elastic net selection, we were left with three lists: 1) Oxidative stress gene variants that were important for DNAm-age; 2) Endothelial function gene variants that were important for DNAm-age; and 3) Metal processing gene variants that were important for DNAm-age.

Returning to the original Bind *et al.* methodology, we then summed the sign of the non-zero coefficients for each of the important variants to construct pathway specific variant scores for all study participants. For instance, say hypothetical oxidative stress variants A1, A2, and B3 had elastic net coefficients of +2.3, -1.7, and 1.6. A participant with all of these variants would have an oxidative stress polymorphism score of +1 (*i.e.* 1 – 1 + 1= 1). Another participant with only variants A1 and B3 would have an oxidative stress score of +2 (*i.e.* 1 + 1 = 2). Final binary pathway polymorphism scores were created by dichotomizing each score as low or high aging-risk using the median of each score's distribution in the study sample.

Again, we used fully adjusted mixed effects models to determine if each pathway polymorphism score was independently associated with DNAm-age. We then included interaction terms of our main

exposures ($PM_{2.5}$, sulfate, and ammonium) with the three respective pathway polymorphism scores to identify genetic pathway-air pollution interactions that are related to DNAm-age.

4.3.e.2. Secondary Analysis:

Although our air pollution-genetic pathway interaction models provided us with some insight to biological pathways that may be genetically relevant to DNAm-age, we further investigated if these same pathways had functional relationships with DNAm-age. Thus, we first looked to see if serum endothelial function markers were correlated with DNAm-age in our study sample. We next constructed linear mixed effects models to see if the serum markers were significantly associated with DNAm-age after adjusting for chronological age, blood cell types, and the age-related diseases of lifetime cancer diagnosis, hypertension, diabetes, and coronary heart disease.

All statistical analyses were performed using R Version 3.1.1 (R Core Team, Vienna, Austria) and we considered a P-value < 0.05 to be statistically significant.

4.4. Results

4.4.a. Descriptive Statistics

The demographics and clinical data of all study participants in each of the three biological pathway subsets are presented in Table 1. Participants in each subset had an average age (SD) and an average DNAm-age (SD) both of approximately 74 ± 7 years. Table S1 lists all of the candidate pathway-specific genetic variants used by Bind *et al.* (2014)[21]. Table S2 lists the genetic variants that were selected by the elastic net as important for DNAm-age. No metal processing variants were selected by the elastic net; thus, no variant score could be calculated for that subset. Participants in the oxidative stress subset had variant scores ranging from -6 to 4, and endothelial function subset participants had variant scores ranging from -3 to 0. Both subsets had variant scores with a median of 1.

Table 1. Characteristics of Study Subjects, 2000 – 2011

Variable	Oxidative Stress Subset (N = 702)*	Endothelial Function Subset (N = 779)*	Metal Processing Subset (N = 744)*
Age, mean (SD)	74.7 (6.89)	74.8 (6.95)	74.7 (6.80)
DNAm-age, mean (SD)	74.2 (7.97)	73.9 (7.92)	73.8 (7.51)
Variant Score, mean (range)	-0.51 (-6, 4)	-1.23 (-3, 0)	-
Temperature, mean (SD)	11.3 (0.99)	11.3 (0.98)	11.3 (1.03)
Pack years, mean (SD)	20.4 (24.7)	20.6 (24.7)	20.9 (24.9)
Smoking Status, N (%)			
Current	35 (5)	37 (5)	34 (5)
Former	450 (64)	512 (65)	480 (65)
Never	217 (31)	230 (30)	230 (30)
Season, N (%)			
Spring	177 (25)	192 (25)	188 (25)
Summer	146 (21)	164 (21)	157 (21)
Fall	238 (34)	267 (34)	251 (34)
Winter	141 (20)	156 (20)	148 (20)
BMI, N (%)			
Healthy/Lean	168 (24)	189 (24)	172 (23)
Overweight	370 (53)	406 (52)	391 (53)
Obese	164 (23)	184 (24)	181 (24)
Alcohol Consumption, N (%)			
< 2 drinks/day	560 (80)	620 (80)	599 (81)
≥ 2 drinks/day	142 (20)	159 (20)	145 (19)
Education, N (%)			
≤ 12 years	192 (27)	207 (27)	192 (26)
12 - 16 years	320 (46)	355 (46)	341 (46)
> 16 years	190 (27)	217 (27)	211 (28)
Lifetime Cancer Diagnosis, N (%)			
Yes	390 (55)	435 (56)	426 (57)
No	312 (45)	344 (44)	318 (43)
Coronary Heart Disease, N (%)			
Yes	221 (31)	261 (34)	257 (35)
No	481 (69)	518 (66)	487 (65)
Diabetes, N (%)			
Yes	120 (17)	138 (18)	131 (18)
No	582 (83)	641 (82)	613 (82)
Hypertension, N (%)			
Yes	514 (73)	571 (73)	541 (73)
No	188 (27)	208 (27)	203 (27)

*From 552 participants (940 visits), we excluded participants missing pathway-specific variant data. This resulted in three distinct, but not mutually exclusive, subsets.

Table 2 summarizes the mean levels of $PM_{2.5}$, sulfate, and ammonium in each of the three subsets. In the oxidative stress subset, 10.3 (2.14) $\mu g/m^3$, 3.40 (0.80) $\mu g/m^3$, and 1.04 (0.29) $\mu g/m^3$ were the average (IQR) levels of $PM_{2.5}$, sulfate, and ammonium respectively. Both the endothelial function and metal processing subsets showed similar levels of these particles. The mean (IQR) levels for $PM_{2.5}$, sulfate, and ammonium in the endothelial function subset were 10.3 (2.09) $\mu g/m^3$, 3.38 (0.80) $\mu g/m^3$, and 1.04 (0.29) $\mu g/m^3$. The mean (IQR) levels for same particles in the metal processing subset were 10.3 (2.09) $\mu g/m^3$, 3.42 (0.84) $\mu g/m^3$, and 1.04 (0.28) $\mu g/m^3$.

Table 2. Mean 1-Year Particulate Matter 2.5 ($PM_{2.5}$), Sulfate, and Ammonium Concentrations, 2000-2011

Particle ($\mu g/m^3$)	Oxidative Stress Subset (N = 702)	Endothelial Function Subset (N = 779)	Metal Processing Subset (N = 744)
$PM_{2.5}$, mean (IQR)	10.3 (2.14)	10.3 (2.09)	10.3 (2.09)
Sulfate, mean (IQR)	3.40 (0.80)	3.38 (0.80)	3.42 (0.84)
Ammonium, mean (IQR)	1.04 (0.29)	1.04 (0.29)	1.04 (0.28)

4.4.b.-Year Particle Levels and Variant Scores as Predictors of DNAm-age

Table 3 summarizes the results of linear mixed effects models where dichotomized variant scores and IQR increases in 1-year particle levels were modeled as independent predictors of DNAm-age. In the endothelial function subset, an IQR increase in 1-year $PM_{2.5}$ (β=0.67, p=0.005), sulfate (β=0.64, p<0.0001), and ammonium (β=0.49, p=0.002) were all significantly, positively associated with DNAm-age. 1-year IQR increases in all three particles were also significant positive predictors of DNAm-age in the metal processing subset. Similar trends were observed in the oxidative stress subset where sulfate (β=0.64, p<0.0001) and ammonium (β=0.58, p=0.0005) were significant positive predictors of DNAm-age and $PM_{2.5}$ (β=0.46, p=0.07) was a marginally significant positive predictor of DNAm-age.

13 oxidative stress variants and 3 endothelial function variants were selected by the elastic net as

important for predicting DNAm-age. These variants were used to calculate the variant scores for these pathways (Table S2). Again, no variants were selected by the elastic net for the metal processing subset; hence, no variant score could be calculated for this biological pathway. In both the oxidative stress and endothelial function subsets, individuals with high aging-risk variant scores (≥ median) on average had at least a 0.62-year higher DNAm-age than their counterparts with low aging-risk variant scores. However, these relationships were not statistically significant (Table 3).

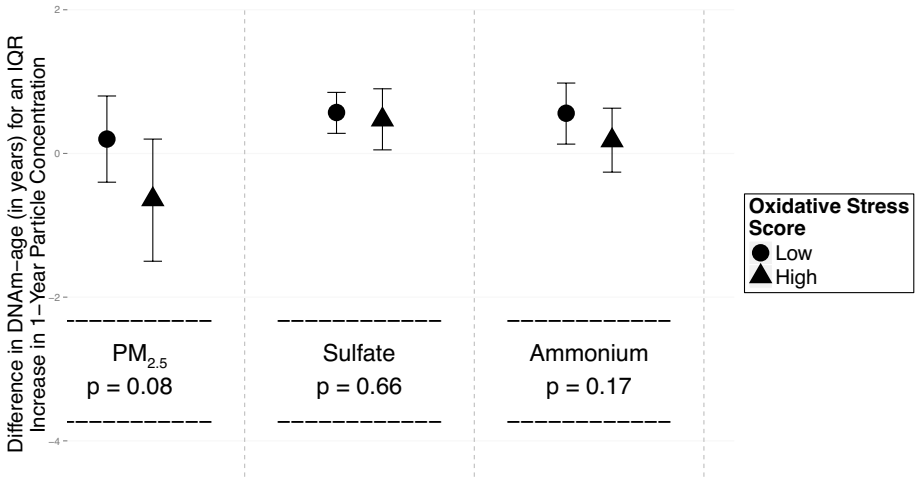

Figure 1 | Difference in DNAm-age for one interquartile range increase in 1-year particle exposure according to oxidative stress score (low versus high) in the fully-adjusted linear mixed effects model.

Table 3. Mean 1-Year Particulate Concentrations and Polymorphism Score as Independent Predictors of DNAm-age

Predictor	Oxidative Stress Subset (N = 702)		Endothelial Function Subset (N = 779)		Metal Processing Subset (N = 744)	
	Difference in DNAm-age (years) for IQR (95% CI)	***P***	**Difference in DNAm- age (years)for IQR (95% CI)**	***P***	**Difference in DNAm-age (years) for IQR (95% CI)**	***P***
$PM_{2.5}$	0.46 (-0.04, 0.97)	0.07	0.67 (0.21, 1.15)	**0.005**	0.48 (0.003, 0.94)	**0.05**
Sulfate	0.64 (0.38, 0.89)	**<0.0001**	0.64 (0.40, 0.88)	**<0.0001**	0.53 (0.29, 0.77)	**<0.0001**
Ammonium	0.58 (0.25, 0.91)	**0.0005**	0.49 (0.18, 0.80)	**0.002**	0.53 (0.16, 0.89)	**0.005**
Variant Score*						
Low	ref	-	ref	-	-	-
High	0.62 (-0.65, 1.89)	0.34	0.67 (-0.46, 1.79)	0.25	-	-

All models adjusted for chronological age, blood cell type, temperature, pack years, smoking status, season, BMI, alcohol consumption, education, lifetime cancer diagnosis, hypertension, diabetes, and coronary heart disease. Sulfate, ammonium and polymorphism score models are additionally adjusted for total $PM_{2.5}$ mass. *Binary pathway polymorphism scores were created by dichotomizing each score using the median of each score's distribution in the study sample. Participants had a low aging-risk if their score was less than (<) the median and a high aging-risk score if their score was greater than or equal to (≥) the median.

4.4.c. Effect Modification by Variant Scores

Figures 1 and 2 depict the modifying role of the oxidative stress (Figure 1) and endothelial function (Figure 2) variant scores on the relationship of 1-year particle exposures with DNAm-age. The association of all three particles on DNAm-age was greater in individuals with a low aging-risk oxidative stress variant score when compared to individuals with a high aging-risk oxidative stress variant score (Figure 1), but none of these differences were statistically significant. The effect of all three particles on DNAm-age was greater in individuals with a high aging-risk endothelial function variant score when compared to individuals with a low aging-risk endothelial function variant score (Figure 2). These differences were significant for sulfate and ammonium exposure but not quite for $PM_{2.5}$. The relationships in Figure 1 are quantified in Table S3.

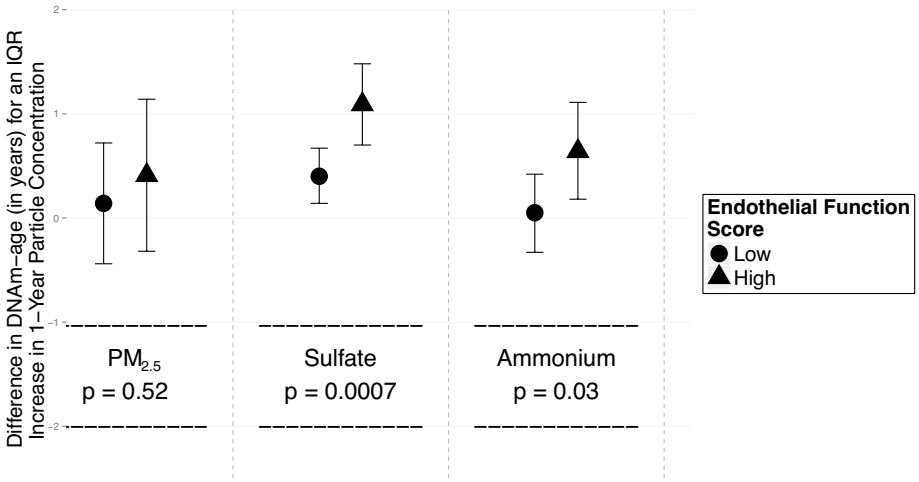

Figure 2 | Difference in DNAm-age for one interquartile range increase in 1-year particle exposure according to endothelial function score (low versus high) in the fully-adjusted linear mixed effects model.

4.4.d. Relationships of Serum Endothelial Functional Markers with DNAm-age

In our secondary analysis, DNAm-age was significantly positively correlated with both ICAM (r=0.13, p=0.0001) and VCAM (r=0.25, p<0.0001) (Table S4). However, DNAm-age was not significantly correlated with VEGF (r=0.02, p=0.54). After adjusting for chronological age, blood cell types, cancer, hypertension, diabetes, and coronary heart disease, ICAM1 (β=0.01, p=0.005) and VCAM1 (β=0.002, p=0.004) were both significant positive predictors of DNAm-age. VEGF (β=-0.00003, p=0.82) was not significantly associated with DNAm-age (Table 4).

Table 4. Associations of Serum Endothelial Function Markers with DNAm-age

	Difference in DNAm-age (95% CI)	***P***	**N**
Marker			
ICAM (ng/mL)	0.01 (0.002, 0.012)	**0.005**	608
VCAM (ng/mL)	0.002 (0.0005, 0.0026)	**0.004**	608
VEGF (pg/mL)	-0.00003 (-0.0003, 0.0003)	**0.82**	608

Models adjusted for chronological age, blood cell type, lifetime cancer diagnosis, hypertension, diabetes, and coronary heart disease.

4.4.e. Endothelial Function Variant Score as a Modifier of the Association of $PM_{2.5}$ with ICAM and VCAM

We performed subsequent analyses, using fully-adjusted mixed effects models, where we found that the positive associations of 1-year particle levels with ICAM (Figure S1) and VCAM (Figure S2) were greater in individuals with high aging-risk endothelial function variant scores when compared to their counterparts. These findings were only statistically significant for ICAM.

4.5. Discussion

The present study employed a large longitudinal cohort of elderly men to: 1) identify pathway-specific genetic variants that were related to DNAm-age and 2) determine if these variants modified the association of $PM_{2.5}$ and $PM_{2.5}$ component levels with DNAm-age. In each of our pathway subsets, we first wanted to ensure that we observed similar relationships of $PM_{2.5}$, sulfate, and ammonium with DNAm-age as previously reported. Indeed, this was the case as sulfate and ammonium were significant positive predictors of DNAm-age in all three subsets. $PM_{2.5}$ levels were significant positive predictors of DNAm-age in the endothelial function and metal processing subsets, while being marginally significant in the oxidative stress subset – potentially due to reduced power attributed to the subset's slightly reduced sample size.

We next used components of a method developed by Bind *et al.* (2014) to calculate variant scores for oxidative stress, endothelial function, and metal processing pathways. The addition of an elastic net selection to the Bind *et al.* method allowed us to optimize the sensitivity of the approach to DNAm-age relationships, while allowing for the identification of genetic-pathway variants that were specifically important for DNAm-age. Given that each of these pathways is known to be associated with $PM_{2.5}$-related disease, we predicted that the elastic net would select important variants from each of the pathways. This was the case for oxidative stress and endothelial function, but not metal processing. Published literature has already demonstrated that these metal processing variants do not modify the effect of $PM_{2.5}$ levels on a panel serum physiological markers including fibrinogen, ICAM, and CRP[21]. Thus, it is possible that metal processing pathways have little or no relationship with DNAm-age physiology especially in the context of $PM_{2.5}$ exposure. Nevertheless, no studies have examined the relationship of $PM_{2.5}$'s metal component species with DNAm-age and no studies have examined the relationship of direct metal exposures with DNAm-age. Such studies will be necessary to confirm our null findings of metal processing physiology with DNAm-age.

A total of 13 oxidative stress variants and 3 endothelial function variants were selected by the elastic net. A subsequent literature review revealed that many of the selected oxidative stress variants, including rs2284367 (CAT), rs2300181 (CAT), and rs1799811 (GSTP1), have already been implicated as

effect modifiers of the relationship of $PM_{2.5}$ and its component species with numerous health endpoints[38,39]. None of the selected endothelial function variants had been implicated in $PM_{2.5}$ relationships, but there is evidence of their role in impacting disease susceptibility following other environmental insults like radiation, cigarette smoke, and pesticide exposures[40-42]. After using the direction of the elastic net coefficients of these variants to construct pathway-specific variant scores, we determined if these scores were associated with DNAm-age. Given that the Bind *et al.* method constructs the scores such that a higher score correlates with a profile of higher risk for increased DNAm-age, we expected that individuals with high oxidative stress or endothelial function scores would on average have higher DNAm-ages than their low score counterparts. After examining the effect estimates, this was the case. High aging-risk oxidative stress score participants on average had a 0.62-year greater DNAm-age and high aging-risk endothelial function participants on average had a 0.67-year greater DNAm-age. Despite these trends, these results were not statistically significant.

When we next explored the modifying role of these variant scores on particle-DNAm-age relationships, we found that the effect of sulfate and ammonium on DNAm-age were on average approximately 0.60-years greater in participants with a high aging-risk endothelial function score when compared to participants with a low aging-risk endothelial function score. A similar trend was observed with $PM_{2.5}$ and endothelial score interactions, but this was not statistically significant. This result suggests that DNAm-age is sensitive to endothelial function physiology and is further supported by our secondary analysis that revealed significant associations of ICAM and VCAM with DNAm-age after adjusting for covariates. Numerous human and animal studies have demonstrated that $PM_{2.5}$ exposure upregulates expression of endothelial factors, which are known to play a role in vascular dysfunction[43,44]. Moreover, vascular physiology is a ubiquitous component of many disease processes and may help explain why DNAm-age has been linked to all-cause mortality, malignancy, cognitive deficits and a host of other diseases[45,46]. Endothelial micro-particles from acute coronary artery patients (a surrogate marker of endothelial dysfunction) have been shown to promote thrombogenecity and aging phenotypes in healthy coronary artery cells[47]. In a cross-sectional study, endothelial VCAM was associated with increased

vascular resistance and lower cognitive performance[48]. On the contrary pharmaceutical agents that are used to treat age-related disease (*e.g.* statins) have been shown to increase endothelial progenitor cells, which may promote endothelial repair and offer benefits like cardio-protection[49].

It is interesting that endothelial variants significantly modified the associations of sulfate and ammonium, but not total $PM_{2.5}$. As it is widely accepted that different $PM_{2.5}$ species have different toxicological effects, this finding may speak to a specific toxicity of these component species via endothelial function pathways[50]. It has already been demonstrated that ammonium and sulfate moieties can impact endothelial function[51,52]. It is also important to highlight that VEGF was not significantly correlated with DNAm-age. VEGF was also not significantly associated with DNAm-age after adjusting for covariates. The differences between DNAm-age's relationship with VEGF compared to its relationship with ICAM or VCAM could possibly be attributed to VEGF gene and protein expression. VEGF production is induced in cells that are hypoxic and circulating VEGF then binds to endothelial cells to promote angiogenesis[30]. ICAM and VCAM are more specifically produced by endothelial cells and play a prominent role in endothelial cell interactions with inflammatory cells[29]. Hence, our data allude to a specific relationship between DNAm-age and endothelial function that may be related to immune regulation. This finding is particularly promising, as the immune system has long been thought to play a role in the adverse effects of $PM_{2.5}$[4]. In all, our findings and the existing literature suggest that the endothelial function pathway is a promising place to begin understanding the molecular relevance of DNAm-age. Future studies including these and other endothelial function markers are necessary to confirm our findings and further define this relationship.

Finally, none of the particle-oxidative stress score interactions were statistically significant, but it is worth noting the direction of the effect estimates for these interaction terms. Like the endothelial function score, individuals with a high aging-risk oxidative stress score had a higher DNAm-age when compared to individuals with a low aging-risk oxidative stress scores (results not statistically significant). However, unlike the endothelial function score, a high aging-risk oxidative stress score appeared to mitigate the effect of particles of DNAm-age. These results were not statistically significant but may suggest competing effects

between high particle exposure and a high aging-risk oxidative stress physiology. When both are simultaneously present (*i.e.* the interaction of both variables is considered) they appear to inhibit or dampen each other's effects. Such a phenomenon is often observed in epidemiologic research and biological systems[53]. One air pollution related example of competing effects is the mitigation of associations of particles with the birth complication preeclampsia when multiple particle sources are considered simultaneously[54].

The strengths of the current study include utilization of novel biomarker and genetic pathway tools, rigorous statistical methods, and a large longitudinal cohort with repeated measures of ambient pollutant exposures, DNA methylation, and potential confounders. This is the first study to use genetic variants to study the relationship of ambient particles with DNAm-age. On the contrary, our study does have some notable limitations. First, given that a majority of NAS participants are retired and are very likely to spend most of their time at home, we use a validated chemical transport model to estimate 1-year ambient levels of $PM_{2.5}$, sulfate, and ammonium at participants' addresses. Such an approach at estimating personal exposures could potentially result in non-differential misclassification, but this is likely to attenuate statistical associations rather than bias them away from the null[55,56]. Secondly, we employ a genetic variant score approach that is somewhat limited because it does not provide genome-wide resolution of the three biological pathways. Nonetheless, the variants that are present are representative of their respective pathways. Third, to maximize statistical power, we used our full cohort to calculate variant scores and test for effect modification of DNAm-age relationships. This could be a source of bias and is a limitation of this study. Still, our subsequent analysis demonstrating that the endothelial function variant score was a significant modifier of the $PM_{2.5}$-ICAM relationship demonstrates that the variation captured by our score impacts the association of $PM_{2.5}$ with endothelial function markers that are independent of DNAm-age. Notwithstanding this evidence, validation of our score and findings in an independent cohort is a future direction of this work. Fourth, all bisulfite-mediated methods used for quantifying DNA methylation are limited in their ability to distinguish between 5-methylcytosine and its oxidation product 5-hydroxymethylcytosine[57]. Lastly, our findings are based on an elderly cohort of Caucasian males that reside

in a lightly-polluted environment. To date, only one study has explicitly examined race and sex differences in DNAm-age and data from that study suggests that men have higher DNAm-ages than women[19]. Furthermore, there is evidence that race and sex differences can impact individual responses to $PM_{2.5}$ exposure. For instance, one study reported that urban $PM_{2.5}$ levels were associated with asthma exacerbations in African Americans, but not Caucasian Americans[58]. Nevertheless, more work must be done to confirm if these, and similar, reported differential effects are truly due to race/sex or if they are instead due to differences in residential characteristics and other social determinants. A limited amount of research has explicitly explored the race or sex differences of the endothelial function variants selected by our elastic net. One study examining gene-gene interactions that influence pulmonary tuberculosis susceptibility reported strong interactions between the rs2248814 (NOS2A) variant and other genes in African Americans but not Caucasians[59]. Another study reported that the rs1800779 (NOS3) variant was positively associated with high tension primary open glaucoma in women, but not in men[60]. With respect to the results of the present study, additional studies involving other demographic groups, in different environments, and using other assessments of endothelial function will be necessary to confirm our findings more broadly.

4.6. Conclusion

In summary, our findings add evidence that genetic variation can impact the association of long-term fine particle levels with DNAm-age. In particular, the effect of 1-year particle levels on DNAm-age was greater in individuals with a high aging-risk endothelial function genetic variant profile when compared to individuals with a low aging-risk variant profile. We also report novel, robust positive associations of serum endothelial markers with DNAm-age. Although the biological relevance of DNAm-age is still greatly undefined, our study makes a valiant, early attempt at addressing this important research gap. Again, future studies in different populations using these and other endothelial markers will be necessary to broaden the understanding of the relationship of endothelial function with DNAm-age.

4.7. Contributions and Support

Funding Information: This work was supported by grants from the National Institute of Environmental Health Sciences (NIEHS) (R01ES021733 and R01ES025225). Other support comes from NIEHS grants ES015172, ES014663, ES020010, P30ES009089; Environmental Protection Agency (EPA) grants RD832416 and RD83587201; and National Heart, Lung, and Blood Institute (NHLBI) grant 2T32HL007118-41. Research reported in this publication was also supported by the Office of the Director, National Institutes of Health of the National Institutes of Health under Award Number DP5OD021412. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The US Department of Veterans Affairs (VA) Normative Aging Study (NAS) is supported by the Cooperative Studies Program/ERIC, US Department of Veterans Affairs, and is a research component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). The views expressed in this paper are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs. Additional support was provided by the US Department of Agriculture, Agricultural Research Service (contract 53-K06-510).

Contributors: JCN and JDS conceived and designed the study. EC, QD, ACJ, LH, and PV gathered data. JCN performed the data analyses and drafted the manuscript. MB, LD, JDS, EC, YO, MGW, and AAB contributed to the analyses. All authors revised and approved the manuscript.

Conflict of interest statement: None declared

Ethics approval: Boston VA Medical Center, Harvard T.H. Chan School of Public Health (protocol 14027-102).

Data Availability

Data are from the Normative Aging Study, from which restricted data are available for researchers who meet the criteria. A subset of the methylation data is deposited at NCBI dbGaP (study accession number: phs000853.v1.p1.

4.8. References

1 Mayor, S. Nine in 10 people are exposed to air pollution over WHO limits, warns report. *BMJ: British Medical Journal (Online)*, 354 (2016).

2 Apte, J. S., Marshall, J. D., Cohen, A. J. & Brauer, M. Addressing Global Mortality from Ambient PM2.5. *Environ. Sci. Technol.* **49**, 8057-8066, doi:10.1021/acs.est.5b01236 (2015).

3 Martinelli, N., Olivieri, O. & Girelli, D. Air particulate matter and cardiovascular disease: a narrative review. *Eur. J. Intern. Med.* **24**, 295-302, doi:10.1016/j.ejim.2013.04.001 (2013).

4 Zhong, J. *et al.* Cardiac Autonomic Dysfunction: Particulate Air Pollution Effects Are Modulated by Epigenetic Immunoregulation of Toll-like Receptor 2 and Dietary Flavonoid Intake. *J Am Heart Assoc* **4**, e001423, doi:10.1161/JAHA.114.001423 (2015).

5 Chen, S. *et al.* Fine Particulate Constituents and Lung Dysfunction: A Time-Series Panel Study. *Environ Sci Technol*, doi:10.1021/acs.est.6b03901 (2017).

6 Schikowski, T. *et al.* Association of air pollution with cognitive functions and its modification by APOE gene variants in elderly women. *Environ. Res.* **142**, 10-16, doi:10.1016/j.envres.2015.06.009 (2015).

7 Power, M. C. *et al.* Traffic-related air pollution and cognitive function in a cohort of older men. *Environ. Health Perspect.* **119**, 682-687, doi:10.1289/ehp.1002767 (2011).

8 Raaschou-Nielsen, O. *et al.* Particulate matter air pollution components and risk for lung cancer. *Environ Int* **87**, 66-73, doi:10.1016/j.envint.2015.11.007 (2016).

9 Lepeule, J., Laden, F., Dockery, D. & Schwartz, J. Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009. *Environ Health Perspect* **120**, 965-970, doi:10.1289/ehp.1104660 (2012).

10 Lane-Cordova, A. D. *et al.* Aging, not age-associated inflammation, determines blood pressure and endothelial responses to acute inflammation. *J Hypertens* **34**, 2402-2409, doi:10.1097/HJH.0000000000001103 (2016).

11 Rowe, J. W. & Kahn, R. L. Successful Aging 2.0: Conceptual Expansions for the 21st Century. *J Gerontol B Psychol Sci Soc Sci* **70**, 593-596, doi:10.1093/geronb/gbv025 (2015).

12 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).

13 Nwanaji-Enwerem, J. C. *et al.* Long-term ambient particle exposures and blood DNA methylation age: findings from the VA normative aging study. *Environ Epigenet* **2**, doi:10.1093/eep/dvw006 (2016).

14 Nwanaji-Enwerem, J. C. *et al.* Associations between Long-term Exposure to PM2.5 Component Species and Blood DNA Methylation Age in the Elderly: The VA Normative Aging Study. *Environ Int* **102**, 57-65 (2017).

15 Ward-Caviness, C. K. *et al.* Long-term exposure to air pollution is associated with biological aging. *Oncotarget*, doi:10.18632/oncotarget.12903 (2016).

16 Wolf, E. J. *et al.* Accelerated DNA methylation age: Associations with PTSD and neural integrity. *Psychoneuroendocrinology* **63**, 155-162, doi:10.1016/j.psyneuen.2015.09.020 (2016).

17 Levine, M. E. *et al.* DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY)* **7**, 690-700 (2015).

18 Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* **8**, 1844-1865, doi:10.18632/aging.101020 (2016).

19 Horvath, S. *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol* **17**, 171, doi:10.1186/s13059-016-1030-0 (2016).

20 Lowe, D., Horvath, S. & Raj, K. Epigenetic clock analyses of cellular senescence and ageing. *Oncotarget* **7**, 8524-8531, doi:10.18632/oncotarget.7383 (2016).

21 Bind, M. A. *et al.* A novel genetic score approach using instruments to investigate interactions between pathways and environment: application to air pollution. *PLoS One* **9**, e96000, doi:10.1371/journal.pone.0096000 (2014).

22 Dai, L. *et al.* Fine particles, genetic pathways, and markers of inflammation and endothelial dysfunction: Analysis on particulate species and sources. *J Expo Sci Environ Epidemiol* **26**, 415-421, doi:10.1038/jes.2015.83 (2016).

23 Mordukhovich, I. *et al.* Exposure to sub-chronic and long-term particulate air pollution and heart rate variability in an elderly cohort: the Normative Aging Study. *Environ Health* **14**, 87, doi:10.1186/s12940-015-0074-z (2015).

24 Mordukhovich, I. *et al.* Association Between Particulate Air Pollution and QT Interval Duration in an Elderly Cohort. *Epidemiology* **27**, 284-290, doi:10.1097/EDE.0000000000000424 (2016).

25 Bell, B., Rose, C. L. & Damon, A. The Veterans Administration Longitudinal Study of Healthy Aging. *The Gerontologist* **6**, 179-184, doi:10.1093/geront/6.4.179 (1966).

26 Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-1369, doi:10.1093/bioinformatics/btu049 (2014).

27 Di, Q., Koutrakis, P. & Schwartz, J. A hybrid prediction model for PM 2.5 mass and components using a chemical transport model and land use regression. *Atmospheric Environment* **131**, 390-399 (2016).

28 Alexeeff, S. E. *et al.* Medium-term exposure to traffic-related air pollution and markers of inflammation and endothelial function. *Environ Health Perspect* **119**, 481-486, doi:10.1289/ehp.1002560 (2011).

29 Schnoor, M., Alcaide, P., Voisin, M. B. & van Buul, J. D. Crossing the Vascular Wall: Common and Unique Mechanisms Exploited by Different Leukocyte Subsets during Extravasation. *Mediators Inflamm* **2015**, 946509, doi:10.1155/2015/946509 (2015).

30 Jaipersad, A. S., Lip, G. Y., Silverman, S. & Shantsila, E. The role of monocytes in angiogenesis and atherosclerosis. *J Am Coll Cardiol* **63**, 1-11, doi:10.1016/j.jacc.2013.09.019 (2014).

31 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86, doi:10.1186/1471-2105-13-86 (2012).

32 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & Team, R. C. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1ñ120 h ttp. *CRAN. R-project. org/package= nlme* (2015).

33 Safran, M. *et al.* GeneCards Version 3: the human gene integrator. *Database (Oxford)* **2010**, baq020, doi:10.1093/database/baq020 (2010).

34 Lever, J., Krzywinski, M. & Altman, N. Points of Significance: Regularization. *Nature Methods* **13**, 803-804 (2016).

35 Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301-320 (2005).

36 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).

37 Lenters, V. *et al.* Prenatal Phthalate, Perfluoroalkyl Acid, and Organochlorine Exposures and Term Birth Weight in Three Birth Cohorts: Multi-Pollutant Models Based on Elastic Net Regression. *Environ Health Perspect* **124**, 365-372, doi:10.1289/ehp.1408933 (2016).

38 Ren, C. *et al.* Air pollution and homocysteine: more evidence that oxidative stress-related genes modify effects of particulate air pollution. *Epidemiology* **21**, 198-206, doi:10.1097/EDE.0b013e3181cc8bfc (2010).

39 Ren, C. *et al.* Effect modification of air pollution on Urinary 8-Hydroxy-2'-Deoxyguanosine by genotypes: an application of the multiple testing procedure to identify significant SNP interactions. *Environ Health* **9**, 78, doi:10.1186/1476-069X-9-78 (2010).

40 Ayala-Haedo, J. A. *et al.* Analysis of single nucleotide polymorphisms in the NOS2A gene and interaction with smoking in age-related macular degeneration. *Ann Hum Genet* **74**, 195-201, doi:10.1111/j.1469-1809.2010.00570.x (2010).

41 Hancock, D. B., Martin, E. R., Vance, J. M. & Scott, W. K. Nitric oxide synthase genes and their interactions with environmental factors in Parkinson's disease. *Neurogenetics* **9**, 249-262, doi:10.1007/s10048-008-0137-1 (2008).

42 Zhang, J. *et al.* Genetic variants in inducible nitric oxide synthase gene are associated with the risk of radiation-induced lung injury in lung cancer patients receiving definitive thoracic radiation. *Radiother Oncol* **111**, 194-198, doi:10.1016/j.radonc.2014.03.001 (2014).

43 Xiao, X., Cao, L., Wang, R., Shen, Z. X. & Cao, Y. X. Airborne fine particulate matter alters the expression of endothelin receptors in rat coronary arteries. *Environ Pollut* **218**, 487-496, doi:10.1016/j.envpol.2016.07.028 (2016).

44 Wu, S. *et al.* Chemical constituents and sources of ambient particulate air pollution and biomarkers of endothelial function in a panel of healthy adults in Beijing, China. *Sci Total Environ* **560-561**, 141-149, doi:10.1016/j.scitotenv.2016.03.228 (2016).

45 Toth, P., Tarantini, S., Csiszar, A. & Ungvari, Z. I. Functional Vascular Contributions to Cognitive Impairment and Dementia (VCID): Mechanisms and Consequences of Cerebral Microvascular Dysfunction in Aging. *Am J Physiol Heart Circ Physiol*, ajpheart 00581 02016, doi:10.1152/ajpheart.00581.2016 (2016).

46 Mendonca, G. V., Pezarat-Correia, P., Vaz, J. R., Silva, L. & Heffernan, K. S. Impact of Aging on Endurance and Neuromuscular Physical Performance: The Role of Vascular Senescence. *Sports Med*, doi:10.1007/s40279-016-0596-8 (2016).

47 Abbas, M. *et al.* Endothelial Microparticles from Acute Coronary Syndrome Patients Induce Premature Coronary Artery Endothelial Cells Ageing and Thrombogenicity: Role of the Ang II/AT1 Receptor/NADPH Oxidase-mediated Activation of MAPKs and PI3-kinase Pathways. *Circulation*, doi:10.1161/CIRCULATIONAHA.116.017513 (2016).

48 Tchalla, A. E. *et al.* Elevated Soluble Vascular Cell Adhesion Molecule-1 Is Associated With Cerebrovascular Resistance and Cognitive Function. *J Gerontol A Biol Sci Med Sci*, doi:10.1093/gerona/glw099 (2016).

49 Ricottini, E. *et al.* Effect of High-Dose Atorvastatin Reload on the Release of Endothelial Progenitor Cells in Patients on Long-Term Statin Treatment Who Underwent Percutaneous

Coronary Intervention (from the ARMYDA-EPC Study). *Am J Cardiol* **117**, 165-171, doi:10.1016/j.amjcard.2015.10.043 (2016).

50 Zanobetti, A., Franklin, M., Koutrakis, P. & Schwartz, J. Fine particulate air pollution and its components in association with cause-specific emergency admissions. *Environ Health* **8**, 58, doi:10.1186/1476-069X-8-58 (2009).

51 Ando, K. & Fujita, T. Inhibitory effect of ammonium chloride on acetylcholine-induced relaxation. *Hypertension* **24**, 189-194 (1994).

52 Lin, C. J. *et al.* Indoxyl Sulfate Impairs Endothelial Progenitor Cells and Might Contribute to Vascular Dysfunction in Patients with Chronic Kidney Disease. *Kidney Blood Press Res* **41**, 1025-1036, doi:10.1159/000452604 (2016).

53 Fan, A. M., Alexeeff, G. & Khan, E. *Toxicology and risk assessment*. (CRC Press, 2015).

54 Dadvand, P. *et al.* Particulate air pollution and preeclampsia: a source-based analysis. *Occup Environ Med* **71**, 570-577, doi:10.1136/oemed-2013-101693 (2014).

55 Kioumourtzoglou, M. A. *et al.* Exposure measurement error in PM2.5 health effects studies: a pooled analysis of eight personal exposure validation studies. *Environ Health* **13**, 2, doi:10.1186/1476-069X-13-2 (2014).

56 Weisskopf, M. G. & Webster, T. F. Trade-offs of personal vs. more proxy exposure measures in environmental epidemiology. *Epidemiology* (2017).

57 Reinders, J. & Paszkowski, J. Bisulfite methylation profiling of large genomes. *Epigenomics* **2**, 209-220, doi:10.2217/epi.10.6 (2010).

58 Glad, J. A. *et al.* The Relationship of Ambient Ozone and PM2.5 Levels and Asthma Emergency Department Visits: Possible Influence of Gender and Ethnicity. *Archives of Environmental & Occupational Health* **67**, 103-108, doi:10.1080/19338244.2011.598888 (2012).

59 Velez, D. R. *et al.* NOS2A, TLR4, and IFNGR1 interactions influence pulmonary tuberculosis susceptibility in African-Americans. *Hum Genet* **126**, 643-653, doi:10.1007/s00439-009-0713-y (2009).

60 Kang, J. H. *et al.* Endothelial nitric oxide synthase gene variants and primary open-angle glaucoma: interactions with sex and postmenopausal hormone use. *Invest Ophthalmol Vis Sci* **51**, 971-979, doi:10.1167/iovs.09-4266 (2010).

Chapter 5:

Impacts of the Mitochondrial Genome on the Relationship of Long-term Ambient Fine Particle Exposure with Blood DNA Methylation Age

Jamaji C. Nwanaji-Enwerem[a*]; Elena Colicino[b]; Lingzhen Dai[a]; Akin Cayir[a,c]; Marco Sanchez-Guerra[a,d]; Hannah E. Laue[a,b]; Vy T. Nguyen[a]; Qian Di[a]; Allan C. Just[e]; Lifang Hou[f]; Pantel Vokonas[g]; Brent A. Coull[h]; Marc G. Weisskopf[a]; Andrea A. Baccarelli[b]; and Joel D. Schwartz[a]

[a]*Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA, 02115*

[b]*Department of Environmental Health Sciences, Columbia Mailman School of Public Health, New York, NY, USA, 10032*

[c]*Vocational Health College, Canakkale Onsekiz Mart University, Canakkale, Turkey, 17100*

[d]*Department of Developmental Neurobiology, National Institute of Perinatology, Mexico City, Mexico, 11000*

[e]*Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA, 10029*

[f]*Center for Population Epigenetics, Department of Preventive Medicine, Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, 60611*

[g]*VA Normative Aging Study, Veterans Affairs Boston Healthcare System and the Department of Medicine, Boston University School of Medicine, Boston, MA, USA, 02118*

[h]*Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA, 02115*

5.1. Abstract

Background: The mitochondrial genome has long been implicated in age-related disease, but no studies have examined its role in the relationship of long-term fine particle ($PM_{2.5}$) exposure and DNA methylation age (DNAm-age) – a novel measure of biological age.

Objective: In this analysis based on 940 observations between 2000 and 2011 from 552 Normative Aging Study participants, we determined the roles of mitochondrial DNA haplogroup variation and mitochondrial genome abundance in the relationship of $PM_{2.5}$ with DNAm-age.

Methods: We used the GEOS-chem transport model to estimate address-specific, one-year $PM_{2.5}$ levels for each participant. DNAm-age and mitochondrial DNA markers were measured from participant blood samples.

Results: Nine haplogroups (H, I, J, K, T, U, V, W, and X) were present in the population. In fully-adjusted linear mixed-effects models, the association of $PM_{2.5}$ with DNAm-age (in years) was significantly diminished in carriers of haplogroup V ($P_{interaction}$= 0.01; β= 0.18, 95%CI: -0.41, 0.78) compared to non-carriers (β= 1.25, 95%CI: 0.58, 1.93). Mediation analysis estimated that decreases in mitochondrial DNA copy number, a measure of mitochondrial genome abundance, mediated 12% of the association of $PM_{2.5}$ with DNAm-age.

Conclusion: Our data suggests that the mitochondrial genome plays a role in DNAm-age relationships particularly in the context of long-term $PM_{2.5}$ exposure.

5.2. Introduction

Research continues to implicate long-term fine particulate air pollution ($PM_{2.5}$) as a major risk factor for aging and age-related disease. For instance, a recent study of over 500 elderly individuals reported a 27% increase in the risk of an individual developing metabolic syndrome for every 1 $\mu g/m^3$ increase in annual $PM_{2.5}$ concentration[1]. In addition to metabolic disease, long-term $PM_{2.5}$ exposure has been associated with an increased likelihood of all-cause mortality[2] and significant deficits in cardiac autonomic function[3], cognitive performance[4], and respiratory ability[5]. Even on a molecular level, researchers have consistently demonstrated relationships of $PM_{2.5}$ with more traditional measures of biological aging including, telomere length[6] and inflammatory markers[7]. These $PM_{2.5}$-biological aging relationships are of particular interest because they often persist independent of age-related diseases and they may offer insight as to how $PM_{2.5}$ interacts with biological systems to adversely impact human health. Recently, researchers reported positive associations of long-term $PM_{2.5}$ with DNA methylation age (DNAm-age)[8], a novel tissue-independent measure of biological age calculated from DNA methylation values at 353 age-correlated CpG dinucleotides[9]. Furthermore, the authors examined the relationships of five major $PM_{2.5}$ component species (ammonium, elemental carbon, organic carbon, nitrate, and sulfate) with DNAm-age, and found that sulfate and ammonium were most associated with DNAm-age[10]. Although the ability of DNAm-age to reflect previous environmental exposures and predict multiple health outcomes makes it a promising biomarker of aging[11-15], it is so novel that the molecular implications of these relationships remain largely unknown. Addressing this research gap is of paramount importance for future aging research involving this biomarker.

Mitochondria are membrane-bound intracellular organelles tasked with energy production and highly involved in the biological aging processes[16-17]. Mitochondria possess their own genomes which exist as circular double-stranded molecules of DNA that code for a number of biological effectors including some major components of the energy-generating electron transport chain (ETC)[18-19]. Due to the mitochondrial genome's proximity to the ETC (the major source of intracellular reactive oxygen species) and its diminished DNA repair capacity (in comparison to nuclear DNA), mitochondrial DNA is particularly susceptible to oxidative damage[20]. There is also convincing evidence that the mitochondrial

genome is susceptible to damage from exogenous oxidative stressors[21-23]. Overall, it is the damage to the mitochondrial genome that has been specifically linked to accelerated aging[24-25]. Given that $PM_{2.5}$ exposure is a known risk factor of systemic oxidative stress[26] and since $PM_{2.5}$ exposure has already been linked to mitochondrial genome integrity[27-28], we hypothesized that the mitochondrial genome may be involved in the $PM_{2.5}$-DNAm-age relationship.

In the present study, we examined the impact of mitochondrial genomic variation and abundance (a compensatory response to poor mitochondrial genome integrity) on the relationship of $PM_{2.5}$ and its component species with DNAm-age. First, we examined if different mitochondrial haplogroups (forms of normal mitochondrial genetic variation that potentially impact ETC capacity)[29-30] modified the association of $PM_{2.5}$ and its component species with DNAm-age. Next, we determined the relationship of one commonly used measure of mitochondrial genome abundance, mitochondrial DNA copy number, with DNAm-age. Mitochondrial copy number is the ratio of a cell's mitochondrial DNA to nuclear DNA. Fluctuations in mitochondrial copy number often occur with normal mitochondrial biogenesis and degradation, but the measure is also sensitive to exogenous stressors and is thought to be an adaptive response to compensate for mitochondrial genome damage[31]. Copy number has already been associated with $PM_{2.5}$ levels[32], but no studies have examined its relationship with DNAm-age. Finally, we determined if copy number mediated and/or modified the association of $PM_{2.5}$ with DNAm-age.

5.3. Materials and Methods

5.3.a. Study Population

Participants in the present analysis were active participants in the Veteran Affairs Normative Aging Study (NAS), a longitudinal cohort study of aging established in 1963[33]. The NAS is a closed cohort of now elderly community-dwelling men living in the Greater Boston area. At enrollment, all participants were free of chronic diseases. Participants return every 3 to 5 years for onsite, follow-up study visits. During these recurring visits, participants receive comprehensive outpatient medical evaluations, bio-specimens (including blood) are collected, and participants provide detailed information about their diets and other

lifestyle factors that may affect their health. All participants provided written informed consent to the VA Institutional Review Board (IRB), and human subjects approval was granted by the VA and Harvard T.H. Chan School of Public Health IRBs (protocol 14027-102).

All NAS men with continued study participation as of the year 2000, when address-specific $PM_{2.5}$ component species levels became available, were eligible for the present study sample. We began with a total of 552 participants with 940 observations between the years 2000 and 2011. This was the study sample that was used in reporting the significant associations between $PM_{2.5}$ component species and DNAm-age in our previous publication[10]. Of these 552 participants, 249 (45%) had one visit, 218 (40%) had two visits, and 85 (15%) had three or more visits. From this sample, we then excluded participants missing mitochondrial haplogroup data. This resulted in a final study sample of 508 participants with 870 total study visits. In the final study sample, 227 participants (45%) had one visit, 200 (39%) had two visits, and 81 (16%) had three or more visits.

5.3.b. DNA Methylation and DNA Methylation Age (DNAm-age)

Whole blood was collected from each participant during each NAS follow-up visit. We performed bisulfite conversion (EZ-96 DNA Methylation Kit, Zymo Research, Orange, CA, USA) on extracted DNA from the buffy coat of the whole blood, and then used the Illumina Infinium HumanMethylation450 BeadChip to measure the DNA methylation of CpG probes. To minimize batch effects and ensure a similar age distribution across chips and plates, we randomized chips across plates and used a two-stage age-stratified algorithm to randomize samples. For quality control, we removed samples where >5% of probes had a beadcount < 3 or > 1% of probes had a detection P-value >0.05. After pre-processing the remaining samples with Illumina-type background correction without normalization and normalizing the samples with dye-bias and BMIQ3 adjustments, we generated methylation beta values[34]. Beta values represent the percentage of methylation for each of the ~480,000 CpG sites in the BeadChip array. In other words, beta = intensity of the methylated signal (M) / [intensity of the unmethylated signal (U) + intensity of the methylated signal (M) + 100].

DNAm age was calculated using Horvath's publically available online calculator (http://labs.genetics.ucla.edu/horvath/dnamage/). DNAm-age was derived from an elastic net penalized regression run on multiple data sets of different cell and tissue types. After 21,369 CpG probes – shared by both the Illumina HumanMethylation27 and HumanMethylation450 BeadChip platforms – were regressed on a calibrated version of chronological age, the elastic net selected 353 CpGs that correlated with age (193 positively and 160 negatively)[9]. The model coefficients from these 353 CpGs were used by the calculator to predict the age of each DNA sample (*i.e.* DNAm-age). The calculator maintains predictive accuracy (age correlation 0.97, error = 3.6 years) across almost all body tissues including blood and brain[9].

5.3.c. Fine Particulate ($PM_{2.5}$) Air Pollution

We used the simulation outputs from GEOS-chem, a chemical transport model[35], fused with land-use variables to generate one-year exposure estimates for $PM_{2.5}$ as well as sulfate and ammonium, the major $PM_{2.5}$ component species demonstrated to be most important in predicting DNAm-age[10]. Ten-fold cross-validation demonstrated that the model performed well for $PM_{2.5}$ mass and its component species with R^2s ranging from 0.70 to 0.88[36]. Existing literature demonstrates that the one-year $PM_{2.5}$ exposure window is robustly associated with DNAm-age[8]. We generated daily estimates at the 1 km x1 km area resolution and one-year total $PM_{2.5}$ and $PM_{2.5}$ component species exposure windows by averaging daily exposures for the 365 days prior to the day of each participants' NAS visit. Given that greater than 90% of NAS participants are retired, home address exposures are expected to be a good proxy for their individual ambient exposures. After geocoding and linking participants' residencies to an area level grid-point; and accounting for address changes and time spent away from home (>7 days), we assigned particle estimates to each participant's address.

5.3.d. Mitochondrial (DNA) Haplogroups

Participant blood was genotyped using Taqman or Sequenom assays (Applied Biosystems, Foster City, CA)[37]. All samples were successfully genotyped and internal blinded quality control samples were

>99% concordant. Hardy-Weinberg equilibrium tests were not assessed because they are not valid for mitochondrial polymorphisms[38]. No heteroplasmy (heterozygous samples) were observed. Nine mitochondrial DNA haplogroups (H, I, J, K, T, U, V, W, and X) were observed in the cohort and all are common to populations of European ancestry. Based on the phylogenetic evolutionary tree and restriction fragment length polymorphisms, these haplogroups can be grouped into four clusters (Cluster 1:J,T; Cluster 2: V,H; Cluster 3: U,K; Cluster 4: I, W, X)[39]. The clusters are widely known, and since overall type I error increases as the number of statistical tests increases, many epidemiologic studies first perform cluster analyses[37, 40]. Following this framework, we perform primary cluster analyses and subsequently explore individual haplogroups of interest.

5.3.e. Mitochondrial DNA Copy Number (Genome Abundance)

As noted, the mitochondrial genome is particularly vulnerable to both endogenous and exogenous (*e.g.* air pollution) oxidative stressors due to its proximity to the ETC, lack of protective barriers (*i.e.* histone proteins, chromatin organization, etc.), and relatively limited DNA damage repair activity[41]. We are utilizing one measurement of mitochondrial genome abundance (copy number) that is sensitive to oxidative stress. Mitochondrial copy number represents the ratio of mitochondrial DNA copy number to the nuclear DNA copy number (mtDNA:nDNA) and was also calculated from whole blood samples collected at every visit. As previously described[31], real-time PCR (RT-PCR) is used to measure the ratio of a mitochondrial gene (mtDNA 12S ribosomal ribonucleic acid) to a nuclear gene (Ribonuclease P gene), which is normalized to a reference DNA sample (a pool of 300 test samples) to obtain relative mitochondrial DNA copy number values controlled for plate effects.

5.3.f. Statistical Analysis

5.3.f.1.Covariates

The relationships of the mitochondrial genome with DNAm-age and its role in the association of $PM_{2.5}$ with DNAm-age were evaluated using linear mixed-effects models including a random participant-

specific intercept to account for correlation between repeated outcome measures (*i.e.* multiple visits for a participant). In the analyses, we controlled for the following covariates *a priori* based on previous analyses[8, 10] and the relevant literature[42-43]: chronological age (continuous), blood cell proportions [plasma cells, CD4+ lymphocytes, CD8+ lymphocytes, natural killer (NK) cells, monocytes, and granulocytes] (continuous) determined via Houseman and Horvath methods[9, 44], average 1-year temperature (continuous address-specific satellite measurements[10]), cumulative cigarette pack years (continuous), smoking status (current, former, or never), season of visit (spring [March-May], Summer [June-August], Fall [September-November], and Winter [December-February]), body mass index (BMI) (lean [<25], overweight [25-30], obese [>30]), alcohol intake (yes/no ≥ 2 drinks daily), maximum years of education (continuous), cancer (yes/no history of lifetime cancer diagnosis), coronary heart disease (yes/no based on electrocardiogram, validated medical records, or physical exam), diabetes (physician diagnosis or a fasting blood glucose > 126 mg/dL), and hypertension (yes/no antihypertensive medication use or systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg).

5.3.f.2. Direct Associations

We first used fully-adjusted linear mixed effects models to evaluate previously published positive associations of one-year $PM_{2.5}$, sulfate, and ammonium levels with DNAm-age. Sulfate and ammonium models were additionally adjusted for $PM_{2.5}$ mass. To limit multiple comparisons and the potential for false positive results, we performed mitochondrial haplogroup cluster analyses – as conducted in a previously published NAS study of haplogroups[37] – evaluating the direct relationships of mitochondrial haplogroup clusters with DNAm-age and mitochondrial DNA copy number. We also used fully-adjusted mixed-effects models to determine the associations of mitochondrial DNA copy number with DNAm-age.

5.3.f.3. Mitochondrial Haplogroup as an Effect Modifier

Since haplogroup is a genetic parameter that does not change during life, it is not on the causal pathway of the exposure and outcome but could potentially impact the relationship of the exposure with the

outcome. For these reasons, it appropriate to consider it as an effect modifier. Specifically, we evaluated if the haplogroup clusters modified the associations of $PM_{2.5}$, sulfate, and ammonium with DNAm-age. In these analyses the reference group was all participants without the cluster of interest. For example, when we evaluated the modifying role of mitochondrial haplogroup cluster 1 on the association of $PM_{2.5}$ with DNAm-age, we compared participants genotyped as having cluster 1 against all other participants (*i.e.* participants genotyped as having clusters 2, 3, and 4). Structuring the analyses this way allows us to compare the findings of each specific haplogroup cluster to a mixed population of haplogroup clusters. This helps with interpreting the results especially since there is no strict biological evidence that defines one particular haplogroup cluster as a control or reference group. After determining clusters with statistically significant modifying effects on the $PM_{2.5}$-DNAm-age relationship, we re-ran the models testing the modifying role of the individual haplogroups within those particular clusters and with individual $PM_{2.5}$ components (sulfate and ammonium) as the predictors. Again, all of these models were fully-adjusted.

5.3.f.4. Mitochondrial DNA Copy Number as an Effect Modifier and/or Mediator

Unlike haplogroups, which are determined at birth and remain the same throughout life, copy number can change throughout life. In fact, empirical evidence exploring the relationships between short-term versus long-term $PM_{2.5}$ exposure and mitochondrial genome abundance suggest that copy number is subject to much change over time[45]. Moreover, there is experimental evidence demonstrating that depletion of the mitochondrial genome results in aberrant methylation of nuclear DNA at promoter CpG islands[46]. Given this evidence, we hypothesized that the association of $PM_{2.5}$ with DNAm-age could be mediated through and/or modified by copy number. To test this hypothesis, we employed a 4-way decomposition mediation method. Standard methods of testing for effect modification operate under the assumption that the modifier is not on the casual pathway between the exposure and outcome. Thus, these results may be misleading if mediation is truly present and the candidate modifier is indeed on the causal pathway[47]. The 4-way decomposition method circumvents the risk of these potentially misleading results by allowing one to simultaneously parse out: 1) the controlled direct effect [the effect of the exposure on the outcome due

neither to mediation nor interaction]; 2) the reference interaction [the effect of the exposure on the outcome due to interaction alone]; 3) the mediated interaction [the effect of the exposure on the outcome due to mediation and interaction]; and 4) the pure indirect effect [the effect of the exposure on the outcome due to mediation alone][48].

As mentioned earlier, there is some risk of simple interaction (effect modification) models resulting in misleading results if the candidate modifier is indeed on the causal pathway between the exposure and the outcome. This is not the case for simple mediation analyses because the goal of mediation is to provide evidence that a variable of interest is or is not on the causal pathway[47]. Due to the newness of the 4-way decomposition method, we performed a sensitivity analysis using a standard, simple mediation approach where we used fully-adjusted linear mixed-effects models and modeled[49-50]: Step 1) $PM_{2.5}$ as a predictor of DNAm-age; Step 2) $PM_{2.5}$ as a predictor of mitochondrial DNA copy number; and Step 3) $PM_{2.5}$ as a predictor of DNAm-age controlling for mitochondrial DNA copy number. The proportion of the effect mediated by mitochondrial DNA copy number was calculated as the percentage of natural indirect effect over the sum of natural direct and natural indirect effect. In other words, [(Step 2 $\beta PM_{2.5}$ * Step 3 βmtDNA CN)] / [(Step 2 $\beta PM_{2.5}$ * Step 3 βmtDNA CN) + (Step 3 $\beta PM_{2.5}$)]. The statistical significance of the mediation effect was assessed via the Sobel Z test.

5.3.f.5. Additional Sensitivity Analyses

Although all the covariates for diabetes, CHD, hypertension, and BMI were categorized using well-known and biologically relevant definitions, we performed sensitivity analyses examining if any resolution on potential confounding was lost by using these discrete categories. Specifically, we re-ran our direct association models and the simple mediation analysis (which specifically lists out mediation steps) using fully adjusted models where the aforementioned categorical variables were replaced with continuous measures of fasting blood glucose, total cholesterol, HDL cholesterol, systolic blood pressure, diastolic blood pressure, and BMI.

Our mediation analyses used fully-adjusted models; thus, we assumed limited exposure-outcome, exposure-mediator, and mediator-outcome confounding. However, due to the nature of this prospective repeated measures study, changes in DNAm-age at one visit could potentially affect copy number at a subsequent visit[51]. Hence, we performed an analysis testing the aforementioned association to check the assumption of time-varying confounding.

5.3.f.6. Analysis Software

The 4-way decomposition mediation analysis was performed with a published SAS macro in SAS, version 9.3 (SAS Institute, Inc., Cary, North Carolina)[48]. All other statistical analyses were performed using R Version 3.1.1 (R Core Team, Vienna, Austria) and we considered a P-value < 0.05 to be statistically significant.

5.4. Results

5.4.a. Descriptive Statistics

Table 1 describes the demographic and clinical data for all participants. Participants had a mean (SD) DNAm-age of 74.1 (7.89) years and mean (SD) age of 74.8 (6.97) years. A majority of the men had completed at least 12 years of formal education (74%), consumed less than 2 drinks a day (81%), were former smokers (65%), and did not have coronary heart disease (65%) or diabetes (82%). The mean (SD) exposure levels for $PM_{2.5}$, sulfate, and ammonium were 10.3 (2.13) $\mu g/m^3$, 3.39 (0.80) $\mu g/m^3$, and 1.04 (0.28) $\mu g/m^3$ respectively. Most participants were genotyped as having mitochondrial haplogroup cluster 2 (51%). 17% of the participants were cluster 1, 23% were cluster 3, and 9% were cluster 4. In regards to individual haplogroup frequencies, a majority of the participants were haplogroup V carriers (42%). Additional individual haplogroup frequencies are reported in Table S1.

Table 1. Characteristics of Study Subjects (2000 – 2011)

Main Variables		**All Visits, N = 870**
Age (years), mean (SD)		74.8 (6.97)
DNAm-age (years), mean (SD)		74.1 (7.89)
One- Year Fine Particle Level ($\mu g/m^3$), mean (IQR)		
	$PM_{2.5}$	10.3 (2.13)
	Sulfate	3.39 (0.80)
	Ammonium	1.04 (0.28)
Mitochondrial Haplogroup Cluster, N (%)		
	1 (JT)	146 (17)
	2(VH)	447 (51)
	3 (UK)	200 (23)
	4 (IWX)	87 (9)
Lifestyle and Environmental Variables		
Alcohol Consumption, N (%)		
	< 2 drinks/day	705 (81)
	≥ 2 drinks/day	165 (19)
BMI, N (%)		
	Healthy/Lean	206 (24)
	Overweight	457 (52)
	Obese	207 (24)
Education, N (%)		
	≤ 12 years	224 (26)
	12 – 16 years	402 (46)
	> 16 years	244 (28)
Pack years, mean (SD)		20.9 (24.8)
Smoking Status, N (%)		
	Current	40 (5)
	Former	566 (65)
	Never	264 (30)
Season, N (%)		
	Spring	219 (25)
	Summer	182 (21)
	Fall	298 (34)
	Winter	171 (20)
Temperature (°C), mean (SD)		11.3 (0.98)
Age-Related Diseases		
Coronary Heart Disease, N (%)		
	Yes	308 (35)
	No	562 (65)
Diabetes, N (%)		
	Yes	159 (18)
	No	711 (82)
Hypertension, N (%)		
	Yes	639 (73)
	No	231 (27)
Lifetime Cancer Diagnosis, N (%)		
	Yes	486 (56)
	No	384 (44)

5.4.b. One-Year Particle Levels and Haplogroup Clusters as Predictors of DNAm-age

Table 2 summarizes the results from fully-adjusted linear mixed-effects models examining the independent relationships of $PM_{2.5}$, sulfate, ammonium, and individual haplogroup clusters with DNAm-age. One-year IQR increases in $PM_{2.5}$ (p=0.007), sulfate (p<0.0001), and ammonium (p=0.0005) were all significantly associated with increases in DNAm-age of at least 0.58 years (approximately 7 months). None of the haplogroup clusters were significantly associated with DNAm-age.

Table 2. Mean One-Year Particulate Concentrations and Mitochondrial Haplogroup Cluster as Independent Predictors of DNAm-age (N = 870)

Predictor	Difference in DNAm-age for IQR (95% CI)	*P*
$PM_{2.5}$	0.64 (0.18, 1.11)	0.007
Sulfate	0.58 (0.35, 0.82)	<0.0001
Ammonium	0.58 (0.26, 0.91)	0.0005
Haplogroup Cluster		
1 (JT)	-0.27 (-1.59, 1.05)	0.69
2 (VH)	-0.42 (-1.43, 0.60)	0.42
3 (UK)	0.93 (-0.28, 2.15)	0.13
4 (IWX)	-0.22 (-1.99, 1.55)	0.81

All models adjusted for chronological age, blood cell type, temperature, pack years, smoking status, season, BMI, alcohol consumption, education, lifetime cancer diagnosis, hypertension, diabetes, and coronary heart disease. Sulfate and ammonium models are additionally adjusted for total $PM_{2.5}$ mass. Haplogroup models are adjusted for all three particles.

5.4.c. Effect Modification by Haplogroup Clusters and Individual Haplogroups

Figure 1 depicts the modifying role of the mitochondrial haplogroup clusters on the association of one-year $PM_{2.5}$ levels with DNAm-age. Only the cluster 2 mitochondrial DNA genotype significantly (p=0.007) modified the association of $PM_{2.5}$ levels with DNAm-age. The effect of $PM_{2.5}$ on DNAm-age was diminished by approximately 1 year when comparing individuals with the cluster 2 genotype to all individuals without the cluster 2 genotype. Figure 2 depicts a subsequent analysis examining the modifying

role of the individual haplogroups in cluster 2 (haplogroups V and H) on the association of $PM_{2.5}$ levels with DNAm-age. The effect of $PM_{2.5}$ on DNAm-age was diminished by approximately 1 year when comparing individuals with the haplogroup V genotype to all individuals without the haplogroup V genotype (p=0.01). Figure 3 depicts the modifying role of haplogroup cluster V on the association of the $PM_{2.5}$ components sulfate and ammonium with DNAm-age. Similar to total $PM_{2.5}$, the association of ammonium with DNAm-age was diminished in individuals with a haplogroup V genotype when compared to individuals without a haplogroup V genotype (p=0.03). This relationship persisted even when we included one-year nitrate levels as a covariate in the mixed-effects model (Figure S1).

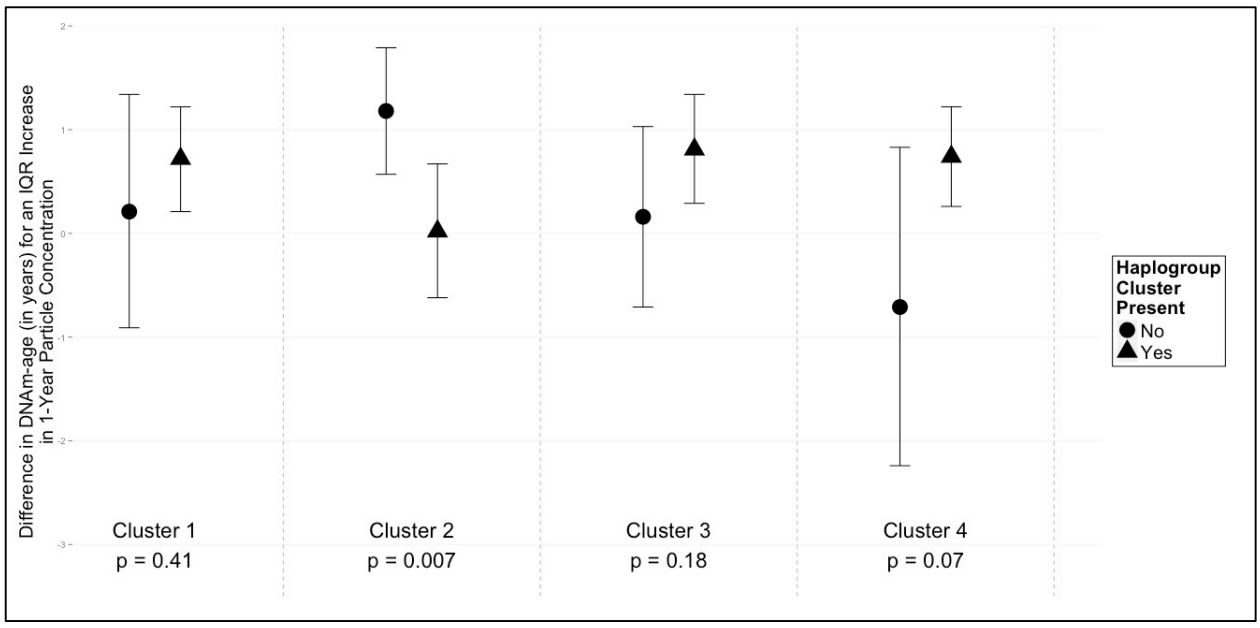

Figure 1 | Difference in DNAm-age for one interquartile range increase in one-year $PM_{2.5}$ exposure comparing participants with and without the respective mitochondrial haplogroup clusters in fully-adjusted mixed-effects models.
Cluster 1 (JT); Cluster 2 (VH); Cluster 3 (UK) and Cluster 4 (IWX).

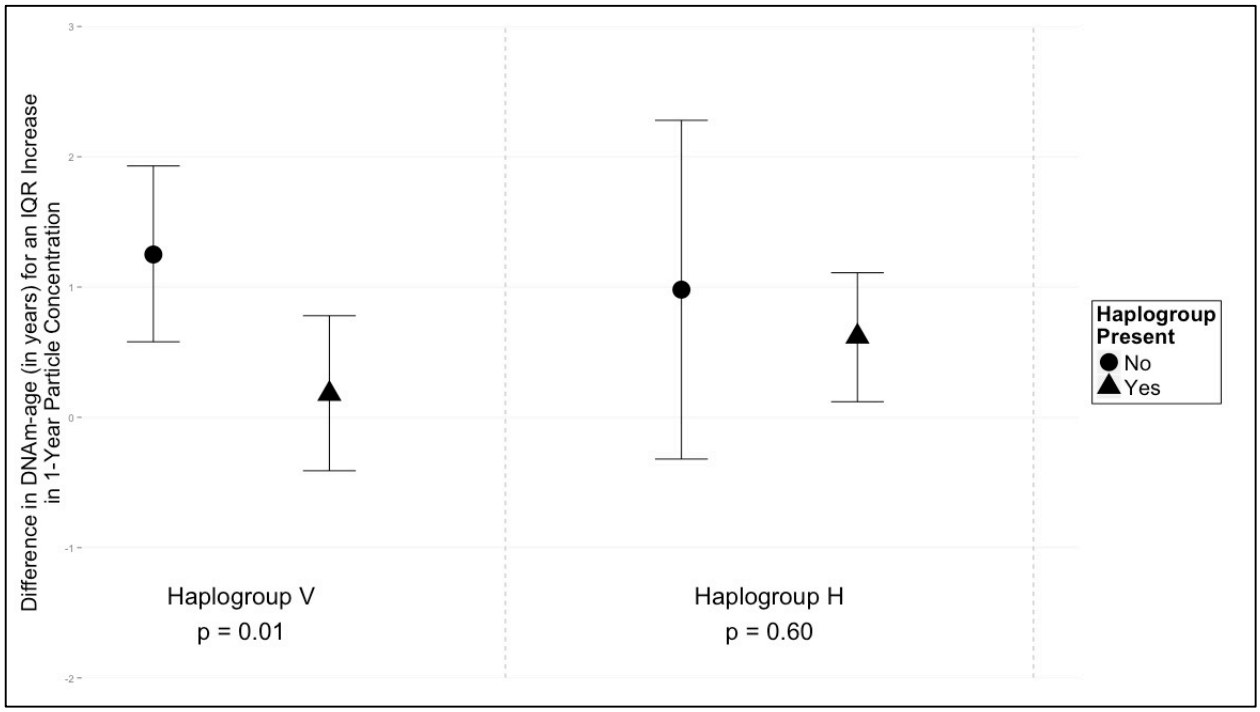

Figure 2 | Difference in DNAm-age for one interquartile range increase in one-year $PM_{2.5}$ exposure comparing participants with and without the respective mitochondrial haplogroups from cluster 2 in fully-adjusted mixed-effects models.

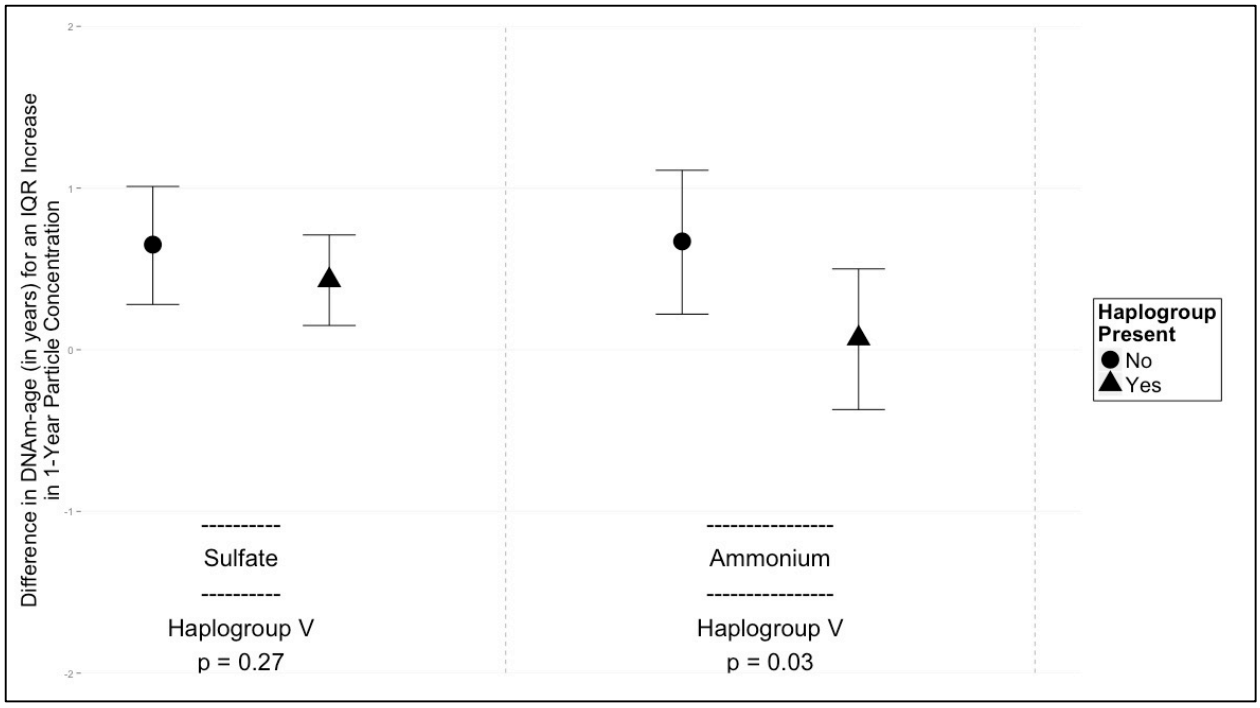

Figure 3 | Difference in DNAm-age for one interquartile range increase in one-year sulfate and ammonium exposure comparing participants with and without the V mitochondrial haplogroup in fully-adjusted mixed-effects models.

5.4.d. Relationships of Mitochondrial DNA Copy Number with DNAm-age

In fully-adjusted linear mixed-effects models examining the relationship of copy number with DNAm-age, we found that copy number (β=-3.31, $p<0.0001$) was significantly, negatively associated with DNAm-age. However, copy number was not significantly associated with chronological age (β=0.57, $p=0.17$) (Table 3). These relationships persisted in sensitivity analysis adjusting for continuous variables instead of disease categories (Table S2).

Table 3. Relationships of Mitochondrial Copy Number with Age and DNAm-age (N=797)

Outcome	Difference in Outcome for IQR (95% CI)	*P*
Age	0.57 (-0.25, 1.39)	0.17
DNAm-age	-3.31 (-4.62, -2.00)	<0.0001

All models adjusted for mitochondrial haplogroup, $PM_{2.5}$, sulfate, ammonium, blood cell type, temperature, pack years, smoking status, season, BMI, alcohol consumption, education, lifetime cancer diagnosis, hypertension, diabetes, and coronary heart disease. DNAm-age model is also adjusted for chronological age.

5.4.e. Mediation Analyses

There was no evidence of time-varying confounding of DNAm-age on mitochondrial DNA copy number (Table S3). Table 4 presents the results of a 4-way decomposition meditation analysis examining the potential role of the mitochondrial DNA copy number as a mediator of the association of $PM_{2.5}$ levels with DNAm-age. The controlled direct effect of one-year $PM_{2.5}$ (due neither to mediation nor interaction) was positive and statistically significant (β=0.81, p=0.02). The pure indirect effect of one year $PM_{2.5}$ (due to mediation alone) was also statistically significant (β=0.22, p=0.02). The percentage of the effect mediated by the copy number was estimated to be 12.2%. There was no evidence of any significant effect modification by copy number. These mediation relationships were consistent with results from the sensitivity analysis using the simple mediation approach (Table S4) and adjusting for continuous variables instead of disease categories (Table S5).

Table 4. Results of 4-Way Decomposition Mediation Analysis of Mitochondrial DNA Copy Number as a Mediator of the Relationship of $PM_{2.5}$ with DNAm-age (N=797)

Effect	Interpretation	β (95% CI)	SE	t	*P*	$P_{mediation}$	% of Effect Mediated by Mediator
Controlled Direct Effect	Due neither to mediation nor interaction	0.81 (0.15, 1.48)	0.34	2.4	0.02	-	-
Reference Interaction	Due to interaction alone	0.12 (-0.04, 0.30)	0.08	1.53	0.13	-	-
Mediated Interaction	Due to mediation and interaction	-0.09 (-0.20, 0.02)	0.06	-1.63	0.10	-	-
Pure Indirect Effect	Due to mediation alone	0.22 (0.07, 0.38)	0.08	2.79	0.01	0.02	12.2

Results based on fully-adjusted models adjusted for chronological age, blood cell type, temperature, pack years, smoking status, season, BMI, alcohol consumption, education, lifetime cancer diagnosis, hypertension, diabetes, and coronary heart disease. Effects were calculated using the published SAS macro where continuous covariates were set to their mean values and categorical variables were set to the category with the greatest proportion of study participants.

5.4.f. Relationships of Haplogroup Clusters with Mitochondrial DNA Copy Number

Table S6 presents the results of fully-adjusted linear mixed-effects models examining the association of each mitochondrial haplogroup cluster with mitochondrial DNA copy number. Only the association with cluster 3 was statistically significant (β=0.06, p=0.02). We also found that the association of $PM_{2.5}$ with mitochondrial DNA copy number was greater in individuals with the haplogroup V genotype when compared to individuals without the haplogroup V genotype (p=0.001) (Figure S2).

5.5. Discussion

In the present study, we used fully-adjusted linear mixed-effects models to investigate the role of the mitochondrial genome in the relationship of long-term $PM_{2.5}$ exposure with DNAm-age in a large longitudinal aging cohort. To our knowledge, this is the first study to demonstrate: 1) that mitochondrial DNA haplogroup V significantly reduces the association of one-year $PM_{2.5}$ and ammonium exposure levels with DNAm-age and 2) that decreases in mitochondrial DNA copy number partially mediate the association of one-year $PM_{2.5}$ exposure levels with DNAm-age. Additionally, we observed novel associations of mitochondrial DNA copy number with DNAm-age and the mitochondrial haplogroup cluster 3 genotype.

The number of studies examining relationships of ambient $PM_{2.5}$ with DNAm-age are limited, but our results are consistent with what they report[8, 10, 52]. Here, we observed comparable, significant positive associations of $PM_{2.5}$, sulfate, and ammonium with DNAm-age. No existing studies have examined the relationships of the mitochondrial genome with DNAm-age; however, we believed that such relationships would exist because both DNAm-age – as previously mentioned – and mitochondrial genome integrity are robustly associated with $PM_{2.5}$ levels[31-32, 53].

With respect to mitochondrial haplogroups specifically, only two studies have examined relationships of mitochondrial haplogroups with air pollution and both studies examined the modifying role that mitochondrial haplogroups may have on the health effects of air pollution. The first study used a panel of 38 subjects with 417 total observations to test if air pollutant exposure-associated inflammation was

stronger in carriers of mitochondrial haplogroup H versus U[54]. The authors justified limiting their study design to haplogroups H and U based on evidence suggesting that genetic variation due to mitochondrial haplogroups impacts the coupling of respiratory chain and the subsequent development of endogenous reactive oxygen species (ROS) by the mitochondria[55]. Haplogroup H has a relatively tightly coupled respiratory chain and has been associated with increased oxidative damage and risk of age-related diseases like Parkinson's[56-57]. On the other hand, haplogroup U has a less tightly coupled respiratory chain and has been shown to be protective against Parkinson's disease[56]. In the end, this study found that air pollutant (black carbon, carbon monoxide, nitric oxides and polycyclic aromatic hydrocarbons) associations with inflammatory markers (IL-6 and TNF-α) were stronger for individuals with the haplogroup H genotype when compared to haplogroup U individuals. Although this study's results were consistent with the aforementioned literature, the study was only based on 38 subjects and only relationships between two haplogroups were explored. In a larger study of 582 subjects with multiple visits, the researchers investigated if 9 different haplogroups (phylogenetically grouped into 4 clusters) resulted in differential susceptibility to cognitive effects of long-term black carbon exposure[37]. These researchers observed impaired cognition in carriers of cluster 1 (J and T) and even worse cognition of carriers of cluster 4 (I, W, and X). No effects were observed in cluster 2 (H and V) or 3 (K and U) carriers. Unlike the first study, these authors did not observe any effect modification in the clusters that contained haplogroups H and V. Another major difference between the two studies is that former reported significant findings with short-term air pollution exposures (≤ 5 days) and the second used a one-year exposure window. In all, the findings of these two studies suggest that the impact of haplogroups on air pollution relationships may be health outcome specific and may vary depending on the duration of air pollution exposure.

Given that DNAm-age has been associated with numerous age-related diseases, we believed that it would be associated with haplogroups that were also associated with age-related diseases, like haplogroup H. Nevertheless, we found no direct associations of haplogroups with DNAm-age in our study sample. However, we did find that haplogroup cluster 2 (V and H) significantly lessened the positive association of $PM_{2.5}$ levels with DNAm-age. Further analyses suggested that this protective effect was predominately due

to haplogroup V and it persisted even when examining the effects of the $PM_{2.5}$ component ammonium. Although cluster 2 haplogroups, like H, have been traditionally thought to be health-adverse, there is also existing evidence that these haplogroups may also offer some health benefits. For instance, haplogroup H carriers were found to have a 2.12 fold increased chance of survival at 180 days following a septic episode compared to non-carriers of haplogroup H[30]. In the sepsis study the researchers did not measure other haplogroups, but it is possible that related haplogroup V could also be protective if explored. Furthermore, in a study that compared the frequency distributions of haplogroups in athletes versus non-athlete controls, researchers found that the V haplogroup was overrepresented in endurance athletes (15.7%) compared with controls (7.5%)[58]. A major issue in existing haplogroup research is that groups being compared are not always the same and often relative findings are being interpreted. Thus, findings of an adverse effect of haplogroup H when it is compared to haplogroup U may not exist when haplogroup H is compared to haplogroup V. In an attempt to remedy future issues with such comparisons, our study always compares carriers of a specific haplogroup or cluster to all other individuals who were not carriers of the haplogroup or cluster. Thus, we are effectively comparing carriers of each haplogroup to a mixed population of haplogroups. Still, future studies using this comparison paradigm will be necessary to confirm our findings of a protective effect of cluster 2 and haplogroup V.

With respect to mitochondrial DNA copy number, our results agree with existing evidence that long-term $PM_{2.5}$ exposure is associated with decreases in mitochondrial DNA copy number[6]. We also report novel evidence that mitochondrial DNA copy number is negatively associated with DNAm-age. Since mitochondrial copy number is viewed as a measure of the mitochondria's ability to respond to and buffer biological stressors, and a reduced copy number can be due to an exhausted mitochondrial buffering capacity (often observed with long-term environmental stresses)[31, 51], it is biologically conceivable that increases in buffering capacity would be associated with less of an "adverse" outcome like aging. Moreover, due to the strong associations of mitochondrial DNA copy number with DNAm-age and the known cross-talk between the nuclear and mitochondrial genomes[59], we believed that one measure may mediate the other's relationship with $PM_{2.5}$. DNAm-age and mitochondrial DNA copy number were measured from

blood taken at the same study visit so we took into account a number of considerations in assessing which would be the most biologically plausible outcome and the most biologically plausible mediator. First, although methylation of some nuclear genes like mitochondrial DNA polymerase γ catalytic subunit (PolgA) have been shown to regulate mitochondrial DNA copy number[60], PolgA methylation does not contribute to the DNAm-age metric[9]. Further, evidence has shown that DNAm-age is not simply the sum of its component CpG DNA methylation levels[8]. Rather, DNAm-age is a biomarker of aging with a unique balance of stability and responsiveness that allows it to simultaneously reflect past exposures[61] and predict future disease risk[62]. This unique balance of stability and responsiveness is best explained by the fact that, aside from the context of induced pluripotent stem cells, DNAm-age appears to only increase with time[63]. In the seminal DNAm-age paper by Horvath (2013), it is hypothesized that "DNAm age measures the cumulative work done by a particular kind of epigenetic maintenance system (EMS), which helps maintain epigenetic stability … This model would explain the high tick rate during organismal development since a high power is required to maintain epigenetic stability during this stressful time. At the end of development, a constant amount of power is sufficient to maintain stability leading to a constant tick rate … DNAm age should be accelerated by many perturbations that affect epigenetic stability[9]." In line with this current understanding of DNAm-age, it is feasible that different environmental exposures or biological microenvironments that affect epigenetic stability could exacerbate the otherwise constant rate of DNAm-age increase[61, 64].

In contrast to DNAm-age, evidence shows that mitochondrial DNA copy number is a more variable metric able to rapidly change (increase or decrease) in response to short-term and long-term exposures but unable to intrinsically record long-term trends because it must be kept within a relatively stable range to maintain optimal physiological function[45]. Moreover, mitochondrial effectors like apoptosis-inducing factor, which are normally localized in the mitochondria, have been shown to translocate to the nucleus where they trigger DNA fragmentation, chromatin condensation, and other DNA changes[59]. Flavin adenine dinucleotide (FAD) and α ketoglutarate (α-KG) are two additional co-factors that are synthesized within the mitochondria, but are actively involved in the processes of nuclear methylation[65]. Thus, for our

mediation analysis, we ultimately found that most evidence supported DNAm-age as an outcome with long-term reporting ability and a mitochondrial copy number as a mediator. Given this analytical framework, we found that mitochondrial DNA copy number significantly mediated about 12% of the observed effect of $PM_{2.5}$ on DNAm-age. Mitochondrial DNA copy number has been shown to decrease with age and copy number has been associated with other age-related outcomes like frailty and mortality[66-67]. Hence, our findings are in agreement with existing aging research but controlled experiments must be performed to confirm if copy number is indeed a mediator of the relationship between $PM_{2.5}$ and DNAm-age.

Finally, we explored the relationships of mitochondrial haplogroups and copy number in our study sample. The weaker association between copy number and DNAm-age in individuals with haplogroup V, may suggest that copy number is not the ultimate source of their protection against the effects of $PM_{2.5}$. This theory is also supported by the findings that 1) copy number is not directly associated with haplogroup cluster 2 (V and H) and 2) copy number only mediates 12% of the association between $PM_{2.5}$ and DNAm-age.

The current study possesses a number of strengths including the use of novel biomarker and a large longitudinal cohort with repeated measures of ambient pollutant exposures, DNA methylation, mitochondrial genome measures, and potential confounders. In fact, this is the first study to use mitochondrial genetic variants and genome abundance to study the relationship of ambient particles with DNAm-age. Still, our study has a few notable limitations. First, we utilized address-specific $PM_{2.5}$ and $PM_{2.5}$ component exposure estimates which could potentially misclassify personal exposure levels. However, the majority of NAS participants are retired and very likely spend most of their time at home. Moreover, any resulting non-differential misclassification is likely to attenuate statistical associations rather than bias them away from the null[68-69]. Secondly, the mitochondrial genotyping technique resulted in haplogroup designations that may not encompass more recent mutations that could potentially impact the relationship of $PM_{2.5}$ with DNAm-age. Nonetheless, the haplogroups that were used have been utilized in many studies and our main objective was to identify common, normal forms of mitochondrial variation that may impact the $PM_{2.5}$-DNAm-age relationship. Larger studies are warranted to evaluate the impact of rarer forms of

mitochondrial genome variation on the $PM_{2.5}$-DNAm-age relationship. Finally, our findings are based on a cohort of elderly Caucasian males that reside in a lightly-polluted environment. Additional studies involving other demographic groups and in different environments will be needed to confirm our findings more broadly.

5.6. Conclusion

Overall, our study supports the premise that mitochondrial physiology is important for DNAm-age relationships, particularly in the context of ambient fine particle air pollution. Our data specifically suggests that mitochondrial haplogroups and copy number appear to be two different – but not necessarily mutually exclusive – ways that the relationship of $PM_{2.5}$ with DNAm-age is impacted by mitochondrial physiology. Future research aimed at further understanding the relationships of mitochondrial physiology with shared $PM_{2.5}$ and aging-related health outcomes will be critical for addressing this important public and environmental health topic.

5.7. Contributions and Support

Data Availability: Data are from the Normative Aging Study, from which restricted data are available for researchers who meet the criteria. A subset of the methylation data is deposited at NCBI dbGaP (study accession number: phs000853.v1.p1.

Acknowledgements: This work was supported by grants from the National Institute of Environmental Health Sciences (NIEHS) (R01ES021733 and R01ES025225). Other support comes from NIEHS grants ES015172, ES014663, ES020010, P30ES009089, P30ES000002; Environmental Protection Agency (EPA) grants RD832416 and RD83587201; and National Heart, Lung, and Blood Institute (NHLBI) grant 2T32HL007118-41. The US Department of Veterans Affairs (VA) Normative Aging Study (NAS) is supported by the Cooperative Studies Program/ERIC, US Department of Veterans Affairs, and is a research component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). The views expressed in this paper are those of the authors and do not necessarily represent

the views of the US Department of Veterans Affairs. Additional support was provided by the US Department of Agriculture, Agricultural Research Service (contract 53-K06-510). AC was supported by The Scientific and Technological Research Council of Turkey (TUBITAK-2219). All graphics were created by the authors.

5.8. References

1. Wallwork, R. S.; Colicino, E.; Zhong, J.; Kloog, I.; Coull, B. A.; Vokonas, P.; Schwartz, J. D.; Baccarelli, A. A., Ambient Fine Particulate Matter, Outdoor Temperature, and Risk of Metabolic Syndrome. *Am J Epidemiol* **2017,** *185* (1), 30-39.

2. Wang, Y.; Shi, L.; Lee, M.; Liu, P.; Di, Q.; Zanobetti, A.; Schwartz, J. D., Long-term Exposure to PM2.5 and Mortality Among Older Adults in the Southeastern US. *Epidemiology* **2017,** *28* (2), 207-214.

3. Mordukhovich, I.; Coull, B.; Kloog, I.; Koutrakis, P.; Vokonas, P.; Schwartz, J., Exposure to sub-chronic and long-term particulate air pollution and heart rate variability in an elderly cohort: the Normative Aging Study. *Environ Health* **2015,** *14*, 87.

4. Schikowski, T.; Vossoughi, M.; Vierkötter, A.; Schulte, T.; Teichert, T.; Sugiri, D.; Fehsel, K.; Tzivian, L.; Bae, I.-s.; Ranft, U.; Hoffmann, B.; Probst-Hensch, N.; Herder, C.; Krämer, U.; Luckhaus, C., Association of air pollution with cognitive functions and its modification by APOE gene variants in elderly women. *Environ. Res.* **2015,** *142*, 10-16.

5. Rice, M. B.; Ljungman, P. L.; Wilker, E. H.; Dorans, K. S.; Gold, D. R.; Schwartz, J.; Koutrakis, P.; Washko, G. R.; O'Connor, G. T.; Mittleman, M. A., Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the Framingham heart study. *Am J Respir Crit Care Med* **2015,** *191* (6), 656-64.

6. Pieters, N.; Janssen, B. G.; Dewitte, H.; Cox, B.; Cuypers, A.; Lefebvre, W.; Smeets, K.; Vanpoucke, C.; Plusquin, M.; Nawrot, T. S., Biomolecular Markers within the Core Axis of Aging and Particulate Air Pollution Exposure in the Elderly: A Cross-Sectional Study. *Environ Health Perspect* **2016,** *124* (7), 943-50.

7. Green, R.; Broadwin, R.; Malig, B.; Basu, R.; Gold, E. B.; Qi, L.; Sternfeld, B.; Bromberger, J. T.; Greendale, G. A.; Kravitz, H. M.; Tomey, K.; Matthews, K.; Derby, C. A.; Jackson, E. A.; Green, R.; Ostro, B., Long- and Short-term Exposure to Air Pollution and Inflammatory/Hemostatic Markers in Midlife Women. *Epidemiology* **2016,** *27* (2), 211-20.

8. Nwanaji-Enwerem, J. C.; Colicino, E.; Trevisi, L.; Kloog, I.; Just, A. C.; Shen, J.; Brennan, K.; Dereix, A.; Hou, L.; Vokonas, P.; Schwartz, J.; Baccarelli, A. A., Long-term ambient particle exposures and blood DNA methylation age: findings from the VA normative aging study. *Environ Epigenet* **2016,** *2* (2).

9. Horvath, S., DNA methylation age of human tissues and cell types. *Genome Biol.* **2013,** *14* (10), R115.

10. Nwanaji-Enwerem, J. C.; Dai, L.; Colicino, E.; Oulhote, Y.; Di, Q.; Kloog, I.; Just, A. C.; Hou, L.; Vokonas, P.; Baccarelli, A.; Weisskopf, M. G.; Schwartz, J., Associations between Long-term Exposure to PM2.5 Component Species and Blood DNA Methylation Age in the Elderly: The VA Normative Aging Study. *Environ Int* **2017,** *102*, 57-65.

11. Levine, A. J.; Quach, A.; Moore, D. J.; Achim, C. L.; Soontornniyomkij, V.; Masliah, E.; Singer, E. J.; Gelman, B.; Nemanim, N.; Horvath, S., Accelerated epigenetic aging in brain is associated with pre-mortem HIV-associated neurocognitive disorders. *J. Neurovirol.* **2015**.

12. Marioni, R. E.; Shah, S.; McRae, A. F.; Ritchie, S. J.; Muniz-Terrera, G.; Harris, S. E.; Gibson, J.; Redmond, P.; Cox, S. R.; Pattie, A.; Corley, J.; Taylor, A.; Murphy, L.; Starr, J. M.; Horvath, S.; Visscher, P. M.; Wray, N. R.; Deary, I. J., The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol* **2015,** *44* (4), 1388-1396.

13. Chen, B. H.; Marioni, R. E.; Colicino, E.; Peters, M. J.; Ward-Caviness, C. K.; Tsai, P. C.; Roetker, N. S.; Just, A. C.; Demerath, E. W.; Guan, W.; Bressler, J.; Fornage, M.; Studenski, S.; Vandiver, A. R.; Moore, A. Z.; Tanaka, T.; Kiel, D. P.; Liang, L.; Vokonas, P.; Schwartz, J.; Lunetta, K. L.; Murabito, J. M.; Bandinelli, S.; Hernandez, D. G.; Melzer, D.; Nalls, M.; Pilling, L. C.; Price, T. R.; Singleton, A. B.; Gieger, C.; Holle, R.; Kretschmer, A.; Kronenberg, F.; Kunze, S.; Linseisen, J.; Meisinger, C.; Rathmann, W.; Waldenberger, M.; Visscher, P. M.; Shah, S.; Wray, N. R.; McRae, A. F.; Franco, O. H.; Hofman, A.; Uitterlinden, A. G.; Absher, D.; Assimes, T.; Levine, M. E.; Lu, A. T.; Tsao, P. S.; Hou, L.; Manson, J. E.; Carty, C. L.; LaCroix, A. Z.; Reiner, A. P.; Spector, T. D.; Feinberg, A. P.; Levy, D.; Baccarelli, A.; van Meurs, J.; Bell, J. T.; Peters, A.; Deary, I. J.; Pankow, J. S.; Ferrucci, L.; Horvath, S., DNA methylation-

based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* **2016,** *8* (9), 1844-1865.

14. Perna, L.; Zhang, Y.; Mons, U.; Holleczek, B.; Saum, K. U.; Brenner, H., Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics* **2016,** *8*, 64.

15. Carroll, J. E.; Irwin, M. R.; Levine, M.; Seeman, T. E.; Absher, D.; Assimes, T.; Horvath, S., Epigenetic Aging and Immune Senescence in Women With Insomnia Symptoms: Findings From the Women's Health Initiative Study. *Biol Psychiatry* **2016**.

16. Balaban, R. S.; Nemoto, S.; Finkel, T., Mitochondria, oxidants, and aging. *Cell* **2005,** *120* (4), 483-95.

17. Finkel, T.; Holbrook, N. J., Oxidants, oxidative stress and the biology of ageing. *Nature* **2000,** *408* (6809), 239-47.

18. Wallace, D. C., A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet* **2005,** *39*, 359-407.

19. Pello, R.; Martin, M. A.; Carelli, V.; Nijtmans, L. G.; Achilli, A.; Pala, M.; Torroni, A.; Gomez-Duran, A.; Ruiz-Pesini, E.; Martinuzzi, A.; Smeitink, J. A.; Arenas, J.; Ugalde, C., Mitochondrial DNA background modulates the assembly kinetics of OXPHOS complexes in a cellular model of mitochondrial disease. *Hum Mol Genet* **2008,** *17* (24), 4001-11.

20. Shaughnessy, D. T.; McAllister, K.; Worth, L.; Haugen, A. C.; Meyer, J. N.; Domann, F. E.; Van Houten, B.; Mostoslavsky, R.; Bultman, S. J.; Baccarelli, A. A.; Begley, T. J.; Sobol, R. W.; Hirschey, M. D.; Ideker, T.; Santos, J. H.; Copeland, W. C.; Tice, R. R.; Balshaw, D. M.; Tyson, F. L., Mitochondria, energetics, epigenetics, and cellular responses to stress. *Environ Health Perspect* **2014,** *122* (12), 1271-8.

21. Birch-Machin, M. A.; Bowman, A., Oxidative stress and ageing. *Br J Dermatol* **2016,** *175 Suppl 2*, 26-29.

22. Lu, M.; Gong, X., Upstream reactive oxidative species (ROS) signals in exogenous oxidative stress-induced mitochondrial dysfunction. *Cell Biol Int* **2009,** *33* (6), 658-64.

23. Deavall, D. G.; Martin, E. A.; Horner, J. M.; Roberts, R., Drug-induced oxidative stress and toxicity. *J Toxicol* **2012,** *2012*, 645460.

24. Vermeij, W. P.; Hoeijmakers, J. H.; Pothof, J., Genome Integrity in Aging: Human Syndromes, Mouse Models, and Therapeutic Options. *Annu Rev Pharmacol Toxicol* **2016,** *56*, 427-45.

25. Payne, B. A.; Chinnery, P. F., Mitochondrial dysfunction in aging: Much progress but many unresolved questions. *Biochim Biophys Acta* **2015,** *1847* (11), 1347-53.

26. Feng, S.; Gao, D.; Liao, F.; Zhou, F.; Wang, X., The health effects of ambient PM2.5 and potential mechanisms. *Ecotoxicol Environ Saf* **2016,** *128*, 67-74.

27. Li, N.; Sioutas, C.; Cho, A.; Schmitz, D.; Misra, C.; Sempf, J.; Wang, M.; Oberley, T.; Froines, J.; Nel, A., Ultrafine particulate pollutants induce oxidative stress and mitochondrial damage. *Environ Health Perspect* **2003,** *111* (4), 455-60.

28. Wei, Y.; Han, I. K.; Shao, M.; Hu, M.; Zhang, O. J.; Tang, X., PM2.5 constituents and oxidative DNA damage in humans. *Environ Sci Technol* **2009,** *43* (13), 4757-62.

29. Torroni, A.; Achilli, A.; Macaulay, V.; Richards, M.; Bandelt, H. J., Harvesting the fruit of the human mtDNA tree. *Trends Genet* **2006,** *22* (6), 339-45.

30. Baudouin, S. V.; Saunders, D.; Tiangyou, W.; Elson, J. L.; Poynter, J.; Pyle, A.; Keers, S.; Turnbull, D. M.; Howell, N.; Chinnery, P. F., Mitochondrial DNA and survival after sepsis: a prospective study. *Lancet* **2005,** *366* (9503), 2118-21.

31. Zhong, J.; Cayir, A.; Trevisi, L.; Sanchez-Guerra, M.; Lin, X.; Peng, C.; Bind, M. A.; Prada, D.; Laue, H.; Brennan, K. J.; Dereix, A.; Sparrow, D.; Vokonas, P.; Schwartz, J.; Baccarelli, A. A., Traffic-Related Air Pollution, Blood Pressure, and Adaptive Response of Mitochondrial Abundance. *Circulation* **2016,** *133* (4), 378-87.

32. Hou, L.; Zhang, X.; Dioni, L.; Barretta, F.; Dou, C.; Zheng, Y.; Hoxha, M.; Bertazzi, P. A.; Schwartz, J.; Wu, S.; Wang, S.; Baccarelli, A. A., Inhalable particulate matter and mitochondrial DNA copy number in highly exposed individuals in Beijing, China: a repeated-measure study. *Part Fibre Toxicol* **2013,** *10*, 17.

33. Bell, B.; Rose, C. L.; Damon, A., The Veterans Administration Longitudinal Study of Healthy Aging. *The Gerontologist* **1966,** *6* (4), 179-184.

34. Aryee, M. J.; Jaffe, A. E.; Corrada-Bravo, H.; Ladd-Acosta, C.; Feinberg, A. P.; Hansen, K. D.; Irizarry, R. A., Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **2014,** *30* (10), 1363-1369.

35. Bey, I.; Jacob, D. J.; Yantosca, R. M.; Logan, J. A.; Field, B. D.; Fiore, A. M.; Li, Q.; Liu, H. Y.; Mickley, L. J.; Schultz, M. G., Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research: Atmospheres* **2001,** *106* (D19), 23073-23095.

36. Di, Q.; Koutrakis, P.; Schwartz, J., A hybrid prediction model for PM 2.5 mass and components using a chemical transport model and land use regression. *Atmospheric Environment* **2016,** *131*, 390-399.

37. Colicino, E.; Power, M. C.; Cox, D. G.; Weisskopf, M. G.; Hou, L.; Alexeeff, S. E.; Sanchez-Guerra, M.; Vokonas, P.; Spiro, A.; Schwartz, J.; Baccarelli, A. A., Mitochondrial haplogroups modify the effect of black carbon on age-related cognitive impairment. *Environ Health* **2014,** *13* (1), 42.

38. Kishida, K.; Tominaga, M.; Matsubara, K.; Taguchi, M.; Noguchi, M.; Tsunawake, N.; Shidoji, Y., An association analysis between mitochondrial DNA A10398G polymorphism and temperament in Japanese young adults. *PLoS One* **2009,** *4* (11), e7763.

39. Finnila, S.; Lehtonen, M. S.; Majamaa, K., Phylogenetic network for European mtDNA. *Am J Hum Genet* **2001,** *68* (6), 1475-84.

40. Kruger, J.; Hinttala, R.; Majamaa, K.; Remes, A. M., Mitochondrial DNA haplogroups in early-onset Alzheimer's disease and frontotemporal lobar degeneration. *Mol Neurodegener* **2010,** *5*, 8.

41. Yakes, F. M.; Van Houten, B., Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. *Proc Natl Acad Sci U S A* **1997,** *94* (2), 514-9.

42. Dai, L.; Mehta, A.; Mordukhovich, I.; Just, A. C.; Shen, J.; Hou, L.; Koutrakis, P.; Sparrow, D.; Vokonas, P. S.; Baccarelli, A. A.; Schwartz, J. D., Differential DNA methylation and PM2.5 species in a 450 K epigenome-wide association study. *Epigenetics* **2016**, 1-10.

43. Panni, T.; Mehta, A. J.; Schwartz, J. D.; Baccarelli, A. A.; Just, A. C.; Wolf, K.; Wahl, S.; Cyrys, J.; Kunze, S.; Strauch, K.; Waldenberger, M.; Peters, 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* **2016,** *124* (7), 983-90.

44. Houseman, E. A.; Accomando, W. P.; Koestler, D. C.; Christensen, B. C.; Marsit, C. J.; Nelson, H. H.; Wiencke, J. K.; Kelsey, K. T., DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **2012,** *13*, 86.

45. Rosa, M. J.; Just, A. C.; Guerra, M. S.; Kloog, I.; Hsu, H. L.; Brennan, K. J.; Garcia, A. M.; Coull, B.; Wright, R. J.; Tellez Rojo, M. M.; Baccarelli, A. A.; Wright, R. O., Identifying sensitive windows for prenatal particulate air pollution exposure and mitochondrial DNA content in cord blood. *Environ Int* **2017,** *98*, 198-203.

46. Smiraglia, D. J.; Kulawiec, M.; Bistulfi, G. L.; Gupta, S. G.; Singh, K. K., A novel role for mitochondria in regulating epigenetic modification in the nucleus. *Cancer Biol Ther* **2008,** *7* (8), 1182-90.

47. Mackinnon, D. P., Integrating Mediators and Moderators in Research Design. *Res Soc Work Pract* **2011,** *21* (6), 675-681.

48. VanderWeele, T. J., A unification of mediation and interaction: a 4-way decomposition. *Epidemiology* **2014,** *25* (5), 749-61.

49. Sobel, M. E., Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological methodology* **1982,** *13*, 290-312.

50. Mackinnon, D. P.; Warsi, G.; Dwyer, J. H., A Simulation Study of Mediated Effect Measures. *Multivariate Behav Res* **1995,** *30* (1), 41.

51. Peng, C.; Bind, M. C.; Colicino, E.; Kloog, I.; Byun, H. M.; Cantone, L.; Trevisi, L.; Zhong, J.; Brennan, K.; Dereix, A. E.; Vokonas, P. S.; Coull, B. A.; Schwartz, J. D.; Baccarelli, A. A., Particulate Air

Pollution and Fasting Blood Glucose in Nondiabetic Individuals: Associations and Epigenetic Mediation in the Normative Aging Study, 2000-2011. *Environ Health Perspect* **2016,** *124* (11), 1715-1721.

52. Ward-Caviness, C. K.; Nwanaji-Enwerem, J. C.; Wolf, K.; Wahl, S.; Colicino, E.; Trevisi, L.; Kloog, I.; Just, A. C.; Vokonas, P.; Cyrys, J.; Gieger, C.; Schwartz, J.; Baccarelli, A. A.; Schneider, A.; Peters, A., Long-term exposure to air pollution is associated with biological aging. *Oncotarget* **2016**.

53. Byun, H. M.; Colicino, E.; Trevisi, L.; Fan, T.; Christiani, D. C.; Baccarelli, A. A., Effects of Air Pollution and Blood Mitochondrial DNA Methylation on Markers of Heart Rate Variability. *J Am Heart Assoc* **2016,** *5* (4).

54. Wittkopp, S.; Staimer, N.; Tjoa, T.; Gillen, D.; Daher, N.; Shafer, M.; Schauer, J. J.; Sioutas, C.; Delfino, R. J., Mitochondrial genetic background modifies the relationship between traffic-related air pollution exposure and systemic biomarkers of inflammation. *PLoS One* **2013,** *8* (5), e64444.

55. Brand, M. D., Uncoupling to survive? The role of mitochondrial inefficiency in ageing. *Exp Gerontol* **2000,** *35* (6-7), 811-20.

56. Khusnutdinova, E.; Gilyazova, I.; Ruiz-Pesini, E.; Derbeneva, O.; Khusainova, R.; Khidiyatova, I.; Magzhanov, R.; Wallace, D. C., A mitochondrial etiology of neurodegenerative diseases: evidence from Parkinson's disease. *Ann N Y Acad Sci* **2008,** *1147*, 1-20.

57. van der Walt, J. M.; Nicodemus, K. K.; Martin, E. R.; Scott, W. K.; Nance, M. A.; Watts, R. L.; Hubble, J. P.; Haines, J. L.; Koller, W. C.; Lyons, K.; Pahwa, R.; Stern, M. B.; Colcher, A.; Hiner, B. C.; Jankovic, J.; Ondo, W. G.; Allen, F. H., Jr.; Goetz, C. G.; Small, G. W.; Mastaglia, F.; Stajich, J. M.; McLaurin, A. C.; Middleton, L. T.; Scott, B. L.; Schmechel, D. E.; Pericak-Vance, M. A.; Vance, J. M., Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. *Am J Hum Genet* **2003,** *72* (4), 804-11.

58. Nogales-Gadea, G.; Pinos, T.; Ruiz, J. R.; Marzo, P. F.; Fiuza-Luces, C.; Lopez-Gallardo, E.; Ruiz-Pesini, E.; Martin, M. A.; Arenas, J.; Moran, M.; Andreu, A. L.; Lucia, A., Are mitochondrial haplogroups associated with elite athletic status? A study on a Spanish cohort. *Mitochondrion* **2011,** *11* (6), 905-8.

59. Dawson, V. L.; Dawson, T. M., Deadly conversations: nuclear-mitochondrial cross-talk. *J Bioenerg Biomembr* **2004,** *36* (4), 287-94.

60. Kelly, R. D.; Mahmud, A.; McKenzie, M.; Trounce, I. A.; St John, J. C., Mitochondrial DNA copy number is regulated in a tissue specific manner by DNA methylation of the nuclear-encoded DNA polymerase gamma A. *Nucleic Acids Res* **2012,** *40* (20), 10124-38.

61. Horvath, S.; Erhart, W.; Brosch, M.; Ammerpohl, O.; von Schonfels, W.; Ahrens, M.; Heits, N.; Bell, J. T.; Tsai, P. C.; Spector, T. D.; Deloukas, P.; Siebert, R.; Sipos, B.; Becker, T.; Rocken, C.; Schafmayer, C.; Hampe, J., Obesity accelerates epigenetic aging of human liver. *Proceedings of the National Academy of Sciences* **2014,** *111* (43), 15538-15543.

62. Levine, M. E.; Hosgood, H. D.; Chen, B.; Absher, D.; Assimes, T.; Horvath, S., DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY)* **2015,** *7* (9), 690-700.

63. Lowe, D.; Horvath, S.; Raj, K., Epigenetic clock analyses of cellular senescence and ageing. *Oncotarget* **2016,** *7* (8), 8524-31.

64. Horvath, S.; Mah, V.; Lu, A. T.; Woo, J. S.; Choi, O.-W.; Jasinska, A. J.; Riancho, J. A.; Tung, S.; Coles, N. S.; Braun, J.; Vinters, H. V.; Coles, L. S., The cerebellum ages slowly according to the epigenetic clock. *Aging (Albany NY)* **2015,** *7* (5), 294-306.

65. Minocherhomji, S.; Tollefsbol, T. O.; Singh, K. K., Mitochondrial regulation of epigenetics and its role in human diseases. *Epigenetics* **2012,** *7* (4), 326-34.

66. Ashar, F. N.; Moes, A.; Moore, A. Z.; Grove, M. L.; Chaves, P. H.; Coresh, J.; Newman, A. B.; Matteini, A. M.; Bandeen-Roche, K.; Boerwinkle, E.; Walston, J. D.; Arking, D. E., Association of mitochondrial DNA levels with frailty and all-cause mortality. *J Mol Med (Berl)* **2015,** *93* (2), 177-86.

67. He, Y. H.; Chen, X. Q.; Yan, D. J.; Xiao, F. H.; Lin, R.; Liao, X. P.; Liu, Y. W.; Pu, S. Y.; Yu, Q.; Sun, H. P.; Jiang, J. J.; Cai, W. W.; Kong, Q. P., Familial longevity study reveals a significant association of mitochondrial DNA copy number between centenarians and their offspring. *Neurobiol Aging* **2016,** *47*, 218 e11-218 e18.

68. Kioumourtzoglou, M. A.; Spiegelman, D.; Szpiro, A. A.; Sheppard, L.; Kaufman, J. D.; Yanosky, J. D.; Williams, R.; Laden, F.; Hong, B.; Suh, H., Exposure measurement error in PM2.5 health effects studies: a pooled analysis of eight personal exposure validation studies. *Environ Health* **2014,** *13* (1), 2.

69. Weisskopf, M. G.; Webster, T. F., Trade-offs of personal vs. more proxy exposure measures in environmental epidemiology. *Epidemiology* **2017**.

Chapter 6:

MicroRNA Processing Gene Polymorphisms, Blood DNA Methylation Age, and Long-term Ambient $PM_{2.5}$ Exposure in Elderly Men

Jamaji C. Nwanaji-Enwerem[a†]; Elena Colicino[b]; Lingzhen Dai[a]; Qian Di[a]; Allan C. Just[c]; Lifang Hou[d]; Pantel Vokonas[e]; Immaculata De Vivo[f]; Bernardo Lemos[a]; Quan Lu[a]; Marc G. Weisskopf[a]; Andrea A. Baccarelli[b]; and Joel D. Schwartz[a]

[a]*Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

[b]*Department of Environmental Health Sciences, Columbia Mailman School of Public Health, New York, NY, USA*

[c]*Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

[d] *Center for Population Epigenetics, Department of Preventive Medicine, Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, IL, USA*

[e]*VA Normative Aging Study, Veterans Affairs Boston Healthcare System and the Department of Medicine, Boston University School of Medicine, Boston, MA, USA*

[f]*Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

This is a pre-copyedited, author-produced version of an article accepted for publication in *Epigenomics* following peer review. This version was reproduced with permission from [Nwanaji-Enwerem, Jamaji C., et al. "miRNA processing gene polymorphisms, blood DNA methylation age and long-term ambient $PM_{2.5}$ exposure in elderly men." *Epigenomics* 9.12 (2017): 1529-1542.] The version of record is available online at: https://www.futuremedicine.com/doi/abs/10.2217/epi-2017-0094. DOI: 10.2217/epi-2017-0094.

6.1. Abstract

Background: The association of long-term $PM_{2.5}$ exposure, an aging risk factor, with DNA methylation age (DNAm-age), an epigenetic biomarker of aging, may involve mediators of gene regulation.

Objective: Since microRNAs are heavily involved in gene regulation, we investigated the modifying role of genetic variation in microRNA-processing genes on the $PM_{2.5}$-DNAm-age relationship.

Methods: We conducted a repeated measures study based on 552 participants from the Normative Aging Study with multiple visits between 2000 and 2011 (n=940 visits). Address-level one-year $PM_{2.5}$ exposures were estimated using the GEOS-chem model. DNAm-age and a panel of 14 SNPs in microRNA-processing genes were measured from participant blood samples.

Results: From an elastic net, four SNPs were identified as important to DNAm-age. In fully-adjusted linear mixed-effects models, having at least one copy of the minor rs4961280 [*AGO2*] allele was associated with a lower DNA methylation age (β=-1.13; 95%CI: -2.26, -0.002). However, only the rs4961280 [*AGO2*] SNP modified the $PM_{2.5}$-DNAm-age relationship. The association of $PM_{2.5}$ with DNAm-age was significantly ($P_{interaction}$=0.01) weaker in homozygous carriers of the major *AGO2* allele (β=0.38; 95%CI: -0.20, 0.96) when compared to all other participants (β=1.58; 95%CI: 0.76, 2.39). Gene network analyses revealed known physical, genetic, and co-expression relationships of *AGO2* with genes that contribute methylation values to the DNAm-age measure including *IPO8* and *TIPARP*.

Conclusion: Our results suggest that microRNA-processing impacts DNAm-age relationships particularly in the context of long-term $PM_{2.5}$ exposure.

6.2. Introduction

Between 2015 and 2050, the global percentage of individuals over the age of 60 is expected to almost double from 12% to 22%[1]. This shift in the age composition of the global population is particularly important because it will likely be accompanied by immense public health and economic burdens due to an unprecedented volume of aging-related diseases. For instance, in the United States alone, the 2017 disease prevalence and economic costs associated with Alzheimer's dementia were 5.5 million and $259 billion. By 2050, those statistics are expected to be 16 million and $1.1 trillion respectively[2]. In an effort to curb these expanding disease and economic burdens, there has been an ever-growing emphasis on research aimed at understanding biological aging and the factors that contribute to adverse aging-related health outcomes.

Ambient fine particle air pollution ($PM_{2.5}$) – often considered the world's largest singular environmental health risk – is one potentially modifiable risk factor for aging-related diseases including cardiovascular disease, cognitive decline, and cancer[3-6]. Of particular interest, is the association of long-term $PM_{2.5}$ exposure levels with DNA methylation (DNAm) age, a novel epigenome-wide DNA methylation-based measure of biological aging[7,8]. Like other biomarkers of biological aging, DNAm-age has been associated with all-cause mortality and aging-related diseases[9,10]. In contrast to other biomarkers of aging, researchers remain highly uncertain about what DNAm-age is capturing on a molecular physiological level[11]. By examining the relationship of DNAm-age with $PM_{2.5}$, a widely studied exogenous exposure and aging risk factor, we can begin to understand more about DNAm-age physiology.

As previously mentioned, DNAm-age is derived from measurements of DNA methylation. DNA methylation is a biological process where methyl groups are added to DNA nucleotides and often result in changes in gene expression[12]. Micro RNAs (miRNAs) are small non-coding RNA molecules that can also regulate gene expression and have been associated with $PM_{2.5}$, aging/aging-related diseases, and DNA methylation[13-15]. miRNAs are produced from nuclear transcripts that form hairpin structures. Following nuclear and cytoplasmic processing by a series of enzymes, miRNAs are incorporated into a structure called the RNA-induced silencing complex (RISC). RISC achieves post-transcriptional gene regulation by using one strand of the incorporated miRNA to target messenger RNAs (mRNAs) via nucleotide complementary

base pairing. Once the relevant mRNA is targeted, RISC inhibits subsequent protein production by inducing mRNA cleavage or by reducing translation of the mRNA molecule[16].

Despite existing knowledge of the role of miRNAs in $PM_{2.5}$ and aging biology, studies have not yet examined the role of miRNAs in the $PM_{2.5}$-DNAm-age relationship. Given the shared role of miRNAs and DNA methylation in gene regulation[17], we hypothesized that miRNA physiology would be related to DNAm-age and may play a role in the relationship of $PM_{2.5}$ with DNAm-age. In the present study, we investigated if single nucleotide polymorphisms (SNPs) in miRNA processing genes modified the associations of long-term $PM_{2.5}$ and $PM_{2.5}$ component species (sulfate and ammonium) exposure with DNAm-age in participants of the elderly Normative Aging Study (NAS). Rather than simply testing all the SNPs in our panel, we utilized a methodical framework to identify and analyze significant SNPs and $PM_{2.5}$–SNP interactions. We first employed an elastic net (penalized regression) selection model to identify SNPs in miRNA processing genes that were specifically important to DNAm-age. Subsequently, we used fully-adjusted linear mixed effects models to test for statistically significant direct associations of the elastic net selected SNPs and $PM_{2.5}$-SNP interactions with DNAm-age. We also conducted a number of secondary analyses to better ascertain if particular $PM_{2.5}$ component species were responsible for the relationships we observed.

6.3. Methods

6.3.a. Study Population

The U.S. Department of Veterans Affairs (VA) Normative Aging Study (NAS) is a longitudinal study of aging that was established in 1963 and recruited male participants from the Greater Boston area that were free of any chronic disease[18]. The NAS is now a closed cohort, but every 3-5 years since recruitment, participants return for onsite, follow-up study visits. During these recurring visits, participants undergo thorough physical examinations, report lifestyle practices via questionnaires, and provide bio-specimens including blood. At recruitment, all participants provided written informed consent to the VA

Institutional Review Board (IRB) and were at least 18 years of age. The VA and Harvard T.H. Chan School of Public Health IRBs granted human subjects approval (protocol 14027-102).

Our study sample is derived from all NAS men with continued study participation since the year 2000, when address-level $PM_{2.5}$ component species estimates became available. We started with a total of 552 participants with 940 study visits (observations) between the years 2000 and 2011[8]. Of these 552 participants, 249 (45%) had one study visit, 218 (40%) had two study visits, and 85 (15%) had three or more study visits. From this sample, we then excluded participants missing miRNA processing gene polymorphism data. This resulted in a final study sample of 471 participants with 808 total study visits. In the final study sample, 208 participants (44%) had one visit, 189 (40%) had two visits, and 74 (16%) had three or more visits.

6.3.b. Measuring DNA Methylation and Computing DNA Methylation (DNAm) Age

We extracted DNA from whole blood provided by participants during NAS visits. After performing bisulfite conversion on the DNA (EZ-96 DNA Methylation Kit, Zymo Research, Orange, CA, USA), we performed methylation analysis using the Illumina HumanMethylation450 BeadChip platform (Infinium HD Methylation protocol, Illumina, San Diego, CA, USA). To ensure a similar age distribution across chips/plates and minimize batch effects, we used a two-stage age-stratified algorithm to randomize samples and randomized chips across plates. For quality control purposes, we removed samples where >5% of probes had a beadcount < 3 or > 1% of probes had a detection P-value >0.05. The remaining samples were pre-processed with Illumina-type background correction without normalization and normalized with dye-bias and BMIQ3 adjustments. Next, we generated methylation beta values, which represent the percentage of methylation for each of the ~480,000 CpG sites in the BeadChip array. Beta = intensity of the methylated signal (M) / [intensity of the unmethylated signal (U) + intensity of the methylated signal (M) + 100].

DNAm age was computed using the publically available online calculator (http://labs.genetics.ucla.edu/horvath/dnamage/). DNAm-age was derived from a penalized regression (an elastic net) run on numerous datasets of diverse cell and tissue types where CpG probes shared by both the

Illumina HumanMethylation27 and HumanMethylation450 BeadChip platforms were regressed on a calibrated version of chronological age. 353 CpGs that correlated with age (193 positively and 160 negatively) were selected by the elastic net[11]. The model coefficients from these 353 CpGs were used by the calculator to predict the age of each DNA sample (*i.e.* DNAm-age). The calculator maintains predictive accuracy (age correlation 0.97, error = 3.6 years) across almost all body tissues including blood, bone, and brain[11].

6.3.c. One-Year Address-Level Ambient Fine Particulate Matter ($PM_{2.5}$) Exposure Estimation

We focused on the one-year $PM_{2.5}$ exposure window because existing literature demonstrates that it is robustly associated with DNAm-age[7]. Furthermore, greater than 90% of NAS participants are retired; thus, home address exposures are expected to be a good proxy for their individual ambient exposures. Using the GEOS-chem chemical transport model (http://www.geoschem.org), we generated daily estimates at the 1 km x1 km area resolution for total $PM_{2.5}$. The GEOS-chem model is particularly useful because it allows us to predict $PM_{2.5}$ component species like ammonium and sulfate at the same 1 km x 1 km area resolution. Sulfate and ammonium are the major $PM_{2.5}$ component species that have been previously shown to be important in predicting DNAm-age[8]. After geocoding and linking participants' residencies to an area level grid-point, we assigned particle estimates to each participant's address. One-year total $PM_{2.5}$ and $PM_{2.5}$ component species exposure estimates were determined by averaging daily exposures for the 365 days prior to the day of each participants' NAS visit. Ten-fold cross-validation demonstrated that the model performed well for $PM_{2.5}$ mass and its component species with R^2s ranging from 0.70 to 0.88[19].

6.3.d. Genotyping Micro RNA Processing Gene Polymorphisms

The panel of 24 microRNA (miRNA) processing gene single nucleotide polymorphisms (SNPs) examined in this study were selected from previous studies that investigated the association of miRNA processing gene SNPs and chronic aging-related diseases[20,21]. Some of these same SNPs have been shown to modify relationships of ambient air pollutants with aging-related disease[22,23]. We performed genotyping

on DNA extracted from participants' blood. Multiplex PCR assays were designed with Sequenom SpectroDESIGNER software (Sequenom, Inc., San Diego, CA). The extension product was subsequently spotted onto a 384-well spectroCHIP and analyzed in the MALDI-TOF mass spectrometer (Sequenom, Inc.). We duplicated the assay for 5% of the samples. Of all the 24 SNPs analyzed for this study, all were successfully detected.

Following genotyping, we excluded ten SNPs for which the number of participants who were homozygous minor variant carriers was less than 10 [rs595961 and rs636832 in *AGO1*; rs197388 and rs197414 in *DDX20*; rs417309 in *DGCR8*; rs3742330 in *DICER1*; rs2740348 and rs3744741 in *GEMIN4*; rs1106042 in *PIWIL1*] and one in which Hardy–Weinberg equilibrium was not met at the 0.05 level [rs10719 in *DROSHA*]. This exclusion criteria has been utilized in already published studies that use this panel of SNPs[22,23]. The remaining 14 SNPs were used in the study analyses. Linkage disequilibrium (LD) of SNPs within the same gene was previously assessed using the LDPlotter tool (https://www.pharmgat.org/Tools/pbtoldplotform)[23].

6.3.e. Statistical Analysis

6.3.e.1. Elastic Net Selection of miRNA Processing Gene SNPs:

The aim of the present study was to examine if SNPs in miRNA processing genes modified the association of long-term $PM_{2.5}$ and $PM_{2.5}$ component species levels with DNAm-age. In an effort to A) limit multiple comparisons and B) identify specific miRNA processing gene SNPs that are important to DNAm-age, we first employed an elastic net (penalized regression) via the glmnet function in the R glmnet package. Our elastic net method was similar to that described by Lenters and colleagues[24] and the full documentation for running all aspects of the elastic net via glmnet is publically available (https://cran.r-project.org/web/packages/glmnet/index.html. We have also used a comparable elastic net strategy in a previous publication[25]. In short, the elastic net regression linear models utilized a hybrid of ridge and LASSO penalty functions to determine which SNPs were important to DNAm-age. By combining both of these penalty functions, the elastic net is able to perform selection while allowing for the inclusion of highly-

related genetic variants[26,27]. In our case, the highly related variants were the panel of miRNA processing gene SNPs. Specifically, in our elastic net selection model, all 14 SNPs and their respective $PM_{2.5}$-SNP interactions were regressed on DNAm-age. Chronological age (continuous), blood cell proportions [plasma cells, CD4+ lymphocytes, CD8+ lymphocytes, natural killer (NK) cells, monocytes, and granulocytes] (continuous, determined via Houseman and Horvath methods[11,28]), average 1-year temperature (continuous address-specific satellite measurements[8]), cumulative cigarette pack years (continuous), smoking status (current, former, or never), and season of visit (spring [March-May], Summer [June-August], Fall [September-November], and Winter [December-February]), were also included in the selection model as unpenalized variables. The existing air pollution, DNA methylation, and DNAm-age literature have identified these variables as important potential confounders[7,8,29,30]. Cross-validation was performed to determine the optimal degree of penalization and the minimum mean-squared error (MSE) of prediction from repeated 10-fold cross-validation was used in the final elastic net selection model. miRNA processing gene SNPs with non-zero model coefficients were considered as "selected" by the elastic net.

6.3.e.2. Covariates:

The direct relationships of the miRNA processing gene SNPs with DNAm-age and the role of these SNPs as modifiers of the association of $PM_{2.5}$ with DNAm-age were examined using fully-adjusted linear mixed-effects models. These models included a random participant-specific intercept to account for correlation between repeated outcome measures resulting from having multiple study visits for participants. In the analyses using fully-adjusted models, we controlled for all the variables used in the elastic net selection model as well as body mass index (BMI) (lean [<25], overweight [25-30], obese [>30]), alcohol intake (yes/no ≥ 2 drinks daily), maximum years of education (continuous), cancer (yes/no history of lifetime cancer diagnosis), ischemic heart disease (yes/no based on electrocardiogram, validated medical records, or physical exam), diabetes (physician diagnosis or a fasting blood glucose > 126 mg/dL), and hypertension (yes/no antihypertensive medication use or systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg).

6.3.e.3. Effect Modification by and Direct Associations of miRNA Processing Gene SNPs:

Using the same covariates from the direct effect fully-adjusted linear mixed effects models, we evaluated if the miRNA processing gene SNPs selected by the elastic net modified the association of $PM_{2.5}$ with DNAm-age. Given a statistically significant modifying effect, we evaluated if the SNP of interest modified the associations of sulfate and ammonium with DNAm-age. In these analyses, the reference group was participants who were homozygous for the major variant of the SNP. For instance, if we evaluated the modifying role of 'SNP A' on the association of $PM_{2.5}$ with DNAm-age, we compared participants genotyped as homozygous for the major variant of 'SNP A' against all other participants (*i.e.* participants genotyped heterozygous or homozygous for the minor variant of 'SNP A'). We conducted secondary analyses exploring trends in significant modifier effects across all three genotypes (homozygous major variant, heterozygous, and homozygous minor variant). When we observed that a miRNA processing gene SNP significantly modified the relationship of a $PM_{2.5}$ component species with DNAm-age, we also conducted an additional sensitivity analysis. In this sensitivity analysis, we subtracted that component species from total $PM_{2.5}$ and reevaluated effect modification by the SNP.

Also using fully-adjusted linear mixed effects models, we determined if elastic net selected miRNA processing gene SNPs had direct associations with DNAm-age when modeled as joint predictors with $PM_{2.5}$ levels. We also performed a sensitivity analysis, examining these direct associations, where we subtracted component species from total $PM_{2.5}$ as previously described above.

6.3.e.4. Network Analysis:

We used the publically available Genemania platform (https://genemania.org) to explore gene network relationships (co-expression, physical interactions, and genetic interactions) between the genes encompassing the elastic net selected SNPs and the 353 genes that contribute CpGs to the DNAm-age metric.

6.3.e.5. Analysis Software:

The gene network analysis was performed using the Genemania plugin for Cytoscape. All remaining statistical analyses were performed using R Version 3.1.1 (R Core Team, Vienna, Austria) and we considered a P-value < 0.05 to be statistically significant.

6.4. Results

6.4.a. Descriptive Statistics

Table 1 summarizes the demographic and clinical characteristics of study participants across all study visits. All participants were Caucasian males with a mean (SD) chronological age and DNAm-age of 75.0 (7.03) and 74.1 (8.02). A majority of the participants had completed at least 12 years of formal education (74%), were former smokers (67%), and were overweight or obese (77%). In this study sample, the prevalence of ischemic heart disease, diabetes, and hypertension were 35%, 18%, and 75% respectively. The lifetime prevalence of a cancer diagnosis was 57%.

The mean (IQR) one-year $PM_{2.5}$, sulfate, and ammonium levels were 10.3 (2.15) $\mu g/m^3$, 3.39 (0.81) $\mu g/m^3$, and 1.05 (0.29) $\mu g/m^3$. Table S1 presents the Pearson correlation coefficients and the proportion of total $PM_{2.5}$ mass of GEOS-chem transport model derived $PM_{2.5}$ component species across all study visits. Sulfate made up the greatest proportion of $PM_{2.5}$ mass (33.2%), and ammonium made up 10.2% of $PM_{2.5}$ mass. The correlation coefficients for sulfate and ammonium with total $PM_{2.5}$ mass were 0.30 ($p<0.0001$) and 0.51 ($p<0.0001$) respectively.

Table 1. Characteristics of Study Participants (2000 – 2011)

Main Variables	All Visits, N = 808
Age (years), mean (SD)	75.0 (7.03)
DNAm-age (years), mean (SD)	74.1 (8.02)
One- Year Fine Particle Level ($\mu g/m^3$), mean (IQR)	
$PM_{2.5}$	10.3 (2.15)
Sulfate	3.39 (0.81)
Ammonium	1.05 (0.29)
Lifestyle and Environmental Variables	
Alcohol Consumption, N (%)	
< 2 drinks/day	647 (80)
≥ 2 drinks/day	161 (20)
BMI, N (%)	
Healthy/Lean	189 (23)
Overweight	427 (53)
Obese	192 (24)
Education, N (%)	
≤ 12 years	206 (26)
12 – 16 years	379 (47)
> 16 years	223 (27)
Pack years, mean (SD)	21.2 (24.7)
Smoking Status, N (%)	
Current	36 (4)
Former	538 (67)
Never	234 (29)
Season, N (%)	
Spring	204 (25)
Summer	175 (22)
Fall	271 (33)
Winter	158 (20)
Temperature (°C), mean (SD)	11.3 (0.99)
Aging-Related Diseases	
Ischemic Heart Disease, N (%)	
Yes	287 (35)
No	521 (65)
Diabetes, N (%)	
Yes	148 (18)
No	660 (82)
Hypertension, N (%)	
Yes	604 (75)
No	204 (25)
Lifetime Cancer Diagnosis, N (%)	
Yes	458 (57)
No	350 (43)

6.4.b. Elastic Net Selected miRNA Processing Gene SNPs

Table S2 lists the 14 miRNA processing SNPs that were included in the elastic net selection model. Of these 14 SNPs, four (rs4961280 [*AGO2*], rs6877842 [*DROSHA*], rs910924 [*GEMIN4*], and rs784567 [*TARBP2*]) were selected by the elastic net with DNA methylation as the outcome.

Table 2. Mean One-Year Fine Particle ($PM_{2.5}$) Concentrations and MicroRNA Processing Gene Single Nucleotide Polymorphisms (SNPs) as Joint Predictors of DNAm-age (N = 808)

Predictor	Difference in DNAm-age for IQR (95% CI)	*P*
$PM_{2.5}$	0.76 (0.24, 1.24)	**0.003**
Elastic Net Selected miRNA SNPs[a]		
rs4961280 (*AGO2*)	-1.13 (-2.26, -0.002)	**0.05**
rs6877842 (*DROSHA*)	-0.78 (-1.92, 0.37)	0.18
rs910924 (*GEMIN4*)	-0.41 (-1.47, 0.65)	0.45
rs784567 (*TARBP2*)	-1.35 (-2.61, -0.09)	**0.04**

Note. Model adjusted for chronological age, blood cell type, temperature, pack years, smoking status, season, BMI, alcohol consumption, education, lifetime cancer diagnosis, hypertension, diabetes, and ischemic heart disease. [a]Values for the miRNA processing SNPs are in reference to participants whose genotypes are homozygous for the major variant. Bold text specifies statistically significant P values (<0.05).

Of the four elastic net selected SNPs, only rs4961280 (*AGO2*) and rs784567 (*TARBP2*) were significantly associated with DNAm-age in fully-adjusted linear mixed effects models that included $PM_{2.5}$ levels as a covariate (Table 2). For rs4961280 (*AGO2*), individuals who had at least one copy of the minor SNP allele on average had a 1.13-year lower DNAm-age than individuals with the homozygous major variant (allele) genotype ($p<0.05$). When we compared all three genotypes, on average, individuals who were homozygous for the minor rs4961280 (*AGO2*) variant (AA) had the lowest DNAm-age. Homozygous major carriers (CC) had the highest DNAm-age and heterozygous individuals (CA) had an intermediate DNAm-age (Figure S1). The trend for this relationship was statistically significant ($p=0.04$). For rs784567 (*TARBP2*), individuals who had at least one copy of the minor SNP allele on average had a 1.35-year lower

DNAm-age than individuals with the homozygous major allele genotype (p=0.04, Table 2). When we looked across all three rs784567 (*TARBP2*) genotypes, a trend similar to rs4961280 (*AGO2*) was observed, but the trend did not reach statistical significance (p=0.08, Figure S1). These relationships persisted in sensitivity analyses where the ammonium component was subtracted from total $PM_{2.5}$ mass (Table S3).

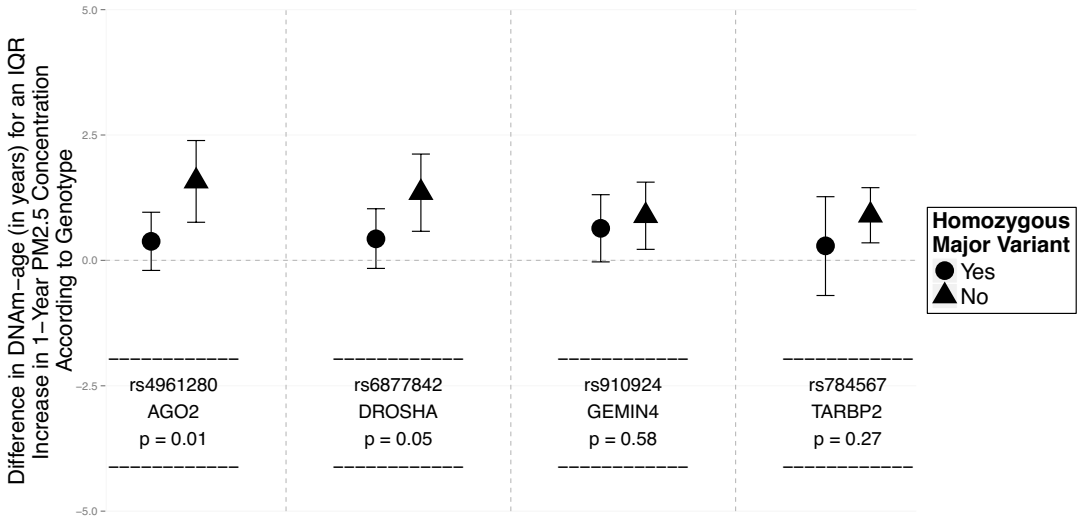

Figure 1 | Difference in DNAm-age for one interquartile range increase in one-year particle exposure levels comparing participants with and without a homozygous major variant genotype for *AGO2, DROSHA, GEMIN4, and TARBP2* in fully-adjusted linear mixed effects models.

Out of all four SNPs, only the rs4961280 (*AGO2*) SNP significantly modified the association of $PM_{2.5}$ with DNAm-age (p=0.01) – although the rs6877842 (*DROSHA*) SNP neared statistical significance (p=0.052, Figure 1). Specifically, the association of $PM_{2.5}$ with DNAm-age was greater in individuals who were not homozygous for the major *AGO2* variant (allele) when compared to individuals who were

homozygous for the major allele. This trend was also observed for the ammonium $PM_{2.5}$ component species. No significant effect modification by the *AGO2* SNP was observed for sulfate levels and the rs6877842 (*DROSHA*) SNP did not significantly modify the relationships of ammonium or sulfate with DNAm-age (Figure 2).

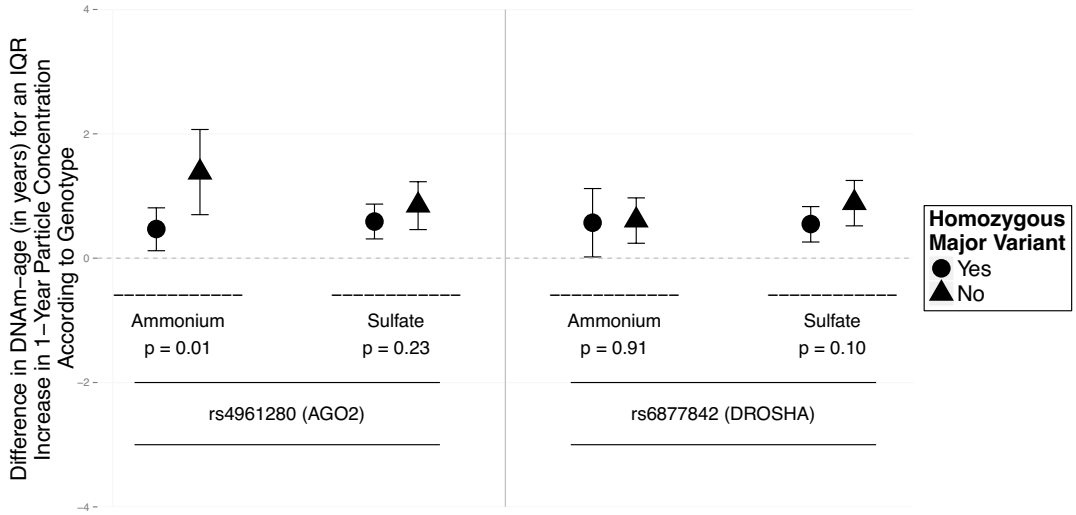

Figure 2 | Difference in DNAm-age for one interquartile range increase in one-year particle exposure (ammonium and sulfate) levels comparing participants with and without a homozygous major variant genotype for *AGO2* and *DROSHA* in fully-adjusted linear mixed effects models.

We observed a significant (p=0.01) increasing trend for the association of $PM_{2.5}$ with DNAm-age when comparing the three rs4961280 (*AGO2*) SNP genotypes (Figure 3). The strongest (magnitude) association was observed in individuals who were homozygous for the minor allele and the smallest association was observed in individuals with the homozygous major allele genotype. An association of intermediate

magnitude was observed in individuals who were heterozygous for the genotype. A similarly significant (p=0.02) trend across genotypes was observed for the relationship of ammonium with DNAm-age (Figure 3). In a sensitivity analysis where the ammonium component was subtracted from total $PM_{2.5}$ mass, we still observed a significant – though slightly attenuated – trend in the $PM_{2.5}$ (less ammonium) and DNAm-age relationship across *AGO2* genotypes (Figure S2).

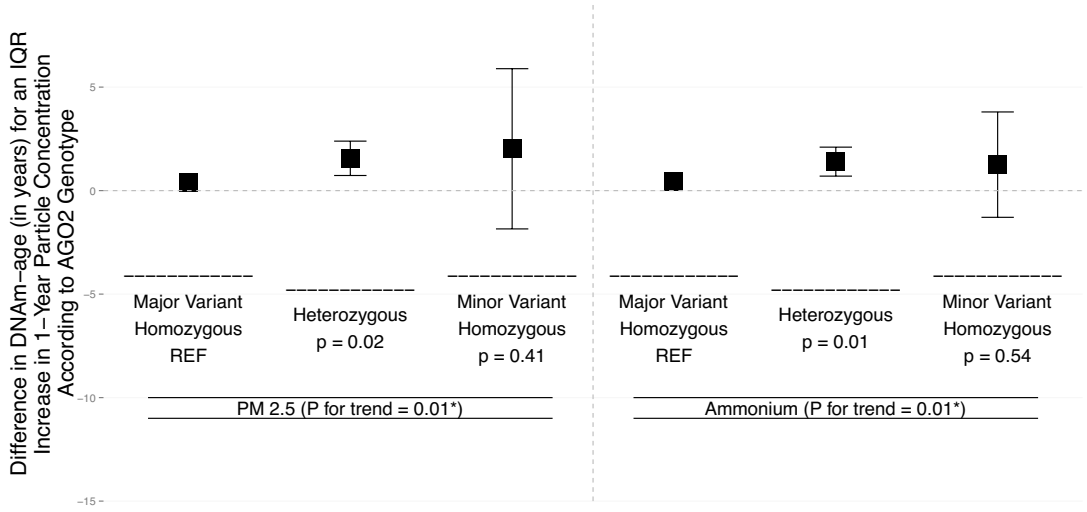

Figure 3 | Difference in DNAm-age for one interquartile range increase in one-year particle exposure ($PM_{2.5}$ and Ammonium) levels comparing participants of homozygous major variant (N=526), heterozygous (N=257), and homozygous minor variant genotypes (N=25) for *AGO2* in fully-adjusted linear mixed-effects models. *P value for the test of linear trend across genotypes was based on a linear mixed-effects regression model where the three *AGO2* genotypes were fit as a continuous measure.

6.4.c. Gene Network Analysis

Figure 4 depicts the results of a network analysis examining relationships of *AGO2*, *DROSHA*, *GEMIN4*, and *TARBP2* with the 353 genes that contribute CpGs to the DNAm-age measure. *IPO8* was the sole DNAm-age CpG contributing gene that had a physical interaction with *AGO2*. Many genes were found

to have genetic interactions or be co-expressed with *AGO2*. *PAPOLG* and *TIPARP* were the only two genes that had both genetic interactions and were co-expressed with *AGO2. PAPOLG* and *TIPARP* were also co-expressed or had a genetic interaction with *DROSHA* (Table S4).

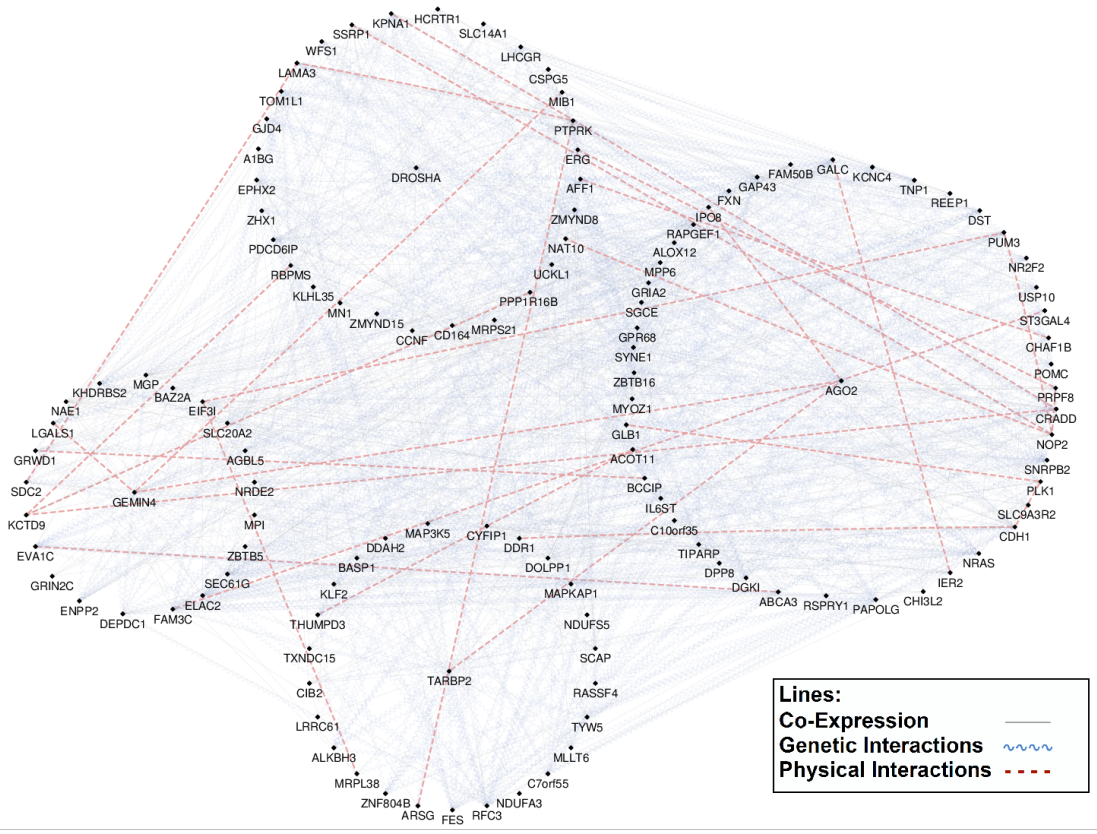

Figure 4 | Curated network map depicting relationships of *AGO2, DROSHA, GEMIN4,* and *TARBP2* with genes that contribute component CpG methylation to DNAm-age. Each of the elastic net selected genes is surrounded by a circle of related genes that contribute CpG methylation to the DNAm-age metric. Solid lines that connect genes represent co-expression. Dashed lines that connect genes represent physical interactions. Squiggly lines that connect genes represent genetic interactions.

6.5. Discussion

The present study utilized a DNAm-age elastic net selection model to identify four SNPs in miRNA processing genes (rs4961280 [*AGO2*], rs6877842 [*DROSHA*], rs910924 [*GEMIN4*] and rs784567 [*TARBP2*]) of which two (rs4961280 [*AGO2*] and rs784567 [*TARBP2*]) were directly associated with DNAm-age in a population of community-dwelling elderly men. Additionally, the study demonstrated a significant modifier effect of the rs4961280 (*AGO2*) SNP on the associations of one-year $PM_{2.5}$ and ammonium (one $PM_{2.5}$ component species) levels with DNAm-age. More specifically, our data suggests that the association of $PM_{2.5}$ with DNAm-age is attenuated in individuals carrying at least one copy of the rs4961280 (*AGO2*) major variant allele. Our results were consistent (though slightly attenuated) in sensitivity analyses where we subtracted ammonium levels from total $PM_{2.5}$ mass. This suggests that the impact of the rs4961280 (*AGO2*) SNP on the relationship between $PM_{2.5}$ and DNAm-age is largely – but not exclusively – due to ammonium is levels. Moreover, a gene network analysis revealed physical interactions, genetic interactions, and co-expression relationships of *AGO2*, *DROSHA*, *GEMIN4*, and *TARBP2* with genes that contribute CpGs to the DNAm-age metric. To our knowledge, this is the first study to examine relationships of miRNA processing physiology with epigenetic age both independently and in the context of long-term $PM_{2.5}$ exposure.

DROSHA is a gene located on human chromosome 5 and it encodes an RNA-specific endoribonuclease that is involved in the initial step of nuclear miRNA processing[31]. *GEMIN4* and *TARBP2* are located on chromosomes 17 and 12 respectively, and they both encode enzymes that are involved in the cytoplasmic processing of miRNAs. After a literature review examining the relationships of the SNPs in these three genes with air pollution, we found only one previous study – also in the Normative Aging Study cohort – demonstrating that in comparison to other participants, individuals heterozygous for the rs910924 (*GEMIN4)* SNP genotype were more likely to have lower global cognition measurements given the same level of black carbon exposure[22]. Nonetheless, we did find a number studies implicating *DROSHA*, *GEMIN4*, and *TARBP2* in numerous aging-related diseases including prostate cancer, and colorectal

cancer[32,33]. Since cancers are often related to changes in DNA methylation, the connections of these genes with cancer may be the reason why their SNPs were selected by the elastic net[34].

In contrast to the other three SNPs, we did observe significant effect modification of the $PM_{2.5}$-DNAm-age relationship by the rs4961280 (*AGO2*) SNP. Furthermore, we also observed a significant direct relationship of the rs4961280 (*AGO2*) SNP with DNAm-age. *AGO2* (Argonaute Protein 2) is a gene located on human chromosome 2. Argonaute (AGO) proteins, including AGO2, form the core of the RNA-induced Silencing Complex (RISC) which is involved in gene silencing via RNA interference. AGO proteins are well-conserved across species and structurally include an amino-terminal, PAS, Piwi, and MID domains. Humans have eight AGO proteins; however, only AGO 1-4 are capable of loading miRNA in RISC. Moreover, only AGO2 appears to have the ability to cleave mRNA targets and achieve transcript instability/silencing[35]. In addition to their role in RISC, it has also been demonstrated that AGO proteins play a role in stabilizing and maintaining proper levels of mature miRNA strands[36].

Our data suggested that individuals with at least one copy of the rs4961280 (*AGO2*) major variant had an attenuated association of $PM_{2.5}$/ammonium with DNAm-age when compared to individuals who were homozygous for the minor variant. However, individuals with at least one copy of the minor variant on average had lower DNAm-ages when compared to individuals that were homozygous for the major variant. Very few studies have explicitly examined relationships of the rs4961280 (*AGO2*) SNP and none of them were in the contexts of $PM_{2.5}$ or aging. However, we did find one study that demonstrated that the minor variant was associated with a reduced risk of benign prostatic hypertrophy (BPH) in a Serbian population[32]. Since BPH is most common in aging men, this study is in alignment with our finding that the minor allele is associated with qualities of being "younger" (*i.e.* a lower DNAm-age or a lower risk of BPH)[37]. Further work will need to be done to understand why it is the major allele that attenuates the positive association of $PM_{2.5}$ with DNAm-age, but the minor allele that is directly associated with a lower biological age. Moreover, it will be helpful for the field to understand why ammonium relationships with DNAm-age were impacted by the rs4961280 (*AGO2*) SNP and sulfate relationships were not.

Of the existing studies that examine relationships of *AGO2* – not simply the SNP, but the gene – with aging in humans, the majority are cell culture based and describe AGO2 as a factor involved in molecular processes related to biological aging such as cellular senescence, stem cell renewal, and endothelial function[38-40]. We also found a few animal studies that showed relationships of *AGO2* with chronological aging. A study examining relationships of miRNAs with aging in *Drosophila* revealed that with age, there was a global increase in miRNAs loaded in AGO2 but not AGO1. Furthermore, mutations in *AGO2* resulted in shorter life span and neurodegeneration. Together, these data suggest that *AGO2* impacts aging-associated processes[41]. Another study looking to elucidate how intermittent fasting increases longevity in *Caenorhabditis elegans* demonstrated that fasting upregulates the expression of miRNA-induced silencing complex (RISC) components including argonautes. In this study, fasting upregulated AGO2 by 2 fold[42].

Our network analyses of *AGO2*, *DROSHA*, *GEMIN4*, *TARBP2*, and genes that contribute CpGs to the DNAm-age metric demonstrated one physical interaction between *AGO2* and a gene called *IPO8*. *IPO8* (Importin 8) is a gene on chromosome 12 that encodes a protein involved in mediating the nuclear import of other proteins with nuclear localization signals. IPO8 has also been shown to mediate the cytoplasm to nucleus transport of mature miRNAs. Moreover, this IPO8 mediated transport of miRNAs is dependent on the physical association of IPO8 with the AGO2 complex[43]. The literature primarily describes IPO8 as an optimum reference gene for micro-array and RT-PCR studies in multiple tissue types including the lung[44]. Two genes (*PAPOLG* and *TIPARP*) had both genetic interactions and co-expression relationships with *AGO2*. These two genes also had genetic interactions with *TARBP2* and *DROSHA*. *PAPOLG* (Poly [A] polymerase gamma) is a gene on chromosome 2 that encodes an exclusively nuclear-localized poly (A) polymerase responsible for catalyzing template-independent extension of the 3' end of a strand of DNA/RNA[45]. To our knowledge, no explicit studies have examined the relationships of *PAPOLG* and biological aging but *PAPOLG* has been implicated in relationships involving aging-related health outcomes. For instance, a 12-week trial examined if a dietary intervention of 400g/week of high-glucoraphanin (HG) broccoli altered plasma metabolites linked to cancer risk when compared to diets of 400g/week of standard

broccoli or 400g/week of peas. No other modifications were made to the participants' diets. The study revealed that the levels of plasma metabolites (including FAD) of individuals receiving HG broccoli were differentiated by *PAPOLG* genotypes. This suggests that *PAPOLG* may interact with diet to impact the levels of metabolites including those that have been implicated in cancer risk[46]. *TIPARP* (TCDD-inducible poly [ADP-ribose] polymerase) is a gene on chromosome 3 that encodes a member of the poly (ADP-ribose) polymerase super family[47]. In a study exposing human aortic endothelial cells to 10 μg/ml of fine and ultrafine ambient particulate matter from California, mRNA levels of enzymes including TIPARP increased[48]. Another study exposed human adenocarcinomic human alveolar basal epithelial (A549) cells to 10 μg/ml of winter and summer $PM_{2.5}$ from Milan and found that $PM_{2.5}$ from both seasons modulated *TIPARP* gene expression[49].

Strengths of our study include the combination of a novel biomarker, rigorous statistical methods, and access to a large cohort with extensive and repeated information regarding $PM_{2.5}$ exposure levels, DNA methylation data, and potential confounders from multiple study visits. This is the first study to use miRNA processing gene variants to study the relationship of ambient fine particles with DNAm-age. However, our study does have some notable limitations. First, we use a validated chemical transport model to generate address-level one-year $PM_{2.5}$, sulfate, and ammonium exposure estimates. Given that most NAS participants are retired, we believe that particle exposure levels at their homes approximately capture their personal exposures. Still, there is some risk of exposure misclassification. Nonetheless, such non-differential misclassification is likely to underestimate any observed associations rather than bias them away from the null[50]. Second, we utilize a panel of a miRNA processing gene SNPs that is somewhat limited because it does not provide genome-wide resolution of all genes involved in miRNA processing. Nevertheless, this panel has been successfully utilized in other environmental health studies and we use a rigorous elastic net approach to identify our variants of interest[22]. Although we did not test for effect modification with all the SNPs in our panel, our targeted approach identified significant interactions that persisted even in sensitivity analyses. These findings will be informative to more comprehensive, future research. Lastly, our study examines the role of the miRNA processing pathway by using genetic variants

of miRNA processing genes and our findings are based on a cohort of elderly Caucasian males who reside in a lightly-polluted environment. Future studies involving other demographic groups, in different environments, and using miRNA expression levels will be necessary to broadly confirm and add to these important but early findings.

6.6. Conclusion

In conclusion, genotypes of the rs4961280 (*AGO2*) miRNA processing SNP were directly associated with DNAm-age and modified the associations of one-year $PM_{2.5}$ and ammonium levels with DNAm-age in this population of community-dwelling Caucasian elderly men. Although our findings need to be confirmed in other individuals of this same demographic group and different populations, they begin to address the important research gap concerning the biological relevance of DNAm-age and the physiology of the $PM_{2.5}$-DNAm-age relationship. Future studies will be necessary to elucidate more nuanced relationships of miRNA physiology with epigenetic aging.

6.7. Contributions and Support

Contributors: JCN and JDS conceived and designed the study. EC, QD, ACJ, LH, and PV gathered data. JCN performed the data analyses and drafted the manuscript. LD, JDS, EC, IDV, BLL, QL, MGW, and AAB contributed to the analyses. All authors revised and approved the manuscript.

Conflict of interest statement: None declared

Ethics approval: Boston VA Medical Center, Harvard T.H. Chan School of Public Health (protocol 14027-102).

Data Availability: Data are from the Normative Aging Study, from which restricted data are available for researchers who meet the criteria. Methylation data is deposited at NCBI dbGaP (study accession number: phs000853.v1.p1).

Funding: This work was supported by grants from the National Institute of Environmental Health Sciences (NIEHS) [R01ES021733 and R01ES025225]. Other support comes from NIEHS grants [ES015172,

ES014663, ES020010, P30ES009089, and P30ES000002]; Environmental Protection Agency (EPA) grants [RD832416 and RD83587201]; and a National Heart, Lung, and Blood Institute (NHLBI) grant [2T32HL007118-41]. Additional support was provided by the US Department of Agriculture, Agricultural Research Service [contract 53-K06-510]. The US Department of Veterans Affairs (VA) Normative Aging Study (NAS) is supported by the Cooperative Studies Program/ERIC, US Department of Veterans Affairs, and is a research component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). The views expressed in this paper are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs. JCN is also supported by a NIH/NIA Ruth L. Kirschstein National Research Service Award (1 F31AG056124-01A1).

6.8. References

1 World Health, O. Ageing and health. (2015).

2 Association, A. s. 2017 ALZHEIMER'S DISEASE FACTS AND FIGURES. *Alzheimer's & Dementia* **13**, 325-373 (2017).

3 World Health, O. Burden of disease from household air pollution for 2012. *WHO, Geneva* (2014).

4 Zhong, J. *et al.* Traffic-Related Air Pollution, Blood Pressure, and Adaptive Response of Mitochondrial Abundance. *Circulation* **133**, 378-387, doi:10.1161/CIRCULATIONAHA.115.018802 (2016).

5 Schikowski, T. *et al.* Association of air pollution with cognitive functions and its modification by APOE gene variants in elderly women. *Environ Res* **142**, 10-16, doi:10.1016/j.envres.2015.06.009 (2015).

6 Raaschou-Nielsen, O. *et al.* Particulate matter air pollution components and risk for lung cancer. *Environ Int* **87**, 66-73, doi:10.1016/j.envint.2015.11.007 (2016).

7 Nwanaji-Enwerem, J. C. *et al.* Long-term ambient particle exposures and blood DNA methylation age: findings from the VA normative aging study. *Environ Epigenet* **2**, doi:10.1093/eep/dvw006 (2016).

8 Nwanaji-Enwerem, J. C. *et al.* Associations between Long-term Exposure to PM2.5 Component Species and Blood DNA Methylation Age in the Elderly: The VA Normative Aging Study. *Environ Int* **102**, 57-65 (2017).

9 Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* **8**, 1844-1865, doi:10.18632/aging.101020 (2016).

10 Horvath, S. & Ritz, B. R. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY)* **7**, 1130-1142 (2015).

11 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).

12 Edwards, J. R., Yarychkivska, O., Boulard, M. & Bestor, T. H. DNA methylation and DNA methyltransferases. *Epigenetics Chromatin* **10**, 23, doi:10.1186/s13072-017-0130-8 (2017).

13 Rodosthenous, R. S. *et al.* Ambient particulate matter and microRNAs in extracellular vesicles: a pilot study of older individuals. *Part Fibre Toxicol* **13**, 13, doi:10.1186/s12989-016-0121-0 (2016).

14 Kumar, S., Vijayan, M., Bhatti, J. S. & Reddy, P. H. MicroRNAs as Peripheral Biomarkers in Aging and Age-Related Diseases. *Prog Mol Biol Transl Sci* **146**, 47-94, doi:10.1016/bs.pmbts.2016.12.013 (2017).

15 Parodi, F. *et al.* Epigenetic dysregulation in neuroblastoma: A tale of miRNAs and DNA methylation. *Biochim Biophys Acta* **1859**, 1502-1514, doi:10.1016/j.bbagrm.2016.10.006 (2016).

16 Morita, S. *et al.* One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation. *Genomics* **89**, 687-696, doi:10.1016/j.ygeno.2007.01.004 (2007).

17 Hodjat, M. *et al.* Environmental toxicants, incidence of degenerative diseases, and therapies from the epigenetic point of view. *Arch Toxicol*, doi:10.1007/s00204-017-1979-9 (2017).

18 Bell, B., Rose, C. L. & Damon, A. The Veterans Administration Longitudinal Study of Healthy Aging. *The Gerontologist* **6**, 179-184, doi:10.1093/geront/6.4.179 (1966).

19 Di, Q., Koutrakis, P. & Schwartz, J. A hybrid prediction model for PM 2.5 mass and components using a chemical transport model and land use regression. *Atmospheric Environment* **131**, 390-399 (2016).

20 Yang, H. *et al.* Evaluation of genetic variants in microRNA-related genes and risk of bladder cancer. *Cancer Res* **68**, 2530-2537, doi:10.1158/0008-5472.CAN-07-5991 (2008).

21 Horikawa, Y. *et al.* Single nucleotide polymorphisms of microRNA machinery genes modify the risk of renal cell carcinoma. *Clin Cancer Res* **14**, 7956-7962, doi:10.1158/1078-0432.CCR-08-1199 (2008).

22 Colicino, E. *et al.* Long-term exposure to black carbon, cognition and single nucleotide polymorphisms in microRNA processing genes in older men. *Environ Int* **88**, 86-93, doi:10.1016/j.envint.2015.12.014 (2016).

23 Wilker, E. H. *et al.* Black carbon exposures, blood pressure, and interactions with single nucleotide polymorphisms in MicroRNA processing genes. *Environ Health Perspect* **118**, 943-948, doi:10.1289/ehp.0901440 (2010).

24 Lenters, V. *et al.* Prenatal Phthalate, Perfluoroalkyl Acid, and Organochlorine Exposures and Term Birth Weight in Three Birth Cohorts: Multi-Pollutant Models Based on Elastic Net Regression. *Environ Health Perspect* **124**, 365-372, doi:10.1289/ehp.1408933 (2016).

25 Nwanaji-Enwerem, J. C. *et al.* Modifying Role of Endothelial Function Gene Variants on the Association of Long-term PM2.5 Exposure with Blood DNA Methylation Age: the VA Normative Aging Study. *Toxicol Sci*, doi:10.1093/toxsci/kfx077 (2017).

26 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).

27 Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301-320 (2005).

28 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86, doi:10.1186/1471-2105-13-86 (2012).

29 Panni, T. *et al.* 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* **124**, 983-990, doi:10.1289/ehp.1509966 (2016).

30 Dai, L. *et al.* Differential DNA methylation and PM2.5 species in a 450 K epigenome-wide association study. *Epigenetics*, 1-10, doi:10.1080/15592294.2016.1271853 (2016).

31 Daugaard, I. & Hansen, T. B. Biogenesis and Function of Ago-Associated RNAs. *Trends Genet* **33**, 208-219, doi:10.1016/j.tig.2017.01.003 (2017).

32 Nikolic, Z. *et al.* Genetic variants in RNA-induced silencing complex genes and prostate cancer. *World J Urol* **35**, 613-624, doi:10.1007/s00345-016-1917-0 (2017).

33 Mullany, L. E., Herrick, J. S., Wolff, R. K., Buas, M. F. & Slattery, M. L. Impact of polymorphisms in microRNA biogenesis genes on colon cancer risk and microRNA expression levels: a population-based, case-control study. *BMC Med Genomics* **9**, 21, doi:10.1186/s12920-016-0181-x (2016).

34 An, J., Rao, A. & Ko, M. TET family dioxygenases and DNA demethylation in stem cells and cancers. *Exp Mol Med* **49**, e323, doi:10.1038/emm.2017.5 (2017).

35 Meister, G. *et al.* Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**, 185-197, doi:10.1016/j.molcel.2004.07.007 (2004).

36 Zhang, X., Azhar, G. & Wei, J. Y. The expression of microRNA and microRNA clusters in the aging heart. *PLoS One* **7**, e34688, doi:10.1371/journal.pone.0034688 (2012).

37 Unnikrishnan, R., Almassi, N. & Fareed, K. Benign prostatic hyperplasia: Evaluation and medical management in primary care. *Cleve Clin J Med* **84**, 53-64, doi:10.3949/ccjm.84a.16008 (2017).

38 Kim, B. S., Im, Y. B., Jung, S. J., Park, C. H. & Kang, S. K. Argonaute2 regulation for K+ channel-mediated human adipose tissue-derived stromal cells self-renewal and survival in nucleus. *Stem Cells Dev* **21**, 1736-1748, doi:10.1089/scd.2011.0388 (2012).

39 Benhamed, M., Herbig, U., Ye, T., Dejean, A. & Bischof, O. Senescence is an endogenous trigger for microRNA-directed transcriptional gene silencing in human cells. *Nat Cell Biol* **14**, 266-275, doi:10.1038/ncb2443 (2012).

40 Lopez, P., Wagner, K. D., Hofman, P. & Van Obberghen, E. RNA Activation of the Vascular Endothelial Growth Factor Gene (VEGF) Promoter by Double-Stranded RNA and Hypoxia: Role of Noncoding VEGF Promoter Transcripts. *Mol Cell Biol* **36**, 1480-1493, doi:10.1128/MCB.01096-15 (2016).

41 Abe, M. *et al.* Impact of age-associated increase in 2'-O-methylation of miRNAs on aging and neurodegeneration in Drosophila. *Genes Dev* **28**, 44-57, doi:10.1101/gad.226654.113 (2014).

42 Kogure, A., Uno, M., Ikeda, T. & Nishida, E. The MicroRNA Machinery Regulates Fasting-Induced Changes in Gene Expression and Longevity in Caenorhabditis elegans. *J Biol Chem*, doi:10.1074/jbc.M116.765065 (2017).

43 Weinmann, L. *et al.* Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell* **136**, 496-507, doi:10.1016/j.cell.2008.12.023 (2009).

44 Nguewa, P. A. *et al.* Identification of importin 8 (IPO8) as the most accurate reference gene for the clinicopathological analysis of lung specimens. *BMC Mol Biol* **9**, 103, doi:10.1186/1471-2199-9-103 (2008).

45 Kyriakopoulou, C. B., Nordvarg, H. & Virtanen, A. A novel nuclear human poly(A) polymerase (PAP), PAP gamma. *J Biol Chem* **276**, 33504-33511, doi:10.1074/jbc.M104599200 (2001).

46 Armah, C. N. *et al.* A diet rich in high-glucoraphanin broccoli interacts with genotype to reduce discordance in plasma metabolite profiles by modulating mitochondrial function. *Am J Clin Nutr* **98**, 712-722, doi:10.3945/ajcn.113.065235 (2013).

47 Katoh, M. & Katoh, M. Identification and characterization of human TIPARP gene within the CCNL amplicon at human chromosome 3q25.31. *Int J Oncol* **23**, 541-547 (2003).

48 Aung, H. H. *et al.* Comparative gene responses to collected ambient particles in vitro: endothelial responses. *Physiol Genomics* **43**, 917-929, doi:10.1152/physiolgenomics.00051.2011 (2011).

49 Gualtieri, M. *et al.* Gene expression profiling of A549 cells exposed to Milan PM2.5. *Toxicol Lett* **209**, 136-145, doi:10.1016/j.toxlet.2011.11.015 (2012).

50 Kioumourtzoglou, M. A. *et al.* Exposure measurement error in PM2.5 health effects studies: a pooled analysis of eight personal exposure validation studies. *Environ Health* **13**, 2, doi:10.1186/1476-069X-13-2 (2014).

Chapter 7:

Conclusion

7.1. Introduction

The work presented in this dissertation represents the first published studies examining relationships of any traditional environmental pollutant with DNAm-age. Since our initial study, we have replicated some biological aging relationships in an independent German cohort and reported relationships of other ambient pollutants (*e.g.* nitric oxide) with DNAm-age[1]. Moreover, other research groups have since explored the relationships of other environmental exposures with DNAm-age[2]. Here, we summarize the existing body of research that describes relationships of DNAm-age with chemical environmental pollutants.

7.2. Air Pollutants

Of the work examining the relationships of chemical exposures with DNAm-age, the literature on air pollutants is the most extensive. *Nwanaji-Enwerem et al. (2016)* was the first group to describe associations of any chemical pollutant with DNAm-age and their study focused on relationships of DNAm-age with long-term ambient fine particle ($PM_{2.5}$) and Black Carbon (BC) exposures[3]. DNAm-age was measured in peripheral blood leukocytes from 1032 samples taken from 589 community-dwelling older men who were participants in the ongoing VA Normative Aging Study (NAS). Using linear mixed effects models adjusted for chronic diseases, lifestyle factors, environmental factors, and white blood cell composition, the group found that a 1 $\mu g/cm^3$ increase in one-year $PM_{2.5}$ exposure was significantly associated with a 6-month increase in DNAm-age ($p < 0.0001$). A 1 $\mu g/cm^3$ increase in one-year BC exposure was also significantly associated with DNAm-age ($\beta = 2.83$, $p = 0.03$). However, when $PM_{2.5}$ and BC were simultaneously modeled as predictors of DNAm-age, only $PM_{2.5}$ remained statistically significant. This prompted the authors to focus their attention to $PM_{2.5}$ relationships including an analysis exploring the associations of $PM_{2.5}$ levels with methylation at the 353 DNAm-age component CpG sites. 20 out of 353 DNAm-age component CpGs had significant associations with one-year $PM_{2.5}$ exposure, but enrichment analysis did not return any significant findings for these 20 sites[3].

Given the results from the combined $PM_{2.5}$ and BC models, the authors hypothesized that $PM_{2.5}$ component species besides the carbonaceous species were responsible for the association with DNAm-age. Hence, they performed a follow-up study in the NAS to identify which of five major $PM_{2.5}$ component species (*i.e.* ammonium, elemental carbon, organic carbon, nitrate, and sulfate) were driving the DNAm-age association. Using the adaptive least absolute shrinkage and selection operator method, the researchers identified ammonium and sulfate as the $PM_{2.5}$ component species most associated with DNAm-age[4]. Importantly, the associations of $PM_{2.5}$, sulfate, and ammonium with DNAm-age all remained statistically significant even when the analyses were limited to one-year $PM_{2.5}$ exposures within US EPA national ambient air quality standards. This suggested that there was a risk of biological aging even at accepted air pollution levels. To further understand the underlying biology of the risks implied by this association, subsequent work by this same group focused on identifying potential modifiers and mediators of the $PM_{2.5}$-DNAm-age relationship in the NAS.

First, they used elastic net penalized regression to identify endothelial function related SNPs most important for DNAm-age and next they calculated a polymorphism score based on the important SNPs[5]. This score, which reflected endothelial function physiology and aging risk, was then explored as a modifier of the $PM_{2.5}$-DNAm-age association. The authors found that the magnitude of the association of $PM_{2.5}$ with DNAm-age was significantly higher in individuals with a high aging-risk endothelial function score (β_{high} = 1.09, 95% CI: 0.70, 1.48) when compared to individuals with a low score (β_{low} = 0.40, 95% CI: 0.14, 0.67, $P_{interaction}$ = 0.0007). Following this finding of significant effect modification, the authors then examined the relationship of DNAm-age with serum endothelial function markers in the same cohort of NAS participants. In these analyses, DNAm-age was positively associated with serum ICAM (β = 0.01, p = 0.005) and VCAM (β = 0.002, p = 0.004). This was the first study to use pathway specific genetic variants to understand the association of $PM_{2.5}$ with DNAm-age. Furthermore, the integrated data from genetic and functional analyses suggests a role in the $PM_{2.5}$-DNAm-age relationship for endothelial function, which is already appreciated as a contributor to $PM_{2.5}$-related processes[6-8].

The second molecularly focused study conducted by the group suggests that mitochondrial physiology may be involved in the $PM_{2.5}$-DNAm-age relationship. Similar to the study involving endothelial function, this study is able to provide a more convincing argument for its conclusion because it utilizes an integration of genetic (mitochondrial DNA haplogroups) and functional (mitochondrial DNA copy number) measures in the NAS. Mitochondrial haplogroups represent normal mitochondrial genetic variation and can potentially impact energy generating capacity by the organelle. Mitochondrial copy number is the ratio of a cell's mitochondrial DNA to nuclear DNA. Changes in copy number can occur normally with mitochondrial biogenesis and degradation, but changes can also be related to exogenous stressors and enable the organelle to compensate for mitochondrial genome damage[9]. Out of the nine haplogroups found and tested (H, I, J, K, T, U, V, W, and X) in the study sample, no haplogroups showed direct associations with DNAm-age. However, carriers of Haplogroup V ($\beta = 0.18$, 95% CI: −0.41, 0.78) demonstrated a diminished magnitude of the $PM_{2.5}$ association with DNAm-age when compared to non-carriers ($\beta = 1.25$, 95% CI: 0.58, 1.93, $P_{interaction} = 0.01$). Copy number was negatively associated with DNAm-age ($\beta = -3.31$, $p < 0.0001$), and was estimated to significantly mediate 12% of the $PM_{2.5}$ association with DNAm-age. Like endothelial function, mitochondrial genome physiology has been previously implicated in $PM_{2.5}$-related processes[10-13]. Hence, these results though they need to be replicated in a different population, were in line with the existing literature.

The final molecularly focused study performed by this group also utilized the NAS cohort and elastic net penalized regression to identify two SNPs in microRNA processing genes that had direct associations with DNAm-age: rs4961280 [*AGO2*] and rs784567 [*TARBP2*]. Individuals with at least one copy of the minor variant of either one of these two SNPs had on average at least a 1.13-year lower DNAm-age than homozygous major individuals. However, significant effect modification was only observed by the *AGO2* SNP. Having at least one copy of the major *AGO2* allele significantly reduced the magnitude of the $PM_{2.5}$ association with DNAm-age[14]. Again, the authors explored these relationships because microRNA physiology has been previously implicated in $PM_{2.5}$ processes[15-17]. Hence, the findings were again broadly in alignment with the existing literature. Overall, each of these molecular endothelial

function, mitochondrial, and microRNA relationships need to be replicated in different populations and demographic groups, but these studies represent some of the most comprehensive data aimed at understanding DNAm-age's molecular relationships with environmental exposures.

Still, efforts have been made to replicate the direct associations of $PM_{2.5}$ and BC with two variations of DNAm-age (IEAA and DNAmAA). Intrinsic epigenetic age acceleration (IEAA) is the residual that results from regressing DNAm-age on chronological age and measures of blood cell counts. Hence, IEAA is independent of both age and cell counts. DNAm-age age acceleration (DNAmAA) is the residual from regressing DNAm-age on chronological age alone. Hence, DNAmAA is independent of chronological age[18]. *Ward-Caviness et al. (2016)* utilized a cohort of 1777 men and women from the German KORA cohort and examined relationships of IEAA and DNAmAA with ambient pollutants: $PM_{2.5}$, BC, coarse particulate matter (PM_{10}), and nitrogen oxide (NO_x)[1]. In a combined sex model (males and females included) $PM_{2.5}$ was not significantly associated with IEAA ($\beta = 0.02$, $p = 0.88$) or DNAmAA ($\beta = 0.04$, $p = 0.77$). BC was also not significantly associated with IEAA or DNAmAA in combined sex models. However, unlike $PM_{2.5}$, BC did show some sex-specific relationships. BC was negatively associated with IEAA in men, but positively associated with IEAA in women. BC was positively associated with DNAmAA in women but showed no association in men. With respect to the other pollutants that were examined, PM_{10} had significant negative associations with IEAA and DNAmAA in men but no relationships in women. NO_x had significant negative associations with IEAA and DNAmAA in men, but positive associations women. This study also examined $PM_{2.5}$ associations with IEAA and DNAmAA in a subset of the NAS cohort and found only the IEAA associations to be statistically significant but negative in direction ($\beta = -0.42$, $p = 0.03$). The data from this study broaden the knowledge of important relationships of air pollutants with DNAm-age-related measures. Although, the findings using the full cohort are predominantly null, important sex differences were identified. This suggests that the impact of pollutants on DNAm-age may differ by sex, which is not farfetched given that it is already appreciated that on average chronological age-matched DNAm-age is lower in women than it is in men[19]. The results also speak to the relationships between the different DNAm-age measures. What is most surprising is that the relationships of IEAA and

$PM_{2.5}$ in the KORA cohort and the subset of the NAS cohort are either not statistically significant or are in the opposite direction of those previously reported by *Nwanaji-Enwerem et al. (2016)*. Using IEAA versus adjusting for white blood cell proportions within statistical models – which was the method employed by *Nwanaji-Enwerem et al. (2016)*[3] – could potentially lead to different results even though this is not expected. The authors also noted that total cholesterol, age[2], and physical activity were included as covariates in their model but were not used in the *Nwanaji-Enwerem et al. (2016)* models. Moreover, there were some differences in the sample sizes used by the two studies given that additional covariates included in the latter study were missing at random for some participants in the initial study sample. The initial *Nwanaji-Enwerem et al. (2016)* study was based on 1032 observations from 589 participants, while this latter work was based on a subset of 734 observations from 496 participants in the NAS. Ultimately, more work is necessary to elucidate the associations of ambient air pollutants with DNAm-age measures, but these future studies will need to pay close attention to consistency in model design.

Smoking can be considered a form of personal air pollution. *Gao et al. (2016)* examined the associations of smoking with DNAmAA in a cohort of 1509 male and female participants from the ESTHER study in Germany[20]. They found that no self-reported smoking related indicators *(i.e.* smoking status, cessation time, and cumulative exposure) were significantly associated with DNAmAA. They next compiled a list of 150 smoking-related CpGs that were independent from the 353 DNAm-age component CpGs. Each of the 150 CpGs had been identified at least twice in previously published active-smoking related epigenome wide association studies. 66 of these 150 smoking related CpGs were associated with DNAmAA after a validation step and were used to create a smoking index. A one standard deviation increase in smoking index was associated with a 1-year increase in DNAmAA. This study doesn't demonstrate direct associations of smoking self-reports with DNAm-age, possibly because self-reports and not the most accurate measure of smoking. However, the study does demonstrate that methylation of separate sites that are sensitive to certain exposures may have important associations with DNAm-age. Thus, DNAm-age may not always be sensitive in itself to an exposure, but it can be used with other loci to

develop a composite marker that is exposure sensitive. These types of integrated methylation analyses will likely be useful in future environmental exposure work.

7.3. Metal Exposures

The work examining the relationships of metal exposures with DNAm-age is far more limited than that examining air pollutants. To the best of our knowledge, only two studies have been published in this area. The first examines associations of age acceleration (here defined as chronological age subtracted from DNAm-age) with urinary cadmium (Cd) in 40 non-smoking women from Thailand[2]. Urinary Cd was first measured in the women and used to split the cohort into a high exposure (mean chronological age = 60.4 years) and a low exposure group (mean chronological age = 58.8 years). Then, linear mixed effects models adjusted for urinary creatinine, age, and white blood cells were used to examine the association of Urinary Cd with blood age acceleration. The study reports no significant differences in mean age acceleration between the groups and no associations with urinary Cd. However, secondary analyses revealed 20 of the 353 DNAm-age CpGs were differentially methylated between the high and low exposure groups. In reviewing the CpGs differentially expressed between the two Cd exposure groups, four of them (*EIF3I*, *TNFRSF13C*, *ZBTB5*, and *ACAP2*) were also noted for being associated with one-year $PM_{2.5}$ levels in the *Nwanaji-Enwerem et al. (2016)* studies[3,4]. *ZBTB5*, which is known to be involved in transcriptional regulation, was identified in both *Nwanaji-Enwerem et al. (2016)* studies (using different prediction models) and this Cd study. Future work that looks for associations of the 353 DNAm-age component CpGs with environmental exposures could also be useful for building a deeper understanding of environmental exposure related DNAm-age biology.

The second study related to metal exposure was a pilot study of 68 subjects examining the potential relationships of chronic cobalt (Co) and chromium (Cr) exposure from metal-on-metal hip implants with DNAmAA[21]. 34 arthritis patients with metal on metal hip replacements and 34 arthritis patients with non-metal hip replacements were recruited from 2009 to 2010 in the United Kingdom. There were 30 men and 4 women in each group and the average chronological age of each group was 59.7 years. Although serum

Co and Cr levels were significantly greater in the group with metal-on-metal hip implants, the study ultimately found no associations between metal exposure and DNAmAA. These findings may suggest that the exposure (metal rubbing off from implants) has a small biological impact with respect to DNA methylation or that a larger sample size is needed to detect statistically significant changes. Regardless, future studies exploring the relationships of metal exposure and DNAm-age will be highly useful and informative.

7.4. Organochlorine Pesticide Exposures

To date, one study has examined the relationships of organochlorine pesticides with DNAm-age[22]. This study measured three organochlorine pesticides – (4-chlorophenyl)-1,1-dichloroethene (DDE), hexa-chlorobenzene (HCB), and transnonachlor (TNC) – in the plasma of 967 participants from the Swedish PIVUS study. In statistical models adjusted for lifestyle and environmental factors, but not white blood cell composition, TNC ($\beta = 0.86$, $p = 0.006$) was significantly associated with DiffAge (defined as the difference between DNAm-age and chronological age). DDE ($\beta = 0.31$, $p = 0.10$) and HCB ($\beta = 0.15$, $p = 0.67$) were not significantly associated with DiffAge. This study highlights the issue of consistency in ongoing DNAm-age research. Consistency across study design will be important for conducting future research and interpreting the results of existing studies. Age acceleration, DNAm-age, and IEAA have all been used as outcomes in DNAm-age research. Age acceleration is at times defined as the difference between DNAm-age and chronological age (or vice versa) or even as the residuals from regressing DNAm-age on chronological age. This organochloride study elects to introduce a new term "DiffAge" which has an overlapping definition as the difference between DNAm-age and chronological age. Consistency in nomenclature will be critical for limiting confusion, replicating findings, and conveying results to larger scientific and lay community.

Consistency in study design is also critical to the interpretation of research findings. Again, it is important to note that this organochloride pesticide study did not to adjust for blood cell counts in its

DNAm-age statistical models. In the majority of studies, researchers adjust for blood cell counts as covariates or they simply use IEAA, which already accounts for blood cell proportions. It is widely accepted that blood cell counts should be adjusted for unless the researchers have a biological reason for why it is inappropriate (*e.g.* when HIV is the exposure and may impact DNAm-age through changes in blood cell count[23,24]). In a perfect world, each of the variations of DNAm-age that do account for blood cells would result in similar findings; however, to simply believe that they would is a major assumption. A future study showing how each of these measures does or does not vary in the context of one outcome or one biological process could be useful for providing this needed methodologic understanding.

7.5. Conclusion

In conclusion, DNAm-age is a novel biomarker that is pertinent to human aging and aging related conditions. Moreover, DNAm-age has been associated with a range of chemical environmental exposures (Figure 1). Establishing a better understanding of DNAm-age's molecular relationships will be critical for actualizing the maximum utility of this biomarker. DNAm-age may prove to be sensitive and specific for a particular disease that is related to environmental exposures. It may also prove to simply be a useful measure of a more general biological process with disease implications. Developing consistency in DNAm-age research communications/methodologies will be important for reaching either of these conclusions.

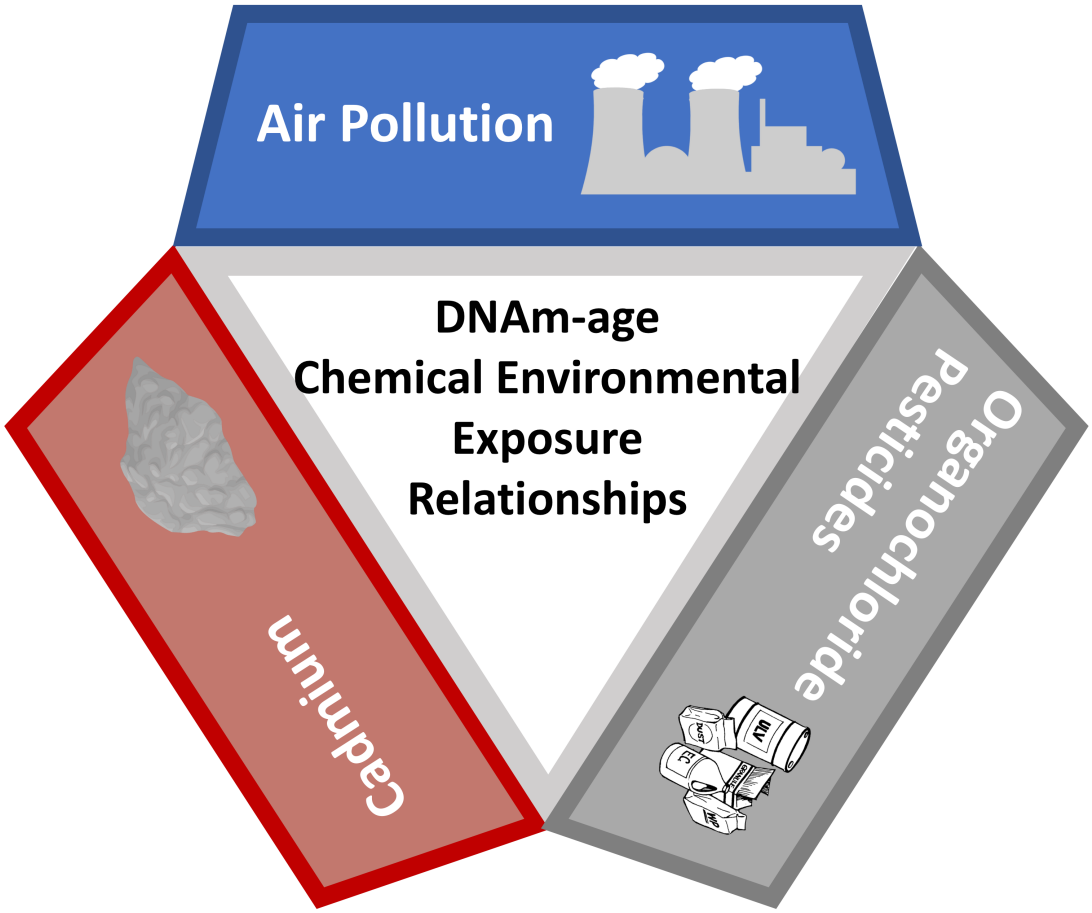

Figure 1 | Reported relationships of DNAm-age with chemical environmental pollutants in the existing literature. To date, DNAm-age has known relationships with air pollutants (*e.g.* black carbon and $PM_{2.5}$), cadmium, and organochloride pesticides (*e.g.* transnonachlor).

7.6. References

1 Ward-Caviness, C. K. *et al.* Long-term exposure to air pollution is associated with biological aging. *Oncotarget*, doi:10.18632/oncotarget.12903 (2016).

2 Demanelis, K. *et al.* Cadmium exposure and age-associated DNA methylation changes in non-smoking women from northern Thailand. *Environmental Epigenetics* **3**, dvx006-dvx006, doi:10.1093/eep/dvx006 (2017).

3 Nwanaji-Enwerem, J. C. *et al.* Long-term ambient particle exposures and blood DNA methylation age: findings from the VA normative aging study. *Environ Epigenet* **2**, doi:10.1093/eep/dvw006 (2016).

4 Nwanaji-Enwerem, J. C. *et al.* Associations between Long-term Exposure to PM2.5 Component Species and Blood DNA Methylation Age in the Elderly: The VA Normative Aging Study. *Environ Int* **102**, 57-65 (2017).

5 Nwanaji-Enwerem, J. C. *et al.* Modifying Role of Endothelial Function Gene Variants on the Association of Long-term PM2.5 Exposure with Blood DNA Methylation Age: the VA Normative Aging Study. *Toxicol Sci*, doi:10.1093/toxsci/kfx077 (2017).

6 Pope, C. A., 3rd *et al.* Exposure to Fine Particulate Air Pollution Is Associated With Endothelial Injury and Systemic Inflammation. *Circ Res* **119**, 1204-1214, doi:10.1161/CIRCRESAHA.116.309279 (2016).

7 Dai, L. *et al.* Fine particles, genetic pathways, and markers of inflammation and endothelial dysfunction: Analysis on particulate species and sources. *J Expo Sci Environ Epidemiol* **26**, 415-421, doi:10.1038/jes.2015.83 (2016).

8 Wu, S. *et al.* Chemical constituents and sources of ambient particulate air pollution and biomarkers of endothelial function in a panel of healthy adults in Beijing, China. *Sci Total Environ* **560-561**, 141-149, doi:10.1016/j.scitotenv.2016.03.228 (2016).

9 Nwanaji-Enwerem, J. C. *et al.* Impacts of the Mitochondrial Genome on the Relationship of Long-Term Ambient Fine Particle Exposure with Blood DNA Methylation Age. *Environ Sci Technol* **51**, 8185-8195, doi:10.1021/acs.est.7b02409 (2017).

10 Pieters, N. *et al.* Biomolecular Markers within the Core Axis of Aging and Particulate Air Pollution Exposure in the Elderly: A Cross-Sectional Study. *Environ Health Perspect* **124**, 943-950, doi:10.1289/ehp.1509728 (2016).

11 Zhong, J. *et al.* Traffic-Related Air Pollution, Blood Pressure, and Adaptive Response of Mitochondrial Abundance. *Circulation* **133**, 378-387, doi:10.1161/CIRCULATIONAHA.115.018802 (2016).

12 Lambertini, L. & Byun, H. M. Mitochondrial Epigenetics and Environmental Exposure. *Curr Environ Health Rep* **3**, 214-224, doi:10.1007/s40572-016-0103-2 (2016).

13 Peng, C. *et al.* Associations of Annual Ambient Fine Particulate Matter Mass and Components with Mitochondrial DNA Abundance. *Epidemiology* **28**, 763-770, doi:10.1097/EDE.0000000000000717 (2017).

14 Nwanaji-Enwerem, J. C. *et al.* miRNA processing gene polymorphisms, blood DNA methylation age and long-term ambient PM2.5 exposure in elderly men. *Epigenomics* **9**, 1529-1542, doi:10.2217/epi-2017-0094 (2017).

15 Wilker, E. H. *et al.* Black carbon exposures, blood pressure, and interactions with single nucleotide polymorphisms in MicroRNA processing genes. *Environ Health Perspect* **118**, 943-948, doi:10.1289/ehp.0901440 (2010).

16 Colicino, E. *et al.* Long-term exposure to black carbon, cognition and single nucleotide polymorphisms in microRNA processing genes in older men. *Environ Int* **88**, 86-93, doi:10.1016/j.envint.2015.12.014 (2016).

17 Rodosthenous, R. S. *et al.* Ambient particulate matter and microRNAs in extracellular vesicles: a pilot study of older individuals. *Part Fibre Toxicol* **13**, 13, doi:10.1186/s12989-016-0121-0 (2016).

18 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).

19 Horvath, S. *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol* **17**, 171, doi:10.1186/s13059-016-1030-0 (2016).

20 Gao, X., Zhang, Y., Breitling, L. P. & Brenner, H. Relationship of tobacco smoking and smoking-related DNA methylation with epigenetic age acceleration. *Oncotarget* **7**, 46878-46889, doi:10.18632/oncotarget.9795 (2016).

21 Steinberg, J., Shah, K. M., Gartland, A., Zeggini, E. & Wilkinson, J. M. Effects of chronic cobalt and chromium exposure after metal-on-metal hip resurfacing: An epigenome-wide association pilot study. *J Orthop Res* **35**, 2323-2328, doi:10.1002/jor.23525 (2017).

22 Lind, P. M., Salihovic, S. & Lind, L. High plasma organochlorine pesticide levels are related to increased biological age as calculated by DNA methylation analysis. *Environ Int* **113**, 109-113, doi:10.1016/j.envint.2018.01.019 (2018).

23 Levine, A. J. *et al.* Accelerated epigenetic aging in brain is associated with pre-mortem HIV-associated neurocognitive disorders. *J. Neurovirol.*, doi:10.1007/s13365-015-0406-3 (2015).

24 Horvath, S. & Levine, A. J. HIV-1 Infection Accelerates Age According to the Epigenetic Clock. *J Infect Dis* **212**, 1563-1573, doi:10.1093/infdis/jiv277 (2015).

Appendices

Appendix 1: Chapter 2 Supplementary Data

Table S1. Pearson Correlations with 1-Year Particulate Matter 2.5 ($PM_{2.5}$) and Black Carbon (BC) Estimates in NAS data (2000-2011)

Particle	Pearson Correlation Coefficient	Number of Observations	*P*
$PM_{2.5}$			
1 Year	**-**	1032	**-**
2 Year	0.97	436	**<0.0001**
3 Year	0.95	321	**<0.0001**
4 Year	0.91	222	**<0.0001**
5 Year	0.90	182	**<0.0001**
BC			
1 Year	-	898	-
2 Year	0.97	881	**<0.0001**
3 Year	0.95	866	**<0.0001**
4 Year	0.97	853	**<0.0001**
5 Year	0.97	829	**<0.0001**

Table S2. Significant Relationships between Mean DNAm-age and Dichotomous/Categorical Variables in All Participant Observations

	Mean DNAm-age in years (SD)
Coronary heart disease	Yes: 75.2 (8.10) No: 73.5 (7.74) ***P* = 0.002**
Hypertension	Yes: 74.6 (7.93) No: 72.5 (7.60) ***P* < 0.0001**
Lifetime cancer diagnosis	Yes: 75.1 (8.19) No: 72.8 (7.34) ***P* < 0.0001**
Smoking Status	Never: 75.7 (8.05) Former: 73.4 (7.80) Current: 73.6 (6.70) ***P* < 0.0001***

* only the p value for T test between never and former smokers was statistically significant

Table S3. Pearson Correlations of Continuous Variables in Participant Observations

	DNAm-age	$PM_{2.5}$	BC	Age	Tempera-ture	Pack years	Telomere	CD4 cells	NK cells	Monocytes	Granulo-cytes	Plasma cells	CD8 cells
$PM_{2.5}$	0.01 *P* = 0.87 N = 1032												
BC	0.03 *P* = 0.33 N = 898	**0.41** ***P* <0.0001** **N = 898**											
Age	**0.60** ***P* <0.0001** **N = 1032**	**-0.18** ***P* <0.0001** **N = 1032**	0.02 *P* = 0.48 N = 898										
Temperature	-0.02 *P* = 0.60 N = 1032	0.04 *P* = 0.20 N = 1032	**0.35** ***P* <0.0001** **N = 898**	-0.03 *P* = 0.34 N = 1032									
Pack years	-0.03 *P* = 0.30 N = 1032	0.06 *P* = 0.07 N = 1032	0.04 *P* = 0.18 N = 898	**-0.12** ***P* = 0.0002** **N = 1032**	-0.03 *P* = 0.33 N = 1032								
Telomere	-0.06 *P* = 0.08 N = 857	0.04 *P* = 0.22 N = 857	0.02 *P* = 0.54 N = 770	-0.03 *P* = 0.44 N = 857	-0.04 *P* = 0.30 N = 857	-0.03 *P* = 0.32 N = 857							
CD4 cells	**-0.17** ***P* <0.0001** **N = 1032**	**0.08** ***P* = 0.008** **N = 1032**	0.05 *P* = 0.10 N = 898	**-0.17** ***P* = 0.0002** **N = 1032**	-0.04 *P* = 0.16 N = 1032	-0.01 *P* = 0.73 N = 1032	0.01 *P* = 0.77 N = 857						
NK cells	**0.25** ***P* <0.0001** **N = 1032**	-0.05 *P* = 0.15 N = 1032	-0.02 *P* = 0.65 N = 898	**0.19** ***P* <0.0001** **N = 1032**	0.03 *P* = 0.33 N = 1032	**-0.10** ***P* = 0.002** **N = 1032**	**-0.10** ***P* = 0.003** **N = 857**	**-0.14** ***P* <0.0001** **N = 1032**					
Monocytes	**-0.10** ***P* = 0.004** **N = 1032**	-0.01 *P* = 0.66 N = 1032	0.02 *P* = 0.50 N = 898	-0.05 *P* = 0.08 N = 1032	**0.12** ***P* = 0.0001** **N = 1032**	0.03 *P* = 0.33 N = 1032	-0.01 *P* = 0.77 N = 857	**-0.28** ***P* <0.0001** **N = 1032**	**-0.18** ***P* <0.0001** **N = 1032**				
Granulocytes	-0.06 *P* = 0.07 N = 1032	0.005 *P* = 0.88 N = 1032	-0.01 *P* = 0.87 N = 898	0.05 *P* = 0.09 N = 1032	-0.04 *P* = 0.23 N = 1032	0.03 *P* = 0.30 N = 1032	0.06 *P* = 0.10 N = 857	**-0.50** ***P* <0.0001** **N = 1032**	**-0.53** ***P* <0.0001** **N = 1032**	**0.08** ***P* = 0.02** **N = 1032**			
Plasma cells	**-0.13** ***P* <0.0001** **N = 1032**	**-0.16** ***P* <0.0001** **N = 1032**	0.01 *P* = 0.82 N = 898	**0.10** ***P* = 0.001** **N = 1032**	0.05 *P* = 0.14 N = 1032	-0.02 *P* = 0.39 N = 1032	0.03 *P* = 0.43 N = 857	**-0.51** ***P* <0.0001** **N = 1032**	**-0.24** ***P* <0.0001** **N = 1032**	**0.29** ***P* <0.0001** **N = 1032**	**0.61** ***P* <0.0001** **N = 1032**		
CD8 cells	**0.19** ***P* <0.0001** **N = 1032**	-0.06 *P* = 0.05 N = 1032	**0.07** ***P* = 0.03** **N = 898**	**0.27** ***P* <0.0001** **N = 1032**	0.03 *P* = 0.31 N = 1032	0.04 *P* = 0.21 N = 1032	**-0.11** ***P* = 0.001** **N = 857**	**-0.39** ***P* <0.0001** **N = 1032**	**0.36** ***P* <0.0001** **N = 1032**	**0.10** ***P* = 0.001** **N = 1032**	-0.004 *P* = 0.90 N = 1032	**0.21** ***P* <0.0001** **N = 1032**	
naïve CD8 cells	**-0.20** ***P* <0.0001** **N = 1032**	0.005 *P* = 0.88 N = 1032	**-0.10** ***P* = 0.004** **N = 898**	**-0.20** ***P* <0.0001** **N = 1032**	-0.06 *P* = 0.06 N = 1032	**-0.07** ***P* = 0.02** **N = 1032**	0.05 *P* = 0.13 N = 857	**0.13** ***P* <0.0001** **N = 1032**	**-0.19** ***P* <0.0001** **N = 1032**	-0.06 *P* = 0.07 N = 1032	**0.29** ***P* <0.0001** **N = 1032**	**0.08** ***P* = 0.01** **N = 1032**	**-0.44** ***P* <0.0001** **N = 1032**

Table S4. 1-Year Particulate Matter 2.5 ($PM_{2.5}$) and Black Carbon (BC) as Predictors of DNA Methylation (DNAm) Age in Models Adjusted for 450k Plate

Particle (1 μg/m^3)	Difference in DNAm-age (95% CI)	*P*	N	AIC
$PM_{2.5}$				
Model 1	0.36 (0.15, 0.58)	**0.001**	1032	6320.22
Model 2	0.34 (0.12, 0.55)	**0.003**	1032	6334.26
Model 3	0.33 (0.11, 0.55)	**0.003**	1032	6334.98
Model 4	0.31 (0.09, 0.53)	**0.006**	1032	6337.17
BC				
Model 1	2.03 (-0.30, 4.36)	0.09	898	5528.47
Model 2	2.28 (-0.23, 4.79)	0.07	898	5541.63
Model 3	2.16 (-0.36, 4.68)	0.09	898	5542.51
Model 4	2.10 (-0.42, 4.62)	0.10	898	5543.39
Two-Particle Model 1			898	5527.19
$PM_{2.5}$	0.33 (0.06, 0.60)	**0.02**		
BC	0.93 (-1.57, 3.43)	0.46		
Two-Particle Model 2			898	5541.51
$PM_{2.5}$	0.31 (0.02, 0.60)	**0.04**		
BC	0.93 (-1.90, 3.76)	0.52		
Two-Particle Model 3			898	5542.71
$PM_{2.5}$	0.30 (0.01,0.59)	**0.04**		
BC	0.88 (-1.95, 3.71)	0.54		
Two-Particle Model 4			898	5543.88
$PM_{2.5}$	0.29 (-0.01,0.58)	0.06		
BC	0.88 (-1.95, 3.71)	0.54		

Model 1: adjusted for chronological age, blood cell type, and 450k plate.

Model 2: Model 1 but additionally adjusted for temperature, pack years, smoking status, season, BMI, alcohol consumption, and education.

Model 3: Model 2 but additionally adjusted for history of cancer, hypertension, chd, and diabetes.

Model 4: Model 2 but additionally adjusted for statins and medications for diabetes and hypertension.

Table S5. 1-Year Particulate Matter 2.5 ($PM_{2.5}$) and Black Carbon (BC) as Predictors of DNA Methylation (DNAm) Age in Participants with Only One NAS Visit

Particle (1 μg/m^3)	Difference in DNAm-age (95% CI)	*P*	N	AIC
$PM_{2.5}$				
Model 1	0.29 (-0.32, 0.90)	0.36	237	1505.79
Model 2	0.38 (-0.25, 1.01)	0.24	237	1509.82
Model 3	0.38 (-0.25, 1.01)	0.24	237	1504.55
Model 4	0.39 (-0.25, 1.02)	0.23	237	1510.27
BC				
Model 1	0.31 (-3.60, 4.23)	0.88	239	1521.24
Model 2	0.46 (-3.94, 4.87)	0.84	239	1525.69
Model 3	0.76 (-3.59, 5.10)	0.73	239	1516.33
Model 4	0.70 (-3.79, 5.19)	0.76	239	1524.62
Two-Particle Model 1			239	1521.52
$PM_{2.5}$	0.52 (-0.26, 1.30)	0.19		
BC	-0.52 (-4.68, 3.64)	0.80		
Two-Particle Model 2			239	1526.55
$PM_{2.5}$	0.42 (-0.39, 1.25)	0.30		
BC	-0.32 (-5.05, 4.39)	0.89		
Two-Particle Model 3			239	1517.14
$PM_{2.5}$	0.44 (-0.36, 1.23)	0.28		
BC	-0.04 (-4.62, 4.54)	0.99		
Two-Particle Model 4			239	1525.38
$PM_{2.5}$	0.45 (-0.37, 1.26)	0.28		
BC	-0.12 (-4.80, 4.55)	0.96		

Model 1: adjusted for chronological age and blood cell type.

Model 2: Model 1 but additionally adjusted for temperature, pack years, smoking status, season, BMI, alcohol consumption, and education.

Model 3: Model 2 but additionally adjusted for history of cancer, hypertension, chd, and diabetes.

Model 4: Model 2 but additionally adjusted for statins and medications for diabetes and hypertension.

Table S6. Seasonal Analysis of 1-Year Particulate Matter 2.5 ($PM_{2.5}$) and Black Carbon (BC) as Predictors of DNA Methylation (DNAm) Age

Particle (1 μg/m^3)	Difference in DNAm-age (95% CI)	*P*	N	AIC
$PM_{2.5}$				
Spring	0.43 (-0.14, 1.00)	0.13	249	1560.31
Summer	1.13 (0.54, 1.71)	**0.002**	245	1541.28
Fall	0.63 (0.22, 1.03)	**0.003**	350	2287.70
Winter	0.37 (-0.36, 1.09)	0.29	188	1184.74
BC				
Spring	2.85 (-3.18, 8.88)	0.30	219	1387.22
Summer	2.85 (-7.61, 13.31)	0.36	214	1361.99
Fall	3.09 (-1.68, 7.87)	0.20	293	1930.10
Winter	2.61 (-4.02, 9.24)	0.37	172	1077.01
Two-Particle Spring			219	1386.70
$PM_{2.5}$	0.59 (-0.27, 1.45)	0.14		
BC	1.00 (-5.77, 7.76)	0.73		
Two-Particle Summer			214	1352.99
$PM_{2.5}$	1.06 (-2.82, 4.93)	0.18		
BC	-0.11 (-32.11, 31.90)	0.97		
Two-Particle Fall			293	1928.58
$PM_{2.5}$	0.62 (-0.002, 1.24)	0.05		
BC	0.33 (-5.22, 5.87)	0.91		
Two-Particle Winter			172	1078.58
$PM_{2.5}$	0.25 (-0.81, 1.31)	0.57		
BC	1.68 (-6.42, 9.77)	0.62		

Models adjusted for chronological age, blood cell type, temperature, pack years, smoking status, BMI, alcohol consumption, and education.

Table S7. Pearson Correlation of Change in Particle Exposure and Change in DNAm-age in Participants with Multiple Visits

	Pearson Correlation Coefficient with Change in DNAm-age	*P*	N
Change in Particle			
$PM_{2.5}$	-0.07	0.19	352
BC	0.08	0.17	296

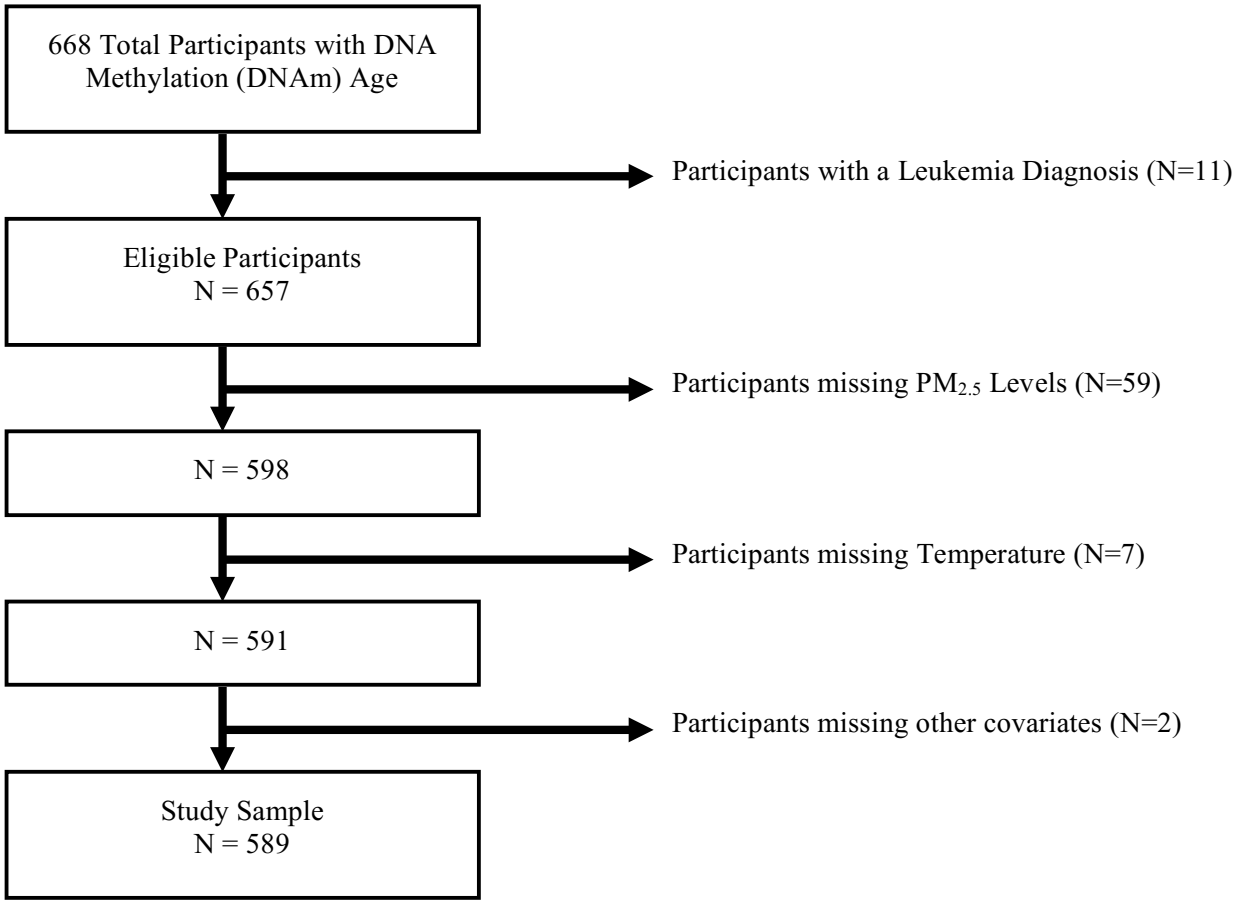

Figure S1 | Eligible and ineligible participants in NAS data (2000-2011). Figure 1 depicts the selection scheme utilized for arriving at the final study sample.

Appendix 2: Chapter 3 Supplementary Data

Table S1. Pearson Correlations of Particulate Matter 2.5 ($PM_{2.5}$) and Component Species Concentrations Across All Study Visits

Particle ($\mu g/m^3$)	EC	OC	Sulfate	Nitrate	Ammonium
OC	0.64				
Sulfate	0.38	0.46			
Nitrate	0.40	0.58	0.25		
Ammonium	0.43	0.49	0.43	0.45	
$PM_{2.5}$	0.62	0.67	0.30	0.46	0.53

All Pearson correlations were significant with $P < 0.0001$

Table S2. Mean 1-Year Particulate Matter 2.5 ($PM_{2.5}$) and Component Species Concentrations Across First Study Visits

Particle ($\mu g/m^3$)	Mean (SD)	IQR	N
$PM_{2.5}$	10.9 (1.41)	1.68	552
$PM_{2.5}$ Component Species			
EC	0.61 (0.18)	0.21	552
OC	3.24 (0.81)	1.00	552
Sulfate	3.87 (1.16)	0.70	552
Nitrate	1.19 (0.27)	0.36	552
Ammonium	1.13 (0.26)	0.19	552

Table S3. 1-Year Particulate Matter 2.5 ($PM_{2.5}$) and Component Species as Joint Predictors of DNA Methylation (DNAm) Age Following LASSO Selection not Accounting for Total $PM_{2.5}$

Particle	Difference in Horvath DNAm-age for IQR (95% CI)	*P*	Difference in Hannum DNAm-age for IQR (95% CI)	*P*	N
Model Framework 4					
$PM_{2.5}$	-0.03 (-0.58, 0.52)	0.97	-	-	940
Sulfate	0.45 (0.21, 0.69)	**0.0003**	-	**-**	940
Ammonium	0.34 (0.002, 0.69)	**0.05**	-	**-**	940
OC	0.42 (-0.17, 1.02)	0.16	-	-	940

Model Framework 4: adjusted for chronological age, blood cell types, temperature, pack years, smoking status, season, BMI, alcohol consumption, and education. No species were selected as predictors of Hannum DNAm-age.

Table S4. Comparison of Estimated Coefficients of $PM_{2.5}$, Sulfate, and Ammonium in the Main Analysis and in the Sensitivity Analysis Where Study Visits with 1-Year $PM_{2.5}$ Levels ≥ 12 µg/m3 were Excluded.

	$PM_{2.5}$		Sulfate		Ammonium	
Analysis (no. of Visits)	**Difference in DNAm-age for IQR (95% CI)**	***P***	**Difference in DNAm-age for IQR (95% CI)**	***P***	**Difference in DNAm-age for IQR (95% CI)**	***P***
Main analysis (n = 940)	0.18 (-0.30, 0.66)	0.45	0.51 (0.28, 0.74)	**<0.0001**	0.36 (0.02, 0.70)	**0.04**
Sensitivity analysis (n = 823)	0.12 (-0.52, 0.75)	0.72	0.50 (0.25, 0.75)	**0.0001**	0.46 (0.06, 0.86)	**0.02**

All models were fully adjusted.

Table S5. Gene Ontology Results from Horvath CpGs Significantly Associated with 1-Year $PM_{2.5}$ Levels

GO ID	GO Term	Genes	FDR Adjusted P
GO:0006446	regulation of translational initiation	RXRA, EIF3M, EIF3I	0.005

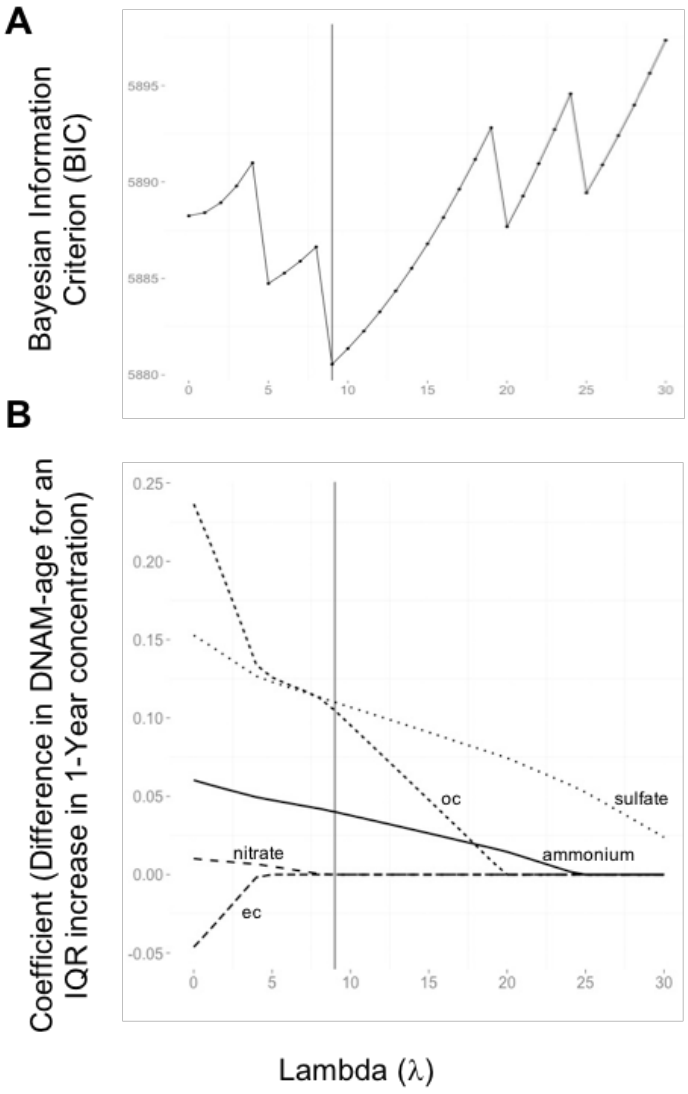

Figure S1 | A) The relationship between BIC, a criterion for model selection and λ (lambda), the adaptive LASSO penalty parameter, for DNAm-age not adjusting for $PM_{2.5}$. The lowest BIC occurs at λ = 9. B) LASSO coefficient paths: plot of coefficient profiles for $PM_{2.5}$ components as a function of λ. The vertical line at λ = 9 denotes the penalty parameter with the lowest BIC. At λ = 9, organic carbon, sulfate, and ammonium are the only $PM_{2.5}$ components with a non-zero coefficient. Again, these models did not adjust for total $PM_{2.5}$ levels.

Appendix 3: Chapter 4 Supplementary Data

Table S1. Candidate Pathway-Specific Genetic Variants

rs number	Gene	Chromosome	Variation	Type
Oxidative Stress				
rs2284367	CAT	11	A/G	Intron
rs1001179	CAT	11	A/G	Promoter
rs2300181	CAT	11	A/G	Intron
rs480575	CAT	11	C/T	Intron
-	HMOX1	22	Short/Long*	Promoter
rs2071746	HMOX1	22	A/T	Promoter
rs5995098	HMOX1	22	C/G	Intron
rs2071749	HMOX1	22	A/G	Intron
rs2071747	HMOX1	22	C/G	coding sequence nonsynonymous
rs1800566	NQO1	16	C/T	coding sequence nonsynonymous
rs1695	GSTP1	11	A/G	coding sequence nonsynonymous
rs1799811	GSTP1	11	Ala/Val	Exon
rs2282679	GC	4	A/C	Intron
rs1155563	GC	4	C/T	Intron
rs2301022	GCLM	1	A/G	Intron
rs3170633	GCLM	1	A/G	3'end
rs4147565	GSTM1	1	Deletion	coding sequence nonsynonymous
-	GSTT1	22	Deletion	coding sequence nonsynonymous
Endothelial Function				
rs12944039	NOS2A	17	A/G	Intron
rs2297516	NOS2A	17	A/C	Intron
rs2072324	NOS2A	17	A/C	Intron
rs2248814	NOS2A	17	A/G	Intron
rs2255929	NOS2A	17	A/T	Intron
rs1137933	NOS2A	17	C/T	coding sequence nonsynonymous
rs1800779	NOS3	7	A/G	Intron
rs1799983	NOS3	7	G/T	coding sequence nonsynonymous
rs2010963	VEGFA	6	C/G	5' Untranslated Region
Metal Processing Genetic Polymorphisms				
rs224572	SLC11A2	12	A/G	Intron
rs422982	SLC11A2	12	A/T	Intron
rs12227734	SLC11A2	12	A/G	Intron
rs11837720	SLC11A2	12	C/G	Intron
rs1005559	SLC11A2	12	A/T	Intron
rs1049296	TF	3	C/T	coding sequence nonsynonymous
rs1799945	HFE	6	C/G	Exon
rs1800562	HFE	6	A/G	Exon
rs1800435	ALAD	9	C/G	coding sequence nonsynonymous

* Short corresponds to less than 25 GT-repeats (0: short/short, 1: short/long, 2: long/long)

Table S2. Pathway-Specific Genetic Variants Selected by the Elastic Net

Oxidative Stress			
rs number	**Variation**	**Type**	**Direction of Model Coefficient**
rs2284367	A/G	HT, HO	+, -
rs1001179	A/G	WT, HO	-, +
rs2300181	A/G	HT, HO	+, -
rs480575	C/T	HT, HO	-, +
rs2071746	A/T	WT, HT	+, -
rs5995098	C/G	WT	+
rs2071749	A/G	HT	-
rs2071747	C/G	WT, HT	+, -
rs1800566	C/T	WT, HO	-, +
rs1799811	Ala/Val	HT, HO	-, +
rs2282679	A/C	WT, HO	+, -
rs2301022	A/G	HO	+
rs3170633	A/G	WT, HO	-, +
Endothelial Function			
rs2248814	A/G	HO	-
rs1137933	C/T	WT	-
rs1800779	A/G	HT	-

WT = Wildtype. HT = Heterozygous for polymorphism. HO = Homozygous for polymorphism.

Table S3. Difference in DNA-age per an IQR Increase in 1-Year Particle Level by Variant Score

Particle Model	Oxidative Stress Cohort (N = 702)			Endothelial Function Cohort (N = 779)		
	Difference in DNAm-age for IQR (95% CI)	***P for Interaction***	**N**	**Difference in DNAm-age for IQR (95% CI)**	***P for Interaction***	**N**
Model 1						
$PM_{2.5}$*Variant Score		0.08			0.52	
Low (< median)	0.20 (-0.40, 0.80)		215	0.14 (-0.44, 0.72)		277
High (≥ median	-0.64 (-1.50, 0.20)		487	0.41 (-0.32, 1.14)		502
Model 2						
Sulfate*Variant Score		0.66			**0.0007**	
Low (< median)	0.57 (0.28, 0.85)		215	0.40 (0.14, 0.67)		277
High (≥ median	0.47 (0.05, 0.90)		487	1.09 (0.70, 1.48)		502
Model 3						
Ammonium*Variant Score		0.17			**0.03**	
Low (< median)	0.56 (0.13, 0.98)		215	0.05 (-0.33, 0.42)		277
High (≥ median	0.18 (-0.26, 0.63)		487	0.64 (0.18, 1.11)		502

All models are adjusted for chronological age, blood cell type, temperature, pack years, smoking status, season, BMI, alcohol consumption, education, lifetime cancer diagnosis, hypertension, diabetes, and coronary heart disease. *Model 1:* Includes an interaction for $PM_{2.5}$ and variant score. *Model 2*: Includes an interaction for sulfate and variant score and is additionally adjusted for $PM_{2.5}$. *Model 3*: Includes an interaction for ammonium and variant score and is additionally adjusted for $PM_{2.5}$.

Table S4. Pearson Correlations of DNAm-age and Serum Endothelial Function Markers

	DNAm-age	Age	ICAM	VCAM
Age	**r = 0.64** ***P* < 0.0001**			
ICAM	**r = 0.13** ***P* = 0.0001**	r = 0.05 *P* = 0.19		
VCAM	**r = 0.25** ***P* < 0.0001**	**r = 0.27** ***P* < 0.0001**	**r = 0.42** ***P* < 0.0001**	
VEGF	r = 0.02 *P* = 0.54	r = -0.005 *P* = 0.90	r = -0.01 *P* = 0.57	r = 0.02 *P* = 0.63

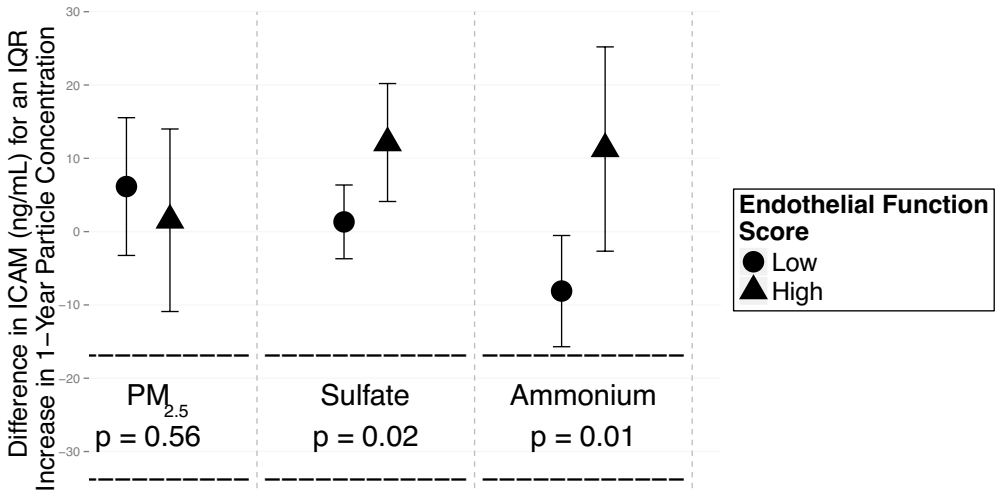

Figure S1 | Difference in Serum ICAM (ng/mL) for one interquartile range increase in 1-year particle exposure according to endothelial function score (low versus high) in the fully-adjusted linear mixed effects model.

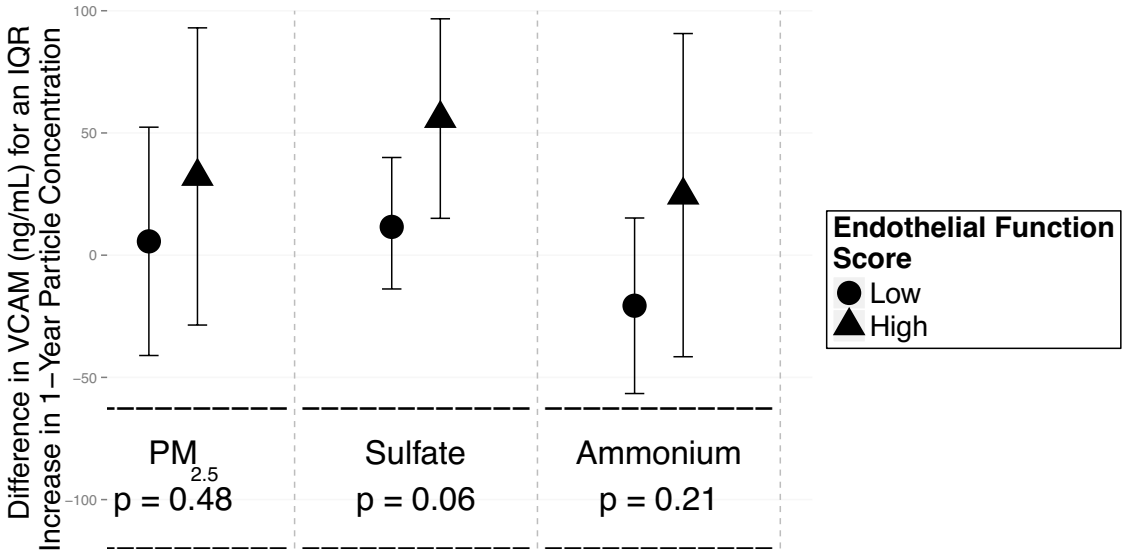

Figure S2 | Difference in Serum VCAM (ng/mL) for one interquartile range increase in 1-year particle exposure according to endothelial function score (low versus high) in the fully-adjusted linear mixed effects model.

Appendix 4: Chapter 5 Supplementary Data

Table S1. Individual Mitochondrial Haplogroups of Study Subjects (2000 – 2011)

Mitochondrial Haplogroup, N (%)	All Visits, N = 870
J	77 (9)
T	69 (8)
V	369 (42)
H	78 (9)
U	99 (11)
K	101 (12)
I	31 (4)
W	20 (2)
X	26 (3)

Table S2. Relationships of Mitochondrial Copy Number with Age and DNAm-age (N=797) – Sensitivity Analysis

Outcome	Difference in Outcome for IQR (95% CI)	*P*
Age	0.49 (-0.31, 1.29)	0.23
DNAm-age	-3.23 (-4.55, -1.92)	<0.0001

All models adjusted for mitochondrial haplogroup, $PM_{2.5}$, sulfate, ammonium, blood cell type, temperature, pack years, smoking status, season, BMI (continuous), alcohol consumption, education, lifetime cancer diagnosis, systolic blood pressure, diastolic blood pressure, fasting blood glucose, total cholesterol and HDL cholesterol. DNAm-age model is also adjusted for chronological age.

Table S3. Estimated Difference in Mitochondrial DNA Copy Number for a One-Year Increase in DNA Methylation Age at the Previous Visit (N = 316)

Model	Difference in Copy Number (95% CI)	*P*
1	-0.005 (-0.013, 0.002)	0.16
2	-0.004 (-0.011, 0.003)	0.29

Model 1 is fully-adjusted for chronological age, blood cell type, temperature, pack years, smoking status, $PM_{2.5}$, season, BMI, alcohol consumption, education, lifetime cancer diagnosis, hypertension, diabetes, and coronary heart disease. Model 2 adjusts for the same covariates as Model 1 in addition to the previous visit's mitochondrial DNA copy number.

Table S4. Mediation Analysis of Mitochondrial DNA Copy Number as a Mediator of the Relationship of $PM_{2.5}$ with DNAm-age (N=797)

Steps	Variable	β (95% CI)	SE	t	*P*	Sobel Z	$P_{mediation}$	% of Effect Mediated by Mediator
Step 1								
Outcome	DNAm-age	-	-	-	-	-	-	-
Predictor	$PM_{2.5}$	0.70 (0.20, 1.20)	0.25	2.78	0.006	-	-	-
Step 2								
Outcome	Copy Number	-	-	-	-	-	-	-
Predictor	$PM_{2.5}$	-0.03 (-0.05, -0.002)	0.01	-2.11	0.04	-	-	-
Step 3								
Outcome	DNAm-age	-	-	-	-	-	-	-
Mediator	Copy Number	-3.41 (-4.74, -2.09)	0.67	-5.07	<0.001	-	-	-
Predictor	$PM_{2.5}$	0.59 (0.10, 1.08)	0.25	2.38	0.02	2.26	0.02	13.6

All steps use fully-adjusted models adjusted for chronological age, blood cell type, temperature, pack years, smoking status, season, BMI, alcohol consumption, education, lifetime cancer diagnosis, hypertension, diabetes, and coronary heart disease.

Table S5. Mediation Analysis of Mitochondrial DNA Copy Number as a Mediator of the Relationship of $PM_{2.5}$ with DNAm-age (N=797) – Continuous Disease Variables

Steps	Variable	β (95% CI)	SE	t	*P*	Sobel Z	$P_{mediation}$	% of Effect Mediated by Mediator
Step 1								
Outcome	DNAm-age	-	-	-	-	-	-	-
Predictor	$PM_{2.5}$	0.70 (0.19, 1.20)	0.26	2.72	0.007	-	-	-
Step 2								
Outcome	Copy Number	-	-	-	-	-	-	-
Predictor	$PM_{2.5}$	-0.03 (-0.05, -0.002)	0.01	-2.12	0.03	-	-	-
Step 3								
Outcome	DNAm-age	-	-	-	-	-	-	-
Mediator	Copy Number	-3.48 (-4.80, -2.15)	0.67	-5.15	<0.001	-	-	-
Predictor	$PM_{2.5}$	0.59 (0.09, 1.09)	0.25	2.33	0.02	2.69	0.01	13.9

All steps use fully-adjusted models adjusted for chronological age, blood cell type, temperature, pack years, smoking status, season, BMI, alcohol consumption, education, lifetime cancer diagnosis, systolic blood pressure, diastolic blood pressure, fasting blood glucose, total cholesterol, and HDL cholesterol.

Table S6. Mitochondrial Haplogroup Clusters as Independent Predictors of Mitochondrial DNA Copy Number (N = 797)

Haplogroup Cluster	Difference in DNAm-age for IQR (95% CI)	*P*
1 (JT)	0.03 (-0.03, 0.08)	0.34
2 (VH)	-0.04 (-0.08, 0.001)	0.06
3 (UK)	0.06 (0.01, 0.11)	0.02
4 (IWX)	-0.05 (-0.12, 0.02)	0.19

All models adjusted for chronological age, blood cell type, temperature, pack years, smoking status, $PM_{2.5}$, season, BMI, alcohol consumption, education, lifetime cancer diagnosis, hypertension, diabetes, and coronary heart disease.

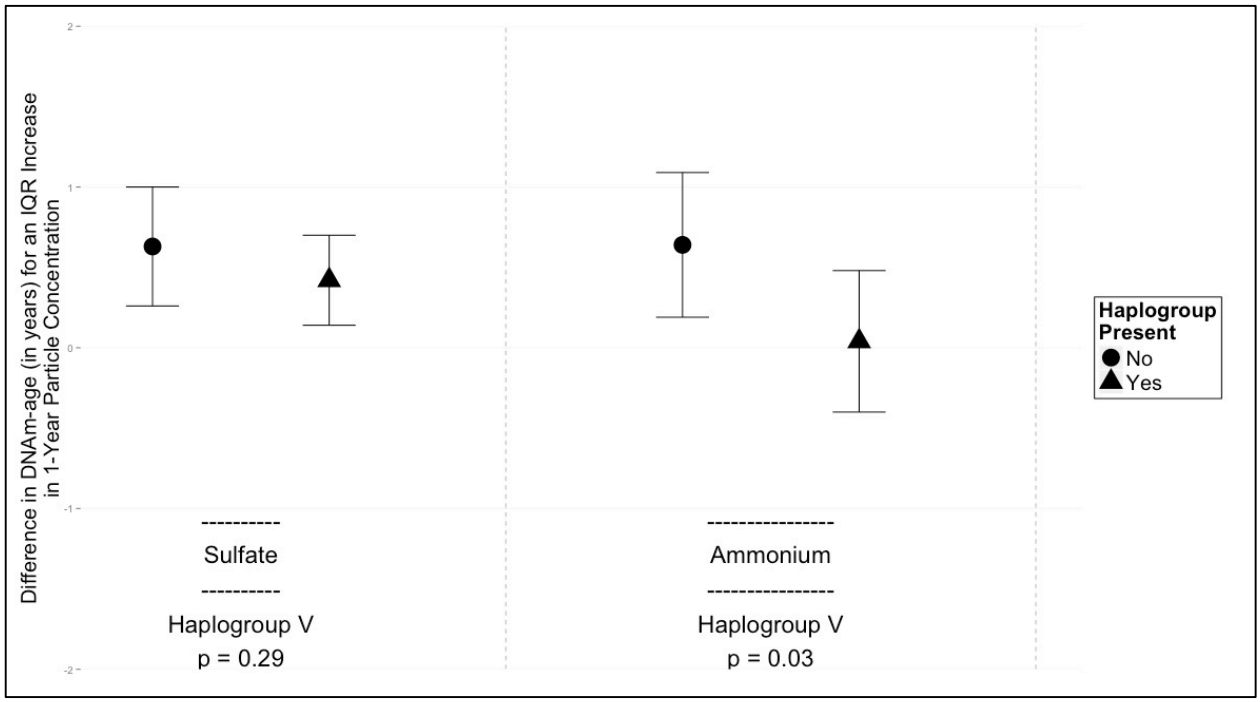

Figure S1 | Difference in DNAm-age for one interquartile range increase in one-year sulfate and ammonium exposure comparing participants with and without the V mitochondrial haplogroup in fully-adjusted models mixed-effects models additionally adjusted for nitrate exposure.

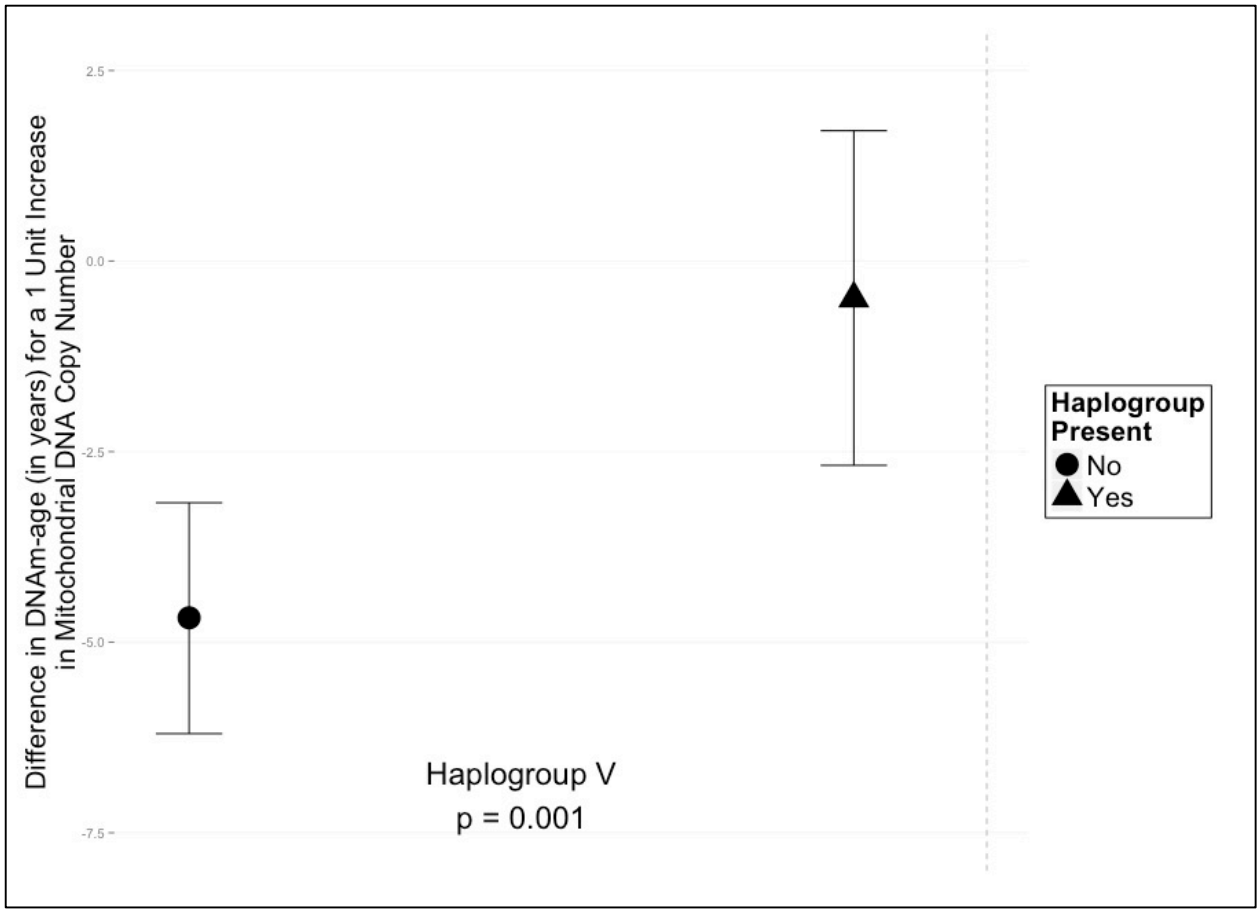

Figure S2 | Difference in DNAm-age for a one-unit increase in mitochondrial DNA copy number comparing participants with and without the V mitochondrial haplogroup in the fully-adjusted mixed-effects model.

Appendix 5: Chapter 6 Supplementary Data

Table S1. Pearson Correlations and Proportion of Particulate Matter 2.5 ($PM_{2.5}$) Mass of $PM_{2.5}$ Component Species Concentrations Across All Study Visits

Particle ($\mu g/m^3$)	Sulfate	Ammonium
Ammonium	0.42	
$PM_{2.5}$ Mass	0.30	0.51
Proportion of $PM_{2.5}$ Mass	33.2%	10.2%

All Pearson correlations were significant with $P < 0.0001$

Table S2. MicroRNA Processing Gene Single Nucleotide Polymorphisms (SNPs) (N=808)

rs number	Gene	Chromosome	Major / Minor Variant	Participants Homozygous Major Variant, N (%)
rs4961280[Ψ]	*AGO2*	8	C/A	526 (65)
rs197412	*DDX20*	1	T/C	321 (40)
rs3757	*DGCR8*	22	G/A	439 (54)
rs1640299	*DGCR8*	22	G/T	233 (29)
rs13078	*DICER1*	14	T/A	540 (67)
rs6877842[Ψ]	*DROSHA*	5	G/C	551 (68)
rs7813	*GEMIN4*	17	C/T	258 (32)
rs910924[Ψ]	*GEMIN4*	17	C/T	403 (50)
rs910925	*GEMIN4*	17	C/G	258 (32)
rs1062923	*GEMIN4*	17	T/C	510 (63)
rs4968104	*GEMIN4*	17	T/A	417 (52)
rs784567[Ψ]	*TARBP2*	12	C/T	189 (23)
rs14035	*RAN*	12	C/T	401 (50)
rs11077	*XPO5*	6	A/C	322 (40)

Note. [Ψ]SNPs selected in DNAm-age Elastic net

Table S3. Mean One-Year Fine Particle ($PM_{2.5}$ less ammonium) Concentrations and MicroRNA Processing Gene Single Nucleotide Polymorphisms (SNPs) as Joint Predictors of DNAm-age (N = 808)

Predictor	Difference in DNAm-age for IQR (95% CI)	*P*
$PM_{2.5}$ (less Ammonium)	0.51 (0.06, 0.97)	**0.03**
Elastic Net Selected miRNA SNPs[a]		
rs4961280 (*AGO2*)	-1.13 (-2.26, -0.004)	**0.05**
rs6877842 (*DROSHA*)	-0.78 (-1.93, 0.37)	0.18
rs910924 (*GEMIN4*)	-0.42 (-1.48, 0.64)	0.43
rs784567 (*TARBP2*)	-1.32 (-2.58, -0.05)	**0.04**

Note. Model adjusted for chronological age, blood cell type, temperature, pack years, smoking status, season, BMI, alcohol consumption, education, lifetime cancer diagnosis, hypertension, diabetes, and ischemic heart disease. [a]Values for the miRNA processing SNPs are in reference to participants whose genotypes are homozygous for the major variant. Bold text specifies statistically significant P values (<0.05).

Table S4. Genemania Curated Gene Interactions

Type of Interaction	Genes Related to *AGO2*	Genes Related to *DROSHA*	Genes Related to *GEMIN4*	Genes Related to *TARBP2*
Physical Interactions				
	IPO8		*LGALS1*	
Co-Expression				
	FXN	*A1BG*	*A1BG*	*BASP1*
	IER2	*C2orf60*	*AGBL5*	*CIB2*
	PUM3	*CCNF*	*ALOX12*	*DDAH2*
	NOP2	*CSPG5*	*BAZ2A*	*DDR1*
	NRAS	*DPP8*	*NRDE2*	*GRWD1*
	PAPOLG	*HCRTR1*	*DEPDC1*	*LRRC61*
	PLK1	*MGC29506*	*EIF3I*	*MLLT6*
	PRPF8	*MN1*	*ELAC2*	*MPI*
	TIPARP	*MRPS21*	*ENPP2*	*MRPL38*
		NR2F2	*FAM3C*	*MYOZ1*
		PAPOLG	*GRIN2C*	*NDUFA3*
		PTPRK	*GRWD1*	*NDUFS5*
		SLC14A1	*IER2*	*SCAP*
		SSRP1	*KCTD9*	*SSRP1*
		UCKL1	*KIAA0020*	*TNP1*
		USP10	*MPI*	
		WFS1	*MPP6*	
		ZMYND15	*NAE1*	
		ZMYND8	*NOP2*	
			PLK1	
			PRPF8	
			SDC2	
			ZBTB5	
Genetic Interactions				
	ABCA3	*AFF1*	*EVA1C*	*ALKBH3*
	ACOT11	*CD164*	*DST*	*ARSG*
	ALOX12	*DPP8*	*IER2*	*C7orf55*
	APOA1BP	*DST*	*KHDRBS2*	*CYFIP1*
	BCCIP	*EPHX2*	*MGP*	*DOLPP1*
	C10orf35	*ERG*	*SDC2*	*ERG*
	CDH1	*GJD4*	*SEC61G*	*FES*
	CHAF1B	*KPNA1*	*SLC20A2*	*KLF2*
	CHI3L2	*LAMA3*	*TNP1*	*MAP3K5*

Table S4. Genemania Curated Gene Interactions (Continued)

	CRADD	*LHCGR*	*MAPKAP1*
	DGKI	*MIB1*	*PAPOLG*
	DPP8	*NAT10*	*RASSF4*
	DST	*PDCD6IP*	*RFC3*
	FAM50B	*PPP1R16B*	*THUMPD3*
	GALC	*RBPMS*	*TXNDC15*
	GAP43	*RSPRY1*	*USP10*
	GLB1	*TIPARP*	*ZNF804B*
	GPR68	*TOM1L1*	
	GRIA2	*WFS1*	
	IL6ST	*ZHX1*	
	KCNC4		
	KLHL35		
	MPP6		
	NAA60		
	NR2F2		
	PAPOLG		
	POMC		
	RAPGEF1		
	REEP1		
	RSPRY1		
	SGCE		
	SLC9A3R2		
	SNRPB2		
	ST3GAL4		
	SYNE1		
	TIPARP		
	TNP1		
	USP10		
	ZBTB16		

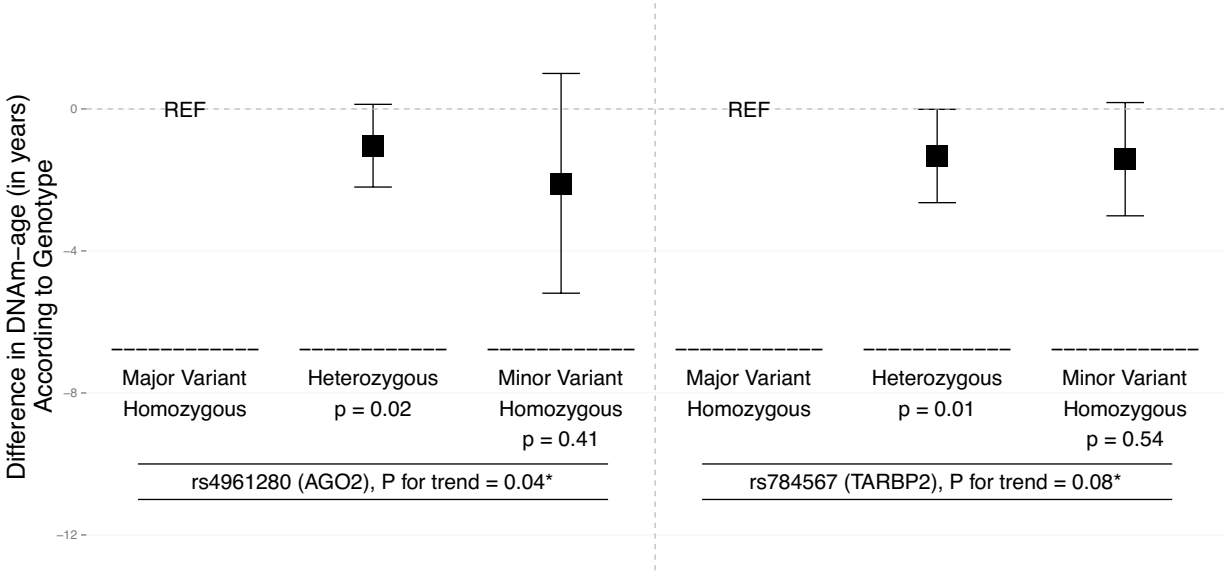

Figure S1 | Difference in DNAm-age according to *AGO2* and *TARBP2* genotypes in fully-adjusted linear mixed-effects models. *AGO2* genotype observation counts: homozygous major variant (N=526), heterozygous (N=257), and homozygous minor variant (N=25). *TARBP2* genotype observation counts: homozygous major variant (N=189), heterozygous (N=434), and homozygous minor variant (N=185). *P value for the test of linear trend across genotypes was based on a linear mixed-effects regression model where the three genotypes for each gene were fit as a continuous measure.

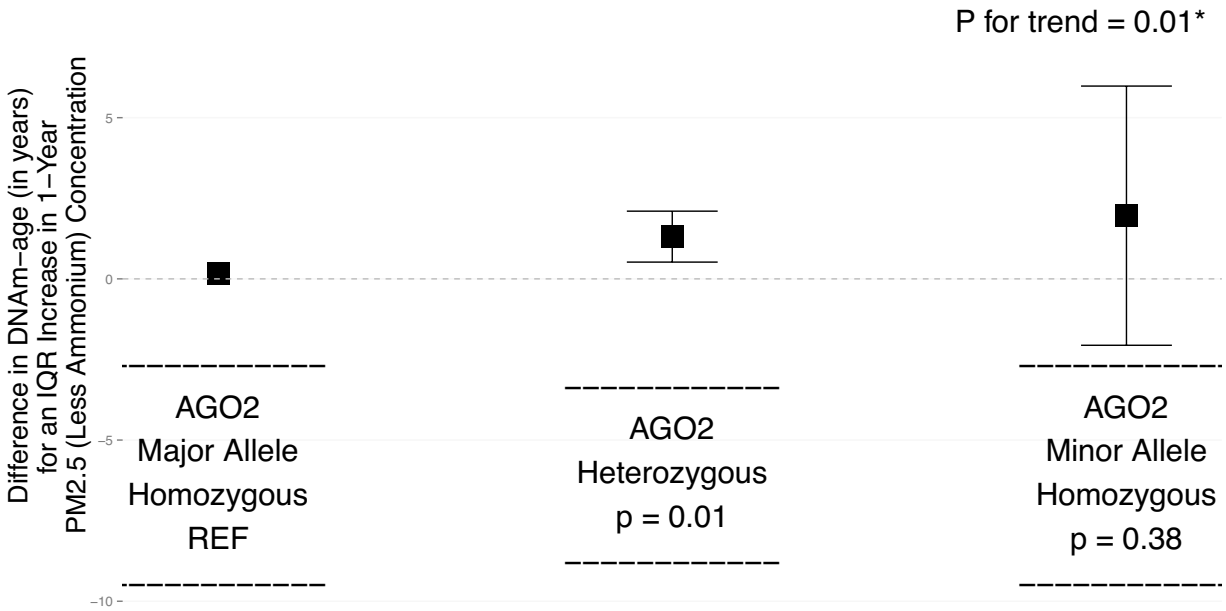

Figure S2 | Difference in DNAm-age for one interquartile range increase in one-year $PM_{2.5}$ (without the ammonium component) levels comparing participants of homozygous major variant (N=526), heterozygous (N=257), and homozygous minor variant genotypes (N=25) for *AGO2* in fully-adjusted linear mixed effects models. *P value for the test of linear trend across genotypes was based on a linear mixed-effects regression model where the three *AGO2* genotypes were fit as a continuous measure.

“Page intentionally left blank”