Unraveling the Role of Epigenetics in Aging and Chronic Disease

by

Alicia Lynn Lazarus Zagel

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiologic Science)
in The University of Michigan
2013

Doctoral Committee:

Professor Sharon L.R. Kardia, Chair
Assistant Research Scientist Lawrence F. Bielak
Assistant Professor Dana Dolinoy
Professor Patricia A. Peyser
Assistant Research Scientist Jennifer A. Smith

Dedication

This dissertation is dedicated to my parents, Linda and Maurie, for supporting me through all of my whimsies, from playing the saxophone to going to grad school. You've always pushed me to be the best woman I can by letting me flourish and follow my dreams. I have an infinite amount of gratitude for the years you've spent guiding me by your living examples of compassion and leadership. You are truly inspirational throughout your calm and excitement.

To my husband, Jonathan, for loving me so selflessly and purely since I first met him. You are one of the most brilliant, kind, and generous people I have ever met, and I'm so blessed to learn from you and to embark on this journey of life with you. Thank you for both celebrating my accomplishments and providing me hugs when I am frustrated or downtrodden. You are so understanding, and I am so glad to have found such a supportive and wonderful man to share my life with. I love you with all of my soul. Go team!

To Sharon for holding my hand when I need it, pushing me out of my comfort zone to teach me lessons, and being such an exceptional role model of strength; to Erin P, Erin W, and Steph for being there with hugs through the hard times and excitement through the great times, and always being the perfect companions for break times; to Jen S. for reminding me that there's a light on the other side, listening to me, and being understanding about this entire process; to Tracy for sharing your years of GENOA knowledge; to Joyce for providing comedic relief and always be willing to lend a hand; to all of the Kardia lab for giving me a helping hand when I needed it and sharing in laughter; and, of course, to my committee (Sharon, Jen, Larry, Dana, and Pat) for sharing their infinite wisdom and always greeting me with open arms.

To my Ami for providing me sunshine and rainbows throughout so many years; to my Lindseebie for always being willing to take my phone calls, whether they're happy or sad – you are a forever friend; to Sauce, Kolby, Ariane, and Sam for many great adventures in Ann Arbor, conversations about life and love, and plenty of delicious fondue and silliness.

To my cousins, aunts, and uncles for being a sea of constant calm and being my cheerleaders since I was born; to all my Minnesota friends for providing me laughter and reminding me where home truly is; to all of my wonderful friends in Michigan, both schoolmates and ultimate players, for you've made Ann Arbor a haven of comfort and fun; and to the rest of my friends around the world for allowing me to share in your experiences.

I love all of you.

Acknowledgements

This dissertation would not have been possible without the help of my advisor, Sharon Kardia, as well as my committee members, Pat Peyser, Larry Bielak, Jen Smith, and Dana Dolinoy, who were very gracious in spreading their wisdom and support during my research endeavors. Further, the GENOA cohort and collaborators, as well as numerous and generous research grants awarded to the GENOA investigators throughout the years, have made this exciting and novel research possible.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Healthy aging is defined as aging in the absence of chronic diseases that limit physical function and mobility. With the increasingly high prevalence of chronic diseases and their risk factors in the United States, it is important to better understand the contributors and predictors of health and longevity. Molecular markers of the epigenome provide one novel set of biomarkers to investigate the intersection of the effects of genetic variation and environmental variation in the initiation and progression of chronic disease. This dissertation focuses on (1) investigating associations between 26,428 DNA methylation markers and chronological age, (2) elucidating the environmental and genetic components of the variation of 26,428 DNA methylation markers, and, finally, (3) predicting aging with DNA methylation markers sites in an African-American population of sibships from the Jackson, MS field center of the Genetic Epidemiology Network of Arteriopathy (GENOA) study. We found that 27% of the genome-wide DNA methylation sites are significantly associated with age. The majority of DNA methylation sites (88%) have significant heritabilities, and there is evidence of an age-related genetic component to this heritability. Finally, when predicting aging, inflammatory biomarkers and DNA methylation markers, together, were found to explain 14% of the variation in aging. This dissertation illustrates that DNA methylation patterns measured in epidemiological studies may be able to provide new insights into the molecular processes underlying aging and chronic disease development.

Chapter I. Introduction

United States population demographics are shifting, and the number of elderly (age 65+) individuals is projected to increase by 135.4% between 2000 and 2050, thus placing a substantial additional burden on public programs for health care and long-term care giving.(*1*) Because of the substantial increase in the senior citizen population, it is important to identify new biomarkers to detect diseases processes at early stages in order to be able to develop interventions that reduce health care expenses and increase well-being before irreversible damage to organ systems has occurred. Healthy aging is characterized by living with "optimal reserve and biological resilience to respond and accommodate daily environmental stressors," meaning an absence of chronic disease risk factors and conditions that diminish physical function.(*2*) Understanding the factors that contribute to healthy and unhealthy aging processes will aid in discovering ways to reduce the burden of chronic disease morbidity.

Chronic diseases, such as hypertension, diabetes, cardiovascular and kidney diseases, are the leading cause of morbidity and mortality in the United States.(*3*) These diseases are a huge economic burden to the United States, with total cardiovascular disease and stroke costs estimated at over $310 billion in direct and indirect costs. Hypertension affects 1 in 3 American adults, with the highest rates among African Americans (44%).(*4*) Hypertension is associated with a wide range of target organ damage to the heart, brain, kidneys, and peripheral arteries resulting in strokes,

cardiovascular events, and end-stage renal disease,(*5*) rates of which are up to 400% in excess within the African American population relative to whites.(*4*)

Over the last several decades, numerous epidemiological studies have been conducted to identify the genetic, environmental, and metabolic factors that predict an individual's or population's risk of chronic diseases such as obesity, hypertension, diabetes, and dyslipidemia which underlie the risk of all vascular diseases.(*6-11*) These epidemiological studies have identified a wide range of lifestyle factors, anthropometric factors, and measures of lipid metabolism, glucose metabolism, blood pressure, and inflammation that now constitute our best predictors of future disease with advancing age.(*12, 13*) All of these risk factors or indicators of disease have significant genetic contributions,(*14-18*) and large international consortia have identified gene regions that have replicated strongly associated relationships with these important risk factors.(*19-21*) Even with the successful discovery of significant genetic contributions to variation in chronic disease risk factors, the amount of variation in chronic disease risk or its major risk factors that can be explained remains modest (e.g. 20-50%).(*15, 17*) As the United States transitions into an unprecedented increase in the number of aging adults over the next few decades, there is an increased need to identify new markers of chronic disease risk that can lead to earlier identification and better treatments.

One new class of biomarkers now available for study in epidemiological cohorts is epigenomic markers (e.g. DNA methylation). The epigenome represents the malleable intersection of genes and environment, wherein environmental factors may modulate gene expression through measureable chemical modifications of the DNA backbone of the genetic code.(*22-24*) Since it is established that chronic diseases have both genetic

and environmental associations,(*12, 13, 15, 17*) epigenetics may help elucidate how environmental signals and genomes interact to alter gene expression profiles underlying disease etiology. Moreover, the study of the epigenome along the age continuum of adulthood may provide more global information about aging and the accumulated effects of healthy and unhealthy lifestyles.(*25*)

The overall goal of this dissertation is to investigate the relationship between variation in DNA methylation and measures of healthy and unhealthy aging. To accomplish this goal, I investigated the relationship between chronological age and 26,428 DNA methylation markers (Aim 1), as well as the role of age-associated genetic factors within the heritability of these DNA methylation markers (Aim 2). Finally, I generated a population-specific prediction model of aging based on traditional and novel risk factors of common chronic diseases to identify a set of DNA methylation markers that best predicts healthy versus unhealthy aging (Aim 3).

Background and Public Health Significance

During the last century, advances in public health and medicine have radically altered the life expectancy of an average United States citizen.(*26*) These health benefits are anticipated to further translate into large shifts in age demographics over the next four decades in the U.S. For instance, while population growth in young adult through middle aged groups (age 16-64) is projected to exhibit increases of 33% between 2000 and 2050, the number of senior citizens (age 65+) is projected to increase from 35 million to 82 million people (135% change) during that time. Additionally, the population aged 85+, which utilizes the most health care and long-term care services, is projected to increase

from 4 million to 19 million (i.e. an increase of 350%). This dramatic increase within senior and elderly populations in the United States is likely to place a substantial additional economic burden on public programs for health and long-term care giving, such as Medicare and Medicaid.(*1*) Although the last several decades of epidemiological studies have identified important risk factors for adult onset diseases with the greatest burden on the public's health, such as heart disease, stroke, diabetes, and hypertension, there is a substantial gap in our understanding of the molecular and cellular processes that characterize the difference between healthy and unhealthy aging processes.

Healthy aging occurs with the absence of chronic diseases and conditions that diminish physical function.(*2*) While genotype appears to contribute 23-50% to life expectancy,(*15, 17*) lifestyle factors, such as dietary choices, exposure to pollutants and toxins, amount of physical activity, and exposure to stress may play a large role in healthy aging.(*27, 28*) As biological risk factors of chronic disease evolve further into pathophysiological disease states, the process of unhealthy aging becomes irreversible.(*2*) It is well-known that the initiation of chronic diseases such as atherosclerosis of the coronary, carotid, or renal arteries often happens decades before a clinical diagnosis. The progression from initial vascular injury to occluded artery often represents pathophysiological processes that evolve from a risk factor into an irreversible state of disease.(*29*) Thus, it is important to understand the processes of healthy and unhealthy aging from their molecular onset through their progression in order to discover where new interventions can occur to alleviate the pain, suffering, and cost of poor health during the aging process. The field of epigenomics offers a unique opportunity to explore new molecular mechanisms that could provide novel insight into aging and disease processes.

Chronic diseases including heart disease, diabetes, and stroke, and their risk factors (e.g., obesity and hypertension) are major players in derailing healthy aging, and they are some of the leading causes of morbidity and mortality in the United States.(*3, 9, 30*) An estimated 85% of senior citizens (age 65+) have at least one chronic condition.(*31*) Over 75% of money spent on health care is spent on individuals with chronic conditions for services such as home health visits, prescriptions, physician visits, and inpatient stays.(*31*) Additionally, 62% of individuals aged 65+ have at least two comorbid chronic conditions, and individuals with these multiple comorbidities are the heaviest users of health care resources.(*31*) Molecular pathways that cascade to create chronic disease states may be important predictors of unhealthy aging processes. These pathways are important to understand and can lead to potential clinical, lifestyle, and pharmaceutical interventions earlier within the disease process to create better disease outcomes, and increase longevity and quality of life over time.

Novel Risk Factors for Chronic Disease

While traditional risk factors for common chronic diseases (e.g. hypertension, dyslipidemia) are important in understanding chronic disease epidemiology and pathophysiology, potential molecular signals that lead toward and/or interact within the processes of these risk factors, such as inflammatory cascades, are beginning to be explored. These molecular signals may be important in the preclinical identification of individuals at risk of developing chronic diseases, which is especially important due to both the dramatic demographic shifts throughout the next half-century that are expected to result in drastic increases in the number of senior citizens (age 65+)(*1*) and the increase in chronic disease prevalence in the United States.(*5, 8, 11, 32-34*) The role of

inflammatory markers has been implicated in some of the major chronic disease risk factors, such as atherosclerosis and type 2 diabetes.(*35, 36*) Further, since DNA methylation is another novel molecular signal that has been found to be associated with a wide range of adverse health outcomes and aging processes,(*37-41*) as well as associated with inflammatory signaling,(*42, 43*) this dissertation will explore the associations of many markers of inflammation within the context of healthy and unhealthy aging, as well as their interaction with DNA methylation of CpG sites. Details of inflammatory and epigenetic biomarkers will be described next, and further details of traditional and novel chronic disease risk factors explored within the course of this dissertation can be found within Appendix 1.

Inflammation. Inflammation occurs in reaction to any type of bodily injury as part of the innate immune response. The inflammatory response involves peripheral blood cells made up of immune cells including monocytes, macrophages, and T-lymphocytes, as well as a number of small molecules such as cytokines, reactive oxygen species, and growth factors. Inflammation is characterized by increased blood flow, elevated cellular metabolism, vasodilation, release of soluble mediators and fluids, and cellular influx. Over time, chronic inflammation can lead to tissue architecture aberrations and, further, to major organ dysfunction.(*44*)

During the aging process, the immune system changes in cellular makeup and reduces in functionality, which increases disease susceptibility. This process is formally known as immunosenescence. The deterioration of immune function increases disease susceptibility, as well as morbidity and mortality rates from infection.(*45*) Biomarkers of immunosenescence include cells involved in innate and humoral immunity, which are

expressed in lymphocytes, and other peripheral blood cells. Lymphocytes are made up of natural killer (NK), T-cells, and B-cells. With older age, naïve T-cells decrease in percentage, while the percentage of memory and effector-memory cells increase. Further, NK-associated receptors show increased expression among older populations.(*46*) Simultaneously, a decrease in B-cell diversity within older ages has also been found to be associated with poor health status.(*47*)

As factors involved in immunity change in makeup with aging, so does the composition of inflammatory markers. Pro-inflammatory mediators such as C-reactive protein (CRP), and cytokines such as interleukin-1 (IL-1), interleukin-6 (IL-6), and tumor necrosis factor-α (TNF-α)are suggested to be chronically up-regulated during the aging process as a backdrop for disease initiation. The elderly exhibit two- to four-fold increases in serum inflammatory markers, which are said to account for the biological mechanisms responsible for decline in physical function and initiation or exacerbated states of inflammatory-related diseases, such as Alzheimer's disease, cardiovascular disease, type 2 diabetes, and sarcopenia.(*46, 48*) Biomarkers of inflammatory response are providing new insight into chronic disease pathophysiology and progression. These factors may allow the ability to detect preclinical symptoms of downstream disease risk.

Furthermore, the process of aging has a molecular component. As centenarians embody a cohort with disease-free aging or delayed disease-onset aging, their offspring also tend to have delayed chronic disease states and markedly reduced prevalence of common chronic diseases, specifically in the context of cardiovascular disease.(*49*) Specifically, polymorphisms in genes involved in immune response and inflammation, such as cytokines, have been suggested to contribute to human longevity.(*50, 51*)

Epigenetic modifications may also be implicated in human longevity and healthy aging phenotypes due to their ability to alter gene expression. Alterations in gene expression within inflammatory pathways may be indicative of how environmental factors play a role in the aging processes.(*46, 51, 52*)

Inflammation has more recently been explored within the causal pathway of common chronic disease processes. In individuals with obesity and type 2 diabetes, chronic overnutrition may propagate oxidative stress and inflammatory changes leading to chronic inflammatory states. Increases in TNF-α and IL-6 due to chronic inflammation may suppress insulin signal transduction, which then promotes chronic inflammation.(*35*) Similarly, the atherosclerotic process, involving plaque formation, growth, and complication over a long period of time, is perceived by the body as injury. Thus, the body initiates an inflammatory response.(*36*)

The inflammatory biomarkers that are used in a prediction model for healthy aging (Aim 3) include 12 measures that are associated with cardiovascular disease and diabetes. These biomarkers include C-reactive protein, fibrinogen, homocysteine, intercellular adhesion molecule, interleukins 6 and 18, monocyte chemotactic protein-1, myeloperoxidase, resistin, serum amyloid A, and tumor necrosis factor receptors 1 and 2. See Appendix 2 for a brief review of the relevance of these markers to cellular processes and aging.

Epigenetic Indicators of Common Chronic Diseases and Aging

Epigenetics is the study of alterations in gene expression caused by biochemical changes that influence DNA and chromatin structure, but do not change the actual DNA

sequence.(*22, 23*) Epigenetic changes are mitotically heritable,(*23*) and are known to occur via DNA methylation, histone modification, chromatin remodeling, and micro-RNA interaction.(*22, 53*) DNA methylation of CpG dinucleotides is the most commonly studied epigenetic mechanism because it is biochemically stable, can be measured using small volumes of DNA, and can be accurately quantified by current biotechnology.(*22*)

Epigenetic modifications begin accruing during fetal development and can continue to change throughout a lifetime. During embryogenesis, the epigenome appears most susceptible to environmental factors. After fertilization, except for within imprinted genes, partially-methylated genomes of the egg and sperm are globally demethylated in order to make the genome available for the developing embryo. Shortly after global demethylation, DNA methylation is reestablished progressively throughout fetal development, as cells are undergoing rapid division, making this a particularly vulnerable developmental stage.(*54*) The epigenome continues to shift throughout the lifetime in response to the accumulation of environment effects and exposures, including diet, pollutants, stress, and other exposures.(*55, 56*) Epigenetic processes are known to link a person's genotype to their expressed phenotype by influencing levels of gene expression.(*57-59*) Since epigenetic changes can alter gene expression, they may begin a cascade of events that result in later-life disease processes, such as cancers and other chronic conditions, as well as more immediate conditions occurring within earlier life stages such as childhood asthma and learning disorders.(*60, 61*)

It is well-known that aging and chronic disease affect cellular mechanisms in a wide range of tissues via epigenetic modifications of the chromosome.(*37-39, 42, 43*) Until recently, the ability to measure these cellular and molecular entities was limited to

small laboratory studies. With the development of new high-throughput chip-based methods of detecting variation in DNA methylation that are both quantitative and highly reproducible,(*62*) we now have the ability to examine how variation in the epigenome is associated with variation in basic human characteristics (e.g. age) and disease risk factors in large epidemiological samples.

Previous studies have found DNA methylation patterns to be associated with age. A landmark study of monozygotic twin pairs has recently demonstrated that genome-wide DNA methylation may result in differing later-life phenotypes, such as chronic disease processes, despite identical genotype, since remarkable epigenetic differences arise between twins as they age.(*40*) Specific pathways in which DNA methylation has been shown to differ across age decades include pathways related to liver development and metabolism,(*39*) inflammation, endothelial function, oxidation,(*41, 63*) and tumor suppression,(*64, 65*) all of which are implicated in chronic disease processes. The relationship between age and epigenetic differences across pathways affecting multiple organ systems indicates that epigenetics may have multiple avenues of influence on health and longevity.

Since epigenetic events provide a modifiable association between a genotype and a resulting phenotype,(*58, 59, 66-70*) unraveling the relationship between epigenetic mechanisms and biological aging processes is crucial to understanding the origins of chronic diseases. Currently, only a few studies have begun to examine the role of DNA methylation in chronic disease etiology. By exploring the relationship between epigenome-wide DNA methylation and age (Aim 1), and the role of age within the heritability of epigenome-wide DNA methylation (Aim 2), we can begin to elucidate the

molecular conversation between age and epigenetics, and evaluate whether novel epigenetic biomarkers will assist in our overall prediction of chronic disease risk and its impact on health aging (Aim 3).

Chapter II. Epigenomic Indicators of Age in African Americans

Introduction

Age is a well-established risk factor for chronic diseases.(*33, 71*) However, the cellular and molecular changes associated with aging processes that are related to chronic disease initiation and progression are not well-understood. While numerous epidemiological studies have identified a wide range of lifestyle factors and clinical indicators that are involved within an individual's or population's risk of chronic diseases and aging,(*6-13*) implications of molecular indicators, such as the amount of genetic variation in longevity and healthy physical aging phenotypes, remains modest (e.g. 20-50%).(*15, 17*) As the United States transitions into an unprecedented increase in the number of aging adults over the next few decades,(*1*) there is an increased need to identify new markers of cellular and molecular changes within aging processes that may pave the road toward earlier identification and better treatments .

Recently, differential DNA methylation patterns that affect gene expression have been shown be associated with aging.(*40*) More specifically, age has been found to be associated with DNA methylation status in pathways related to liver development and metabolism,(*39*) inflammation, endothelial function, oxidation,(*41, 63*) and tumor suppression.(*64, 65*) Since epigenetic events provide a modifiable link between a gene's expression and a resulting phenotype,(*58, 59, 66, 67*) unraveling the relationship between

epigenetic mechanisms and cellular aging processes is crucial to understanding the origins of chronic diseases.

Many previous studies that have investigated the relationship between DNA methylation and aging processes have either focused on specific genomic regions, such as genes in a single biological pathway,(*41, 63*) or have investigated average whole-genome DNA methylation.(*32, 40*) In this study, we use genome-wide DNA methylation information from 26,428 CpG sites in 13,877 genes to investigate the relationship between age and epigenetic variation in 972 African-American adults from the Genetic Epidemiology Network of Arteriopathy (GENOA) study. Further, to investigate the cumulative epigenomic capacity to represent chronological age, we estimated principal components of the measurements of DNA methylation at CpG sites.

Methods

Sample

The Genetic Epidemiology Network of Arteriopathy (GENOA) study is a community-based study investigating the genetics of hypertension and its arteriosclerotic complications in non-Hispanic whites from Rochester, MN and African-Americans from Jackson, MS.(*72*) In the current study, we investigated the relationship between DNA methylation and age in GENOA African-Americans. African-American sibships were recruited such that ≥2 siblings were diagnosed with primary hypertension before the age of 60 years, while other siblings within the sibship were invited to participate independent of hypertension status (N=1,854). The initial examination (Phase I: 1996-1999) included standardized interviews concerning prescription drug usage, cigarette

smoking, physical activity, history of hypertension, diabetes, and cardiovascular disease events; physical examination for blood pressure, height, weight, and waist and hip circumferences; and fasting blood samples for creatinine, total cholesterol, high-density lipoprotein (HDL)-cholesterol, triglycerides, glucose, and insulin. The second examination (Phase II: 2000-2004) included 1,482 participants returning from Phase I, and included re-measurement of interview, physical examination, and blood characteristics, as well as additional measurements of arteriosclerotic target organ damage, including heart, kidney and peripheral artery traits. DNA methylation was quantified on 1,008 Phase II participants using stored blood samples collected during the second examination. Comparisons between Phase II individuals within and outside of the DNA methylation study are available in Supplementary Table 1. Written informed consent was obtained from all subjects and approval was granted by participating Institutional Review Boards. Complete information on the GENOA study population and sample measurements can be found detailed in Appendix 3.

Measurement of DNA methylation

Samples were prepared and DNA methylation was measured according to previously published methods (*73*) which are summarized as follows. For comprehensive information on these measurements, refer to Appendix 4.

Sample Preparation and Methylation Assay. DNA was isolated from peripheral blood leukocytes from GENOA Phase II stored samples, and bisulfite converted with the EZ DNA Methylation Gold Kit (Zymo Research, Orange CA). Bisulfite-converted DNA samples were whole-genome amplified, enzymatically fragmented, and purified, then hybridized to Illumina Infinium HumanMethylation27K BeadChips, which contained

locus-specific DNA oligomers and a set of 56 control probes. The array was then fluorescently stained, scanned using the Illumina BeadXpress reader, and assessed for fluorescence intensities across the methylated and unmethylated bead types at 27,578 CpG sites.

Data processing and methylation quantification: At each CpG site, fluorescent signals were measured from the site-specific **M** (methlyated) and **U** (unmethylated) bead types. The raw fluorescence data from the scanner was processed using Illumina BeadStudio software. To reduce batch and chip effects, the correlation structure among all 56 control probes was evaluated within channel to identify the most parsimonious subset of probes that explained the maximum amount of batch and chip variation across samples (5 probes in the red channel and 8 probes in the green channel; Supplementary Table 2). Quality control standards were employed by linearly regressing the 13 selected probes onto the intensity signals from the methylated and unmethylated bead types separately across each CpG site.

For each individual and across every CpG site, the M-Value was calculated from the **M** and **U** values as a continuous measure of methylation. The M-Value is a commonly used measurement in microarray analysis that was more recently adapted for use in DNA methylation array data due to its ability to equalize the variance across the CpG sites.(*74, 75*) The M-Value for each individual *i* at a single site, *k*, is calculated as: $\text{M-Value}_{ik} = \log_2[(\max(\mathbf{M}_{ik},0) + 1) / (\max(\mathbf{U}_{ik},0) + 1)]$, where a constant is added to prevent large-scale changes caused by small intensity estimation errors.(*75*) While many studies using the Illumina Infinium HumanMethylation27K array present results of the Beta Value, which are continuous variables between 0 and 1 that are calculated from the

measured **M** and **U** intensities and represent the percent methylation at a particular CpG site within an individual, we will present results of the M-Value since we feel that it has more desirable statistical properties which allow for more precise conformation to modeling assumptions. The relationship between Beta and M-Values represents a logit transformation, where $\text{M-Value}_{ik} = \log_2 [\text{Beta Value}_{ik} / (1 - \text{Beta Value}_{ik})]$. Unmethylated M-Values are considered to be < -2.0, methylated M-Values are > +2.0, and semi-methylated M-Values are between -2 and +2.

Before statistical analysis, samples were checked for data quality. From the 1,008 study participants, 7 individuals were removed due to poor bisulfite conversion control efficiency, measured by bisulfite conversion control intensity of < 4,000. An additional 29 individuals were removed from the analysis due to extreme control probe values, assessed as having at least one control probe with a value of greater than 4 standard deviations from its mean value. This resulted in a total sample size of 972 individuals.

For this study, we analyzed only autosomal CpG sites. A total of 58 CpG sites were removed from the analysis because they were found to be multimodal based on the Dip Test proposed by Hartigan and Hartigan (*76*) using a cut-off of $p<0.001$ on the signal intensities of the methylated and/or unmethylated bead types, which clearly violated the statistical modeling assumptions. This resulted in a total number of 26,428 CpG sites included in our analysis. Finally, 2,984 non-specific binding probes and 908 polymorphic probes overlapping with single nucleotide polymorphisms (SNPs) (*77*) were identified and denoted in the result tables. Though these sites were not removed from the analysis, we have interpreted the results from these sites with caution. As a final method of quality control, DNA methylation values greater than ±4 standard deviation from each

M-Value's mean were removed from each of the 26,428 CpG sites. (See Appendix 4 for more information.)

Statistical Analyses

Linear mixed modeling. Using the *nlme* package in the statistical package R,(*78*) we used a linear mixed modeling approach to evaluate the cross-sectional associations between epigenetic variation and age variation, while accounting for the familial relationships among study participants. In order to examine the association between DNA methylation and age, we considered each of the 26,428 individual CpG sites separately as outcomes, with participant age as a covariate: $E_{ijk} = \beta_0 + \beta_1 \text{Age}_{ij0} + W_{0jk} + \varepsilon_{ijk}$, for participant *i* in sibship *j* at CpG site *k*. Age_{ijk} represents participant age at Phase II exam, E_{ijk} is the value of an M-Value, and W_{0jk} is the random effect for each sibship. After performing this modeling, 4 CpG sites exhibited convergence issues when M-Values were used as the outcome. These non-converging sites were subsequently removed from the analysis.

Due to the strong associations between age and many of the CpG sites that we observed after performing the modeling described above, we wanted to assess the joint effects of CpG sites with age. To assess this, we first used a second set of models to evaluate how well each of the DNA methylation markers predicted age. This modeling considered age as an outcome and tested for association with each of the 26,428 CpG sites individually as covariates in a linear mixed model: $\text{Age}_{ijk} = \beta_0 + \beta_1 E_{ijk} + W_{0jk} + \varepsilon_{ijk}$, for participant *i* in sibship *j* at the k^{th} M-Value. For both modeling strategies, the

Bonferroni method was used to assess experiment-wise statistical significance of the *p*-values and was 1.89×10^{-6} at a significance level of $\alpha = 0.05$.

In order to better understand the joint effects and correlation structure of the large number of CpG sites associated with age, we performed principal component (PC) analysis. The top 5 PCs were calculated using all CpG site M-Values that were significantly associated with age at 1.89×10^{-6}. Next, the top 5 PCs for the significant M-Values were modeled separately in univariate models with age, with each of the *c* PCs as a single predictor, such that $Age_{ijc} = \beta_0 + \beta_1 \cdot PC_{ijc} + W_{0jc} + \varepsilon_{ijc}$, and then within a multivariable mixed model, $Age_{ij} = \beta_0 + \beta_1 \cdot PC1_{ij} + \beta_2 \cdot PC2_{ij} + \beta_3 \cdot PC3_{ij} + \beta_4 \cdot PC4_{ij} + \beta_5 \cdot PC5_{ij} + W_{0j} + \varepsilon_{ij}$, for participant *i* in sibship *j*. R^2 values based on likelihood ratio models (R^2_{LR}) were calculated for linear mixed models using the R package *lmmfit.(79)*

Results

Description of data. After exclusions, this study contained phenotype and methylation data for 972 African Americans within 296 sibships and including 197 singletons across 26,428 CpG sites. The sample was predominantly female (70.7%) and hypertensive (82.5%), with mean age of 66.3 years and mean body mass index of 31.2 kg/m^2. Further descriptive statistics are presented in Table 1.

Within this population, the 26,428 CpG sites were predominantly unmethylated, where 15,221 (57.6%) CpG sites had a mean M-Value of <-2.0. Across all CpG sites, the mean fluorescence intensities across the methylated bead types ranged from 482 to 39,810 (mean = 28,25), while mean fluorescence intensities on the unmethylated bead types ranged from 497 to 37,310 (mean= 6,865). CpG site Beta Value means ranged

from 0.025 to 0.97, with an average mean Beta Value of 0.31, and M-Value means ranged from -5.37 to 5.07 with an average mean M-Value of -1.58. Due to outlier removal, the mean number of individuals measured across the 26,428 CpG sites is 970.9 (standard deviation: 1.7), with a minimum of 935. (Figure 1).

Age Associations with CpG sites. Upon modeling age as a predictor of methylation, age was statistically significantly associated with 7,601 (28.8%) CpG site M-Values (after Bonferroni correction for α=0.05). Of the sites significant in the association between age and M-Values, 671 (8.8%) contained non-specific binding probes, 159 (2.1%) contained polymorphic probes, and 9 sites (0.12%) were indicated to have both non-specific binding and polymorphic probes, as defined by Chen, et al. (2011). Table 2 shows the 30 CpG sites with the most significant associations between age and M-Value. A striking finding of this analysis is that age has an inverse association with all but two of the top 30 CpG sites, as shown in Table 2. In fact, within the 7,601 CpG site M-Values that were significantly predicted by age (after Bonferroni correction for α=0.05), 309 (4.1%) had positive estimates for the regression coefficient for age. Figure 2 shows the relationship between the average M-Value at each site and the t-statistic corresponding to the regression coefficient for age from the linear mixed model. Of the 7,601 CpG sites with which age was significant, 7,292 (95.9%) have negative t-statistics, and 309(4.1%) have positive t-statistics. Further, within the 7,601 CpG sites with which age was significant, negative t-statistics were observed in 5,589 (73.5%) unmethylated sites, 1,675 (22.0%) semimethylated sites, and 28 (0.37%) methylated sites, while positive t-statistics were observed in 34 (0.45%) unmethylated sites, 106 (1.4%) semimethylated sites, and 169 (2.2%) methylated sites.

Given the very large number of highly significant age associations with DNA methylation at CpG sites, we became interested in how well the DNA methylation markers could predict age. In the second set of models which examined CpG site M-Values as predictors of age, 2,095 (7.9%) sites were significant predictors of age after Bonferroni correction at α=0.05. Nearly all (2,086, 99.6%) of these 2,095 CpG sites were also significant (after Bonferroni correction at α=0.05) in the previously evaluated regression of M-Values on age, and had the same direction of effect. The other 9 CpG sites within this set of 2,095 sites all were positive in direction of effect, and their p-value within the previously reported association (where CpG M-Value was the outcome) was $<10^{-5}$.

Principal component (PC) analysis reduces the dimensionality of data by finding orthogonal linear combinations of the predictors that fit the most variation in the data. This method is helpful to condense data with many outcomes that are potentially correlated. PCs were estimated based on the 2,095 CpG sites found to be significant (after Bonferroni correction at α=0.05) in the association of age on M-Value, to examine initial features of the multivariable distribution of significant epigenetic predictors of age (Table 3). The top 5 PCs for the M-Value analysis accounted for 69.3% of the variability in the 2,095 significant DNA methylation sites, and the next 5 PCs accounted for 4.7% more of the variability in these sites. When the top 5 PCs were used as predictors of age, they explained 26.8% of the variation in age, while the top 10 PCs explained 36.5% of the variation in age.

Discussion

Our findings suggest that age and DNA methylation are very strongly associated at many CpG sites across the genome. Notably, we found that age is significantly associated with 7,601 (28.8%) of the CpG sites as measured by the M-Value. The associations between the methylation markers are so ubiquitous and strong across the age spectrum that we hypothesize that DNA methylation patterns are an important measure of cellular aging that underlies the association between chronic disease and chronological age.

Consistent with previous studies in humans and other vertebrates,(*80-82*) we found that the majority of CpG sites tended to become less methylated with age. However, observing the t-statistics of the association between DNA methylation M-Values and age (Figure 2) shows that this is not always the case. The t-statistic on the y-axis provides two types of information: a) the magnitude of the association with age, and b) the direction of the association with age. For example, a t-statistic of -5.0 represents a *p*-value=$5x10^{-7}$ and indicates that as age increases, methylation decreases. The increased density of negative t-statistics for unmethylated markers (M-Values < -2) indicates that these unmethylated markers are increasingly *less methylated* with older age (i.e., the regression coefficient for age is negative). In contrast, the increased density of positive t-statistics for methylated markers (M-Values > +2) indicates that these methylated markers are increasingly *more methylated* with older age. Semimethylated markers (M-Values between -2 and +2) show the most significant decreases in methylation with age. These changes in methylation may contribute to chronic diseases through a variety of mechanisms. For example, it has been found that loss of methylation in CpG

dinucleotides over time may activate silenced retrotransposons and lead to genomic instability,(*83, 84*) while increases in methylation at CpG dinucleotides may prevent the binding of transcription factors and potentially suppress gene expression.(*85*)

Within our data, principal components of the M-Values of the 2,095 CpG sites most strongly associated with age showed that 69.3% of the total variation within these sites can be explained by the top 5 PCs, and only an additional 4.7% more can be explained further by PCs 6 through 10. Thus, 10 principal components explain the majority (74.0%) of the variation within all 2,095 CpG sites, indicating a highly correlated set of epigenetic biomarkers. Further, the top 5 PCs, together, are able to explain 26.8% of the variation within chronological age. Because previous research has indicated that DNA methylation is a molecular representation of the environment and our findings indicate that the joint effects of 2,095 CpG sites are able to explain a moderate amount of the variation within age, future research may be interested to determine the genetic and environmental components within the epigenetic processes that may contribute to aging processes, and may discern important epigenetic contributions to the aging processes.

Several other studies have examined the association between age and DNA methylation using the same Illumina Infinium HumanMethylation27K microarray platform that was used in this study. For example, a study of 34 monozygotic twin pairs aged 21-55 years found 88 CpG sites to be correlated with age.(*86*) Of these significant sites, 87 were included within GENOA analysis, and, of those, 77 (88.5%) sites replicated the direction of effect within the GENOA study. Further, of sites replicating direction of effect, 65 (84.4%) CpG sites were significant at α=0.05 without Bonferroni

correction, and 33 (42.9%) sites were significant after Bonferroni correction at α=0.05 within the GENOA study.

Secondly, a pediatric population of 398 healthy males aged 3 through 17 found 2,078 site Beta Values to be associated with age, where 477 sites had a positive association and 1,601 sites had a negative association with age.(*87*) Of the 2,078 significant sites, the GENOA study analyzed 2,022 sites and found 1,529 (73.6%) significant at α=0.05, and 904 (43.5%) were significant after Bonferroni correction at α=0.05. Further, 1,465 (72.5%) CpG sites replicated the direction of association in our study and were significant at α=0.05, and 899 (44.5%) replicated the direction of effect in our study and were significant after Bonferroni correction at α=0.05. Of the 1,465 CpG sites significant at α=0.05 and maintaining directional agreement in the GENOA sample, 1,240 (84.6%) were less methylated with increasing age in both studies, and 225 (15.4%) were more methylated with increasing age.

Another study partitioned age into three categories (fetal, where N=30; childhood (age 0-10 years), where N=15; and beyond childhood (age > 10 years), where N=63) to assess methylation patterns throughout different phases of development.(*88*) The authors found 868 CpG sites significant with age in the fetal period, 5,506 sites significant with age during childhood, and 10,578 sites significant with age beyond childhood. We compared our results to their top 99 sites from each age category that were listed in supplementary tables. GENOA evaluated 96 of the reported top 99 most significant CpG sites in the fetal period, and we found 67 (69.8%) of them to be significant at α=0.05, and 27 (28.1%) to be significant after Bonferroni correction of α=0.05. A total of 24 of these 96 sites replicated in direction of association, while 12 (50.0%) were also significant in

our sample at α=0.05, and 6 (25.0%) were significant after Bonferroni correction. GENOA found 57 of their top 99 (57.6%) sites significant with age within the childhood (age 0-10) period significant at α=0.05, and 23 (23.2%) significant after Bonferroni correction. Seventy-six (76.8%) of the top 99 sites were replicated in direction of effect in GENOA, while 49 of these (64.5%) were also significant at α=0.05, and 15 (19.7%) were significant after Bonferroni correction. Finally, GENOA replicated 65 of the top 99 (65.7%) significant sites in the beyond childhood (age > 10) period at α=0.05, and 31 (31.3%) after Bonferroni correction. Of the top 99 sites, GENOA replicated the direction of effect of 94 (95.0%) sites, and, of these, 63 (67.0%) were significant at α=0.05, and 10 (10.6%) were significant after Bonferroni correction.

Finally, we compared our results to a study by Teschendorff, et al., which examined the association between age and DNA methylation within polycomb group protein target genes.(*89*) Within the 589 sites they found to be associated with age in a case-control study of ovarian cancer and DNA methylation, 583 sites were analyzed within the GENOA sample, and 516 of those (87.6%) were significant at α=0.05, while 361 (61.3%) were significant after Bonferroni correction for α=0.05. Of the 583 sites analyzed within GENOA, 539 (91.5%) indicated directional agreement with the GENOA sample, whereas 499 (92.6%) were significant at α=0.05, and 361 (67.0%) were significant after Bonferroni correction. Further, of the 499 sites significant in GENOA at α=0.05 that indicated directional agreement with respect to their association with age, 153 (30.7%) were positive, and 346 (69.3%) were negative. This group of genes indicates a much larger proportion of positive significant associations than within the

overall genome-wide analysis in GENOA, likely due to the group of genes selected for study by Teschendorff, et al.

Additional studies have found significant associations between age and DNA methylation within specific chronic disease-related genetic pathways using alternative methods of methylation measurement (e.g. PCR-based methods). Promoter regions of genes involved in liver development and metabolism, inflammation, endothelial function, and oxidation such as *INS*, *KCNQ1OT1*, and *IGF2*,(*90*) *iNOS, TLR2,* and *GCR*,(*41*) and *TNF* (*63*) were found to have significant decreases in methylation with increasing age. Our data supports significant age-associated decreases in methylation of *INS* (all 4 CpG sites in GENOA demonstrate age-related decreases in methylation, with 3 sites significant ranging from 1.96×10^{-12} to 2.64×10^{-3}), *IGF2* (2 of 5 CpG sites measured in GENOA indicated significant decreases, $p = 5.09 \times 10^{-7}$ and 2.55×10^{-4}), and *TNF* (both CpG sites measured in GENOA highly significant; cg04425624, $p=3.63 \times 10^{-12}$; cg11484872, $p=2.58 \times 10^{-9}$). Methylation at *KCNQ1OT1*, which is regulated by an imprinting control region in *KCNQ1*, was not measured in this sample, however methylation at *KCNQ1* was measured in the GENOA sample and indicated mixed results, with 3 CpG sites showing statistically significant increases with age and 9 CpG sites indicating statistically significant decreases with age. CpG sites in GENOA measured in *iNOS, TLR2,* and *GCR* did not exhibit significant results within the CpG sites measured in GENOA.

Conversely, promoter regions of inflammatory genes, such as *LEP*, *ABCA1*, and *GNASAS*,(*90*) and *IFNγ*, *F3*, *CRAT* and *OGG* (*41*) have been found to have increases in DNA methylation with increasing age. Our data supports significant age-associated

increases in methylation of *LEP* (both CpG sites in GENOA significant at $p=7.42x10^{-3}$ and p=0.022). *GNASAS* was not measured within our study, as it is encoded on the minus-strand of chromosome 20q13.32, though *GNAS*, which slightly overlaps *GNASAS* and is encoded on the plus-strand of chromosome 20q13.3, indicated that 6 of the 29 CpG sites within the *GNAS* region measured within GENOA that exhibited significant ($p<0.05$) age-related increases, while 11 CpG sites indicated significant age-related decreases. Neither *ABCA1* nor *IFNγ* had CpG sites that reached significance within GENOA, but *F3*, *CRAT*, and *OGG* actually indicated significant decreases within GENOA. These differences in results may be due to the location of the measured methylation site, since not all of the CpG sites in our study are located in regions with large numbers of CpG sites, or CpG islands. In addition, different methylation sites within the same gene may exhibit contrasting relationships with age.

There are a number of factors that may lead to the differences in the significance of CpG sites with age observed between previous studies and our own. Tissue samples from which DNA was extracted and analyzed can create differing methylation profiles. For instance, Bocklandt, et al.(*86*) extracted DNA from saliva samples, while Numata, et al.(*88*) extracted DNA from dorsolateral prefrontal cortex tissue, both of which may exhibit differing methylation from our sample of peripheral blood cells. Also, population demographics may create differing epigenetic profiles between other studies and the current study, such as the investigation of much younger(*86-88*) or older(*41*, *90*) populations than our own. Since the GENOA cohort is an older African American population with a high prevalence of hypertension, it is likely that the differing chronic disease indicators create a unique DNA methylation signature. Further, differing

statistical techniques, such as the use of correlation by Bocklandt, et al.(*86*) versus linear mixed modeling in our study, creates differences in assessment of significance levels of CpG sites. Sample sizes also will create discrepancies between statistically significant findings. As a more extreme comparison, the sample size within each of the developmental stages was very small in the study by Numata, et al.(*88*), with 30 fetal participants, 15 children, and 63 participants ranging in age from 10 to 84, with only 38 participants falling into the same age group as our sample of 972 individuals (ages 39-95 years). Of note, our sample was much larger (N=972) than all other studies we discussed.

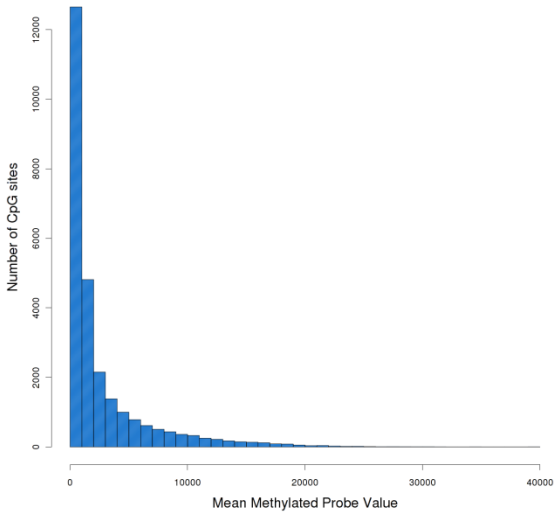
Since we assessed a cell population of peripheral blood leukocytes that consists largely of neutrophils (40-75%) and lymphocytes (16-48%),(*91*) we hypothesize that we may be exploring the action of these cell types in promoting chronic inflammation, such as by promoting the formation of atherosclerotic plaques in our predominantly hypertensive population. This may result in different methylation patterns than in other cell types. However, despite differences in demographics and health status, tissue sample types, statistical techniques, and sample size between our study and previously reported studies, it is important to recognize that 1) there are many unique epigenetic factors that are associated with age across a variety of studies, and that, 2) when unified, these studies may indicate groupings of CpG sites that are important indicators of age and developmental stage across a variety of populations.

This study indicates that the relationship between age and epigenome-wide DNA methylation levels is intricately intertwined, implicating involvement from across the genome during the aging process. Since the GENOA sample is a predominantly hypertensive population, we can hypothesize that chronic disease processes, such as

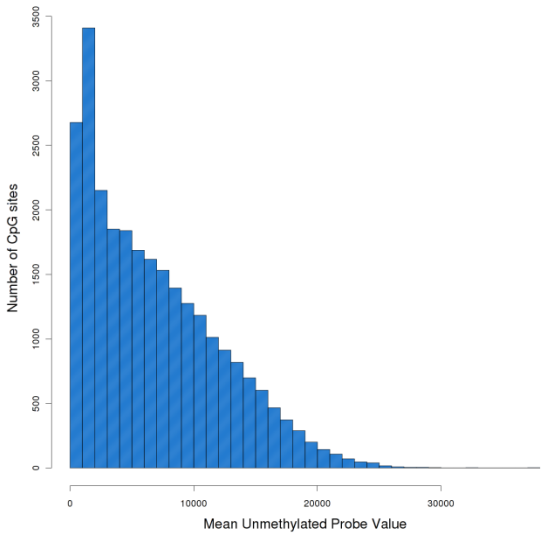
inflammatory pathways that lead to elevated blood pressure, may also be contributing to the association between age and epigenetic variation in this population. Further studies within this cohort will help to decipher the role of aging along with genetic and environmental components that are contributing to variation in DNA methylation. Since the epigenome is a mediator of environmental and genetic effects, as we look toward a more thorough understanding of healthy aging during the later stages of life, it may provide a new set of biomarkers of health that embody genome, environment, and health outcomes.

Figure 1. Distribution of means across 26,428 markers of methylation

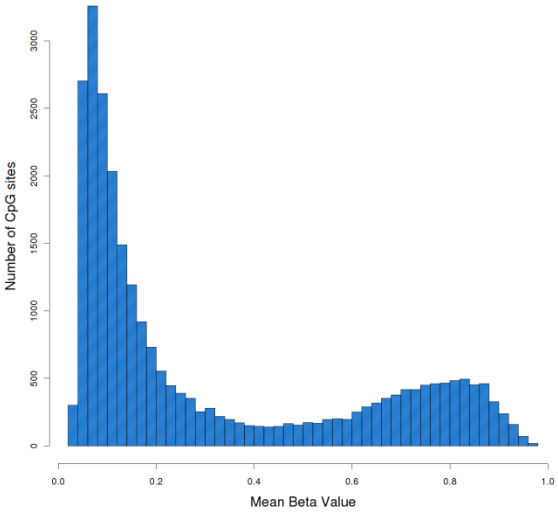
A) Methylated probe signal intensities (Range: 482 to 39,810, mean = 2,825)



B) Unmethylated probe signal intensities (Range: 497 to 37,310, mean = 6,865)



C) Beta Values (Range: 0.02 to 0.97, mean = 0.31)



D) M-Value (Range: -5.37 to 5.07, mean = -1.58)

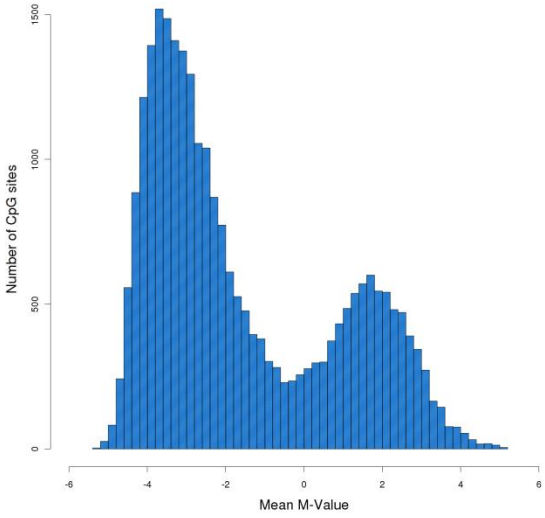


Table 1. Baseline characteristics of GENOA participants after outlier removal: A) Continuous variables, and B) Categorical variables

A) Continuous Variables

	N	Count of Outliers Removed	Count of Missing Values	Range	Mean (SD)
Age, years	972	0	0	39-95	66 (8)
BMI, kg/m^2	965	0	7	16.4-55.1	31.1 (6.1)
Systolic BP, mm Hg	970	0	2	79-221	140(21)
Diastolic BP, mm Hg	972	0	0	45-121	78 (11)
Pulse Pressure, mm Hg	971	0	1	26-127	62 (18)
Systolic BP (Adj), mm Hg**	970	0	2	89-231	148 (23)
Diastolic BP (Adj), mm Hg**	972	0	0	50-126	82 (11)
Pulse Pressure (Adj), mm Hg**	969	0	3	28-132	66 (18)
Total cholesterol, mg/dL	972	0	0	73.5-354.5	203.7 (42.1)
Triglycerides, mg/dL	963	5	4	37-345	116.6 (53.8)
HDL cholesterol, mg/dL	967	2	3	21.7-122.25	57.9 (17.1)
LDL cholesterol, mg/dL	972	0	0	24.85-272.1	123.6 (39.7)
Glucose, mg/dL	951	11	10	49.5-245	108.6 (29.6)
Insulin, mU/mL	953	16	3	0.22-58.29	9.23 (8.25)
Serum creatinine, mg/dL	961	11	0	0.42-2.16	0.92 (0.25)

** If a participant is listed as taking hypertensive medications, 10 mm Hg is added to systolic blood pressure, and 5 mm Hg is added to diastolic blood pressure, else recorded blood pressure is used in this variable.

B) Categorical Variables

	Count	Total	Percent
Female sex	687	972	70.7%
Ever smoker	266	909	29.3%
Hypertensive	802	972	82.5%
Diabetic	298	972	30.7%

Table 2. Top 30 methylation sites most strongly associated with age

Outcome	Chr	Gene	Mean (SD) M-Value	Probe Type**	N	β(Age)	p-value
cg19761273	17	CSNK1D	-1.98 (0.3)	0	972	-0.018	8.45E-43
cg15538427	11	LOC221091	-0.11 (0.22)	0	969	-0.013	3.24E-40
cg01820374	12	LAG3	-0.67 (0.31)	0	970	-0.016	6.23E-33
cg17471102	19	FUT3	0.67 (0.29)	0	969	-0.015	1.64E-31
cg15804973	6	MAP3K5	-0.63 (0.34)	0	972	-0.017	1.14E-30
cg03996822	4	RASSF6	-0.21 (0.33)	0	972	-0.016	2.67E-29
cg25538571	8	FLJ46365	-0.67 (0.31)	1	972	-0.015	7.08E-29
cg00451635	16	EMP2	0.62 (0.33)	0	969	-0.016	2.34E-28
cg19722847	12	IPO8	-1.78 (0.32)	0	971	-0.015	8.15E-28
cg14244577	16	DDX19B	-1.7 (0.28)	0	971	-0.013	8.99E-28
cg08888956	12	NTS	0.04 (0.27)	0	972	-0.013	2.13E-27
cg05442902	22	P2RXL1	-1.71 (0.25)	0	971	-0.012	3.53E-27
cg17034109	1	CYB561D1	0.16 (0.25)	0	971	-0.011	9.10E-27
cg16744741	4	PRKG2	-0.46 (0.35)	0	972	-0.016	5.44E-26
cg15037004	5	ZNF366	-0.15 (0.23)	0	970	-0.011	9.61E-26
cg00431114	20	C20orf121	-1.02 (0.27)	0	972	-0.013	1.60E-25
cg22736354	6	NHLRC1	-1.6 (0.39)	0	972	0.018	2.00E-25
cg00168942	10	CX40.1	0.05 (0.26)	0	971	-0.012	4.31E-25
cg07158339	9	FXN	-1.19 (0.32)	0	972	-0.014	5.93E-25
cg04474832	3	ABHD14A	-1.72 (0.28)	0	972	-0.013	6.08E-25
cg27015931	16	MGC50721	-2.72 (0.29)	0	971	-0.013	6.75E-25
cg04662594	8	EPB49	-0.81 (0.38)	1	972	-0.017	2.75E-24
cg03172991	19	NFIX	0.53 (0.16)	0	970	-0.0073	3.86E-24
cg08587542	5	KIAA0141	-2.42 (0.28)	0	971	-0.012	4.75E-24
cg05724065	7	PHKG1	1.52 (0.28)	0	970	-0.012	5.03E-24
cg08090640	17	IFI35	-1.15 (0.36)	0	971	-0.016	6.62E-24
cg21232015	12	CHFR	2.49 (0.32)	0	970	0.014	8.36E-24
cg08319238	19	BCAM	-1.97 (0.24)	0	970	-0.011	8.66E-24
cg09706243	11	POLD4	-0.97 (0.27)	0	969	-0.012	1.35E-23
cg03143849	11	CDKN1C	-0.22 (0.24)	0	970	-0.010	1.47E-23

Model: $E_{ij} = \beta_0 + \beta_1 \cdot Age_{ij} + W_{0j}$

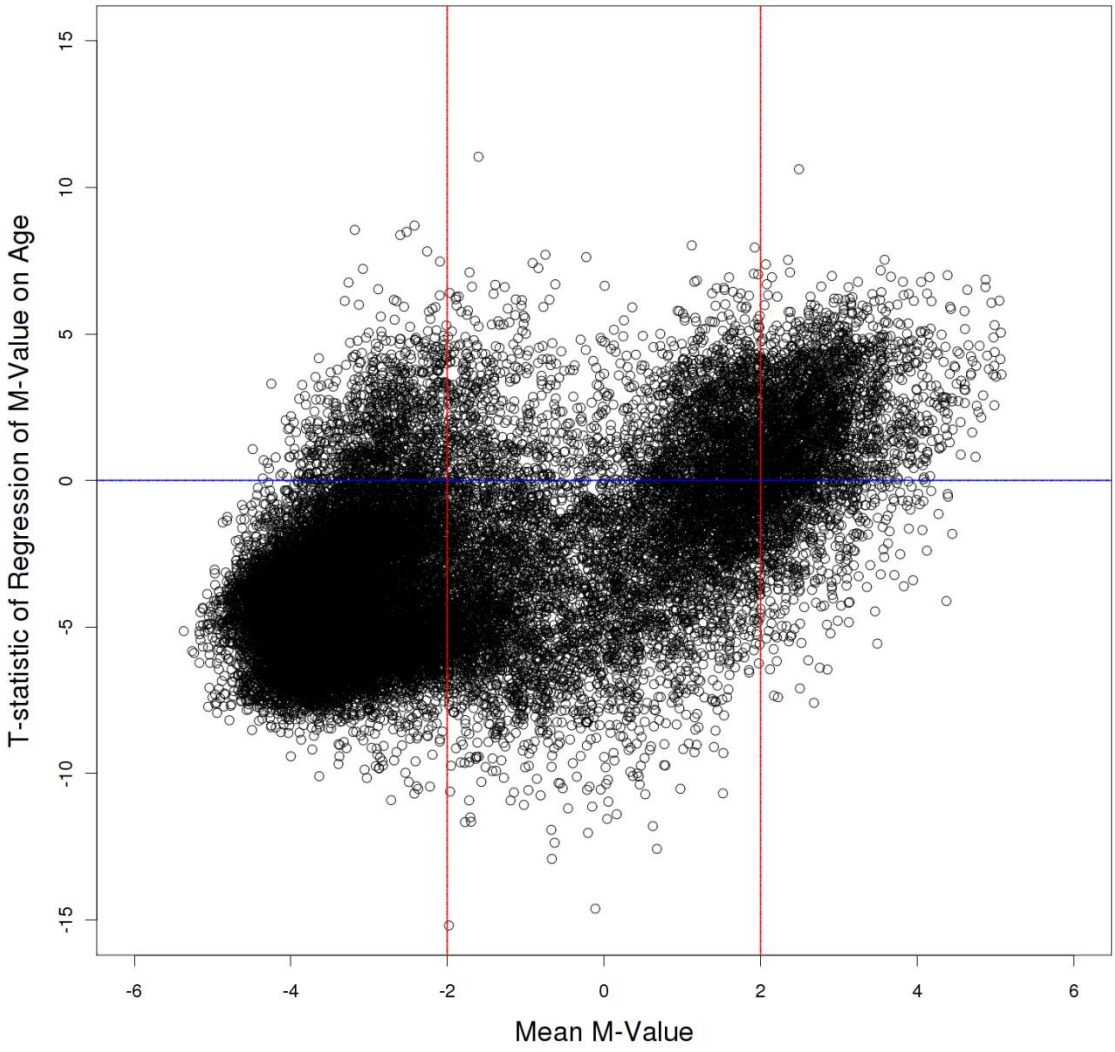
Yellow highlight denotes positive values for regression β_1.

Polymorphic and Non-Specific Probes: (Chen, et al. 2011)

** 0 = Neither, 1 = Polymorphic.

CpG sites listed within this table were not among those with non-specific binding probes.

Figure 2. T-statistic distribution of regression of M-Value on age vs. mean M-Value of corresponding CpG site for 26,428 DNA methylation markers



Red vertical lines at -2 and 2 represent delineation of unmethylated and methylated levels of CpG sites, such that sites having mean M-Value < -2 are unmethylated, and sites having mean M-Value > +2 are methylated.

Blue horizontal lines at ±4.8 represent the t-statistics equivalent to p = 1.86x10-6, which is the Bonferroni-correction for α=0.05

Table 3. Association between top 5 principal components (estimated from 2,095 site M-Values significant with age, after Bonferroni correction for α=0.05) and Age

		Univariate Models			Multivariable Model		
PC	**% Variation Explained**	**β**	**p-value**	**R^2_{LR} x100**	**β**	**p-value**	**R^2_{LR} x100**
1	50.65%	-0.12	6.63E-06	12.72	-0.13	5.60E-08	
2	9.53%	0.15	0.014	10.34	0.16	3.34E-03	
3	4.52%	-0.69	8.84E-14	18.95	-0.72	1.03E-15	
4	2.47%	-0.41	4.79E-04	11.39	-0.43	7.45E-05	
5	2.15%	0.16	0.21	9.58	0.14	0.22	26.76
6-10	4.68%						36.54
Total	74.00%						

Model: $Age_{ij} = \beta_0 + \beta_1 \cdot PC_{ij} + W_{0j} + \varepsilon_{ij}$.

Supplementary Table 1. Comparison of baseline characteristics of Phase II Non-Epigenetics Participants with Phase II Epigenetics Participants. A) Continuous clinical variables, B) Continuous biomarkers, C) Categorical descriptive variables.

A) Continuous clinical variables

	Phase II Non-Epigenetics Participants				Phase II Epigenetics Participants				
Variable	**N**	**% Missing**	**Range**	**Mean (SD)**	**N**	**% Missing**	**Range**	**Mean (SD)**	**Pr > |t|**
Age, years	474	0.0%	26.41-81.52	56.07 (9.21)	1008	0.00%	39.26-94.74	66.34 (7.6)	6.6E-49
BMI, kg/m²	472	0.4%	18.1-57.9	32.6 (7.4)	1001	0.69%	16.41-55.09	31.1 (6.09)	0.013
Systolic BP, mm Hg	473	0.2%	96-208	134.8 (19.3)	1006	0.20%	79-221	139.89 (21.12)	2.2E-06
Diastolic BP, mm Hg	474	0.0%	53-122	81.57 (10.2)	1008	0.00%	45-121	78.29 (11.01)	2.3E-05
Pulse Pressure, mm Hg	474	0.0%	19-127	53.41 (15.11)	1006	0.20%	26-127	61.64 (17.72)	6.0E-17
Adj. Systolic BP, mm Hg (1)	473	0.2%	97-218	142.05 (21.47)	1006	0.20%	89-231	148.15 (22.74)	5.5E-08
Adj. Diastolic BP, mm Hg (1)	474	0.0%	58-127	85.2 (10.95)	1008	0.00%	50-126	82.42 (11.43)	4.6E-04
Adj. Pulse Pressure, mm Hg (1)	473	0.2%	24-132	56.89 (15.72)	1004	0.40%	28-132	65.65 (18.28)	2.2E-18
Total Cholesterol, mg/dL	459	3.2%	72-348.5	197.93 (40.02)	1008	0.00%	73.5-354.5	203.98 (41.95)	7.8E-05
Triglycerides, mg/dL	454	4.2%	28.5-419.5	111.95 (58.94)	1004	0.40%	37-402.5	117.9 (56.68)	9.8E-04
ln (Triglycerides+1)	457	3.6%	3.38-6.45	4.62 (0.49)	1007	0.10%	3.64-6.27	4.69 (0.43)	0.69
HDL-C, mg/dL	457	3.6%	23.8-125.8	55.42 (16.66)	1005	0.30%	21.7-130.35	58.08 (17.48)	7.0E-03
ln (HDL+1)	460	3.0%	3.21-5.17	4 (0.3)	1008	0.00%	3.12-5.05	4.04 (0.29)	0.75
LDL-C, mg/dL	459	3.2%	23.6-253.75	121.1 (36.88)	1008	0.00%	24.85-272.1	123.84 (39.58)	0.064
Glucose, mg/dL	457	3.6%	43.5-296	108.28 (38.68)	998	0.99%	49.5-290	110.56 (34.31)	0.11
ln (Glucose+1)	457	3.6%	3.8-5.69	4.65 (0.28)	1004	0.40%	3.92-5.82	4.69 (0.27)	0.77
Insulin, mU/mL	163	65.6%	1.14-52.46	9.34 (8.53)	1005	0.30%	0.22-115.76	10.39 (12.45)	0.30
ln (Insulin+1)	163	65.6%	0.76-3.98	2.12 (0.62)	1006	0.20%	0.2-5.02	2.12 (0.74)	1.00
Menopause Age, years (2)	270	43.0%	23-59	42.6 (8.26)	699	30.65%	25-62	43.72 (7.85)	0.14

1 If a participant is listed as taking hypertensive medications, 10 mm Hg is added to systolic blood pressure, and 5 mm Hg is added to diastolic blood pressure, else recorded blood pressure is used in this variable.

2 Menopause age is represented for females only.

B) Measures of inflammation

	Phase II Non-Epigenetics Participants				Phase II Epigenetics Participants				
Variable	**N**	**% Missing**	**Range**	**Mean (SD)**	**N**	**% Missing**	**Range**	**Mean (SD)**	**Pr > |t|**
Serum Creatinine, mg/dL	460	2.95%	0.44-3.64	0.89 (0.29)	1008	0.00%	0.42-2.98	0.94 (0.3)	0.69
CRP, mg/L	347	26.79%	0.21-29.9	5.48 (5.8)	971	3.67%	0.21-29.9	6.05 (6.77)	0.37
ln (CRP+1)	347	26.79%	0.08-1.49	0.68 (0.33)	971	3.67%	0.08-1.49	0.7 (0.34)	0.87
Homocysteine, µmol/L	430	9.28%	4.5-25.2	9.38 (2.97)	1002	0.60%	4.7-31	10.57 (3.59)	7.1E-03
ln (Hycs+1)	430	9.28%	1.7-3.27	2.3 (0.26)	1002	0.60%	1.74-3.47	2.41 (0.28)	0.41
Fibrinogen, mg/dL	440	7.17%	123-684	362.44 (88.37)	1007	0.10%	120-680	369.35 (81)	1.6E-03
ICAM, ng/mL	199	58.02%	101-584	296.97 (83.09)	578	42.66%	17-625	275.38 (80.82)	3.2E-16
IL-6, pg/mL	241	49.16%	1.92-29.78	8.71 (5.06)	856	15.08%	1.42-36.45	9.09 (5.31)	0.54
ln (IL-6+1)	245	48.31%	0.46-1.58	0.95 (0.22)	878	12.90%	0.38-1.58	0.97 (0.22)	0.86
IL-18, pg/mL	240	49.37%	6.14-235.97	72.17 (38.66)	863	14.38%	1.86-252.42	69.92 (39.97)	0.19
ln (IL-18+1)	241	49.16%	0.85-2.49	1.81 (0.23)	872	13.49%	0.84-2.49	1.8 (0.25)	0.92
MPO, ng/mL	193	59.28%	11.74-141.47	47.43 (27.67)	954	5.36%	4-160.58	40.26 (25.04)	5.1E-07
ln (MPO+1)	196	58.65%	1.1-2.22	1.63 (0.23)	968	3.97%	0.7-2.22	1.56 (0.24)	0.60
MCP1, pg/mL	205	56.75%	247.44-2124.82	970.81 (319.89)	835	17.16%	159.81-2704.78	1053.43 (369.78)	2.3E-49
ln (MCP+1)	205	56.75%	2.4-3.33	2.96 (0.15)	849	15.77%	2.46-3.49	3.01 (0.16)	0.69
Resistin, ng/mL	312	34.18%	1-12.13	3.77 (1.96)	895	11.21%	1-12.32	4.07 (2.03)	0.40
ln (Resistin+1)	315	33.54%	0.3-1.2	0.65 (0.18)	902	10.52%	0.3-1.2	0.68 (0.17)	0.798
SAA, µg/mL	297	37.34%	2.35-146	26.75 (28.98)	731	27.48%	2.35-146	32.94 (30.88)	2.0E-05
ln (SAA+1)	297	37.34%	0.53-2.17	1.27 (0.39)	731	27.48%	0.53-2.17	1.38 (0.38)	0.51
TNFR-1, pg/mL	219	53.80%	102.78-3310.18	1113.11 (562.23)	872	13.49%	19.89-3844.07	1185.29 (546.83)	1.3E-27
ln (TNFR1+1)	217	54.22%	2.15-3.52	3 (0.21)	882	12.50%	2.17-3.68	3.04 (0.22)	0.76
TNFR-2, pg/mL	240	49.37%	388.3-4998.98	1751.03 (649.43)	849	15.77%	286.83-5333.39	1950.04 (831.08)	4.8E-117
ln (TNFR2+1)	241	49.16%	2.59-3.8	3.22 (0.16)	854	15.28%	2.59-3.8	3.26 (0.17)	0.70

C) Categorical variables

	Non-Epigenetics Participants			Epigenetics Participants		
	Count	**Total**	**% of Total**	**Count**	**Total**	**% of Total**
Female	335	474	70.68%	715	1,008	70.93%
Hypertensive	344	474	72.57%	833	1,008	82.64%
Has Diabetes	128	474	27.00%	308	1,008	30.56%
Ever Smoker	174	474	36.71%	421	1,008	41.77%

Supplementary Table 2. List of probes used to standardize methylated and unmethylated signals.

Red Channel Probes	Green Channel Probes
EXTENSION 1190050	BISULFITE CONVERSION 4670278
EXTENSION 360446	BISULFITE CONVERSION 4670484
NON-POLYMORPHIC 1740025	BISULFITE CONVERSION 5290048
STAINING 4200736	EXTENSION 1190050
STAINING 4570020	EXTENSION 360446
	EXTENSION 520537
	NON-POLYMORPHIC 1740025
	STAINING 5340168

Chapter III. The Role of Age within the Heritability of Epigenetic Markers

Introduction

Age is a well-known risk factor for many common chronic diseases.(*33, 71*) However, the cellular and molecular changes associated with aging processes that are related to chronic disease initiation and progression are not well-understood. More recently, differential DNA methylation patterns that affect gene expression have been found be associated with aging.(*40*) Specifically, age has been found to be associated with methylation status in chronic disease-related pathways such as those involved within liver development and metabolism,(*39*) inflammation, endothelial function, oxidation,(*41, 63*) and tumor suppression.(*64, 65*) Since epigenetic events provide a modifiable link between a gene's expression and a resulting phenotype,(*58, 59, 66, 67*) understanding the relative importance of genes to environment within epigenetic signatures with respect to age is crucial to understanding the role of epigenetics within the origins of chronic diseases.

The total phenotypic variation within a trait is due to the effects of both genetics and environment. In order to determine the amount of phenotypic variation in a trait due to genetics, genetic epidemiologists estimate the heritability of that trait. Narrow-sense heritability compares the trait variation due to additive genetic effects, at a specific time and within a specific population, to the total phenotypic variation. This quantity allows the comparison of a trait across generations within a population, or between different

populations, and may lead to novel insights about the underlying biology of a phenotype.(*92, 93*)

The influence of environmental factors such as nutrition, infection, and exposure to pollutants on DNA methylation has been well-characterized.(*94-97*) It is also known that genetic factors affect DNA methylation. DNA methyltransferases are involved in the process and maintenance of methylation at CpG sites throughout the genome, and genetic factors are also involved in the process of methylation throughout the methioninine cycle,(*98, 99*) as well as processing life course events into cellular and molecular responses that may affect methylation throughout the genome.(*100-102*) In general, more attention has been paid to environmental modifiers of epigenetic profiles (e.g. stress, toxins) rather than their genetic modifiers, despite research indicating that the methylation at some CpG sites likely depends on single nucleotide polymorphisms (SNPs) in the region.(*77*)

Due to the deficiency of research concerning the influence of genetic factors on DNA methylation levels, we examined the additive contribution of genes to the total variation of DNA methylation levels by estimating the heritabilities of DNA methylation levels at 26,428 CpG sites in 13,877 autosomal genes throughout the genome. In addition, because our previous studies illustrate that age is so strongly involved in epigenetic processes,(*103*) we investigated the age-related genetic contributions to these heritabilities. This research can provide a new method of understanding the potential mechanisms underlying associations of DNA methylation and chronic disease outcomes (i.e. the relative strength of the influence of genetic or environmental factors on DNA methylation).

Methods

Sample

The Genetic Epidemiology Network of Arteriopathy (GENOA) study is a community-based study investigating the genetics of hypertension and its arteriosclerotic complications in non-Hispanic whites from Rochester, MN and African-Americans from Jackson, MS.(*72*) In the current study, we investigated the heritability of DNA methylation at 26,428 CpG sites in GENOA African-Americans. African-American sibships were recruited such that ≥2 siblings were diagnosed with primary hypertension before the age of 60 years, while other siblings within the sibship were invited to participate independent of hypertension status (N=1,854). The initial examination (Phase I: 1996-1999) included standardized interviews concerning prescription drug usage, cigarette smoking, physical activity, history of hypertension, diabetes, and cardiovascular disease events; physical examination for blood pressure, height, weight, and waist and hip circumferences; and fasting blood samples for creatinine, total cholesterol, high-density lipoprotein (HDL)-cholesterol, triglycerides, glucose, and insulin. The second examination (Phase II: 2000-2004) included 1,482 participants returning from Phase I, and included re-measurement of interview, physical examination, and blood characteristics, as well as additional measurements of arteriosclerotic target organ damage, including heart, kidney and peripheral artery traits. DNA methylation was quantified on 1,008 Phase II participants using stored blood samples collected during the second examination. Comparisons between Phase II individuals within and outside of the DNA methylation study are available in Supplementary Table 3). Written informed consent was obtained from all subjects and approval was granted by participating

Institutional Review Boards. Complete information on the GENOA study population and sample measurements can be found detailed in Appendix 3.

Measurement of DNA methylation

Samples were prepared and DNA methylation was measured according to previously published methods (*73*) which are summarized as follows. For comprehensive information on these measurements, refer to Appendix 4.

Sample Preparation and Methylation Assay. DNA was isolated from peripheral blood leukocytes from GENOA Phase II stored samples, and bisulfite converted with the EZ DNA Methylation Gold Kit (Zymo Research, Orange CA). Bisulfite-converted DNA samples were whole-genome amplified, enzymatically fragmented, and purified, then hybridized to Illumina Infinium HumanMethylation27K BeadChips, which contained locus-specific DNA oligomers and a set of 56 control probes. The array was then fluorescently stained, scanned using the Illumina BeadXpress reader, and assessed for fluorescence intensities across the methylated and unmethylated bead types at 27,578 CpG sites.

Data processing and methylation quantification: At each CpG site, fluorescent signals were measured from the site-specific **M** (methlyated) and **U** (unmethylated) bead types. The raw fluorescence data from the scanner was processed using Illumina BeadStudio software. To reduce batch and chip effects, the correlation structure among all 56 control probes was evaluated within channel to identify the most parsimonious subset of probes that explained the maximum amount of batch and chip variation across samples (5 probes in the red channel and 8 probes in the green channel; Supplementary

Table 4). Quality control standards were employed by linearly regressing the 13 selected probes onto the intensity signals from the methylated and unmethylated bead types separately across each CpG site.

Two measures of methylation were calculated for each individual, across every CpG site, from the adjusted **M** and **U** signal intensities: the Beta Value and the M-Value. Beta Values are continuous variables ranging from 0 to 1 that are proportional to the percent methylation at each particular CpG site within each individual. The Beta Value for each individual *i* at a single site *k*, is calculated as: Beta $\text{Value}_{ik} = \max(\mathbf{M}_{ik},0) / [\max(\mathbf{U}_{ik},0) + \max(\mathbf{M}_{ik},0) + 100]$.(*104*) Beta Values between 0 and 0.2 are considered to be unmethylated, and those between 0.8 and 1 are considered methylated. Beta Values between 0.2 and 0.8 are considered semi-methylated. The M-Value is a commonly used measurement in microarray analysis that was more recently adapted for use in DNA methylation array data due to its ability to equalize the variance across the CpG sites.(*74, 75*) The M-Value for each individual *i* at a single site, *k*, is calculated as: $\text{M-Value}_{ik} = \log_2[(\max(\mathbf{M}_{ik},0) + 1) / (\max(\mathbf{U}_{ik},0) + 1)]$, where a constant is added to prevent large-scale changes caused by small intensity estimation errors.(*75*) The relationship between Beta and M-Values represents a logit transformation, where $\text{M-Value}_{ik} = \log_2 [\text{Beta Value}_{ik} / (1 - \text{Beta Value}_{ik})]$. Unmethylated M-Values are considered to be < -2.0, methylated M-Values are > +2.0, and semi-methylated M-Values are between -2 and +2. In the following sections, we present results from both the M-Values and the Beta Values for many of the analyses. However, we feel that the M-Value is a better measure of methylation for modeling purposes due to its more desirable statistical properties, which allow better conformation to modeling assumptions.

Before statistical analysis, samples were checked for data quality. From the 1,008 study participants, 7 individuals were removed due to poor bisulfite conversion control efficiency, measured by bisulfite conversion control intensity of < 4,000. An additional 29 individuals were removed from the analysis due to extreme control probe values, assessed as having at least one control probe with a value of greater than 4 standard deviations from its mean value. This resulted in a total sample size of 972 individuals.

For this study, we analyzed only autosomal CpG sites. A total of 58 CpG sites were removed from the analysis because they were found to be multimodal based on the Dip Test proposed by Hartigan and Hartigan (*76*) using a cut-off of $p<0.001$ on the signal intensities of the methylated and/or unmethylated bead types, which clearly violated the statistical modeling assumptions. This resulted in a total number of 26,428 CpG sites included in our analysis. Finally, 2,984 non-specific binding probes and 908 polymorphic probes overlapping with single nucleotide polymorphisms (SNPs) (*77*) were identified and denoted in the result tables. Though these sites were not removed from the analysis, we have interpreted the results from these sites with caution. As a final method of quality control, and since heritability estimates are quite sensitive to outliers, DNA methylation values greater than ±4 standard deviation from each CpG site mean were removed from each of the 26,428 CpG sites. Please refer to Appendix 4 for more detailed information.

Statistical Analyses

Estimating heritability for epigenetic markers: In order to better understand the genetic and environmental contributions to the epigenetic variation within the GENOA cohort, we estimated the heritabilities of DNA methylation levels within each of the

26,428 CpG sites using the t-distribution option in SOLAR (Sequential Oligogenic Linkage Analysis Routines) version 4.2.7.(*105*) The t-distribution option was used in order to ensure better estimates of heritability despite potential kurtosis within a CpG site distribution.

Given the novelty of this set of phenotypes, we estimated heritabilities for a) methylated signal intensity, b) unmethylated signal intensity, c) Beta Value, and d) M-Value at each CpG site. Traditionally, heritabilities are estimated before and after adjusting for age and sex; however, when we performed the adjustment for age, we found that the heritabilities decreased after adjustment (i.e. the genetic variance in the numerator was being reduced by the age adjustment). This finding was the motivation for partitioning the heritability into age-related and residual heritability components by mathematically utilizing the variance component information from raw and age-adjusted output from SOLAR.

Method of partitioning the heritability. After examining unadjusted and age-adjusted models that estimated heritability in SOLAR, we utilized the output in order to calculate the age-adjusted genetic variance for each epigenetic marker. To do so, we followed the subsequent rationale:

Let **P** indicate the phenotype, or the sum of genotype and environment, where **G** indicates Genotype and **E** indicates Environment. The total phenotypic variance can be decomposed into genetic and environmental variance, assuming the absence of covariance between **G** and **E**, such that:

[1] $\mathbf{P} = \mathbf{G} + \mathbf{E}$

[2] $\text{Var}(\mathbf{P}) = \text{Var}(\mathbf{G}) + \text{Var}(\mathbf{E})$

Age designates age-related effects, to be expressed as a subset of phenotype, genotype, or environment; and **NoAge** designates age-independent effects (or non-age associated effects), to be expressed as a subset of phenotype, genotype, or environment.

The total genetic variance can be decomposed into age-related genetic effects ($\mathbf{G_{Age}}$) and age-independent genetic effects ($\mathbf{G_{NoAge}}$). The total environmental variance can be decomposed similarly.

[3] $\text{Var}(\mathbf{G}) = \text{Var}(\mathbf{G_{Age}}) + \text{Var}(\mathbf{G_{NoAge}})$

[4] $\text{Var}(\mathbf{E}) = \text{Var}(\mathbf{E_{Age}}) + \text{Var}(\mathbf{E_{NoAge}})$

If we consider $\text{Var}(\mathbf{E_{Age}})$, the age-related environmental effects, to be the effect of age itself, we can write equation [4] as:

[5] $\text{Var}(\mathbf{E}) = \text{Var}(\mathbf{Age}) + \text{Var}(\mathbf{E_{NoAge}})$

Substituting equations [3] and [5] into equation [2], we have:

[6] $\text{Var}(\mathbf{P}) = \text{Var}(\mathbf{G_{Age}}) + \text{Var}(\mathbf{G_{NoAge}}) + \text{Var}(\mathbf{Age}) + \text{Var}(\mathbf{E_{NoAge}})$

Heritability, or $\mathbf{h^2_{Unadj,}}$ is defined as the proportion of phenotypic variance due to genetic variance. Thus, from equation [2], unadjusted heritability is written as:

[7] $\mathbf{h^2_{Unadj}} = \text{Var}(\mathbf{G}) / \text{Var}(\mathbf{P})$

Substituting equations [3] and [6] into [7], unadjusted heritability is written as:

[8] $\mathbf{h^2_{Unadj}} = [\text{Var}(\mathbf{G_{Age}}) + \text{Var}(\mathbf{G_{NoAge}})] / [\text{Var}(\mathbf{G_{Age}}) + \text{Var}(\mathbf{G_{NoAge}}) + \text{Var}(\mathbf{Age}) + \text{Var}(\mathbf{E_{NoAge}})]$

Age-adjusted heritability is the proportion of phenotypic variance due to genetic variance after removing the variability due to age, and will be referred to as $\mathbf{h^2_{Age\text{-}adj}}$.

Here, since phenotypic variation due to age has been removed from **P**, both Var (**Age**) and Var ($\mathbf{G_{Age}}$) = 0. Thus, the decomposition of the phenotypic variance is as follows:

[9] $\text{Var}(\mathbf{P_{NoAge}}) = \text{Var}(\mathbf{G_{NoAge}}) + \text{Var}(\mathbf{E_{NoAge}})$

From equation [9], age-adjusted heritability is written as:

[10] $\mathbf{h^2_{Age\text{-}adj}} = \text{Var}(\mathbf{G_{NoAge}}) / [\text{Var}(\mathbf{G_{NoAge}}) + \text{Var}(\mathbf{E_{NoAge}})]$

SOLAR outputs the following quantities:

- The phenotypic standard deviation, or **SD(P)**, from which we can calculate Var (P) = $[SD(P)]^2$
- The total heritability estimated within the sample, before adjustment for age, or $\mathbf{h^2_{Unadj}}$
- Age-adjusted heritability estimated within the sample, designated $\mathbf{h^2_{NoAge}}$
- The **proportion of Var (P) explained by age**

Using the proportion of Var (**P**) explained by age, we calculate Var ($\mathbf{P_{NoAge}}$) as follows:

[11] Var ($\mathbf{P_{NoAge}}$) = Var (**P**) – [Var (**P**)*Proportion of Var(P) explained by age]

Rearranging equations [8] and [10] gives:

[12] $\text{Var}(\mathbf{G_{Age}}) + \text{Var}(\mathbf{G_{NoAge}}) = \mathbf{h^2_{Unadj}} * [[\text{Var}(\mathbf{G_{Age}}) + \text{Var}(\mathbf{G_{NoAge}}) + \text{Var}(\mathbf{Age}) + \text{Var}(\mathbf{E_{NoAge}})] = \mathbf{h^2_{Unadj}} * \text{Var}(\mathbf{P})$

[13] $\text{Var}(\mathbf{G_{NoAge}}) = \mathbf{h^2_{Age\text{-}adj}} * [\text{Var}(\mathbf{G_{NoAge}}) + \text{Var}(\mathbf{E_{NoAge}})] = \mathbf{h^2_{Age\text{-}adj}} * \text{Var}(\mathbf{P_{NoAge}})$

Subtracting equation [13] from equation [12] solves for Var (G_{Age}).

[14] $\text{Var}(\mathbf{G_{Age}}) = [\mathbf{h^2_{Unadj}} * \text{Var}(\mathbf{P})] – [\mathbf{h^2_{NoAge}} * \text{Var}(\mathbf{P_{NoAge}})]$

The heritability due to age-related genetic effects is estimated using:

[15] $[\text{Var}(\mathbf{G_{Age}}) / \text{Var}(\mathbf{P})] = \mathbf{h^2_{Age}}$

The proportion of heritability due to age-related genetic effects is:

[16] $\Psi = \mathbf{h^2_{Age}} / \mathbf{h^2_{Unadj}}$

Results

Description of data. After exclusions and quality control, this study contained phenotype and methylation data for 972 African Americans across 26,428 CpG sites. The sample was predominantly female (70.7%) and hypertensive (82.5%), with mean age of 66.3 years and mean body mass index of 31.2 kg/m^2. Further descriptive statistics are presented in Table 4. This study population consisted of 197 singletons and 296 sibships ranging in size from 2 to 10 siblings, with a mean of 2.6 siblings per sibship (see Supplementary Table 5).

Within this population, the 26,428 CpG sites were predominantly unmethylated, where 15,221 (57.6%) CpG sites had a mean M-Value of <-2.0. Across all CpG sites, the mean fluorescence intensities across the methylated bead types ranged from 482 to 39,810 (mean = 28,25), while mean fluorescence intensities on the unmethylated bead types ranged from 497 to 37,310 (mean= 6,865). CpG site Beta Value means ranged from 0.025 to 0.97, with an average mean Beta Value of 0.31, and M-Value means ranged from -5.37 to 5.07 with an average mean M-Value of -1.58. Due to outlier

removal, the mean number of individuals measured across the 26,428 CpG sites is 970.9 (standard deviation: 1.7), with a minimum of 935. (Figure 3.)

Heritability. Heritability was estimated across four measures of DNA methylation: (1) methylated and (2) unmethylated signal intensities, (3) Beta Values, and (4) M-Values. The vast majority of the measures of DNA methylation were significantly heritable at $p<0.05$ (methylated probes: 20,646 (78.1%); unmethylated probes: 24,851 (90.0%); Beta Values: 23,183 (87.7%), M-Values: 23,372 (88.4%)). The estimates of mean heritabilities for each of the four DNA methylation measures in this sample were above 0.30 (methylated probe intensity, mean= 0.41; unmethylated probe intensity mean=0.38; Beta Value, mean = 0.30; M-Value mean =0.30). The distributions of significant heritabilities for the four measures of methylation are shown in Figure 4. It is noticeable that the heritabilities of the methylated and unmethylated signal intensities are distributed bimodally, while the distributions for the heritabilities of the Beta and M-Values are heavily right-tailed.

In order to better understand why the heritabilities of the methylated and unmethylated signals followed bimodal distributions, we used t-tests to investigate whether any CpG site attributes were associated with being in the lower vs. higher mode (cut point of $h^2= 0.41$ for methylated signals, and cut point of $h^2= 0.38$ for unmethylated signals). Significantly associated with mode of methylated and unmethylated signals were probe GC content (methylated probe, p-value = 1.04×10^{-13}; association with unmethylated probe, p-value 7.37×10^{-15}), distance to transcription start (methylated probe, p= 6.32×10^{-3}; unmethylated probe, p=6.20×10^{-31}), and the number of SNPs within a probe (methylated probe, p=9.11×10^{-4}; unmethylated probe, p=1.00×10^{-20}). Further, the

color channel of which the CpG site's intensity levels were measured and whether or not the CpG site is in a CpG island were significant at $p< 0.0001$ across both signal types. This may be important because each color channel represents a different nucleotide following the CpG site, where the red channel represents a final extended base of A or T and the green channel represents a final extended base of G or C, which may be indicative of probe GC content. A t-test of the difference in means of probe GC content by color channel indicated a highly statistically significant association between these two variables ($p=9.83x10^{-9}$). Results from t-tests examining differences in modes of heritabilities and CpG site attributes can be found in Table 5.

For each of the four measures of DNA methylation, (1) methylated and (2) unmethylated signal intensities, (3) Beta Values, and (4) M-Values, the genetic variance was calculated from the heritability and total phenotypic variance. The genetic variance was then decomposed into partitions that were due to age-related genetic factors or genetic factors not associated with age. From this, the percentage of the total genetic variability due to age-related genetic factors and the proportion of age-related genetic effects within the total phenotypic variance were calculated. The mean percentage of total genetic variability due to age-related genetic factors was highest for significant (p-value < 0.05) unmethylated probe intensities (28.3%), while the mean percentage for the significant methylated probe intensity, Beta Value, and M-Value were 17.3%, 18.6%, and 20.3% respectively (Table 6). Each of the four measures of DNA methylation had between 53% and 66% of the significant CpG sites with at least 10% of their total genetic variability explained by age-related genetic factors. Further, age-related genetic effects explained up to 100% of the total genetic variability for some methylated or

unmethylated signal intensities, Beta Values, and M-Values. The sites with highest proportion of variance explained by age-related genetic effects for the M-Value are presented in Table 7. These sites closely corresponded to those with the highest proportion of variance explained by age-related genetic effects for Beta Values.

Utilizing heritabilities to understand epigenetic associations. To better understand the genetic and environmental mechanisms underlying epigenetic associations, we examined the heritability of CpG sites relative to their association with age. Upon examining the epigenetic markers as predictors of age using linear mixed modeling to account for sibship structure, 2,095 sites were significant predictors of age (after Bonferroni correction at α=0.05). Figure 5 integrates information on heritability with the results of the linear mixed modeling of CpG sites as predictors of age to illustrate whether the CpG sites that have strong age associations vary predominantly due to environmental factors, genetic factors, or both.

In Panel A, the raw heritability for each M-value was plotted against the strength of the association between age and CpG site represented by the –log(p-value) of the regression of age on each of the 26,428 CpG sites. The blue horizontal line indicates significant association at Bonferroni corrected α=0.05 (here, the -log(p-value) of the regression of age on methylation is 5.72). The red line at h^2=0.3 denotes the cut-off for low and high heritabilities. Points in the upper left quadrant of Panel A (434 CpG sites, 1.6%) denote methylation markers that have a strong association with age but low heritability, suggesting that the effect of age on methylation at these sites may be moderated through environmental effects that accumulate with age. Points in the upper right quadrant (1,661 CpG sites, 6.3%) denote methylation markers that have a strong

association with age and high heritability, suggesting that the effect of age on methylation may be moderated through age-related genetic factors. The lower left quadrant contains 16,695 (63.2%) CpG sites, and the lower right quadrant contains 7,635 (28.9%) CpG sites. These quadrants indicate weak associations with age and low and high heritabilities, respectively.

Figure 5, Panel B examines the age-adjusted heritabilities for each CpG site M-Value versus the -log(p-value) of the regression of age on methylation. After dividing this plot by heritability of 0.3 and significance of –log(p-value) of the regression of age on DNA methylation at 5.72 (Bonferroni corrected α=0.05), the lower left quadrant contains 18,690 (70.7%) CpG sites, the upper left quadrant contains 1,268 (4.8%) CpG sites, the lower right quadrant has 5,640 (21.3%) CpG sites, and the upper right quadrant has 827 (3.1%) CpG sites. This quadrant of 827 CpG sites indicates sites that are both significantly associated with age, while simultaneously having a high heritability after age-adjustment. Thus, these CpG sites may act as mediators between genetics and age, since they have high age-independent genetic variance components that may indicate a set of age-independent genetic factors that are also highly associated with age.

Figure 5, Panel C indicates the percent of heritability due to age-related genetic factors compared to the –log(p-value) of the regression of age on each of the 26,428 CpG sites. This plot is divided where 10% of the heritability is explained by age-related genes, and the where –log(p-value) of the regression of age on methylation is 5.72 (at Bonferroni corrected α=0.05). In this plot, there are 10,849 (41.3%) CpG sites in the lower left quadrant, 118 (0.5%) CpG sites in the upper left quadrant, 13,317 (50.7%) CpG sites in the lower right quadrant, and 1,975 (7.5%) CpG sites in the upper right

quadrant. Of these 1,975 sites in the upper right quadrant, 1,554 (78.7%) had high unadjusted heritabilities ($h_{unadj}^2 > 0.3$), and 720 (36.5%) had high age-adjusted heritabilities ($h_{Age}^2 > 0.3$). All of the 720 sites with high age-adjusted heritabilities were contained within the set of 1,554 with high unadjusted heritabilities. These 1,975 sites in the upper right quadrant indicate a set of sites that both have a large proportion of their heritability explained by age-related genetic factors while also having strong associations with age, which could indicate a set of CpG sites whose levels are affected by genes that have effects later in life that may play a role in their association with age.

Finally, Figure 5 Panels D and E indicate age-related (Panel D) or age-independent (Panel E) portions of heritabilities relative to the –log(p-value) of the regression of age on each of the 26,428 CpG sites. Each plot was divided by a line indicating the significance of –log(p-value) of the regression of age on DNA methylation at 5.72 (Bonferroni corrected α=0.05). Panel D indicates 24,330 (92.1%) of the sites in the lower section and 2,095 (7.9%) of sites in the upper section. Only 12 sites (0.05%) had age-related heritabilities > 0.3, and these sites were all significant in the regression of age on DNA methylation. Panel E shows 24,330 (92.1%) of sites in the lower section, where 5,520 (22.7%) of them had age-independent heritabilities > 0.3, and 2,095 sites in the upper section, where 767 (36.5%) had age-independent heritabilities > 0.3.

Discussion

The epigenome is a known mediator of environmental effects and gene expression.(*23, 57-59*) While many studies have illuminated the importance of the environment on epigenetics, indicating strong associations between DNA methylation

and environmental conditions such as diet,(*106-109*) stress,(*110-112*) and air pollution.(*113-115*), only a few have begun to explore the importance of genetics within the process of DNA methylation.(*116, 117*) The current study underscores the importance of the contribution of both environmental and genetic factors to DNA methylation, and specifically sheds light on age-related genetic and environmental components to the variation in many DNA methylation markers. First, it is noted that > 75% of CpG sites are significantly heritable, and that their average heritability ranges from 0.30 to 0.41, depending on the specific measurement (methylated or unmethylated signal intensity, Beta Value, M-Value). Second, these heritabilities can be partitioned into age-related and age-independent genetic factors, and we identified multiple epigenetic sites that have strong age-related genetic components. Finally, we used heritabilities to begin to understand the potential mechanisms that may be underlying epigenetic associations with epidemiological risk factors such as age.

Heritability estimates in the GENOA sample indicated mean heritabilities of 0.30 for CpG site M-Values and Beta Values. These estimates are generally greater than those reported elsewhere in the literature (h^2<0.18).(*116-118*) Other studies may have different heritability estimates than those in GENOA due to different population demographics and structures and/or differences in the type of methylation markers assayed. For example, many studies focused on epigenetic variation within specific genomic regions,(*116*) while our study focused on variation across all measured Illumina Infinium sites. Our heritabilities were estimated in a study 972 African-American individuals from hypertensive sibships, while other studies have used a much smaller number of European twins.(*116, 117*) Further, the estimated heritabilities of many of the CpG sites within the

GENOA sample have a sizable age-related component, which may also contribute to the increased heritabilities. Lastly, sib-sib plots, found in Supplementary Figure 1, indicate that some sites with higher heritabilities may be influenced by the sibship clusterings of high leverage values.

Our previous work demonstrated a strong relationship between age and epigenetic variation.(*103*) This finding is consistent with the idea that epigenetic variation represents a lifelong accumulation of environmental exposures. However, when we adjusted for age in our estimation of heritability, we noticed a differential change in the magnitude of heritability that was not expected. This led us to postulate that age-dependent penetrance in some genes could be giving rise to an age-related genetic component that is distinct from age-independent genetic components. In partitioning the genetic variance within our study into age-related and age-independent components, we found that 13,260 (56.7%) of the significantly heritable methylation sites as measured by the M-Value had at least 10% of their total heritability explained by age-related genetic effects. Of these sites, 2,402 (10.3) had at least 50% of their total heritability explained by age-related genetic effects. While age, itself, is known to have a high heritability in many family studies, we would expect that the contribution of age heritability to the overall epigenetic heritability of the CpG sites would be relatively uniform, especially for higher age-related sites, but it is not (See Figure 5 A and B).

The set of 2,402 CpG sites that had at least 50% of their total heritability explained by age-related genetic effects is especially interesting because these CpG sites may be in genomic regions whose roles within the aging processes are not well understood and are, in fact, mediated through epigenetic markers. In order to determine

whether these sites were related to a particular cellular pathway or function, we performed a gene enrichment analysis. Using the Expression Analysis Systematic Explorer (EASE),(*119*) we found that genes related to cell proliferation (Bonferroni-corrected p-value = 0.0107) was overrepresented in this set of 2,402 CpG sites. These findings suggest that methylation sites with a strong age-related heritability component may be particularly important for key pathways related to cancer initiation or progression,(*120, 121*) which is consistent with prior research that has shown how methylation plays a key role in multiple types of cancer.(*122, 123*) Further, these CpG sites may possibly be involved in atherosclerotic processes since excessive hyperplastic cell growth within vascular lesions is known to be a key component of inflammatory response within this disease.(*124*) For the top 30 CpG sites with the largest contribution of age-related genetic factors (Table 7), one logical future study would be to investigate whether *cis*-acting DNA variation has an age-dependent penetrance relationship with CpG site variation.

By examining the heritabilities of DNA methylation markers and their relationship to the strength of the association with age, we can begin to understand whether the cellular aging process at the epigenomic level is more a reflection of genetic or environmental factors. Using EASE to analyze the organismal roles of the 1,661 sites that were found to be significantly associated with age (at Bonferroni-corrected $\alpha = 0.05$ and have high unadjusted heritabilities (> 0.3), we found that anti-pathogen response (Bonferroni-corrected p-value = 0.0022) and response to injury (Bonferroni-corrected p-value = 0.0047) were significant. We also found that the set of 827 sites that were significantly associated with age and had high age-adjusted heritability (> 0.3) were also

significantly associated with anti-pathogen response (Bonferroni-corrected p-value = $1.17x10^{-4}$) and response to injury (Bonferroni-corrected p-value = $9.14x10^{-4}$). Further, of the 1,975 sites that were significantly associated with age and had a large percentage of their heritability due to age-related genetic factors (> 10%), anti-pathogen response (Bonferroni-corrected p-value = $1.09x10^{-4}$) and response to injury (Bonferroni-corrected p-value = 0.042) were significant in EASE analysis. This suggests that heritable sites that are associated with age may be involved in pathways related to immune response and tissue repair. This finding is congruent with rising evidence that immune responses, such as inflammation, may be a significant factor within adult aging processes, especially in African-American populations. Given these results, the epigenome may be a very important layer of the biological role of heritability in which to explore the impact of cellular aging on chronic disease processes that is due to both genetic and environmental influences. Our heritability findings indicate that direct genetic studies of the epigenome (e.g. genome-wide association studies, or quantitative trait loci studies) might also provide new insight into the actual mechanisms of aging or other diseases in which the epigenome is a mediator of environment and genetic effects.

Table 4. Baseline characteristics of GENOA participants after outlier removal: A) Continuous variables, B) Categorical Variables

A) Continuous Variables

	N	Count of Outliers Removed	Count of Missing Values	Range	Mean (SD)
Age, years	972	0	0	39-95	66 (8)
BMI, kg/m^2	965	0	7	16.4-55.1	31.1 (6.1)
Systolic BP, mm Hg	970	0	2	79-221	140(21)
Diastolic BP, mm Hg	972	0	0	45-121	78 (11)
Pulse Pressure, mm Hg	971	0	1	26-127	62 (18)
Systolic BP (Adj), mm Hg**	970	0	2	89-231	148 (23)
Diastolic BP (Adj), mm Hg**	972	0	0	50-126	82 (11)
Pulse Pressure (Adj), mm Hg**	969	0	3	28-132	66 (18)
Total cholesterol, mg/dL	972	0	0	73.5-354.5	203.7 (42.1)
Triglycerides, mg/dL	963	5	4	37-345	116.6 (53.8)
HDL cholesterol, mg/dL	967	2	3	21.7-122.25	57.9 (17.1)
LDL cholesterol, mg/dL	972	0	0	24.85-272.1	123.6 (39.7)
Glucose, mg/dL	951	11	10	49.5-245	108.6 (29.6)
Insulin, mU/mL	953	16	3	0.22-58.29	9.23 (8.25)
Serum creatinine, mg/dL	961	11	0	0.42-2.16	0.92 (0.25)

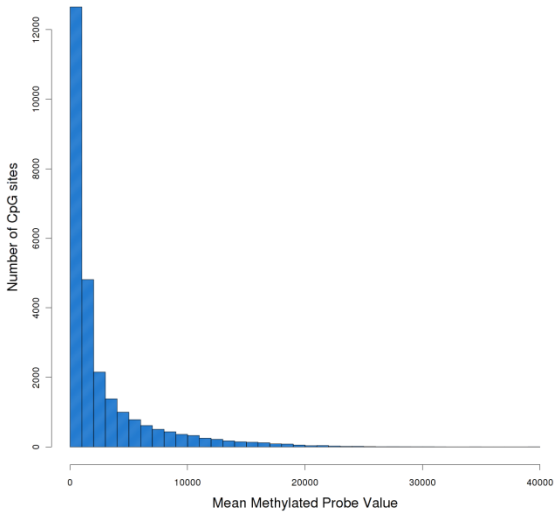
** If a participant is listed as taking hypertensive medications, 10 mm Hg is added to systolic blood pressure, and 5 mm Hg is added to diastolic blood pressure, else recorded blood pressure is used in this variable.

B) Categorical Variables

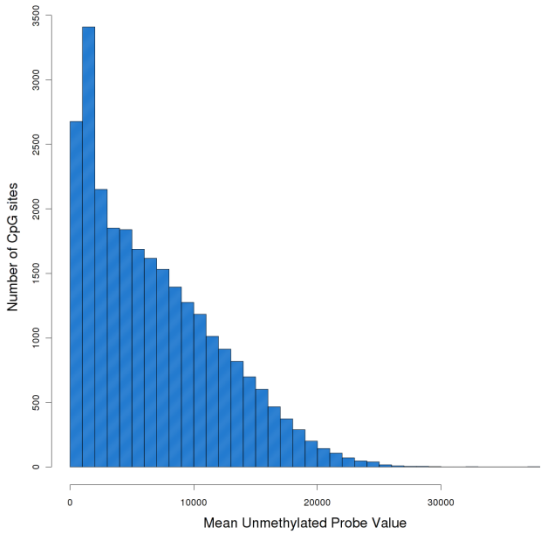
	Count	Total	Percent
Female sex	687	972	70.7%
Ever smoker	266	909	29.3%
Hypertensive	802	972	82.5%
Diabetic	298	972	30.7%

Figure 3. Distribution of means across 26,428 markers of methylation

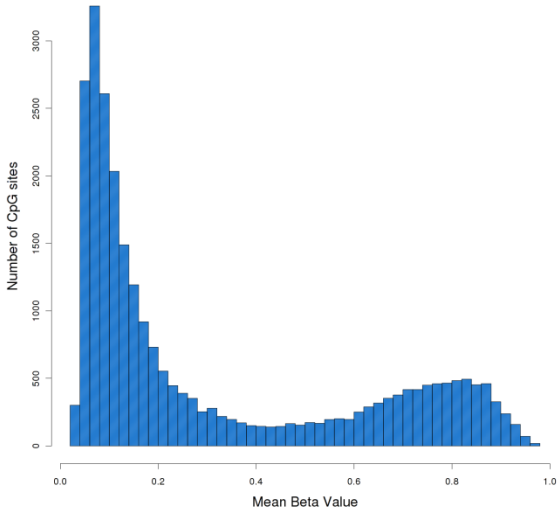
A) Methylated probe signal intensities (Range: 482 to 39,810, mean = 2,825)



B) Unmethylated probe signal intensities (Range: 497 to 37,310, mean = 6,865)



C) Beta Values (Range: 0.02 to 0.97, mean = 0.31)



D) M-Value (Range: -5.37 to 5.07, mean = -1.58)

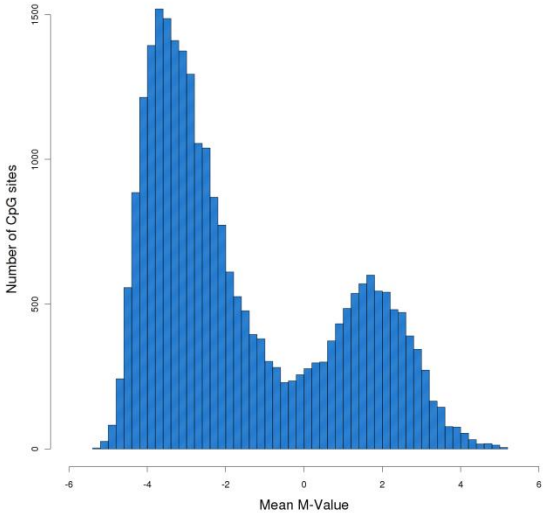
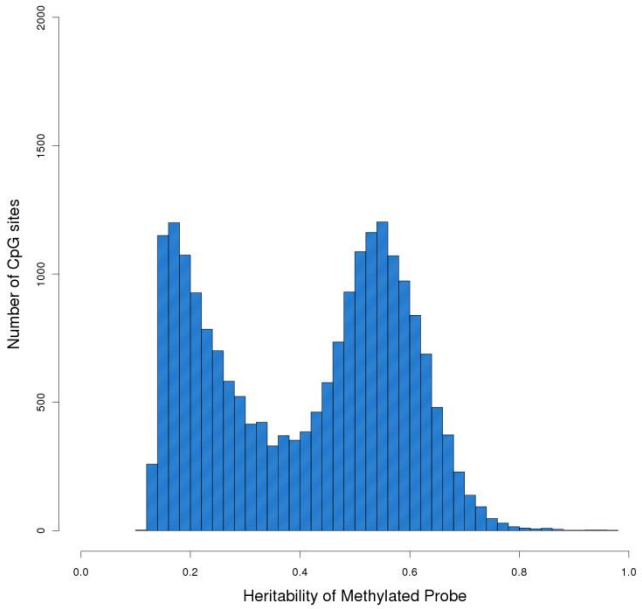
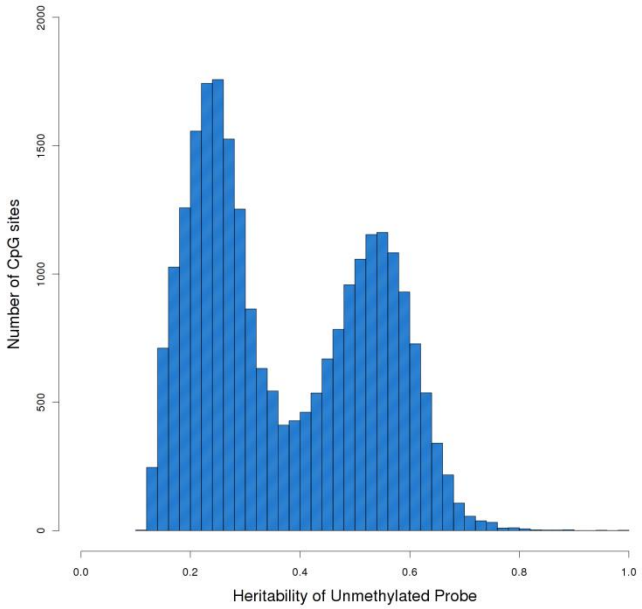


Figure 4. Distribution of significant (p<0.05) heritabilities of methylation markers

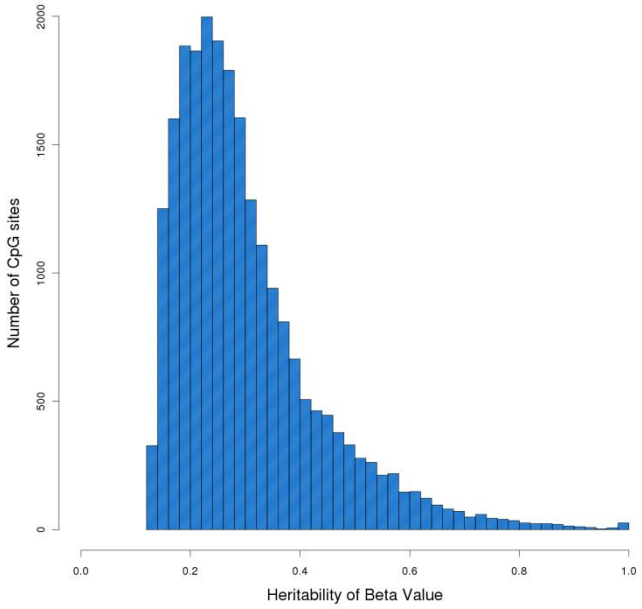
a) Methylated Probe. 20,646 probes. Mean (sd) = 0.41 (0.17). Range, (0.12. 0.97)



b) Unmethylated Probe. 24,851 probes. Mean (sd) = 0.38 (0.16). Range, (0.11, 1)



c) Beta Value. 23,183 probes. Mean (sd) = 0.30 (0.13). Range, (0.12, 1).



d) M-Value. 23,372 probes. Mean (sd) = 0.30 (0.13). Range (0.11, 1).

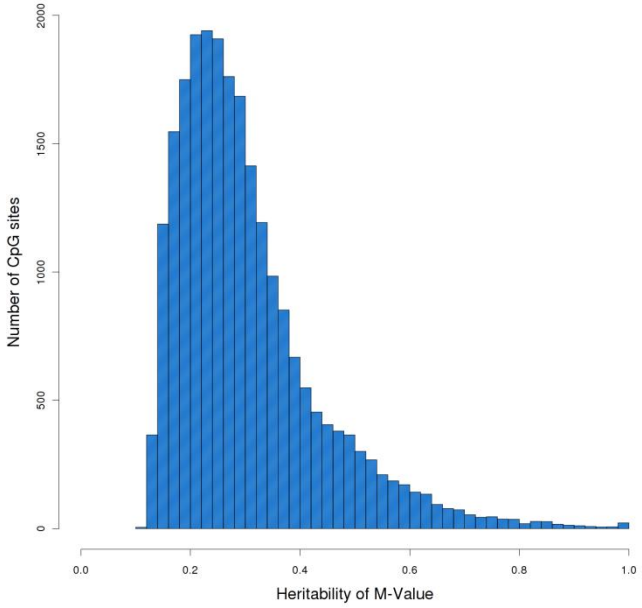


Table 5. Potential contributors to the bimodality of significant methylated and unmethylated signal heritabilities (h^2 significant at $p < 0.05$), A) Methylated Signal, B) Unmethylated Signal

A) Methylated Signal

	< mean h^2, 0.41	≥ mean h^2, 0.41	p-value
Count	9,277	11,369	
Probe GC Content Mean (SD)	28.70% (3.07%)	28.38% (3.22%)	1.04E-13
Distance to Transcription Start Mean (SD)	365.12 (332.27)	378.05 (343.09)	6.32E-03
No. SNPs within Probe Mean (SD)	2.04 (1.93)	2.14 (2.39	9.11E-04
Color Channel Counts	Red: 876 Green: 8401	Red: 10,693 Green: 676	<0.0001
In CpG Island Counts	Yes: 6,823 No: 2454	Yes: 8,429 No: 2940	0.33

B) Unmethylated Signal

	< mean h^2, 0.38	≥ mean h^2, 0.38	p-value
Count of Probes	13,532	11,319	
Probe GC Content, % Mean (SD)	28.70% (2.84%)	28.40% (3.24%)	7.37E-15
Distance to Transcription Start, bp Mean (SD)	329.53 (295.89)	376.47 (343.05)	6.20E-31
No. SNPs within Probe Mean (SD)	1.92 (1.801)	2.17 (2.42)	1.00E-20
Color Channel Counts	Red: 888 Green: 12,644	Red: 10,655 Green: 664	<0.0001
In CpG Island Counts	Yes: 9937 No: 3,595	Yes: 8412 No: 2,907	0.11

Table 6. Heritability and age-related genetic factors

Epigenetic measure	Heritability Mean (SD)	% of total genetic variance due to age-related genetic factors* Mean (SD)	Number of epigenetic markers with age-related genetic factors explaining ≥10% of their total heritability
Methylated Probe	0.41 (0.17)	17.34% (16.24)	12,744
Unmethylated Probe	0.38 (0.16)	28.34% (15.81%)	16,413
Beta Value	0.30 (0.13)	18.62% (18.88%)	12,510
M-Value	0.30 (0.13)	20.29% (20.01)	13,260

* Based on Methylated Probes, Unmethylated Probes, Beta Values, and M-Values with significant heritability ($p<0.05$)

Table 7. Top 30 sites (M-Values) with largest portion of phenotypic variance explained by age-related genetic factors

Trait	Probe Type†	Chr.	Gene	h^2	Phenotypic Var. explained by age-related genetic factors	% of h2 explained by age-related genetic factors	Ranking within regression of age on DNA
cg04662594	1	8	EPB49	0.8660 ♦	0.4317	49.85%	13
cg15538427	0	11	LOC221091	0.5293**	0.3538	66.84%	2
cg15804973	0	6	MAP3K5	0.5897□	0.3353	56.86%	9
cg00451635	0	16	EMP2	0.7521 ♦	0.3312	44.04%	7
cg11860203	0	17	CCL2	0.7519 ♦	0.3182	42.32%	115
cg15037004	0	5	ZNF366	0.5356**	0.3131	58.45%	28
cg19722847	0	12	IPO8	0.3851**	0.3125	81.16%	38
cg00431114	0	20	C20orf121	0.6793□	0.3113	45.83%	15
cg23799313	0	1	PAQR7	0.6319□	0.3091	48.92%	45
cg19761273	0	17	CSNK1D	0.4408**	0.3082	69.92%	1
cg17034109	0	1	CYB561D1	0.6174□	0.3030	49.08%	16
cg01820374	0	12	LAG3	0.7834 ♦	0.3024	38.61%	3
cg08090640	0	17	IFI35	0.5824□	0.2974	51.06%	37
cg14244577	0	16	DDX19B	0.5693□	0.2914	51.18%	12
cg07313155	0	17	THRA	0.4509**	0.2913	64.61%	219
cg25538571	1	8	FLJ46365	0.6571□	0.2889	43.97%	8
cg07211259	0	9	PDCD1LG2	0.5601□	0.2817	50.30%	49
cg26954174	0	16	CARD15	0.6429□	0.2802	43.59%	80
cg13302154	0	12	MGP	0.5167**	0.2798	54.16%	72
cg07158339	0	9	FXN	0.3349*	0.2787	83.23%	74
cg02945646	0	14	AP1G2	0.5186**	0.2779	53.58%	68
cg24801210	0	3	PCNP	0.6772 ♦	0.2777	41.00%	124
cg08116137	2	2	NAT8	0.4828**	0.2767	57.30%	204
cg09997082	0	19	GIPR	0.4893**	0.2749	56.18%	61
cg02142461	2	4	LYAR	0.7660 ♦	0.2726	35.59%	141
cg05037688	1	9	EGFL7	0.6561 ♦	0.2707	41.26%	82
cg16435601	2	17	FALZ	0.3927**	0.2682	68.29%	574
cg09172980	2	2	NAT8	0.5672□	0.2679	47.23%	181
cg20870362	0	9	CCIN	0.6214□	0.2644	42.55%	30
cg03996822	0	4	RASSF6	0.8192 ♦	0.2607	31.83%	5

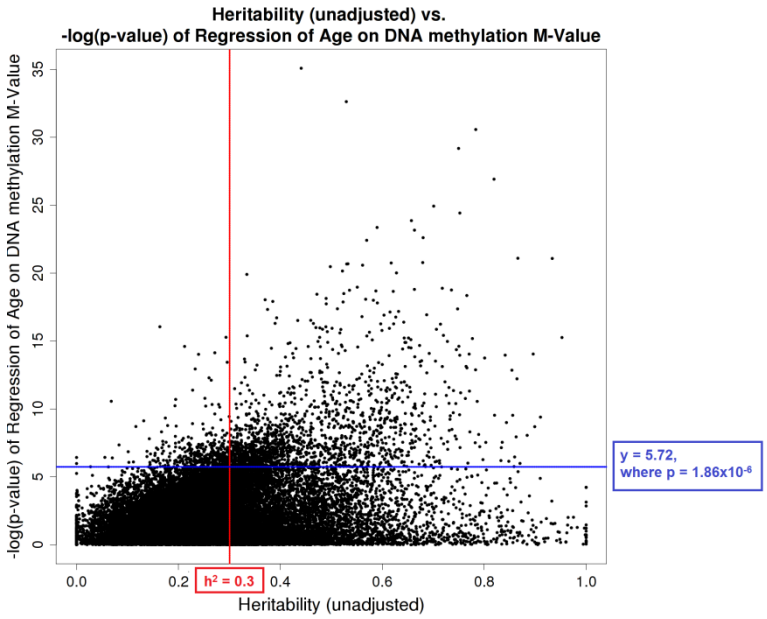
Significance levels: ♦p< 10^{-15}, □p< 10^{-10}, **p< 10^{-5}, *p< 10^{-3}Based on M-Values with significant heritability (p<0.05)

Polymorphic and Non-Specific Probes: (Chen, et al. 2011)

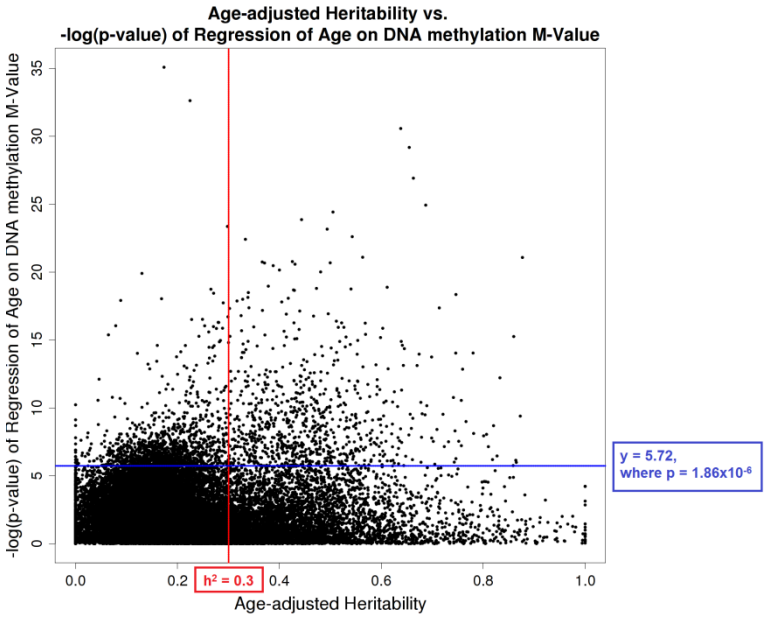
† 0 = Neither, 1 = Polymorphic, 2 = Non-Specific Binding

Figure 5. Heritability (or variance partitions) vs. Significance of regression of Age on DNA methylation M-Values (A) Unadjusted heritability of the M-Value, (B) Age-adjusted heritability of the M-Value, (C) Percent of heritability explained by age-related genetic factors, D) Age-related heritability, E) Age-independent heritability

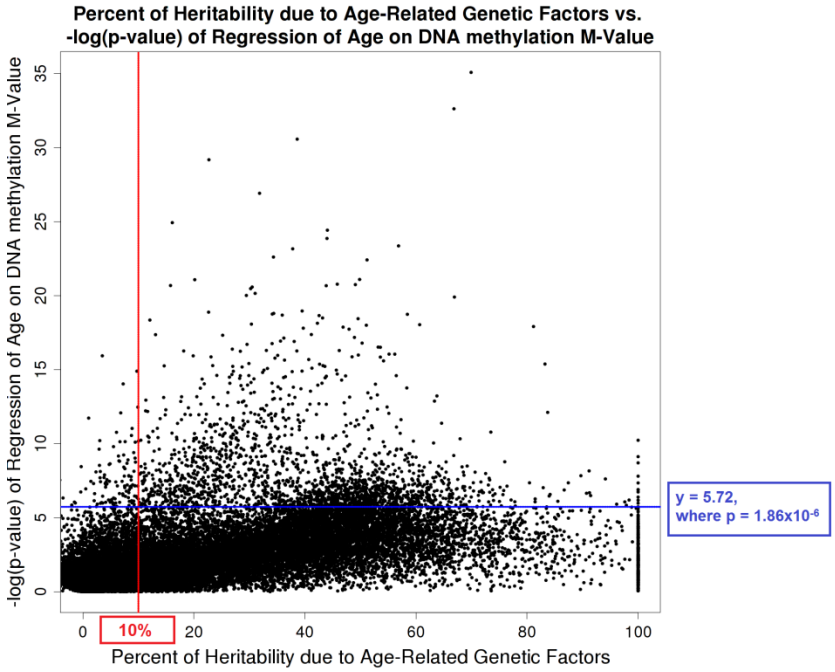
A) Heritability (unadjusted) vs. –log(pvalue) of Regression of Age on DNA methylation M-Value*



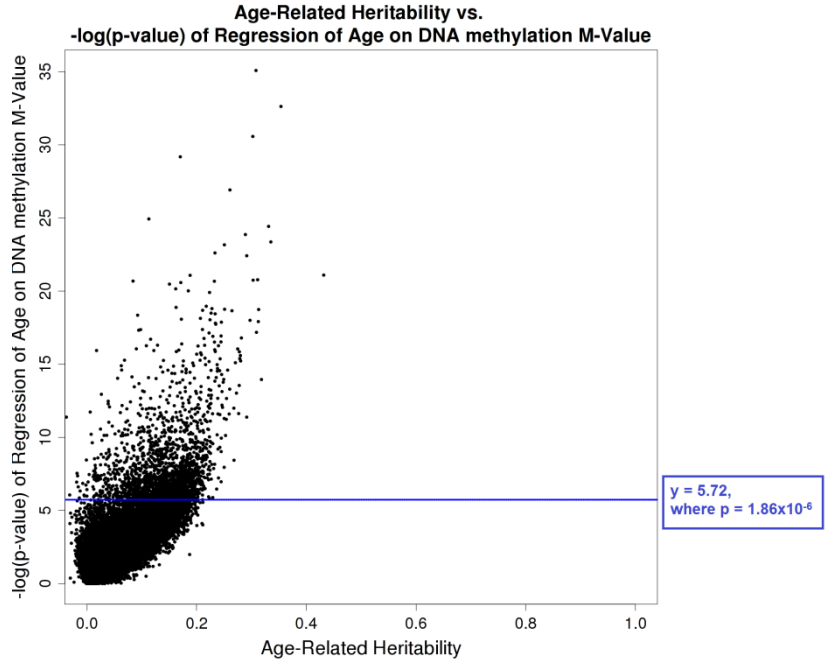
B) Age-adjusted heritability vs. –log(pvalue) of Regression of Age on DNA methylation M-Value*



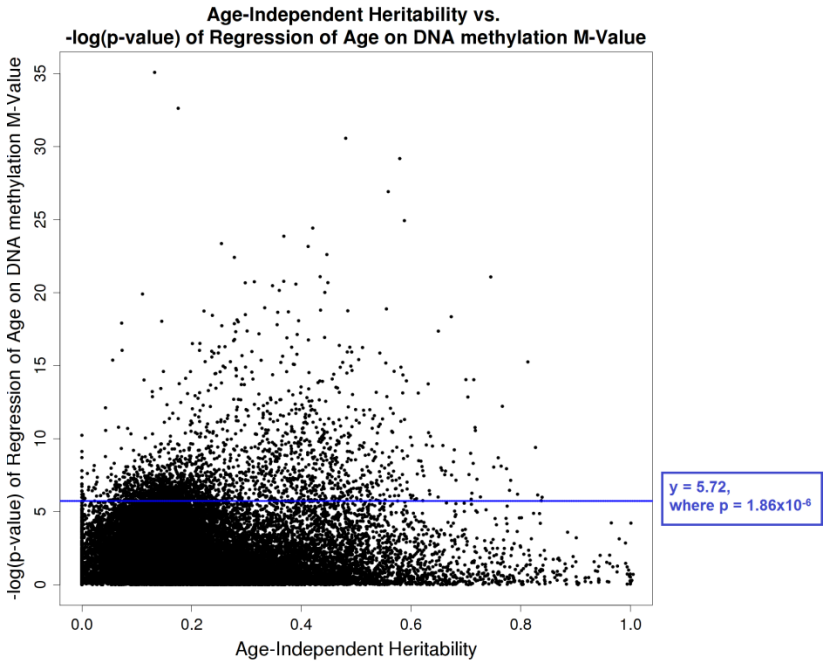
C) Proportion of Heritability due to Age-Related Genetic Factors vs. –log(pvalue) of Regression of Age on DNA methylation M-Value*



D) Age-Related Heritability vs. –log(pvalue) of Regression of Age on DNA methylation M-Value*



E) Age-Independent Heritability vs. –log(pvalue) of Regression of Age on DNA methylation M-Value*



* Linear Mixed Model from which p-value is obtained: $\text{Age}_{ij} = \beta_0 + \beta_1 \cdot \text{E}_{ij} + W_{0j} + \varepsilon_{ij}$ where Age is participant age at GENOA Exam 2, E is the M-Value of the CpG site, W is the random effect for each sibship, and ε is the error term for the i^{th} individual in the j^{th} sibship

Supplementary Table 3. Comparison of baseline characteristics of Phase II Non-Epigenetics Participants with Phase II Epigenetics Participants. A) Continuous clinical variables, B) Continuous biomarkers, C) Categorical descriptive variables.

A) Continuous clinical variables

	Phase II Non-Epigenetics Participants				Phase II Epigenetics Participants				
Variable	**N**	**% Missing**	**Range**	**Mean (SD)**	**N**	**% Missing**	**Range**	**Mean (SD)**	**Pr > |t|**
Age, years	474	0.00%	26.41-81.52	56.07 (9.21)	1008	0.00%	39.26-94.74	66.34 (7.6)	6.6E-49
BMI, kg/m²	472	0.42%	18.1-57.9	32.6 (7.4)	1001	0.69%	16.41-55.09	31.1 (6.09)	0.013
Systolic BP, mm Hg	473	0.21%	96-208	134.8 (19.3)	1006	0.20%	79-221	139.89 (21.12)	2.2E-06
Diastolic BP, mm Hg	474	0.00%	53-122	81.57 (10.2)	1008	0.00%	45-121	78.29 (11.01)	2.3E-05
Pulse Pressure, mm Hg	474	0.00%	19-127	53.41 (15.11)	1006	0.20%	26-127	61.64 (17.72)	6.0E-17
Adj. Systolic BP, mm Hg (1)	473	0.21%	97-218	142.05 (21.47)	1006	0.20%	89-231	148.15 (22.74)	5.5E-08
Adj. Diastolic BP, mm Hg (1)	474	0.00%	58-127	85.2 (10.95)	1008	0.00%	50-126	82.42 (11.43)	4.6E-04
Adj. Pulse Pressure, mm Hg (1)	473	0.21%	24-132	56.89 (15.72)	1004	0.40%	28-132	65.65 (18.28)	2.2E-18
Total Cholesterol, mg/dL	459	3.16%	72-348.5	197.93 (40.02)	1008	0.00%	73.5-354.5	203.98 (41.95)	7.8E-05
Triglycerides, mg/dL	454	4.22%	28.5-419.5	111.95 (58.94)	1004	0.40%	37-402.5	117.9 (56.68)	9.8E-04
ln (Triglycerides+1)	457	3.59%	3.38-6.45	4.62 (0.49)	1007	0.10%	3.64-6.27	4.69 (0.43)	0.69
HDL-C, mg/dL	457	3.59%	23.8-125.8	55.42 (16.66)	1005	0.30%	21.7-130.35	58.08 (17.48)	7.0E-03
ln (HDL+1)	460	2.95%	3.21-5.17	4 (0.3)	1008	0.00%	3.12-5.05	4.04 (0.29)	0.75
LDL-C, mg/dL	459	3.16%	23.6-253.75	121.1 (36.88)	1008	0.00%	24.85-272.1	123.84 (39.58)	0.064
Glucose, mg/dL	457	3.59%	43.5-296	108.28 (38.68)	998	0.99%	49.5-290	110.56 (34.31)	0.11
ln (Glucose+1)	457	3.59%	3.8-5.69	4.65 (0.28)	1004	0.40%	3.92-5.82	4.69 (0.27)	0.77
Insulin, mU/mL	163	65.61%	1.14-52.46	9.34 (8.53)	1005	0.30%	0.22-115.76	10.39 (12.45)	0.30
ln (Insulin+1)	163	65.61%	0.76-3.98	2.12 (0.62)	1006	0.20%	0.2-5.02	2.12 (0.74)	1.00
Menopause Age, years (2)	270	43.04%	23-59	42.6 (8.26)	699	30.65%	25-62	43.72 (7.85)	0.14

1 If a participant is listed as taking hypertensive medications, 10 mm Hg is added to systolic blood pressure, and 5 mm Hg is added to diastolic blood pressure, else recorded blood pressure is used in this variable.

2 Menopause age is represented for females only.

B) Biomarkers of inflammation

	Phase II Non-Epigenetics Participants				Phase II Epigenetics Participants				
Variable	**N**	**% Missing**	**Range**	**Mean (SD)**	**N**	**% Missing**	**Range**	**Mean (SD)**	**Pr > |t|**
Serum Creatinine, mg/dL	460	2.95%	0.44-3.64	0.89 (0.29)	1008	0.00%	0.42-2.98	0.94 (0.3)	0.69
CRP, mg/L	347	26.79%	0.21-29.9	5.48 (5.8)	971	3.67%	0.21-29.9	6.05 (6.77)	0.37
ln (CRP+1)	347	26.79%	0.08-1.49	0.68 (0.33)	971	3.67%	0.08-1.49	0.7 (0.34)	0.87
Homocysteine, µmol/L	430	9.28%	4.5-25.2	9.38 (2.97)	1002	0.60%	4.7-31	10.57 (3.59)	7.1E-03
ln (Hycs+1)	430	9.28%	1.7-3.27	2.3 (0.26)	1002	0.60%	1.74-3.47	2.41 (0.28)	0.41
Fibrinogen, mg/dL	440	7.17%	123-684	362.44 (88.37)	1007	0.10%	120-680	369.35 (81)	1.6E-03
ICAM, ng/mL	199	58.02%	101-584	296.97 (83.09)	578	42.66%	17-625	275.38 (80.82)	3.2E-16
IL-6, pg/mL	241	49.16%	1.92-29.78	8.71 (5.06)	856	15.08%	1.42-36.45	9.09 (5.31)	0.54
ln (IL-6+1)	245	48.31%	0.46-1.58	0.95 (0.22)	878	12.90%	0.38-1.58	0.97 (0.22)	0.86
IL-18, pg/mL	240	49.37%	6.14-235.97	72.17 (38.66)	863	14.38%	1.86-252.42	69.92 (39.97)	0.19
ln (IL-18+1)	241	49.16%	0.85-2.49	1.81 (0.23)	872	13.49%	0.84-2.49	1.8 (0.25)	0.92
MPO, ng/mL	193	59.28%	11.74-141.47	47.43 (27.67)	954	5.36%	4-160.58	40.26 (25.04)	5.1E-07
ln (MPO+1)	196	58.65%	1.1-2.22	1.63 (0.23)	968	3.97%	0.7-2.22	1.56 (0.24)	0.60
MCP1, pg/mL	205	56.75%	247.44-2124.82	970.81 (319.89)	835	17.16%	159.81-2704.78	1053.43 (369.78)	2.3E-49
ln (MCP+1)	205	56.75%	2.4-3.33	2.96 (0.15)	849	15.77%	2.46-3.49	3.01 (0.16)	0.69
Resistin, ng/mL	312	34.18%	1-12.13	3.77 (1.96)	895	11.21%	1-12.32	4.07 (2.03)	0.40
ln (Resistin+1)	315	33.54%	0.3-1.2	0.65 (0.18)	902	10.52%	0.3-1.2	0.68 (0.17)	0.798
SAA, µg/mL	297	37.34%	2.35-146	26.75 (28.98)	731	27.48%	2.35-146	32.94 (30.88)	2.0E-05
ln (SAA+1)	297	37.34%	0.53-2.17	1.27 (0.39)	731	27.48%	0.53-2.17	1.38 (0.38)	0.51
TNFR-1, pg/mL	219	53.80%	102.78-3310.18	1113.11 (562.23)	872	13.49%	19.89-3844.07	1185.29 (546.83)	1.3E-27
ln (TNFR1+1)	217	54.22%	2.15-3.52	3 (0.21)	882	12.50%	2.17-3.68	3.04 (0.22)	0.76
TNFR-2, pg/mL	240	49.37%	388.3-4998.98	1751.03 (649.43)	849	15.77%	286.83-5333.39	1950.04 (831.08)	4.8E-117
ln (TNFR2+1)	241	49.16%	2.59-3.8	3.22 (0.16)	854	15.28%	2.59-3.8	3.26 (0.17)	0.70

C) Categorical variables

	Non-Epigenetics Participants			Epigenetics Participants		
	Count	**Total**	**% of Total**	**Count**	**Total**	**% of Total**
Female	335	474	70.68%	715	1,008	70.93%
Hypertensive	344	474	72.57%	833	1,008	82.64%
Has Diabetes	128	474	27.00%	308	1,008	30.56%
Ever Smoker	174	474	36.71%	421	1,008	41.77%

Supplementary Table 4. List of probes used to standardize methylated and unmethylated signals.

Red Channel Probes	Green Channel Probes
EXTENSION 1190050	BISULFITE CONVERSION 4670278
EXTENSION 360446	BISULFITE CONVERSION 4670484
NON-POLYMORPHIC 1740025	BISULFITE CONVERSION 5290048
STAINING 4200736	EXTENSION 1190050
STAINING 4570020	EXTENSION 360446
	EXTENSION 520537
	NON-POLYMORPHIC 1740025
	STAINING 5340168

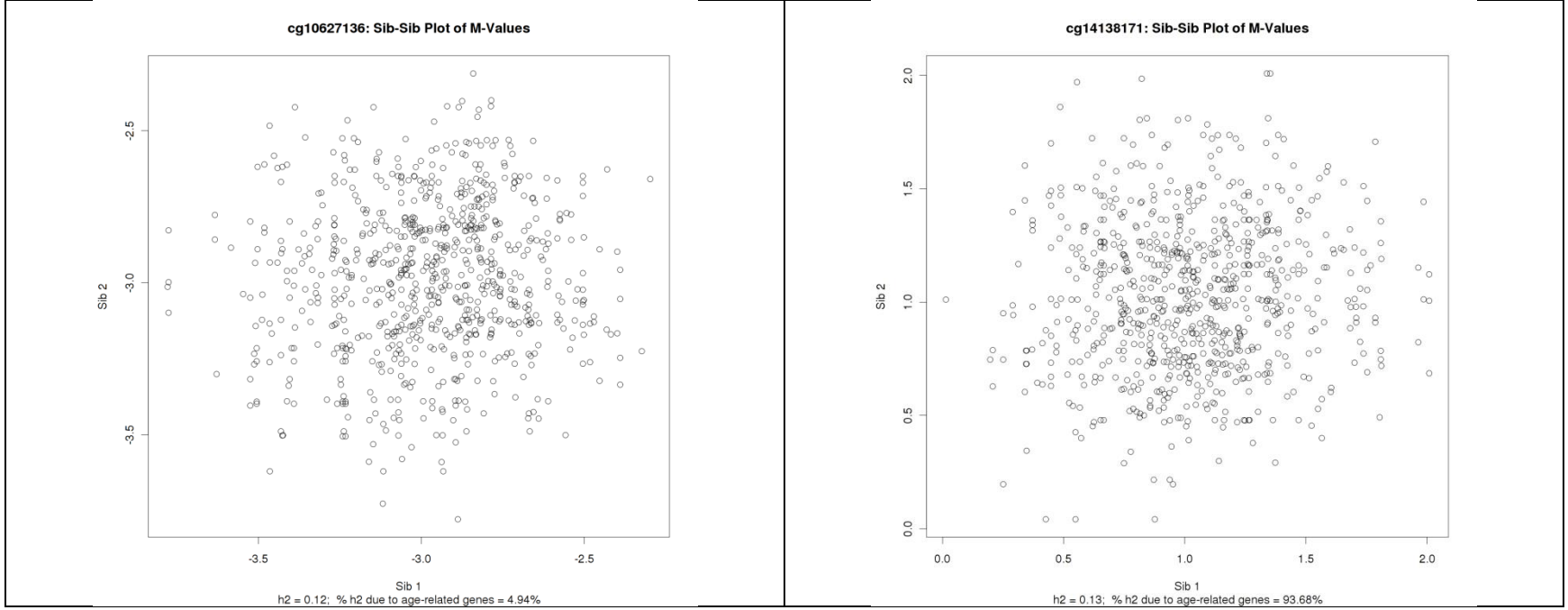
Supplementary Table 5. Sibship sizes in GENOA Epigenetic population.

Sibship Size	Sibship Count	Number of Participants
10	1	10
9	0	0
8	0	0
7	1	7
6	4	24
5	7	35
4	31	124
3	71	213
2	181	362
1	197	197
TOTAL:	**296 sibships 197 singletons**	**972**

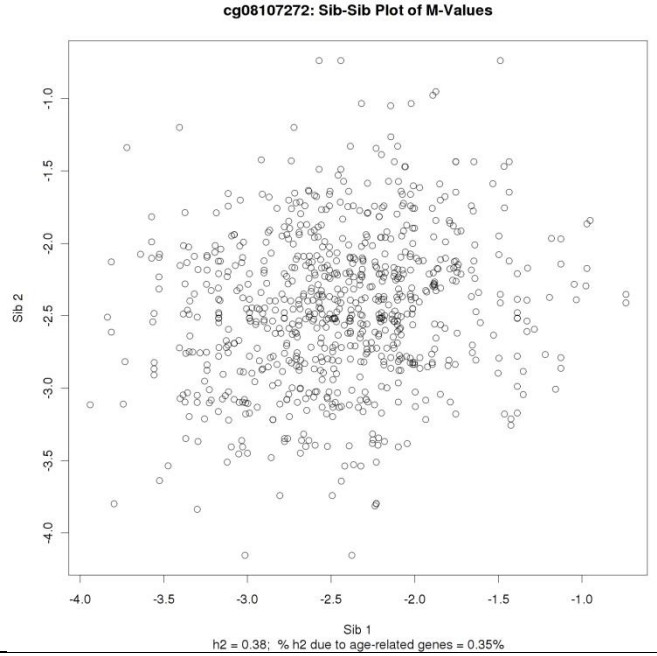
Average sibship size: 2.6

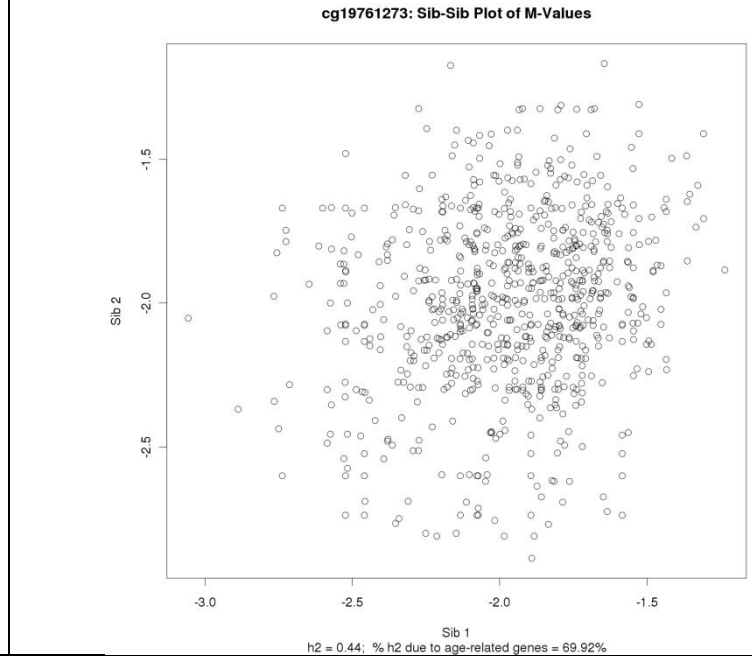
Supplementary Figure 1. Sib-Sib plots of a 6 CpG sites with significant (p<0.05) heritabilities. A) 2 sites with low h^2; B) 2 sites with medium h^2; C) 2 sites with high h^2 (h^2=1)

A) Sites with low heritability

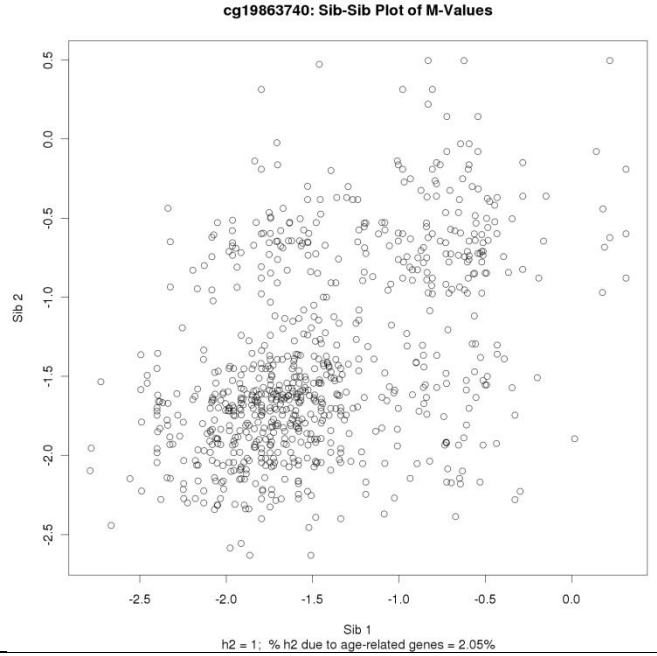


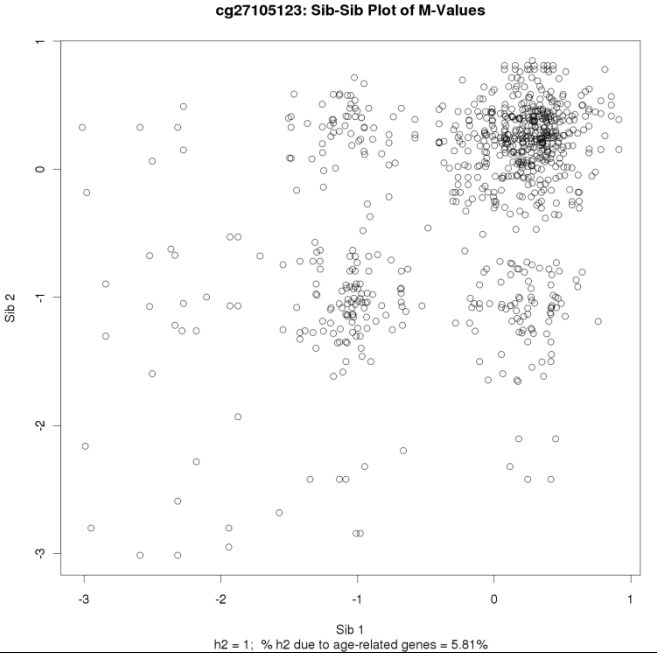
B) Sites with medium heritability





C) Sites with high heritability = 1





Chapter IV. Predicting Healthy and Unhealthy Aging with Epigenetic and Inflammatory Biomarkers

Introduction

Chronic diseases, such as cardiovascular diseases, chronic kidney disease, and diabetes, have long been considered as diseases of older age. Due to improved and increased health care technologies and pharmaceuticals, the United States population of senior citizens is projected to increase by 135.4% between 2000 and 2050, creating a larger proportion of the population suffering from chronic diseases and, thus, in need of health care services and long-term caregiving.(*1*) Moreover, increases in sedentary lifestyle, decreases in physical activity, and increases in caloric consumption have become more prevalent throughout the second half of the 20th Century and into the 21st Century, and represent a kind of unhealthy aging that will have longer term effects on individual and population health. For example, the demographics of the population suffering from chronic diseases are expanding to include younger and younger age groups.(*125*) Chronic diseases are among the leading causes of morbidity and mortality in the United States,(*3*) and place a substantial burden on activities of daily living and physical function.(*126*) Since chronic diseases are typically compounded (e.g. metabolic syndrome, hypertensive kidney disease), interventions at the earliest stages of disease progression are important in order to stop and possibly reverse the damage before it complicates physical function and quality of life.

Over the last several decades, numerous epidemiological studies have been conducted to identify the genetic, environmental, and metabolic factors that predict an individual's or population's risk of chronic diseases such as obesity, hypertension, diabetes, and dyslipidemia, which underlie the risk of vascular diseases.(*6-11*) These epidemiological studies have identified a wide range of lifestyle factors, anthropometric factors, and measures of lipid metabolism, glucose metabolism, blood pressure, and inflammation that now constitute our best predictors of future disease with advancing age.(*12, 13*) Since many common chronic diseases share a myriad of age-related risk factors, clinical measurements of these risk factors, such as blood pressure, body mass index (BMI), and cholesterol, may represent a way to assess the body's biological age (with respect to chronic disease outcomes). The difference between a person's biological age and their chronological age could then be used to assess their level of healthy or unhealthy aging. For instance, using chronic disease risk factors to estimate biological age for an individual may be of clinical importance to guide the direction and extent of treatment regimes in order to manage an individual's chronic disease risk.

Beyond traditional risk factors, novel biomarkers have been implicated in chronic disease and aging processes. Biomarkers of inflammation have been found to be associated with many common chronic diseases (e.g. Alzheimer's disease, cardiovascular disease, type 2 diabetes), and may be telltale signs of preclinical stages of these diseases.(*46, 48*) Further, epigenetics is known as a potential mediator of age and gene expression,(*40, 87*) that may be specifically involved in aging processes through pathways of chronic disease and inflammation.(*39, 41, 63-65*) Utilizing inflammatory markers and epigenetic markers of DNA methylation as biomarkers of preclinical

disease, we may be able to boost the efficacy of a predictive equation for biological age in order to ascertain the origins of healthy and unhealthy aging processes.

This study will use data from a longitudinal study of African American sibships from the Genetic Epidemiology Network of Arteriopathy (GENOA) to accomplish three goals: 1) estimate biological age using traditional chronic disease risk factors, and compare it to chronological age, 2) estimate the change in these traits as measures of healthy and unhealthy aging, and 3) investigate novel epigenetic and inflammatory predictors of aging processes.

Methods

Sample

The Genetic Epidemiology Network of Arteriopathy (GENOA) study is a community-based study investigating the genetics of hypertension and its arteriosclerotic complications in non-Hispanic whites from Rochester, MN and African-Americans from Jackson, MS.(*72*) In the current study, we investigated the capability of predicting aging using DNA methylation in GENOA African-Americans. African-American sibships were recruited such that ≥2 siblings were diagnosed with primary hypertension before the age of 60 years, while other siblings within the sibship were invited to participate independent of hypertension status (N=1,854). The initial examination (Phase I: 1996-1999) included standardized interviews concerning prescription drug usage, cigarette smoking, physical activity, history of hypertension, diabetes, and cardiovascular disease events; physical examination for blood pressure, height, weight, and waist and hip circumferences; and fasting blood samples for creatinine, total cholesterol, high-density

lipoprotein (HDL)-cholesterol, triglycerides, glucose, and insulin. Further, systolic and diastolic blood pressures (SBP and DBP) were adjusted for antihypertensive medication usage by adding 10 mm Hg to SBP (represented as adj. SBP) and 5 mm Hg to DBP (represented as adj. DBP) in order to more closely represent blood pressures as they would be without use of BP-lowering medications. Pulse pressure (PP) was calculated as the difference between SBP and DBP, and adjusted PP was the difference between adj. SBP and adj. DBP. For statistical purposes, each continuous variable was assessed for its approximate normality by examining histograms and estimating skewness and kurtosis. If the variable distribution is heavily skewed, natural log transformation of the variable plus a constant is made, as ln(variable+1). For a list of variables and their transformations, see Supplementary Table 7.

The second examination (Phase II: 2000-2004) included 1,482 participants returning from Phase I, and included re-measurement of interview, physical examination, and blood characteristics, as well as additional measurements of arteriosclerotic target organ damage, including heart, kidney and peripheral artery traits. DNA methylation was quantified on 1,008 Phase II participants using stored blood samples collected during the second examination. (Comparisons between Phase II individuals within and outside of the DNA methylation study are available in Supplementary Table 6). The third examination (Phase III: 2008-2011) included 752 African-American Phase II participants, and included re-measurement of interview, physical examination, and fasting blood characteristics from Phases I and II, as well as measurements of cognition and physical function, and coronary artery calcification (computed tomography). Written informed consent was obtained from all subjects and approval was granted by participating

Institutional Review Boards. Complete information on the GENOA study population and sample measurements can be found detailed in Appendix 3.

Measurement of DNA methylation

Samples were prepared and DNA methylation was measured according to previously published methods (*73*) which are summarized as follows. For comprehensive information on these measurements, refer to Appendix 4.

Sample Preparation and Methylation Assay. DNA was isolated from peripheral blood leukocytes from GENOA Phase II stored samples, and bisulfite converted with the EZ DNA Methylation Gold Kit (Zymo Research, Orange CA). Bisulfite-converted DNA samples were whole-genome amplified, enzymatically fragmented, and purified, then hybridized to Illumina Infinium HumanMethylation27K BeadChips, which contained locus-specific DNA oligomers and a set of 56 control probes. The array was then fluorescently stained, scanned using the Illumina BeadXpress reader, and assessed for fluorescence intensities across the methylated and unmethylated bead types at 27,578 CpG sites.

Data processing and methylation quantification: At each CpG site, fluorescent signals were measured from the site-specific **M** (methlyated) and **U** (unmethylated) bead types. The raw fluorescence data from the scanner was processed using Illumina BeadStudio software. To reduce batch and chip effects, the correlation structure among all 56 control probes was evaluated within channel to identify the most parsimonious subset of probes that explained the maximum amount of batch and chip variation across samples (5 probes in the red channel and 8 probes in the green channel; Supplementary

Table 8). Quality control standards were employed by linearly regressing the 13 selected probes onto the intensity signals from the methylated and unmethylated bead types separately across each CpG site.

For each individual and across every CpG site, the M-Value was calculated from the **M** and **U** values as a continuous measure of methylation. The M-Value is a commonly used measurement in microarray analysis that was more recently adapted for use in DNA methylation array data due to its ability to equalize the variance across the CpG sites.(*74, 75*) The M-Value for each individual *i* at a single site, *k*, is calculated as: $\text{M-Value}_{ik} = \log_2[(\max(\mathbf{M}_{ik},0) + 1) / (\max(\mathbf{U}_{ik},0) + 1)]$, where a constant is added to prevent large-scale changes caused by small intensity estimation errors.(*75*) While many studies using the Illumina Infinium HumanMethylation27K array present results of the Beta Value, which are continuous variables between 0 and 1 that are calculated from the measured **M** and **U** intensities and represent the percent methylation at a particular CpG site within an individual, we will present results of the M-Value since we feel that it has more desirable statistical properties which allow for more precise conformation to modeling assumptions. The relationship between Beta and M-Values represents a logit transformation, where $\text{M-Value}_{ik} = \log_2 [\text{Beta Value}_{ik} / (1 - \text{Beta Value}_{ik})]$. Unmethylated M-Values are considered to be < -2.0, methylated M-Values are > +2.0, and semi-methylated M-Values are between -2 and +2.

Before statistical analysis, samples were checked for data quality. From the 1,008 study participants, 7 individuals were removed due to poor bisulfite conversion control efficiency, measured by bisulfite conversion control intensity of < 4,000. An additional 29 individuals were removed from the analysis due to extreme control probe values,

assessed as having at least one control probe with a value of greater than 4 standard deviations from its mean value. This resulted in a total sample size of 972 individuals.

For this study, we analyzed only autosomal CpG sites. A total of 58 CpG sites were removed from the analysis because they were found to be multimodal based on the Dip Test proposed by Hartigan and Hartigan (*76*) using a cut-off of $p<0.001$ on the signal intensities of the methylated and/or unmethylated bead types, which clearly violated the statistical modeling assumptions. This resulted in a total number of 26,428 CpG sites included in our analysis. Finally, 2,984 non-specific binding probes and 908 polymorphic probes overlapping with single nucleotide polymorphisms (SNPs) (*77*) were identified and denoted in the result tables. Though these sites were not removed from the analysis, we have interpreted the results from these sites with caution. As a final method of quality control, DNA methylation values greater than ±4 standard deviation from each M-Value's mean were removed from each of the 26,428 CpG sites.

Measurement of Inflammatory Biomarkers Measurement

Twelve protein markers of vascular disease were measured in either plasma or serum from Phase II blood samples, using commercially available solid-phase immunoassays and immunoturbidometric assays.(*127, 128*) The following nine markers were measured in plasma: fibrinogen, serum amyloid A (SAA), interleukin 6 (IL6), interleukin 18 (IL18), tumor necrosis factors 1 and 2 (TNFR1 and TNFR2), intercellular adhesion molecule (ICAM), monocyte chemotactic protein-1 (MCP1), and resistin. The three markers measured in serum were C-reactive protein (CRP), myeloperoxidase (MPO), and homocysteine (Hycs). Additional information on biomarker assays, including precision, accuracy, stability, and methods of quality control have been

previously described,(*128*) , and further information on these methods may be found in Appendix 2.

Statistical Analyses

Building predictive models for age using traditional chronic disease risk factors.

Using the *nlme* package in the statistical package R,(*78*) a stepwise regression approach using linear mixed modeling was used to select the model that best predicted age of GENOA Phase I participants using traditional chronic disease risk factors, listed in Table 8. This modeling schema can be found pictorially in Supplementary Figure 2, and is expressed as $Age_{ijk} = \beta_0 + \sum \beta_k \cdot RF_{ijk} + W_{0jk} + \varepsilon_{ijk}$ such that RF is the k^{th} risk factor, higher order term of a risk factor, or interaction between two risk factors of the i^{th} individual from the j^{th} sibship, W is the random effect for each sibship, and ε is the error.

Traditional chronic disease risk factors in Phase I were evaluated for potential collinearity (correlations available in Supplementary Table 9). For pairs of variables with high correlations ($r > 0.8$ or $r < -0.8$) we excluded the second variable from further analysis if the first variable was already in the regression model.

To begin the traditional stepwise regression method using linear mixed modeling with sibships as the random effect, we started with an empty model in which the outcome was Phase I Age. All traditional risk factors were available in a pool of available variables, and a univariate regression between age and each variable was performed. The variable with the lowest regression coefficient p-value was added to the model. After a variable was added to the model, it was removed from the pool of available variables, and the next higher order term of that variable was added into the pool. Also added into the

pool of available variables were the two-way interaction terms for the newly added variable and any variable already expressed as a predictor in the current model. Then, each of the variables within the pool was examined as an addition to the currently selected model to identify which variable had the most significant p-value. After each new variable was added to the model, the p-values of the regression terms were re-examined in the new model. If any previously added variable within the model had a p-value > 0.1, that variable was removed from the model and returned to the pool of available variables, as was any interaction term with that variable.

At each step of the stepwise regression, the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and LR R^2 were recorded. Stepwise regression was halted after no further variables could be added while retaining $p<0.10$. By using AIC and BIC in parallel, we selected the final model that best described the data (AIC) and best described the true model that generated the observed data (BIC)(*129*) with the fewest number of variables. The final best-fitting model was chosen to have the lowest BIC of all models evaluated, while retaining a very low AIC, albeit not the lowest of all the models. The final model contained fewer predictors than the model with the absolute lowest AIC. Information about all steps calculated during stepwise regression can be found in **Error! Reference source not found.**.

Measures of Healthy Biological Age.

When the difference (Δ) between chronological (actual) age and biological (predicted) age was calculated, a positive Δ value indicated healthy biological age, whereas a negative Δ value indicated an unhealthy biological age. Using the Phase I (baseline, N=1,854) data in GENOA, the stepwise regression method outlined above was

used to determine the risk factors and their estimated regression relationship with age (as the outcome), which was then applied to Phase II epigenetic participants (N=972) and epigenetics participants who were followed to Phase III (N=484). The average within-family age difference between Phases I and the Phase being predicted was also calculated and added to the predicted age to maintain family effects in the predicted biological age. For each GENOA Phase of data collection, the residual between actual and predicted age was calculated (Δ_1, Δ_2, or Δ_3). Positive measures of Δ represent healthiness at that time point, while negative measures of Δ represent unhealthiness. For example, if an individual were actually 40 years old, but was predicted as younger than their actual age, say 30 years, then Δ = 40-30 = 10 years younger, biologically.

Measures of Healthy Biological Aging.

The longitudinal nature of the GENOA study also allowed us to examine changes in the measurement of biological age relative to changes in chronological age (ζ) to estimate healthy or unhealthy *aging*. Given three phases of data collection, we could estimate two measures of aging. For example, the changes in predicted age relative to changes in chronological age between Phase I and Phase II is $\zeta_1 = \Delta_2 - \Delta_1$. Similarly, for the time period of Phase II to Phase III, $\zeta_2 = \Delta_3 - \Delta_2$ = (Phase III Chronological Age – Phase II Chronological Age) – (Phase III Predicted Age – Phase II Predicted Age).

A value of ζ_1 is **positive**, indicating *unhealthy aging*, under three conditions: if the latter time point exhibits health ($\Delta_2 > 0$) and either 1) the earlier time point also exhibits health, but is closer to 0 (less healthy) than the latter time point ($\Delta_2 > \Delta_1 > 0$), or 2) the earlier time point is unhealthy, such that $\Delta_2 > 0 > \Delta_1$, or 3) the latter time point is unhealthy, but not as unhealthy as the earlier time point ($0 > \Delta_2 > \Delta_1$). Similarly, a value

of ζ_1 is **negative**, indicating *healthy aging*, under three conditions: if the latter time point is unhealthy ($\Delta_2 < 0$) and either 1) the earlier time point also unhealthiness, but is closer to 0 (healthier) than the latter time point ($\Delta_2 < \Delta_1 < 0$), or 2) the earlier time point is healthy and the latter time point is unhealthy, such that $\Delta_2 < 0 < \Delta_1$, or 3) the latter time point is healthy, but not as healthy as the earlier time point ($0 < \Delta_2 < \Delta_1$). The same lemmas exist when describing ζ_2.

Build the best predictive models for each of ζ_1 and ζ_2 with measures of inflammation.

Since one of the main goals of this study is to predict healthy aging, we used a similar stepwise regression technique outlined above to develop multivariable models predicting each of ζ_1 and ζ_2 using 12 measures of inflammation. For this modeling schema, each of the measures of healthy aging, ζ_1 and ζ_2, were adjusted by its baseline measure of health (Δ_1 or Δ_2, respectively on the model), then 12 measures of inflammation were added to the available in the variable pool. This model can be written as: $\zeta_{1ijk} = \beta_0 + \sum \beta_k \cdot B_{ijk} + W_{0jk} + \varepsilon_{ijk}$ such that B is the k^{th} inflammatory biomarker, higher order term of an inflammatory biomarker, or interaction between two inflammatory biomarkers of the i^{th} individual from the j^{th} sibship, W is the random effect for each sibship, and ε is the error, and is written similarly for ζ_2.

Build the best predictive models for each of ζ_1 and ζ_2 with measures of epigenetics.

In our previous studies,(*103*) we identified 2,095 DNA methylation sites that were significantly associated with age (after Bonferroni correction). To reduce multiple testing burden, we only considered these 2,095 sites as potential predictors of the Δs and ζs. Using univariate linear mixed models that allow us to take into account sibship structure, we identified the CpG sites (Bonferroni corrected p-values = 2.71×10^{-5}) that were

associated with Δ_1, Δ_2, and Δ_3 such that $\Delta_{1ijk} = \beta_0 + \beta_1 E_{ijk} + W_{0jk} + \varepsilon_{ijk}$ such that E is the k^{th} CpG site M-Value of the i^{th} individual from the j^{th} sibship, W is the random effect for each sibship, and ε is the error, and is written similarly for Δ_2 and Δ_3. Then, only sites that were significant (after Bonferroni correction) for Δ_1 and Δ_2 were tested against ζ_1, while sites significant (after Bonferroni correction) for Δ_2, and Δ_3 were tested against ζ_2 written as $\zeta_{1ijk} = \beta_0 + \beta_1 E_{ijk} + W_{0jk} + \varepsilon_{ijk}$, and similarly for ζ_2. Multivariable models of CpG sites were also examined with either ζ_1 and ζ_2 using similar stepwise regression as described previously.

Further, two sets of principal components were calculated based on (1) the joint set of CpG sites that were significant in the associations with Δ_1 and Δ_2, and (2) the joint set of sites that were significant in association with Δ_2 and Δ_3. The corresponding top 1, 5, and 10 principal components were then used as explanatory variables of ζ_1 and ζ_2 in linear mixed multivariable modeling. Lastly, multivariable models including the inflammatory biomarkers modeled in the previous step and CpG sites and PCs were investigated to determine the most predictive model for each of ζ_1 and ζ_2.

Results

Description of data.

The population on which the initial aging model was built was from Phase I (1996-1999) of the GENOA sample, and then applied to Phases II (2000-2004) and III (2008-2011) samples. Descriptive statistics are provided in Table 8. In brief, Phase I participants were mostly female (69.1%), and hypertensive (72.6%). The average age increased 13.41 years from Phase I to the Phase III epigenetics participants. Diabetes

prevalence increased from 21.6% to 35.7%, while hypertension prevalence increased from 72.6% to 87.4% between Phase I and Phase III epigenetics participants. Body mass index (BMI), systolic blood pressure (SBP), and pulse pressure (PP) also increased over time. Total cholesterol, triglycerides, and low density lipoprotein cholesterol (LDL-C) decreased while high-density lipoprotein cholesterol (HDL-C) increased slightly between Phases I and III. Glucose increased slightly, while insulin and serum creatinine (SCr) decreased slightly.

Biological age, healthy age, and healthy aging phenotypes.

Initially, a model was built using traditional chronic disease risk factors to best represent chronological age, so that a biological age could be estimated. The best fit model (Table 10) represented risk factors from all of the major chronic diseases – for example, it contained a measure of obesity (BMI), measures of hypertension (pulse pressure, adjusted diastolic BP), measures of diabetes (glucose, insulin), measures of dyslipidemia (HDL, triglycerides), and a measure of chronic kidney disease (serum creatinine), as well as interaction terms (diastolic BP and BMI, diastolic BP and serum creatinine, pulse pressure and glucose), and higher order terms (pulse pressure squared, diastolic BP squared, serum creatinine squared, and ln glucose squared). The likelihood ratio (LR) R^2 for the model was 0.515.

The regression model from Phase I was applied to data from Phases II and III, with a family-based addition to each of the intercept terms equal to the average within-family age difference between the Phase being calculated and Phase I. The mean average within-family age difference between Phases I and II was 4.66, and was 11.07 years between Phases I and III. Biological age ranged from 26.83 to 80.91 (mean = 58.19),

41.17 to 91.29 (mean = 66.08), and 48.27 to 92.85 (mean = 74.14) in Phases I, II, and III, respectively. The difference between chronological and biological age ranged from -15.79 to 15.15 (mean = 0), -12.76 to 19.64 (mean = 0.16), and -18.03 to 16.84 (mean = -2.52) for Δ_1, Δ_2, and Δ_3, respectively, and ranges of healthy aging were -15.37 to 22.57 (mean = -0.39), and -15.45 to 14.04 (mean = -2.25) for ζ_1 and ζ_2, respectively. Healthy aging between Phases I and II had a heritability of 0.43, even after adjusting for Phase I age, and healthy aging between Phases II and III had a heritability of 0.30, even after adjusting for Phase II age. The distributions of chronological age, biological age and residuals across all three GENOA phases, and ζ_1 and ζ_2 are listed in Supplementary Table 11.

Plots of chronological versus biological age for all three Phases are presented in Supplementary Figure 3. In brief, the proportion of variation in chronological age explained by biological age was 80.3% in Phase I, 58.1% in Phase II, and 56.3% in Phase III. It can be noted that those in the youngest age groups (chronological age < 40) were predominantly being predicted as older than their actual age, and those in the oldest age groups (chronological age > 70) were being predicted as younger than their actual age. Further, plots of the measure of health, Δ, for each participant at Phase I vs. Phase II (Supplementary Figure 4) show that 364 individuals (38.9%) have positive values at both time points indicating they exhibit a healthier than expected age at both time points, 307 individuals (32.8%) have negative values at both time points indicating they exhibit a less healthy than expected age at both time points, 104 individuals (11.1%) were unhealthy at Phase I and healthy at Phase II, and 161 individuals (17.2%) were healthy at Phase I and unhealthy at Phase II. A paired t-test indicated that there is a significant difference in

healthy age at Phase I versus II (Δ_1 vs. Δ_2, $p=4.42x10^{-3}$). Evaluating individuals at Phase II and Phase III, 69 (20.0%) individuals had positive Δs (healthy) at both time points, 156 individuals (45.2%) were had negative Δs (unhealthy) at both time points, 27 individuals (7.8%) were unhealthy at Phase II and then healthy at Phase III, and 93 individuals (27.0%) were healthy at Phase II and unhealthy at Phase III. A paired t-test indicated that there is a significant difference in healthy age at Phase II versus III (Δ_2 vs. Δ_3, $p=2.96x10^{-19}$). Finally, when examining the difference in healthy aging over the two time periods (Phase I to II and Phase II to III), in Supplementary Figure 5, we see that 124 (35.6%) experienced healthy aging throughout both time periods, 20 (5.8%) experienced unhealthy aging throughout both time periods, 80 (23.2%) experienced healthy aging from Phases I to II and unhealthy aging between Phases II to III, and 121 (35.1%) experienced unhealthy aging between Phases I and II followed by healthy aging between Phases II and III. A paired t-test indicated that there is a significant difference in healthy aging between Phases I and II versus between Phases II and III (ζ_1 vs. ζ_2, $p=3.18x10^{-5}$).

Explaining Healthy Aging with Inflammatory Markers.

There were three inflammatory markers that were significant (with p-values < 0.10) in predicting the period of aging between Phases I and II in GENOA (ζ_1). Fibrinogen, ln (TNFR1+1), and ln (Hycs+1) were able to explain an additional 0.4% of the variation in ζ_1 beyond that of health at Phase I (Δ_1), which was already able to explain 11.5% of the variation in ζ_1. Aging over the period between Phases II and III in the GENOA population (ζ_2) only was significantly predicted (at $p<0.10$) by ln (MCP1+1) and ln $(MCP1+1)^2$. These variables were able to explain an additional 3.8% of the variation in ζ_2 beyond the health of individuals at Phase II (Δ_2), which already explained

22.0% of the variation in ζ_2. Thus, for each of ζ_1 and ζ_2, there is only a small amount of additional variation explained by inflammatory markers on top of health status at the period's baseline (see Table 11).

Explaining Healthy Aging with DNA Methylation Markers.

Of the 2,095 CpG sites that were significantly associated with age (at Bonferroni corrected α=0.05, or 0.05/26,428), 244 were significant with Δ_1, 147 were significantly associated with Δ_2, and 6 were significantly associated with Δ_3 at $p< 2.37x10^{-5}$ (Bonferroni corrected α=0.05, or 0.05/2,095). Among the 244 sites significantly associated with Δ_1 or Δ_2 (where 94 sites were significant with both Δ_1 and Δ_2), 3 CpG sites were significantly associated with ζ_1 at $p< 2.05x10^{-4}$ (Bonferroni corrected α=0.05, or 0.05/244). Among the 150 sites significant with Δ_2 or Δ_3 (where 3 sites were significantly associated with both Δ_2 and Δ_3), 7 CpG sites were significantly associated with ζ_2 at $p< 3.33x10^{-4}$ (Bonferroni corrected α=0.05, or 0.05/150). Figure 6 contains information about the number of sites that were significant after regression on age, Δ_1, Δ_2, Δ_3, ζ_1, and ζ_2. Descriptions of the 3 CpG sites significantly associated with ζ_1, and 7 sites significantly associated with ζ_2 are listed in Table 12.

Principal components (PCs) were estimated using the 244 sites significant with either Δ_1 and Δ_2, and the top 10 PCs were modeled univariately, and in a multivariable model against ζ_1. The top 5 of these PCs were able to explain 56.3% of the variation in all 244 sites, and the top 10 PCs were able to explain 62.2% of the variation in all 244 sites. The top 5 principal components were able to explain an additional 2.2% of the variation in ζ_1 beyond that explained by the health of individuals at Phase I (Δ_1). The top

10 PCs, together, were able to explain an additional 3.0% of the variation in ζ_1 beyond that explained by Δ_2 (Table 13, A).

A second set of PCs was estimated on the 150 sites significant with either Δ_2 and Δ_3, and the top 5 PCs were able to explain 51.0% of the variation in those 150 sites, while the top 10 were able to explain 59.8% of the variation. When modeled in a multivariable model to explain ζ_2, the top 5 PCs were able to explain an additional 4.8% of the variation in ζ_2 beyond that explained by health of individuals at Phase II (Δ_2), and the top 10 PCs were able to explain an additional 6.1% of the variation in ζ_2 beyond that explained by health of individuals at Phase II (Δ_2) (Table 13, B). To note, when modeling PC5 against ζ_2 in a univariate model, the model was not able to converge.

Explaining Healthy Aging with Inflammatory and DNA Methylation Markers.

Given what our previous findings of (1) inflammatory markers and (2) CpG sites that best predicted ζ_1 and ζ_2, we then tried to further predict healthy aging over the time periods of Phase I to II (ζ_1) and Phase II to III (ζ_2) using these sets of novel biomarkers collectively. We found that the best predictive model for ζ_1 added 2 CpG sites to the model with the inflammatory biomarkers fibrinogen, ln(TNFR1+1) and ln(Hycs+1), and health of individuals at Phase I (Δ_1). The two significant CpG sites ($p<0.10$) within the model were cg25494227 and cg04474832 which are within the genes *C12orf59* and *ABHD14A*, respectively. These two CpG sites were able to explain an additional 4.64% of the variation in ζ_1 above the model with the three inflammatory biomarkers, and the model with the 3 inflammatory biomarkers and 2 CpG sites was able to explain an additional 5.06% of the variation in ζ_1 than the model with only Δ_1. The final model for

ζ_1 is represented below in Table 14, with stepwise selection steps available in Supplementary Table 12 A.

The best fit model for ζ_2 contained the inflammatory biomarkers, ln (MCP1+1) and ln $(MCP1+1)^2$, as well as 3 of the 7 CpG sites (cg05501357, cg02533173, and cg16005443) previously found to be significant for ζ_2, an interaction term between ln(MCP1+1) and cg05501357, one of the PCs (PC2) estimated on the 150 CpG sites that were significant after Bonferroni correction with Δ_2 and Δ_3, and the health of individuals at Phase II (Δ_2). These DNA methylation and inflammatory markers were able to explain an additional 14.5% of the variation in ζ_2 beyond that explained by the health of individuals at Phase II (Δ_2). Further, adding the 3 CpG sites, PC2, and the interaction term to these inflammatory markers was able to explain an additional 10.6% of the variation in ζ_2. The final model for ζ_2 is represented in Table 15, with model selection steps available in Supplementary Table 12 B.

Discussion

Healthy aging is not a simple concept to measure. Other investigators have categorized healthy aging by using self-reported measures of disease status, disability, physical function, and cognitive function, as in the Health and Retirement Study.(*130*) In the study presented here, we develop continuously distributed representations of healthy aging using clinical measurements of chronic disease risk factors in order to maximize our ability to identify new predictors of aging. Further, as we explore the phenomenon of healthy aging within a specific population, rather than attempting to compare it to a completely healthy population, we are able to understand healthy aging relative to the

overarching aging status of the population average. Since health status, in general, deteriorates throughout older age, comparing an individual's level of healthy or unhealthy aging within its population may allow the segregation of distinct groups that can be targeted for greater medical and social interventions. Thus, in this study, models to predict biological age (from which healthy aging was later assessed) in GENOA were purposely constructed using Phase I of the population, since this population was the baseline community-based sample of African American sibships that have hypertension or family history of hypertension.

Our study found that the correlation between chronological and biological ages was relatively high throughout the three phases of measurement. The R^2 between chronological and biological ages as measured in the GENOA sample was 0.80 at Phase I, 0.58 at Phase II, and 0.56 at Phase III. The decreases in R^2 between phases is explainable by the fact that the initial model to represent healthy aging was constructed on the healthier GENOA cohort from Phase I before being applied to Phase II, where many more people exhibited signs of chronic disease, and in Phase III where the cohort exhibited biases due to survival and morbidity. Thus, the shift to lower correlations between chronological and biological ages after Phase I is likely due to the shift toward unhealthiness with age into Phase II, which is expected in this aging population.

Studies conducted by other research groups relied on some variables not available within the course of the GENOA study, which were found to have predictive power in their models. For example, Jee, Jeon, Kim, Kim, Choe, Park and Jin (*131*) were able to attain a predictive capability for chronological age of 42.3% in men and 46.8% in women when they employed models using measures of physical function, including physical

measures such as vertical jump and cardiopulmonary measures such as forced expiratory volume. Another recent study used hormonal and biochemical signatures to increase predictive value of chronological age, such as prostate specific antigen (PSA), testosterone, dehydroepiandrosterone sulfate (DHEA-S), total antioxidant status and others, which aided in achieving an R^2 of 0.66 for males and 0.62 for females.(*132*) Though these models indicated lower or similar correlations between chronological and biological age to those found within GENOA, future studies may attain a better prediction of biological age that is more representative of the holistic aging process if measures of physical and cognitive function, hormonal regulation, psychosocial or lifestyle factors, and other suggested indicators of healthy aging(*133, 134*) were incorporated. However, variables representing physical and cognitive function were not available within the GENOA cohort at all three phases.

Healthy and unhealthy aging (ζ) was established as the change across time points of the differences between chronological age and biological age (Δ), upon which we were then able to examine the predictive capabilities of epigenetic and inflammatory biomarkers. Both the epigenetic and inflammatory markers were measured using Phase II samples. Consequently, they were measured after the aging process from Phase I to Phase II was measured, and before the aging process from Phase II to Phase III was measured.

Within the GENOA sample, we found that 3 inflammatory markers, homocysteine, TNFR1, and fibrinogen, as well as 2 CpG sites, cg25494227 and cg04474832, had statistically significant associations with ζ_1, while monocyte chemotactic protein-1 significantly predicts ζ_2 within the GENOA population.

Regression relationships of ln(Hycs+1), ln(TNFR1+1), and fibrinogen as predictors for ζ_1 were positive (Table 14), indicating that as levels of these inflammatory biomarkers increased the magnitude of ζ_1 also increased, demonstrating a shift toward unhealthier aging with increases in these biomarkers. Previous works have shown inconsistent results in the association between homocysteine and cardiovascular disease,(*135-140*) with conclusions that homocysteine may be a marker of unhealthy lifestyle and poor dietary choices more so than an independent risk factor for cardiovascular disease.(*138*) The positive association between homocysteine and ζ_1 identified in our analysis may actually indicate the importance of dietary factors with respect to DNA methylation within aging processes. In brief detail, dietary methyl donors and cofactors carrying 1-carbon units enter the methionine pathway, wherein homocysteine is an intermediary amino acid formed during the conversion of methionine to cytosine, and methyl groups needed for DNA methylation and other biological processes are imminently derived.(*141*) Since deficiencies in vitamins B6, B12, and folate, can increase blood levels of homocysteine(*142-145*), it is possible that poor dietary choices act within the context of DNA methylation to be indicative of inflammatory-related disease processes. Further insight into the relationship between ζ_1 and inflammation can be illuminated with respect to TNFR-1 and fibrinogen. TNFR-1 binds TNF-α to cells and mediates activation-induced cytotoxic effects of T-cells.(*146*) TNF-α is central to the inflammatory response, as it regulates leukocyte activation, maturation, cytokine and chemokine release, and production of reactive oxygen and nitrogen intermediates.(*147*) Further, TNF-α may play a central role in atherosclerotic processes due to its interaction with leukocytes as well as endothelial cells, due to recruitment of activated leukocytes, and adipocytes, due to the

perturbation of lipid metabolism.(*147*) Fibrinogen is also known to be central to the development and progression of atherosclerotic plaques.(*148*) High levels of fibrinogen may cause platelet aggregation and increased blood viscosity alterations leading to hypercoagulation, which may slow circulation and increase arterial damage.(*149*) Adding two CpG sites (significant at $p<0.10$) was able to explain an extra 4.6% explanatory power to the model predicting ζ_1. The CpG sites were within *C12orf59* on chromosome 12, and *ABHD14A* on chromosome 3. *C12orf59* is known to interact with *ELAVL1* on chromosome 9.(*150*) *ELAVL1* is a ubiquitous RNA-binding protein that promotes translation of mRNAs. It binds to many proteins known involved in cancer phenotypes, such as cell proliferation, increased cell survival, elevated local angiogenesis, and metastasis. It also promotes expression of pro-inflammatory cytokines such as TNF-α and IL6, and its expression is suggested to be involved in rheumatoid arthritis, inflammatory bowel disease, asthma, and atherosclerosis.(*151*) Expression of *ABHD14A* is suggested to be positively regulated by *ZIC1* during cerebellar development.(*152*) Further, *ZIC1* is found to bind to the *APOE* gene promoter.(*153*) Variants in *APOE*, of course, are known to be major genetic risk factors for late onset Alzheimer's Disease.(*154*) The negative association of CpG sites within these regions and ζ_1 could imply that aging processes may be related to demethylation of gene regions that affect important adverse aging processes such as inflammation and neurodegeneration.

The analysis of key predictors of aging over the time period from Phase II to Phase III (ζ_2) showed that one inflammatory marker (monocyte chemotactic protein-1), 3 CpG sites (cg05501357, cg02533173, cg16005443), an interaction between

cg05501357 and MCP1, and a principal component (estimated from 150 CpG sites) were significant (at $p<0.10$) in the prediction of healthy aging over this time period. Monocyte chemotactic protein-1 has a unique relationship with ζ_2, where its linear term is negatively associated, while its squared term is positively associated. Further, ln(MCP1+1) also has a statistically significant interaction with cg05501357, which is found in the *HIPK3* gene. MCP1 recruits monocytes to sites of active inflammation, and stimulates their maturity into macrophages,(*155, 156*) and elevated levels of MCP1 has been implicated in the role of many disease processes throughout various life stages, including insulin resistance and obesity.(*157, 158*) Further, MCP1 may play a fundamental role in the initiation and progression of atherosclerosis due to its role in recruitment of macrophages to vascular lesions,(*159, 160*) is one of the predominant chemokines involved in central nervous system (CNS) inflammatory process, plays a large role in bone remodeling, and plays a pivotal role in the genesis of kidney damage and renal dysfunction, especially within diabetic nephropathy.(*161*) Polymorphisms within the gene encoding MCP1 have been found to be associated with increased risk of individuals suffering coronary artery disease.(*162*)

The three CpG sites found to be significantly associated with ζ_2 ($p<0.10$), cg05501357, cg02533173, and cg16005443, are found in genes *HIPK3, BRD4,* and *LILRB3*, respectively. Overexpression of *HIPK3* may be associated with decreased sensitivity to Fas-mediated apoptosis, which has the primary role of regulating immune response and tissue-function, including regulating T-cell and B-cell development, maturation, and deletion.(*163*) *BRD4* is known to play a crucial role in maintaining the cell cycle. It has been shown to be involved in cancer, and is thought to co-activate pro-

inflammatory genes.(*164*) *LILRB4* is also shown to be involved in immune function through encoding the *LIR* protein expressed on immune cells on which MHC class I molecules bind to inhibit immune response.(*165*) Further, to evaluate the importance of the PC found significant within this analysis, we used the Expression Analysis Systematic Explorer (EASE) to identify enriched biological pathways(*119*) within the 150 CpG sites from which principal components were estimated. While no significant results were found that passed Bonferroni correction, the biological role of neurogenesis indicated an EASE score of 0.04. This indicates the potential relationship of DNA methylation and cognitive diseases, such as Alzheimer's disease, within the scope of the aging process. Previous research has indicated the importance of the APOE ε4 allele within Alzheimer's disease risk,(*166*) specifically within African Americans,(*167*) and, importantly, the allele falls within a CpG Island that, when hypermethylated, indicates higher risk of late-onset Alzheimer's disease.(*168*) Furthermore, hypertension has been linked to the manifestations of Alzheimer's disease and vascular dementia, though the pathophysiology of this relationship is not well-understood.(*169-173*) This potential relationship between hypertension and dementia may explain why the current study of a predominantly hypertensive population (82.5% in Phase II, 87.3% in Phase III) has uncovered a set of DNA methylation markers within genes that are over-representative of those with roles in neurogenesis.

Using CpG sites to predict biological age is a relatively new concept. A recent study by Hannum, et al.(*174*) was able to predict biological age, and create an "apparent methylomic aging rate" (ratio of biological age to chronological age) in a sample of 656 individuals ranging from 19 to 101 years, using 71 CpG sites from the Illumina 450K

platform and clinical measures such as gender, diabetes status, and BMI. Biological age was able to explain 96% of the variation in chronological age in this initial sample. Further, they were able to validate their prediction model of age in a secondary sample that resulted in an R^2 of 0.91 between chronological and biological age. Of the 71 sites they found predictive within this model of age, only 7 were measured on the Illumina HumanMethylation 27K chip and used within the scope of the GENOA study. As a summary, this set of 7 sites was highly significant (p-value range: 1.20×10^{-25} to 2.53×10^{-15}), agreed on direction of effect with estimates for GENOA age at Phase II, and also agreed on direction of effect for all three measures of health estimated in this study (Δ_1, Δ_2, and Δ_3) with significant p-values (Δ_1: $p = 3.34 \times 10^{-15}$ to 2.05×10^{-5}; Δ_2: $p = 4.61 \times 10^{-11}$ to 4.21×10^{-4}; Δ_3: $p = 9.55 \times 10^{-5}$ to 0.017). They also agreed on the direction of effect with both ζ_1 and ζ_2 within GENOA, though the p-values were not all significant (ζ_1: $p = 1.86 \times 10^{-4}$ to 0.65; ζ_2: $p = 5.24 \times 10^{-3}$ to 0.78). Full details on comparison of these 7 sites across both studies are available in Table 16.

In order to reduce multiple testing burden, we made the decision to utilize only the 2,095 sites that were found to significantly predict age (at Bonferroni-correction for α=0.05) as predictors of health (Δ_1, Δ_2, and Δ_3). We then used the 244 sites significant in the associations with Δ_1 and/or Δ_2 as predictors of healthy aging, ζ_1, and the 150 sites significant in the associations for Δ_1 and/or Δ_2 as predictors of healthy aging, ζ_2. By using these smaller sets of sites as potential predictors of health (Δ_1, Δ_2, and Δ_3) and healthy aging (ζ_1 or ζ_2) instead of the entire set of 26,428 CpG sites we increased our power to detect associations between variables, but may also have introduced type II errors into our study and limited our ability to detect associations that would have

otherwise been significant. The appropriateness of the decision to begin our analysis of aging with the reduced number (2,095) CpG sites was discerned within the context of additional studies examining heritability in the GENOA cohort. The confirmation for our decision was indicated in that the majority of CpG sites significantly associated with Δ_1, Δ_2, or Δ_3 do fall within this set of 2,095 sites, while only some of the significant sites in these analyses are not contained within the set of 2,095 sites (191 of 249 (76.7%) sites significant for Δ_1; 147 of 157 (93.6%) of sites significant for Δ_2, and 6 of 11 (54.6%) of sites significant for Δ_3). When plotting unadjusted heritability versus the –log(p-value) of the associations between 26,428 CpG sites and each of Δ_1, Δ_2, and Δ_3, we see that most of the sites significant for age and the Δ of interest are highly heritable (refer to orange dots in plots within Figure 7). Of the 191 CpG sites that are of Bonferroni corrected significance of α=0.05 for age and Δ_1, the average h^2 is 0.48, and 174 sites (91.1%) have $h^2 > 0.3$. Of the 147 CpG sites that are of Bonferroni corrected significance of $p<0.05$ for age and Δ_2, the average h^2 is 0.51, and 136 sites (92.5%) have $h^2 > 0.3$. Of the 6 CpG sites that are of Bonferroni corrected significant fore α=0.05 for age and Δ_3, the average h^2 is 0.43, and 4 sites (66.7%) have $h^2 > 0.3$ (Figure 7).

Other studies that have attempted to predict biological age using chronic disease risk factors have noted similar relationships between biological and chronological ages as within the GENOA study, where the oldest and youngest age groups were under and over predicted, respectively, compared to middle age groups that were more accurately predicted.(*175-177*) While this issue was addressed using exponential and interaction terms of traditional chronic disease risk factors in our modeling schema, it was unable to be completely resolved. Thus, it is noted that the best prediction of biological age lies

approximately between ages 40 and 75, as is seen in Phase I data in Supplementary Figure 3. This may be in part due to the small sample sizes at the extremes of the GENOA data. When initially predicting biological age, only 85 of the 1,822 participants (4.7%) used in this prediction were under age 40, and only 70 (3.8%) were over age 75. Further, our prediction model may not hold well into Phase III due to the limited sample size that is likely to be biased due to survival and better morbidity status than those from Phase II who are not measured in Phase III. Since Phase III participants are participants of earlier phases who aged well enough to remain participants in our study and maintain their clinical exam visits, these participants are likely to be healthier than GENOA participants who are not measured at Phase III. Having a sample that is larger and includes equal numbers of people within different age strata may aid in the ability to predict biological age more precisely across the age spectrum.

Additionally, GENOA is a unique study that began by the recruitment of hypertensive sibships, indicating that chronic diseases processes had already begun in many individuals at the beginning of the study and often occurred earlier than within the general population. There may be distinct inflammatory and epigenetic biomarkers acting within this cohort that represent the unique cellular aging process over the time periods (ζ_1 and ζ_2). While it is important to understand the mechanisms of how inflammation and DNA methylation work together within the context of healthy and unhealthy aging in narrowly defined subpopulations in order to identify the mechanisms at play within high risk populations, larger and more generalizable populations are needed in order to paint a more complete picture of the interplay between these important biomarkers.

Use of the GENOA study to examine the explanatory power of inflammatory and epigenetic biomarkers as factors in a predictive model for healthy aging is the first of its kind of which we are aware. Though there is still room for better predictive capabilities in order to achieve models that can properly assess health factors involved in later life aging processes based on current biomarker levels, modeling schema similar to this may be useful within the clinic to provide personalized medicine tailored to the individual. With the upcoming demographic increase within the senior citizen population, and the shift of chronic disease patterns to earlier ages within the United States, understanding preclinical symptoms of healthy and unhealthy aging processes will allow for better and earlier medical interventions and new therapeutics for those exhibiting signs of unhealthy aging. Thus, identification of biological indicators involved in aging processes, such as epigenetic and inflammatory biomarkers, is an important advancement in our predictive capability.

Table 8. Description of traditional chronic disease risk factors in GENOA participants, A) Continuous variables, and B) Categorical variables

A) Continuous variables

Variable	Phase 1			Phase 2 Epigenetics Sample			Phase 3 (participants within Phase 2 Epigenetics sample)		
	N	Range	Mean (SD)	N	Range	Mean (SD)	N	Range	Mean (SD)
Age, years	1854	21-91	58 (10)	972	39-95	66 (8)	484	47-98	72 (7)
BMI, kg/m2	1852	14-57	31 (6)	965	16-55	31 (6)	482	15-54	32 (7)
Systolic BP, mm Hg	1852	80-223	136 (23)	971	79-223	140 (21)	481	88-215	139 (21)
Diastolic BP, mm Hg	1853	35-126	78 (12)	972	45-121	78 (11)	481	45-116	72 (11)
Pulse Pressure, mm Hg	1847	19-130	58 (18)	972	26-132	62 (18)	480	22-125	67 (17)
Systolic BP (Adj), mm Hg**	1852	80-233	142 (24)	971	89-233	148 (231)	481	93-225	147 (21)
Diastolic BP (Adj), mm Hg**	1854	40-131	81 (13)	972	50-126	82 (11)	481	50-116	76 (11)
Pulse Pressure (Adj), mm Hg**	1849	19-138	61 (19)	972	26-132	62 (18)	480	27-130	71 (17)
Total cholesterol, mg/dL	1845	76-385	204 (45)	972	74-355	204 (42)	476	84-308	189 (40)
Triglycerides, mg/dL	1834	42-498	141 (65)	966	37-376	117 (55)	473	30-319	97 (46)
ln (triglycerides + 1)	1846	3.8-6.5	4.9 (0.42)	971	3.6-6.3	4.7 (0.44)	476	3.4-6.3	4.5 (0.44)
HDL cholesterol, mg/dL	1844	23-127	55(17)	968	22-130	58 (17)	476	19-115	57 (17)
ln (HDL+1)	1850	3.2-5.2	4.0 (0.3)	972	3.1-5.1	4.0 (0.29)	477	3.0-4.9	4.0 (0.29)
LDL cholesterol, mg/dL	1843	6-290	123 (42)	972	25-272	124 (40)	476	22-236	113 (35)
Glucose, mg/dL	1830	41-309	109 (41)	961	50-282	110 (34)	470	36-277	109 (31)
ln (Glucose + 1)	1847	3.7-5.9	4.7 (0.31)	967	3.9-5.8	4.7 (0.26)	473	3.6-5.8	4.7 (0.25)
Insulin, mU/mL	1831	1-93	12 (12)	969	0-116	10 (13)	465	1-69	10 (9)
ln (Insulin + 1)	1849	0.6-5.3	2.4 (0.73)	970	0.2-5.0	2.1 (0.75)	475	0.88-4.8	2.3 (0.7)
Serum creatinine, , mg/dL	1853	0.5-9.4	0.83 (0.37)	961	0.42-2.2	0.92 (0.25)	486	0.48-2.5	0.92 (0.25)

** If a participant is listed as taking hypertensive medications, 10 mm Hg is added to systolic blood pressure, and 5 mm Hg is added to diastolic blood pressure, else recorded blood pressure is used in this variable.

B) Categorical variables

Variable	Phase 1 Count (%)	Phase 2 Count (%)	Phase 3 Count (%)
Female sex	1,281 (69%)	687 (71%)	353 (73%)
Ever smoker	784 (42%)	266 (29%)	198 (41%)
Hypertensive	1,346 (73%)	802 (83%)	423 (87%)
Diabetic	400 (22%)	298 (31%)	173 (36%)

Table 9. Description of inflammatory markers in GENOA participants

	N	Count of Outliers	Count of Missing Values	Range	Mean (SD)
C-Reactive Protein (CRP), mg/L	936	0	36	0.21-29.9	6.0 (6.8)
Homocysteine (Hycs), µmol/L	964	8	0	4.7-28.7	10.5 (3.5)
Fibrinogen, mg/dL	971	1	0	120-680	369 (81)
Intercellular adhesion molecular (ICAM), ng/mL	556	3	413	17-543	274 (77)
Interleukin-6 (IL6), pg/mL	818	8	146	1.4-29.4	8.8 (4.7)
Interleukin-18 (IL18), pg/mL	829	4	139	2-215	69 (38)
Monocyte chemotactic protein-1 (MCP1), pg/mL	804	2	166	160-2,515	1,048 (359)
Myeloperoxidase (MPO), ng/mL	911	8	53	4-141	40 (23)
Resistin, ng/mL	859	1	112	1.0-12.2	4.1 (2.0)
Serum amyloid A (SAA), µg/mL	704	0	268	2-146	33 (31)
Tumor necrosis factor receptor-1 (TNFRI), pg/mL	840	3	129	112-3,108	1,167 (513)
Tumor necrosis factor receptor-2 (TNFRII), pg/mL	818	1	153	287-5,170	1,945 (822)

Table 10. Predictors of final model of Age with Phase I GENOA Participants (N=1,822)

	Coefficient Value	Std. Error	DF	t-value	p-value
(Intercept)	-185.70	32.80	1126	-5.66	1.90E-08
BMI	0.27	0.15	1126	1.78	7.50E-02
Adj. Diastolic BP	0.51	0.12	1126	4.15	3.60E-05
ln (Glucose+1)	60.99	12.70	1126	4.80	1.77E-06
ln (HDL+1)	3.02	0.64	1126	4.70	2.97E-06
ln (Insulin+1)	-0.87	0.26	1126	-3.30	9.97E-04
ln (Triglycerides+1)	1.22	0.44	1126	2.78	5.54E-03
Adj. Pulse Pressure	1.05	0.13	1126	8.25	4.40E-16
Serum Creatinine	36.47	6.20	1126	5.88	5.46E-09
Adj. Pulse Pressure* ln (Glucose+1)	-0.12	0.03	1126	-4.42	1.10E-05
Adj. Diastolic BP*BMI	0.00	0.00	1126	-2.43	1.52E-02
Adj. Diastolic BP*Serum Creatinine	-0.19	0.06	1126	-3.11	1.92E-03
Adj. Diastolic BP ^{2}	0.00	0.00	1126	-2.83	4.80E-03
ln (Glucose+1) ^{2}	-5.26	1.27	1126	-4.13	3.87E-05
Adj. Pulse Pressure ^{2}	0.00	0.00	1126	-6.55	8.46E-11
Serum Creatinine^2	-6.85	1.85	1126	-3.70	2.28E-04

LR R^2 for Phase I data = 0.515
Model built upon 1,822 Phase I participants who had no missing data, out of 1,848 GENOA Phase I participants

Table 11. Models of (A) ζ_1 and (B) ζ_2 vs. significant inflammatory biomarkers ($p<0.10$)

A) ζ_1 vs. inflammatory biomarkers

Variable	Beta Estimate	p-value
Intercept	-8.047	1.22E-03
Δ_1	-0.276	2.11E-16
Fibrinogen	0.00424	1.73E-02
ln (TNFR1+1)	1.240	6.03E-02
ln (Homocysteine + 1)	1.056	5.73E-02

LR R^2 = 11.95%
N=785

B) ζ_2 vs. inflammatory biomarkers,

Variable	Beta Estimate	p-value
Intercept	68.144	7.45E-02
Δ_2	-0.448	8.84E-15
ln (MCP1 + 1)	-44.729	8.06E-02
ln (MCP1 + 1)^2	7.025	9.75E-02

LR R^2 = 25.84%
N=289

Figure 6. Counts of significant sites for Age, Delta, and Zeta regressions.

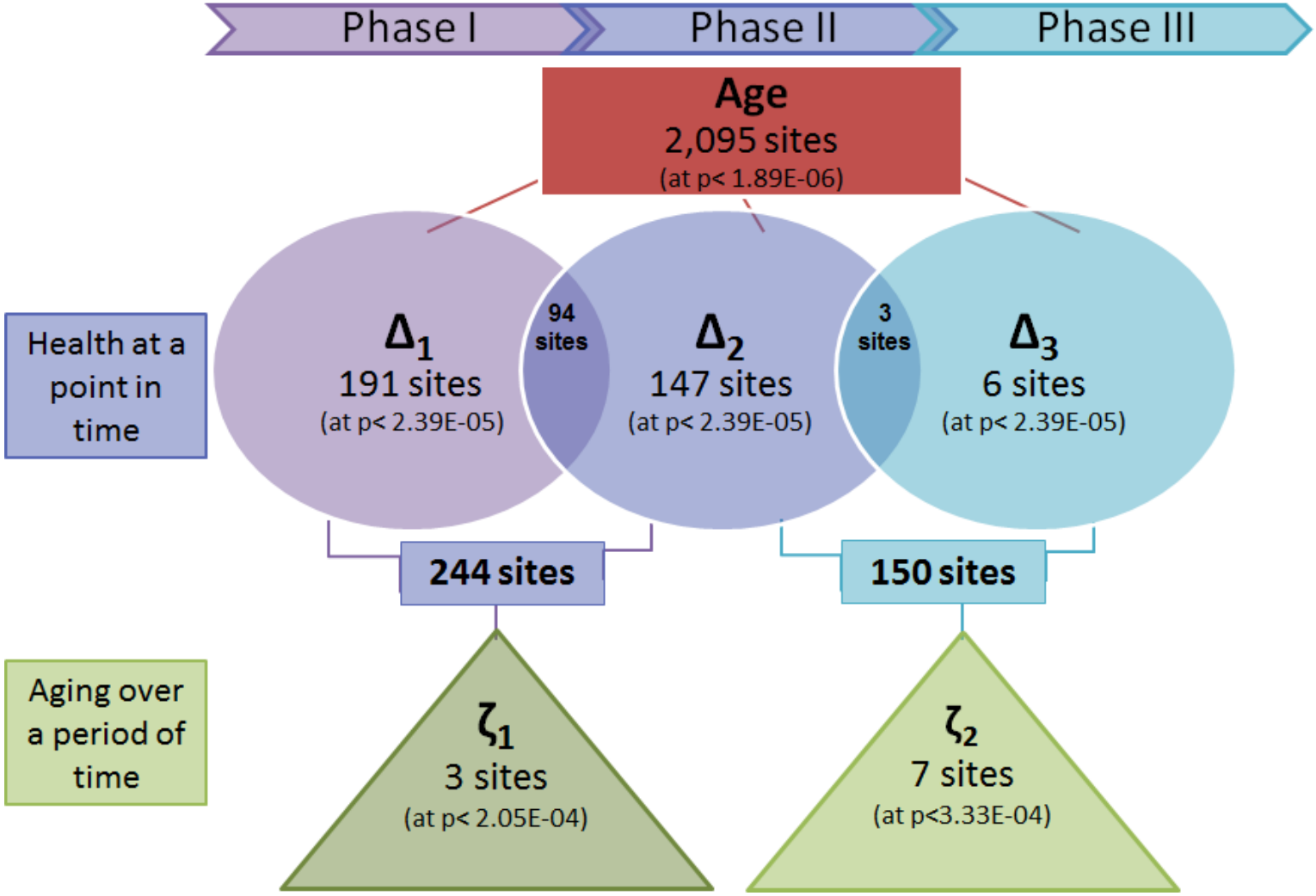
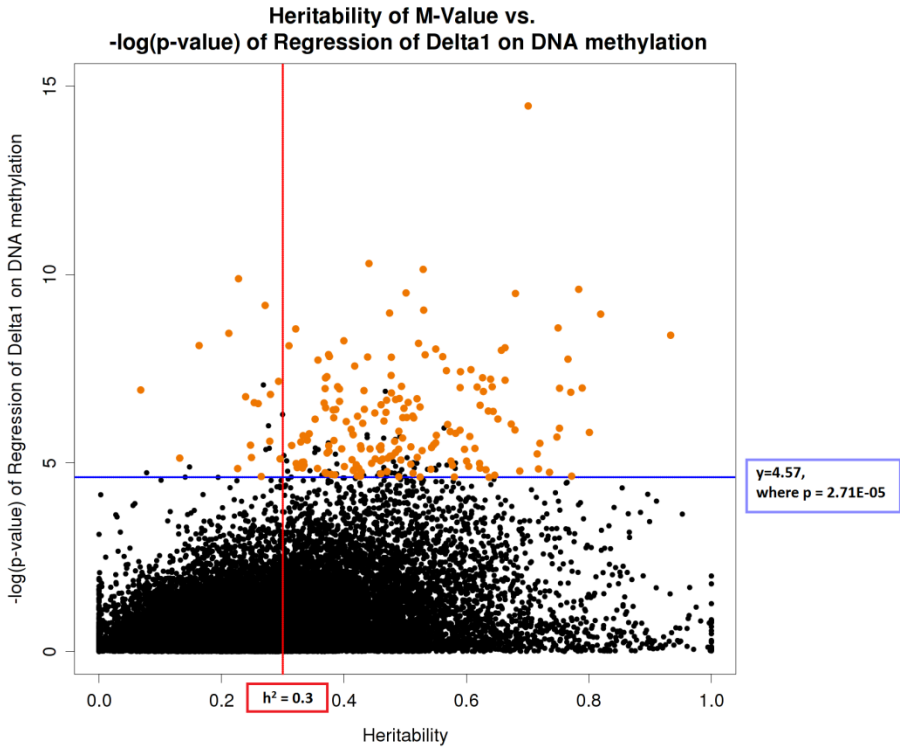


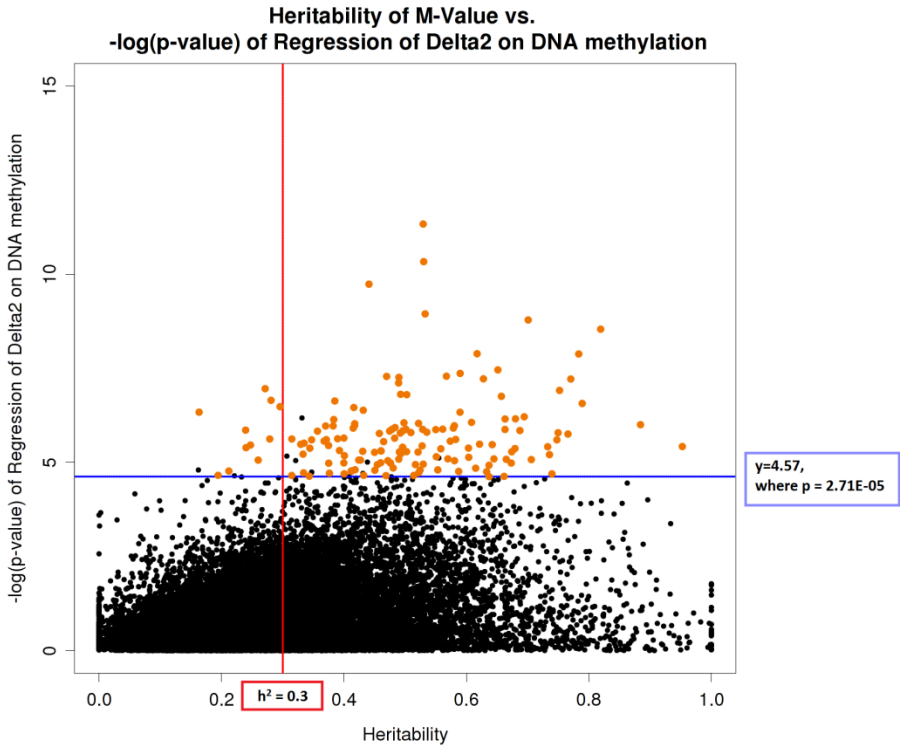
Figure 7. Heritability of M-Value (Raw) vs. -log(p-value) of Association between 26,428 CpG sites and A) Δ_1, B) Δ_2, and C) Δ_3

A) Heritability of M-Value (Raw) vs. -log(p-value) of Association between 26,428 CpG sites and Δ_1



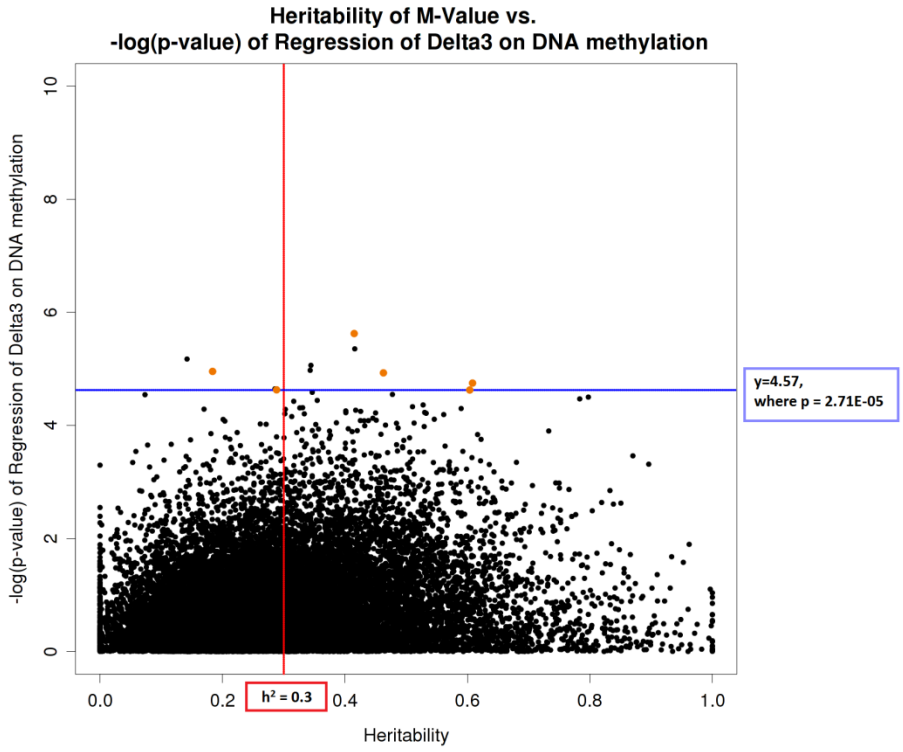
- Orange dots represent 191 CpG sites significant for Age at Bonferroni corrected p < 0.05 (1.86×10^{-6}) and Δ_1 at Bonferroni corrected p<0.05 (2.71×10^{-5})
- Red vertical line at x=0.3 indicates high heritability to its right and low heritability to its left.
- Blue horizontal line at y=4.57 indicates sites that reached Bonferroni-corrected significance above it, and sites that did not reach Bonferroni-corrected significance below it.

B) Heritability of M-Value (Raw) vs. -log(p-value) of Association between 26,428 CpG sites and Δ_2



- Orange dots represent 147 CpG sites significant for Age at Bonferroni corrected p < 0.05 ($1.86x10^{-6}$) and Δ_2 at Bonferroni corrected p<0.05 ($2.71x10^{-5}$)
- Red vertical line at x=0.3 indicates high heritability to its right and low heritability to its left.
- Blue horizontal line at y=4.57 indicates sites that reached Bonferroni-corrected significance above it, and sites that did not reach Bonferroni-corrected significance below it.

C) Heritability of M-Value (Raw) vs. -log(p-value) of Association between 26,428 CpG sites and Δ_3



- Orange dots represent 6 CpG sites significant for Age at Bonferroni corrected p < 0.05 (1.86×10^{-6}) and Δ_3 at Bonferroni corrected p<0.05 (2.71×10^{-5})
- Red vertical line at x=0.3 indicates high heritability to its right and low heritability to its left.
- Blue horizontal line at y=4.57 indicates sites that reached Bonferroni-corrected significance above it, and sites that did not reach Bonferroni-corrected significance below it.

Table 12. CpG sites significant (after Bonferroni correction) in association with ζ_1 or ζ_2

Outcome	CpG Site	Beta Est. (CpG)	p-value (CpG)	Model LR R^2	Chr	Gene	Mean M-Value	Probe Type**	Product
ζ_1	cg25494227	-2.35	4.79E-05	1.66%	12	*C12orf59*	00.77	0	hypothetical protein LOC120939
ζ_1	cg15538427	-2.28	1.29E-04	1.62%	11	*LOC221091*	-0.11	0	hypothetical protein LOC221091
ζ_1	cg04474832	-1.78	1.86E-04	1.32%	3	*ABHD14A*	-1.72	0	abhydrolase domain containing 14A
ζ_2	cg05501357	-2.28	5.31E-05	3.85%	11	*HIPK3*	-0.80	0	homeodomain interacting protein kinase 3
ζ_2	cg16360372	-2.78	6.92E-05	3.73%	5	*SPINK1*	1.70	1	serine protease inhibitor; Kazal type 1
ζ_2	cg02533173	-4.04	1.46E-04	3.38%	19	*BRD4*	0.32	0	bromodomain-containing protein 4 isoform long
ζ_2	cg08290628	-3.39	1.94E-04	3.25%	15	*CORO2B*	0.40	0	coronin; actin binding protein; 2B
ζ_2	cg16005443	-2.24	2.27E-04	3.20%	19	*LILRB3*	-0.90	1	leukocyte immunoglobulin-like receptor; subfamily B (with TM and ITIM domains); member 3
ζ_2	cg15804973	-2.50	3.01E-04	3.05%	6	*MAP3K5*	-0.63	0	mitogen-activated protein kinase kinase kinase 5
ζ_2	cg14123992	-2.93	3.20E-04	3.05%	19	*APOE*	10.52	0	apolipoprotein E precursor

Model: $\zeta_{1ijk} = \beta_0 + \beta_1 E_k + W_{0jk}$

$\zeta_{2ijk} = \beta_0 + \beta_1 E_k + W_{0jk}$

where E is the CpG M-Value for the i^{th} individual in the k^{th} sibship, and W is the random effect for each sibship

Polymorphic and Non-Specific Probes: (Chen, Choufani et al. 2011)

** 0 = Neither, 1 = Polymorphic

Table 13. (A) Association between ζ_1 and top 5 and 10 PCs of sites significant with Δ_1 and Δ_2 based on M-Value (n=244); (B) Association between ζ_2 and top 5 and 10 PCs of sites significant with Δ_2 and Δ_3 based on M-Value (n=150)

(A) ζ_1 and top 5 and 10 PCs

		Univariate Models			Multivariable Model		
PC	**% Variation Explained**	**β(PC)**	**p-value**	**R^2_{LR} x100**	**β(PC)**	**p-value**	**R^2_{LR} x100**
1	34.56	0.04	0.35	0.71	0.05	0.21	
2	12.83	0.28	7.80E-05	2.28	0.30	3.05E-05	
3	3.84	-0.10	0.41	0.69	-0.14	0.25	
4	3.18	0.06	0.66	0.64	0.07	0.59	
5	1.92	-0.20	0.26	0.76	-0.25	0.16	2.81
6-10	5.90						3.63
Total	62.24						

Model: $\zeta_{1ij} = \beta_0 + \beta_1 \cdot PC_{ij} + \beta_2 \cdot \Delta_{1ij} + W_{0j} + \varepsilon_{ij}$.

(B) ζ_2 and top 5 and 10 PCs

		Univariate Models			Multivariable Model		
PC	**% Variation Explained**	**β(PC)**		**R^2_{LR} x100**	**β(PC)**		**R^2_{LR} x100**
1	21.83	0.50	9.68E-05	4.57	0.52	5.76E-05	
2	16.59	0.07	0.65	0.76	0.06	0.67	
3	5.15	-0.03	0.92	0.71	-0.0023	0.99	
4	4.54	0.18	0.49	0.82	0.27	0.30	
5	2.92	NA	NA	NA	-0.62	0.08	5.55
6-10	8.77						6.80
Total	59.81						

Model: $\zeta_{2ij} = \beta_0 + \beta_1 \cdot PC_{ij} + \beta_2 \cdot \Delta_{2ij} + W_{0j} + \varepsilon_{ij}$.

Table 14. Best fitting model for ζ_1 using inflammatory and DNA methylation markers. A) ζ_1 with inflammatory markers and CpG sites

Outcome	Predictor	Beta Est.	p-value	LR R2
ζ_1	Intercept	-10.04	4.52E-04	16.59%
	Δ_1	-0.30	7.80E-17	
	Fibrinogen	0.01	2.06E-03	
	ln (TNFR1+1)	1.44	0.034	
	ln (Hycs + 1)	1.23	0.036	
	cg25494227	-2.48	8.27E-04	
	cg04474832	-1.29	0.031	

N=683

Table 15. Best fitting model for ζ_2 using inflammatory and DNA methylation markers. A) ζ_2 with inflammatory markers. CpG sites, and PCs

Outcome	Predictor	Beta Est.	p-value	LR R2
ζ_2	Intercept	84.91	0.036	36.43%
	Δ_2	-0.50	1.22E-15	
	ln (MCP1+1)	-52.52	0.048	
	ln $(MCP1+1)^2$	7.49	0.084	
	cg05501357	21.24	0.071	
	cg02533173	-2.50	0.042	
	cg16005443	-1.34	0.041	
	PC2	0.36	0.043	
	ln (MCP1+1)* cg05501357	-8.04	0.041	

N=270

Table 16. Comparison of GENOA results to 7 CpG sites found significant within the study by Hannum, Guinney, Zhao, Zhang, Hughes, Sadda, Klotzle, Bibikova, Fan, Gao, Deconde, Chen, Rajapakse, Friend, Ideker and Zhang (*174*)

Marker	Coef *	GENOA Coef (Age)	GENOA p-val (Age)	GENOA Coef (Δ_1)	GENOA p-val (Δ_1)	GENOA Coef (Δ_2)	GENOA p-val (Δ_2)	GENOA Coef (Δ_3)	GENOA p-val (Δ_3)	GENOA Coef (ζ_1)	GENOA p-val (ζ_1)	GENOA Coef (ζ_2)	GENOA p-val (ζ_2)
cg05442902	-22.7	-8.51	1.26E-20	-2.46	9.36E-06	-2.79	1.88E-05	-2.24	0.041	-1.02	0.054	-1.39	0.127
cg04474832	-7.1	-7.75	2.17E-21	-3.01	8.74E-10	-3.80	4.61E-11	-3.73	5.89E-05	-1.78	1.86E-04	-2.17	5.24E-03
cg19722847	-5.66	-6.19	1.24E-18	-1.80	2.05E-05	-2.55	2.32E-07	-1.97	0.017	-1.23	2.11E-03	-1.40	0.041
cg09809672	-0.74	-5.27	8.51E-22	-1.85	4.05E-09	-1.30	4.21E-04	-1.37	0.021	-0.14	0.647	-0.55	0.265
cg22736354	4.42	6.26	1.20E-25	2.79	3.34E-15	2.50	1.63E-09	1.88	7.65E-03	0.70	0.047	0.17	0.775
cg21296230	8.39	4.65	7.47E-15	2.29	6.55E-10	2.31	1.09E-07	2.88	9.55E-05	0.71	0.046	1.17	0.061
cg06493994	9.42	5.69	2.53E-15	2.64	3.60E-09	2.28	1.69E-05	3.35	1.75E-04	0.49	0.257	1.62	0.032

* Coefficient reported by Hannum, Guinney, Zhao, Zhang, Hughes, Sadda, Klotzle, Bibikova, Fan, Gao, Deconde, Chen, Rajapakse, Friend, Ideker and Zhang (*174*)

Supplementary Table 6. Comparison of baseline characteristics of Phase II Non-Epigenetics Participants with Phase II Epigenetics Participants. A) Continuous clinical variables, B) Continuous biomarkers, C) Categorical descriptive variables.

A) Continuous clinical variables

	Phase II Non-Epigenetics Participants				Phase II Epigenetics Participants				
Variable	**N**	**% Missing**	**Range**	**Mean (SD)**	**N**	**% Missing**	**Range**	**Mean (SD)**	**Pr > |t|**
Age, years	474	0.00%	26.41-81.52	56.07 (9.21)	1008	0.00%	39.26-94.74	66.34 (7.6)	6.6E-49
BMI, kg/m²	472	0.42%	18.1-57.9	32.6 (7.4)	1001	0.69%	16.41-55.09	31.1 (6.09)	0.013
Systolic BP, mm Hg	473	0.21%	96-208	134.8 (19.3)	1006	0.20%	79-221	139.89 (21.12)	2.2E-06
Diastolic BP, mm Hg	474	0.00%	53-122	81.57 (10.2)	1008	0.00%	45-121	78.29 (11.01)	2.3E-05
Pulse Pressure, mm Hg	474	0.00%	19-127	53.41 (15.11)	1006	0.20%	26-127	61.64 (17.72)	6.0E-17
Adj. Systolic BP, mm Hg (1)	473	0.21%	97-218	142.05 (21.47)	1006	0.20%	89-231	148.15 (22.74)	5.5E-08
Adj. Diastolic BP, mm Hg (1)	474	0.00%	58-127	85.2 (10.95)	1008	0.00%	50-126	82.42 (11.43)	4.6E-04
Adj. Pulse Pressure, mm Hg (1)	473	0.21%	24-132	56.89 (15.72)	1004	0.40%	28-132	65.65 (18.28)	2.2E-18
Total Cholesterol, mg/dL	459	3.16%	72-348.5	197.93 (40.02)	1008	0.00%	73.5-354.5	203.98 (41.95)	7.8E-05
Triglycerides, mg/dL	454	4.22%	28.5-419.5	111.95 (58.94)	1004	0.40%	37-402.5	117.9 (56.68)	9.8E-04
ln (Triglycerides+1)	457	3.59%	3.38-6.45	4.62 (0.49)	1007	0.10%	3.64-6.27	4.69 (0.43)	0.69
HDL-C, mg/dL	457	3.59%	23.8-125.8	55.42 (16.66)	1005	0.30%	21.7-130.35	58.08 (17.48)	7.0E-03
ln (HDL+1)	460	2.95%	3.21-5.17	4 (0.3)	1008	0.00%	3.12-5.05	4.04 (0.29)	0.75
LDL-C, mg/dL	459	3.16%	23.6-253.75	121.1 (36.88)	1008	0.00%	24.85-272.1	123.84 (39.58)	0.064
Glucose, mg/dL	457	3.59%	43.5-296	108.28 (38.68)	998	0.99%	49.5-290	110.56 (34.31)	0.11
ln (Glucose+1)	457	3.59%	3.8-5.69	4.65 (0.28)	1004	0.40%	3.92-5.82	4.69 (0.27)	0.77
Insulin, mU/mL	163	65.61%	1.14-52.46	9.34 (8.53)	1005	0.30%	0.22-115.76	10.39 (12.45)	0.30
ln (Insulin+1)	163	65.61%	0.76-3.98	2.12 (0.62)	1006	0.20%	0.2-5.02	2.12 (0.74)	1.00
Menopause Age, years (2)	270	43.04%	23-59	42.6 (8.26)	699	30.65%	25-62	43.72 (7.85)	0.14

1 If a participant is listed as taking hypertensive medications, 10 mm Hg is added to systolic blood pressure, and 5 mm Hg is added to diastolic blood pressure, else recorded blood pressure is used in this variable.

2 Menopause age is represented for females only.

B) Measures of inflammation

	Phase II Non-Epigenetics Participants				Phase II Epigenetics Participants				
Variable	**N**	**% Missing**	**Range**	**Mean (SD)**	**N**	**% Missing**	**Range**	**Mean (SD)**	**Pr > |t|**
Serum Creatinine, mg/dL	460	2.95%	0.44-3.64	0.89 (0.29)	1008	0.00%	0.42-2.98	0.94 (0.3)	0.69
CRP, mg/L	347	26.79%	0.21-29.9	5.48 (5.8)	971	3.67%	0.21-29.9	6.05 (6.77)	0.37
ln (CRP+1)	347	26.79%	0.08-1.49	0.68 (0.33)	971	3.67%	0.08-1.49	0.7 (0.34)	0.87
Homocysteine, µmol/L	430	9.28%	4.5-25.2	9.38 (2.97)	1002	0.60%	4.7-31	10.57 (3.59)	7.1E-03
ln (Hycs+1)	430	9.28%	1.7-3.27	2.3 (0.26)	1002	0.60%	1.74-3.47	2.41 (0.28)	0.41
Fibrinogen, mg/dL	440	7.17%	123-684	362.44 (88.37)	1007	0.10%	120-680	369.35 (81)	1.6E-03
ICAM, ng/mL	199	58.02%	101-584	296.97 (83.09)	578	42.66%	17-625	275.38 (80.82)	3.2E-16
IL-6, pg/mL	241	49.16%	1.92-29.78	8.71 (5.06)	856	15.08%	1.42-36.45	9.09 (5.31)	0.54
ln (IL-6+1)	245	48.31%	0.46-1.58	0.95 (0.22)	878	12.90%	0.38-1.58	0.97 (0.22)	0.86
IL-18, pg/mL	240	49.37%	6.14-235.97	72.17 (38.66)	863	14.38%	1.86-252.42	69.92 (39.97)	0.19
ln (IL-18+1)	241	49.16%	0.85-2.49	1.81 (0.23)	872	13.49%	0.84-2.49	1.8 (0.25)	0.92
MPO, ng/mL	193	59.28%	11.74-141.47	47.43 (27.67)	954	5.36%	4-160.58	40.26 (25.04)	5.1E-07
ln (MPO+1)	196	58.65%	1.1-2.22	1.63 (0.23)	968	3.97%	0.7-2.22	1.56 (0.24)	0.60
MCP1, pg/mL	205	56.75%	247.44-2124.82	970.81 (319.89)	835	17.16%	159.81-2704.78	1053.43 (369.78)	2.3E-49
ln (MCP+1)	205	56.75%	2.4-3.33	2.96 (0.15)	849	15.77%	2.46-3.49	3.01 (0.16)	0.69
Resistin, ng/mL	312	34.18%	1-12.13	3.77 (1.96)	895	11.21%	1-12.32	4.07 (2.03)	0.40
ln (Resistin+1)	315	33.54%	0.3-1.2	0.65 (0.18)	902	10.52%	0.3-1.2	0.68 (0.17)	0.798
SAA, µg/mL	297	37.34%	2.35-146	26.75 (28.98)	731	27.48%	2.35-146	32.94 (30.88)	2.0E-05
ln (SAA+1)	297	37.34%	0.53-2.17	1.27 (0.39)	731	27.48%	0.53-2.17	1.38 (0.38)	0.51
TNFR-1, pg/mL	219	53.80%	102.78-3310.18	1113.11 (562.23)	872	13.49%	19.89-3844.07	1185.29 (546.83)	1.3E-27
ln (TNFR1+1)	217	54.22%	2.15-3.52	3 (0.21)	882	12.50%	2.17-3.68	3.04 (0.22)	0.76
TNFR-2, pg/mL	240	49.37%	388.3-4998.98	1751.03 (649.43)	849	15.77%	286.83-5333.39	1950.04 (831.08)	4.8E-117
ln (TNFR2+1)	241	49.16%	2.59-3.8	3.22 (0.16)	854	15.28%	2.59-3.8	3.26 (0.17)	0.70

C) Categorical variables

	Non-Epigenetics Participants			Epigenetics Participants		
	Count	**Total**	**% of Total**	**Count**	**Total**	**% of Total**
Female	335	474	70.68%	715	1,008	70.93%
Hypertensive	344	474	72.57%	833	1,008	82.64%
Has Diabetes	128	474	27.00%	308	1,008	30.56%
Ever Smoker	174	474	36.71%	421	1,008	41.77%

Supplementary Table 7. List of variables and transformations

Variable	Disease Measurement	Units	Description/Transformation
Age	Age	years	none
Body Mass Index (BMI)	Obesity	kg/m2	$BMI = weight\ (kg)\ \frac{weight\ (kg)}{height^2\ (m^2)}$
Systolic Blood Pressure (BP)	Hypertension	mm Hg	none
Diastolic BP	Hypertension	mm Hg	none
Pulse Pressure	Hypertension	mm Hg	none
Adj. Systolic BP	Hypertension	mm Hg	If Hypertension = Yes, Adj Sys BP = Systolic BP + 10 mm Hg Else if Hypertension = N then Adj Sys BP = Systolic BP
Adj. Diastolic BP	Hypertension	mm Hg	If Hypertension = Yes, Adj Dia BP = Diastolic BP + 5 mm Hg Else if Hypertension = N, then Adj Dia BP = Diastolic BP
Adj. Pulse Pressure	Hypertension	mm Hg	Adj PP = Adj Sys BP – Adj Dia BP
Total Cholesterol	Dyslipidemia	mg/dL	Total molar mass of lipoproteins in blood
Triglycerides (trig)	Dyslipidemia	mg/dL	Trig is represented as: ln(Trig + 1)
High density lipoproteins (HDL-C)	Dyslipidemia	mg/dL	HDL is represented as: ln(HDL + 1)
Low density lipoproteins (LDL-C)	Dyslipidemia	mg/dL	If triglycerides < 200 mg/dL, LDL-C = $Total\ Cholesterol - (HDL\text{-}C + triglycerides/5)$ Else if triglycerides ≥ 200 mg/dL, $LDL\text{-}C = Total\ Cholesterol - (HDL\text{-}C + 30)$
Glucose	Diabetes	mg/dL	Glucose is represented as: ln(Glucose + 1)
Insulin	Diabetes	mU/mL	Insulin is represented as: ln(Insulin + 1)
Serum Creatinine (SCr)	Chronic Kidney Disease	mg/dL	Lab value if measured by Isotope Dilution Mass Spectrometry (most of Ph 2,all of Ph 3). Else: Phase 1: SCr=(0.807*(0.8134*SCr_Ph1-0.167)+0.1738); Phase 2: SCr =(0.807*SCr_Ph2)+0.1738.
C-Reactive Protein (CRP)	Inflammation	mg/L	CRP is represented as: ln(CRP+1)
Fibrinogen	Inflammation	mg/dL	None
Homocysteine	Inflammation	µmol/L	Hycs is represented as: ln(Hycs+1)
Intercellular Adhesion Molecule (ICAM)	Inflammation	ng/mL	None
Interleukin-6 (IL-6)	Inflammation	pg/mL	IL-6 is represented as: ln(IL-6+1)
Interleukin-18 (IL-18)	Inflammation	pg/mL	IL-18 is represented as: ln(IL-18+1)

Monocyte Chemotactic Protein-1 (MCP1)	Inflammation	pg/mL	MCP is represented as: ln(MCP1+1)
Myeloperoxidase (MPO)	Inflammation	ng/mL	MPO is represented as: ln(MPO+1)
Resistin	Inflammation	ng/mL	Resistin is represented as: ln(Resistin+1)
Serum Amyloid A (SAA)	Inflammation	µg/mL	SAA is represented as: ln(SAA+1)
Tumor Necrosis Factor Receptor-1 (TNFR-1)	Inflammation	pg/mL	TNFR1 is represented as: ln(TNFR-1+1)
Tumor Necrosis Factor Receptor-2 (TNFR-2)	Inflammation	pg/mL	TNFR2 is represented as: ln(TNFR-2+1

Supplementary Table 8. List of probes used to standardize methylated and unmethylated signals.

Red Channel Probes	Green Channel Probes
EXTENSION 1190050	BISULFITE CONVERSION 4670278
EXTENSION 360446	BISULFITE CONVERSION 4670484
NON-POLYMORPHIC 1740025	BISULFITE CONVERSION 5290048
STAINING 4200736	EXTENSION 1190050
STAINING 4570020	EXTENSION 360446
	EXTENSION 520537
	NON-POLYMORPHIC 1740025
	STAINING 5340168

Supplementary Table 9. Correlations between traditional risk factors measured at Phase I Exam.

	Age	BMI	SysBP Adj	DiaBP Adj	PulsePress Adj	Serum Creatinine	Cholesterol	ln (Trig+1)	ln (HDL+1)	LDL	ln (Glucose+1)	ln (Insulin+1)
Age	1.00											
BMI	-0.09	1.00										
SysBP Adj	0.32	0.12	1.00									
DiaBP Adj	-0.10	0.05	0.64	1.00								
PulsePress Adj	0.48	0.12	0.85	0.14	1.00							
Serum Creatinine	0.19	-0.12	0.11	0.08	0.09	1.00						
Cholesterol	0.09	0.01	0.08	0.06	0.06	0.07	1.00					
ln (Trig+1)	0.10	0.09	0.10	0.03	0.11	0.15	0.29	1.00				
ln (HDL+1)	0.11	-0.09	0.04	-0.02	0.06	-0.17	0.22	-0.33	1.00			
LDL	0.03	0.04	0.05	0.08	0.02	0.11	0.93	0.29	-0.12	1.00		
ln (Glucose+1)	0.11	0.23	0.15	0.01	0.19	0.02	0.02	0.24	-0.17	0.06	1.00	
ln (Insulin+1)	-0.04	0.41	0.06	-0.01	0.09	0.05	-0.02	0.32	-0.28	0.04	0.32	1.00

Cells indicated in red represent variables that we considered to be in high correlation (>0.8)

Supplementary Table 10. Steps of Stepwise Modeling to Obtain Final Model

					Model with new predictor added in:		
Step	**Outcome**	**Last Predictor in Model**	**New Predictor to be added**	**p-value of new of predictor**	**AIC**	**BIC**	**LR R2**
Step 1	ph1_age	null	Adj Pulse Pressure	2.82E-62	12,771.58	12,793.67	0.4386
Step 2	ph1_age	Adj Pulse Pressure	Adj Diastolic BP	3.56E-09	12,738.63	12,766.24	0.4491
Step 3	ph1_age	Adj Diastolic BP	Adj Pulse Pressure^2	3.86E-11	12,696.61	12,729.75	0.4621
Step 4	ph1_age	Adj Pulse Pressure^2	Serum Creatinine	4.47E-12	12,572.83	12,611.44	0.4779
Step 5	ph1_age	Serum creatinine	ln (HDL + 1)	4.41E-06	12,539.36	12,583.49	0.4842
Step 6	ph1_age	ln (HDL + 1)	Adj Diastolic BP^2	4.21E-05	12,524.49	12,574.13	0.4889
Step 7	ph1_age	Adj Diastolic BP^2	Serum creatinine ^2	1.89E-04	12,512.59	12,567.75	0.4928
Step 8	ph1_age	Serum creatinine ^2	BMI	1.57E-03	12,489.84	12,550.50	0.4963
Step 9	ph1_age	BMI	Adj Diastolic BP*BMI	0.0144	12,485.82	12,551.99	0.4980
Step 10	ph1_age	Adj Diastolic BP*BMI	Adj Diastolic BP*Serum creatinine	0.0066	12,480.38	12,552.07	0.5000
Step 11	ph1_age	Adj Diastolic BP*Serum creatinine	ln (trig + 1)	0.0246	12,449.34	12,526.51	0.5020
Step 12	ph1_age	ln (trig + 1)	ln (insulin+1)	0.0144	12,424.80	12,507.46	0.5038
Step 13	ph1_age	ln (insulin+1)	Adj Pulse Pressure * ln (trig+1)	0.0188	12,421.24	12,509.40	0.5053
Step 14	ph1_age	Adj Pulse Pressure * ln (trig+1)	ln (glucose+1)	0.0404	12,381.97	12,475.60	0.5073
Step 15	ph1_age	ln (glucose+1)	ln (glucose+1) ^2	3.88E-05	12,366.90	12,466.03	0.5119
Step 16	ph1_age	ln (glucose+1) ^2	Adj Pulse Pressure * ln (glucose+1)	1.36E-04	12,354.16	12,458.81	0.5158
Step 17	ph1_age	REMOVE Adj Pulse Pressure * ln (trig+1)	Adj Pulse Pressure * ln (glucose+1)	1.10E-05	12,353.85	12,452.99	0.5154
Step 18	ph1_age	Adj Pulse Pressure * ln (glucose+1)	ln (trig + 1)^2	0.0685	12,352.50	12,457.15	0.5163
Step 19	ph1_age	ln (trig + 1)^2	Adj Pulse pressure * Adj Diastolic BP	0.0905	12,351.61	12,461.76	0.5170

					Model with new predictor added in:		
Step	**Outcome**	**Last Predictor in Model**	**New Predictor to be added**	**p-value of new of predictor**	**AIC**	**BIC**	**LR R2**
Step 20	ph1_age	REMOVE BMI and all BMI-interactions	BMI	0.0016	12,356.90	12,456.04	0.5145
Step 21	ph1_age	BMI	Adj Diastolic BP * BMI	0.0119	12,352.50	12,457.15	0.5163
Step 22	ph1_age	Adj Diastolic BP * BMI	Adj Pulse Pressure * Adj Diastolic BP	0.0905	12,351.61	12,461.76	0.5170
Step 23	ph1_age	Adj Pulse Pressure * Adj Diastolic BP	ln (HDL+1) * ln (trig+1)	0.0979	12,350.84	12,466.50	0.5178
Step 24	ph1_age	ln (HDL+1) * ln (trig+1)	Adj Diastolic BP ^3	0.1368	12,350.60	12,471.77	0.5183
Max Model	ph1_age	all terms, interaction, and squared terms	n/a	n/a	12,368.22	12,693.04	0.5265
				Minimum:	**12,350.60**	**12,452.99**	

Cells with lowest AIC and BIC are highlighted in blue.

Model chosen for analysis is highlighted in green.

Supplementary Table 11. Distributions of ζ_1 and ζ_2, and Age, biological age, and Δ for Phases I, II, and III,

	Count	h^2 unadj (h^2 adj**)	Min	Median	Mean	Max
Ph1 Chronological Age	1822	1 (NA)	20.5	58.7	58.2	88.21
Ph1 Biological Age	1822	1 (0)	26.83	58.33	58.19	80.91
Δ_1	1822	0 (0)	-15.79	1.28E-01	-3.98E-13	15.15
Ph2 Chronological Age	963	0.9783 (NA)	39.26	65.96	66.27	91.58
Ph 2 Biological Age	936	1 (0.1141)	41.17	65.83	66.08	91.29
Δ_2	936	0 (0.1142)	-12.76	0	0.16	19.64
Ph3 Chronological Age	479	0.9708 (NA)	47.38	71.2	71.68	98.12
Ph 3 Biological Age	346	1 (0)	48.27	73.76	74.14	92.85
Δ_3	346	0 (0)	-18.03	-2.32	-2.52	16.84
ζ_1	936	0.4314 (0.4314)	-15.37	-0.38	-0.39	22.58
ζ_2	345	0.2968 (0.3022)	-15.45	-2.19	-2.25	14.04

** Adjusted h^2 is adjusted for age within SOLAR models. The variables Ph1 Biological Age, Δ_1, and ζ_1 were adjusted for chronological age at Phase 1. The variables Ph2 Biological Age, Δ_2, and ζ_2 were adjusted for chronological age at Phase 2. Lastly, Ph3 Biological Age and Δ_3 were adjusted for chronological age at Phase 3.

Supplementary Table 12. Stepwise selection steps of final modeling for Aim 3, using inflammatory and methylation markers to predict ζ_1 and ζ_2

A) ζ_1 modeled with inflammatory biomarkers, CpG sites, and PCs (estimated from 244 CpG sites significant for Δ_1 and Δ_2 after Bonferroni correction for α=0.05)

Step	Outcome	Model	New Predictor to be added	p-value of new of predictor	LR R^2 x100 (after new predictor added)
Step 1	ζ_1	Δ_1 + Fibrinogen + ln (TNFR1+1) +ln (Hycs+1)	cg25494227	4.50E-05	16.01%
Step 2	ζ_1	Δ_1 + Fibrinogen + ln (TNFR1+1) +ln (Hycs+1) + cg25494227	cg04474832	0.0314	16.59%
Step 3	ζ_1	Δ_1 + Fibrinogen + ln (TNFR1+1) +ln (Hycs+1) + cg25494227+cg04474832	X_fibrinogen:cg15538427	0.122*	
Final Model	ζ_1	Δ_1 + Fibrinogen + ln (TNFR1+1) +ln (Hycs+1) + cg25494227+cg04474832			16.59%

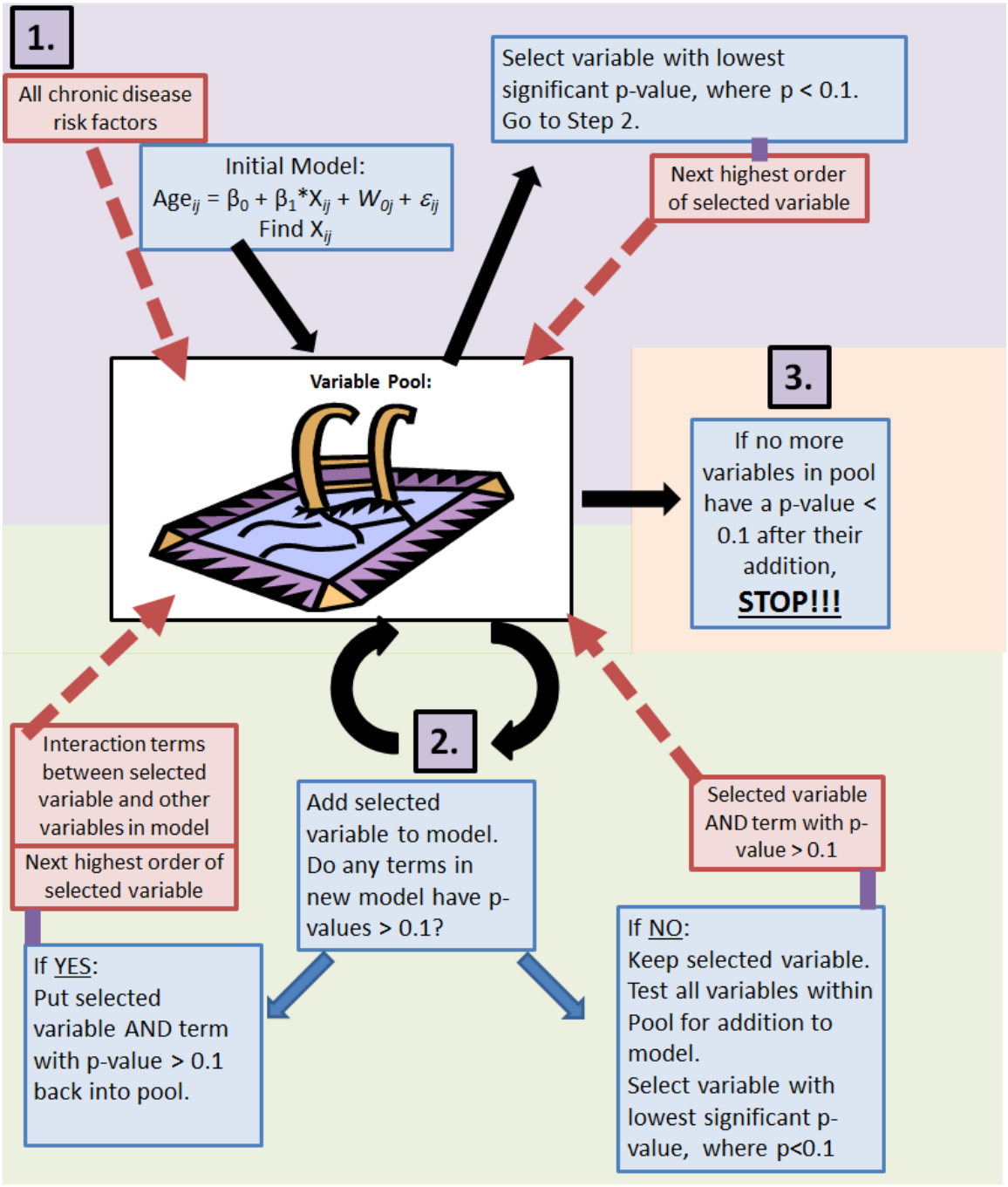
* Yellow indicates where p-value of additional variable was > 0.10

B) ζ_2 modeled with inflammatory biomarkers and PCs (estimated from 148 CpG sites significant for Δ_2 and Δ_3 after Bonferroni correction for α=0.05)

Step	Outcome	Model	New Predictor to be added	p-value of new of predictor	LR R^2 (after new predictor added)
Step 1	ζ_2	Δ_2+ ln (MCP1+1)+ ln $(MCP1+1)^2$	cg05501357	2.50E-05	31.36%
Step 2	ζ_2	Δ_2+ ln (MCP1+1)+ ln $(MCP1+1)^{2+}$ cg05501357	cg02533173	7.05E-03	33.33%
Step 3	ζ_2	Δ_2+ ln (MCP1+1)+ ln $(MCP1+1)^{2+}$ cg05501357 + cg02533173	ln_MCP1:cg05501357	0.0386	34.44%
Step 4	ζ_2	Δ_2+ ln (MCP1+1) + ln $(MCP1+1)^{2+}$ cg05501357 + cg02533173 + ln (MCP1+1)*cg05501357	cg16005443,	0.0552	35.39%
Step 5	ζ_2	Δ_2+ ln (MCP1+1) + ln $(MCP1+1)^{2+}$ cg05501357 + cg02533173 + ln (MCP1+1)*cg05501357 + PC2	PC2	0.041	36.43%
Step 6	ζ_2	Δ_2+ ln (MCP1+1) + ln $(MCP1+1)^{2+}$ cg05501357 + cg02533173 + ln (MCP1+1)*cg05501357 + PC2 + PC6	PC6	0.117*	
Final Model	ζ_2	Δ_2+ ln (MCP1+1) + ln $(MCP1+1)^{2+}$ cg05501357 + cg02533173 + ln (MCP1+1)*cg05501357 + PC2			36.43%

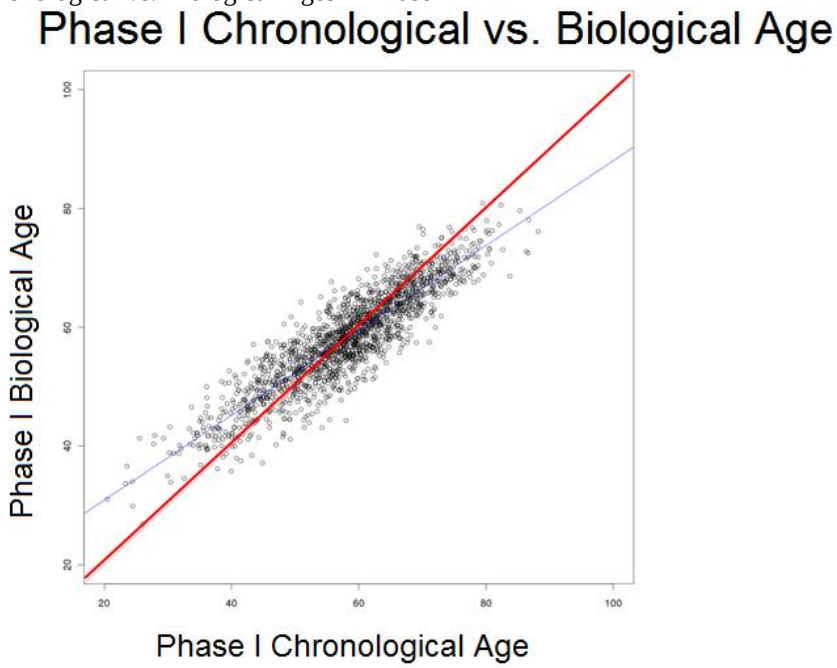
* Yellow indicates where p-value of additional variable was > 0.10

Supplementary Figure 2. Stepwise variable selection method utilized to build models predicting biological age.



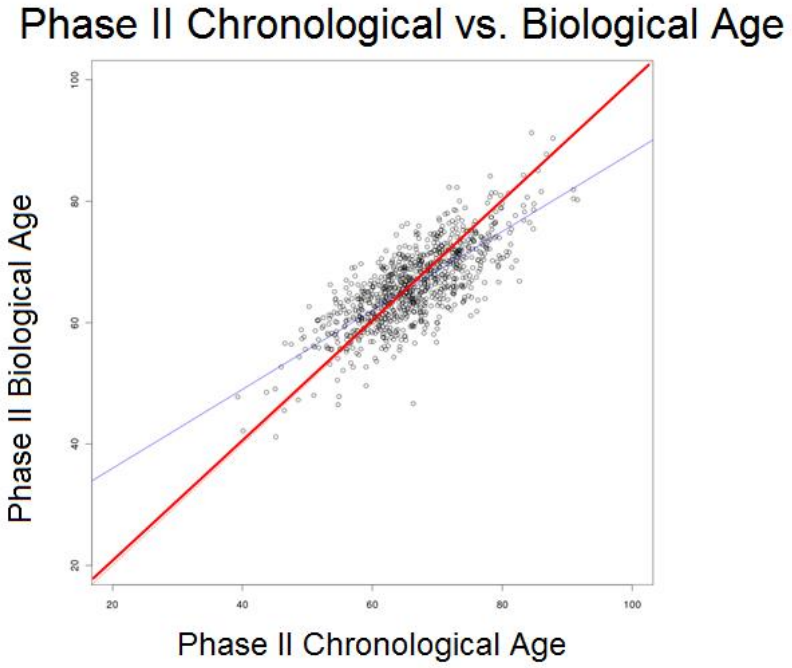
Supplementary Figure 3. Chronological vs. biological Ages in A) Phase I, B) Phase II, and, C) Phase III

A) Chronological vs. Biological Ages in Phase I



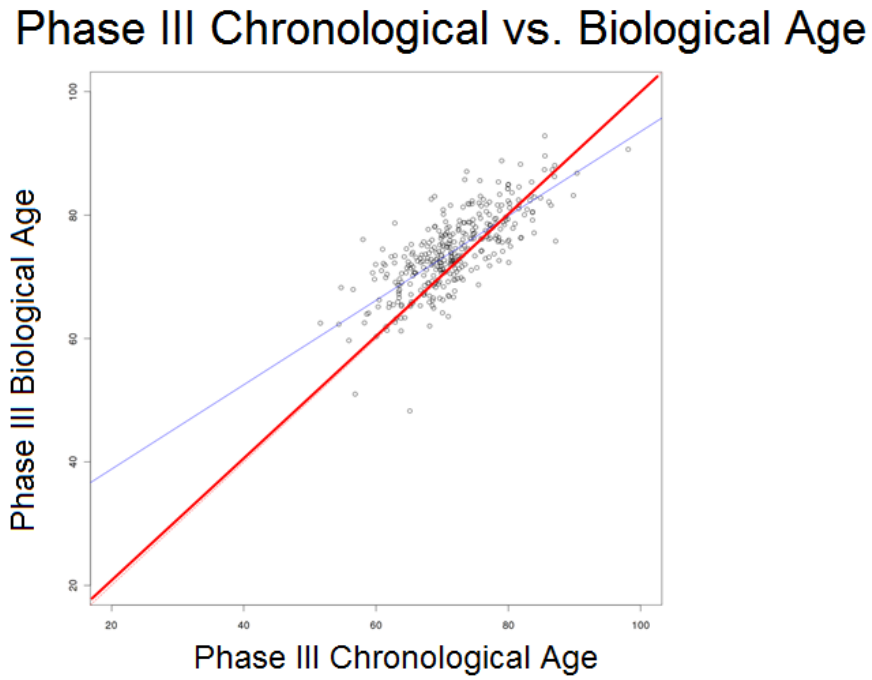
R^2 between chronological and biological ages = 0.80

B) Chronological vs. Biological Ages in Phase II



R^2 between chronological and biological ages = 0.58

C) Chronological vs. Biological Ages in Phase III

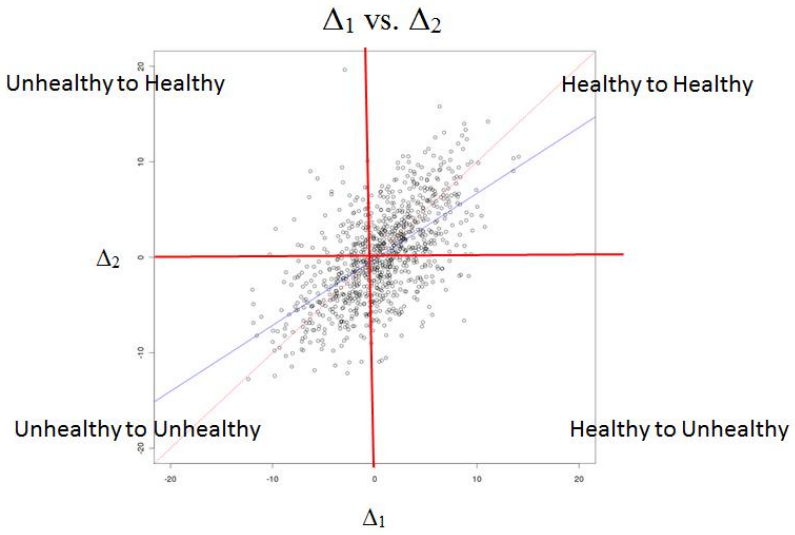


R^2 between chronological and biological ages = 0.56

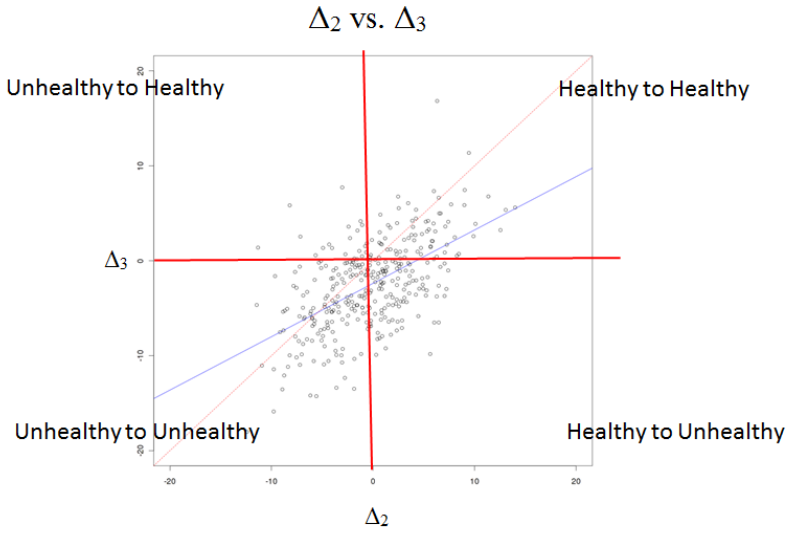
Red Line represents y=x
Blue line represents regression line.

Supplementary Figure 4. Measures of healthy age at different time points. A) Δ_1 vs. Δ_2 B) Δ_2 vs. Δ_3

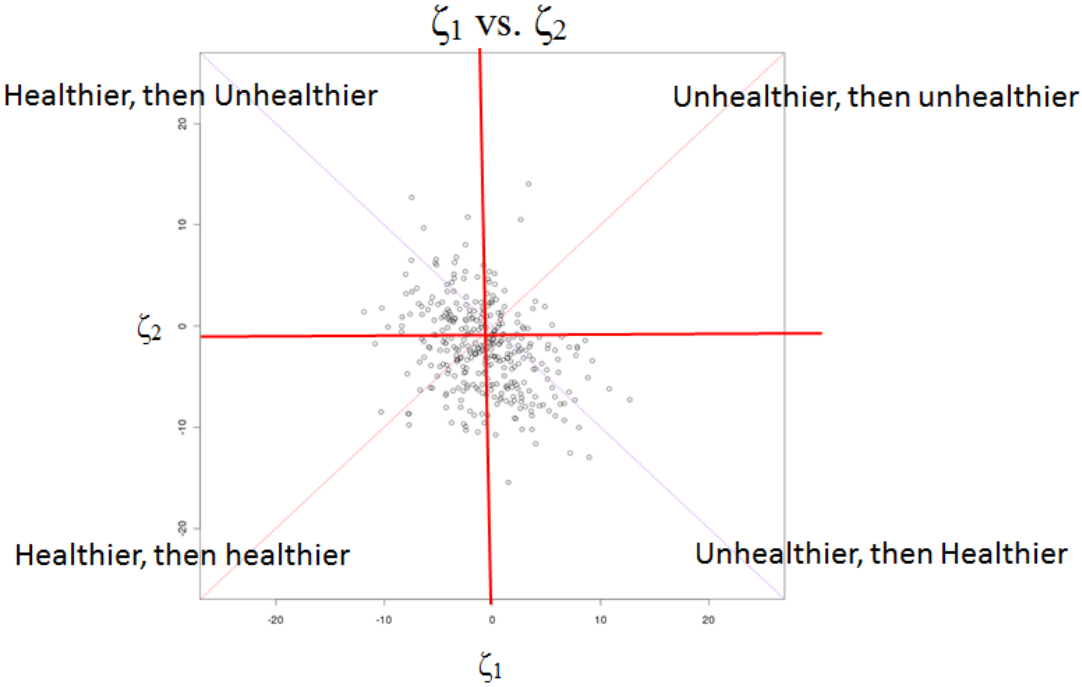
A) Δ_1 vs. Δ_2



B) Δ_2 vs. Δ_3



Supplementary Figure 5. Healthy Aging Over Time: ζ_1 vs. ζ_2



Chapter V. Conclusions and Future Directions

Summary of Findings

The GENOA sample is unique for epigenetic analyses since it contains genome-wide information about methylation on a large number of participants (N=972, after quality control). Due to the large sample size, this study has power to detect smaller effect sizes than many previously published epigenetic studies in smaller samples. Having family data also allows us to explore the genetic and environmental components of DNA methylation within adulthood by estimating heritabilities. Though this study is cross-sectional with respect to measures of inflammation and methylation, and does not allow us to examine DNA methylation information along the different life stages of individuals within our sample, participant ages span five decades which allows us to evaluate the relationship between middle and older age spectrums and DNA methylation values cross-sectionally. This aspect of the data set helps to elucidate the role of DNA methylation within and as a part of the aging process.

In this dissertation, the interplay of age and DNA methylation was explored first to assess how variation in the epigenome is associated with age (Aim 1), then by investigating the role of age within the genetic and environmental variations in the epigenome (Aim2), and, finally, in the context of chronic disease and inflammation, to generate and evaluate a population-specific prediction model of aging in order to identify a set of epigenetic markers that best predicts healthy versus unhealthy aging (Aim 3)..

First, in Aim 1 of this dissertation, the associations between chronological age and 26,428 genome-wide DNA methylation sites were investigated in a community-based sample of African Americans and found to be ubiquitous and strong across the age spectrum. In fact, 7,601 (28.8%) CpG sites were found to be significantly predicted by age (at Bonferroni corrected α=0.05), and, further, 10.3% of the variation in age was explained by the top 10 principal components estimated by the 2,095 (7.9%) sites that significantly predicted age. Of the 7,601 CpG sites that were significantly predicted by chronological age (at Bonferroni corrected α=0.05), 73.5% of them were unmethylated (mean M-Value < -2) and became *less* methylated with older age, while 2.2% were methylated (mean M-Value >2) and became *more* methylated with older age. Previous studies have established that loss of methylation in CpG dinucleotides over time may activate silenced retrotransposons and lead to genomic instability,(*83, 84*) while increases in methylation at CpG dinucleotides may prevent the binding of transcription factors and potentially suppress gene expression.(*85*) Thus, the sets of sites that are significantly associated within the scope of the age continuum may be important mediators between age and chronic disease processes through these mechanisms, and future studies may provide insight as to the processes involved in chronic disease initiation and progression, and cellular aging within the scope of methylation surrounding these genomic regions.

Because the epigenome is a known mediator of environmental and genetic effects,(*23, 57-59*) Aim 2 sought to elucidate the genetic and environmental variation within epigenetic markers by estimating the additive heritabilities of 26,428 genome-wide DNA methylation sites in a community-based sample of African-American sibships. When we adjusted for age in the estimation of heritabilities, we noticed a

differential change in the magnitude of heritability that was not expected, and led us to postulate that age-dependent gene penetrance in certain genes may be giving rise to an age-related genetic component that is distinct from age-independent genetic components. After partitioning the genetic variance within our study into age-related and age-independent components, we found that 13,260 (56.7%) of the significantly heritable methylation sites had at least 10% of their total heritability explained by age-related genetic effects and 2,402 (10.3%) had at least 50% of their total heritability explained by age-related genetic effects. This set of 2,402 sites that had at least 50% of their total heritability explained by age-related genetic effects were located within proximal promoter regions of transcription start sites of genes that were significantly enriched with sites that encoded control of cell proliferation (Bonferroni-corrected p-value = 0.0107). This indicates that methylation sites with strong age-related heritability components may be particularly important for key pathways related to cancer initiation or progression, or atherosclerotic processes due to excessive hyperplastic cell growth within vascular lesions.(*120, 121, 124*)

Further, because our previous studies illustrated that age is so strongly involved in epigenetic processes,(*103*) we examined the age-related genetic contributions to these heritabilities to identify whether the cellular aging process at the epigenomic level is more of a reflection of genetic or environmental factors. There were 1,661 CpG sites found to be significantly associated with age and have high unadjusted heritabilities ($h^2 > 0.3$). Using the Expression Analysis Systematic Explorer (EASE)(*119*) to analyze the organismal roles of these 1,661 sites, anti-pathogen response (Bonferroni-corrected p-value = 1.17×10^{-4}) and response to injury (Bonferroni-corrected p-value = 9.14×10^{-4})

were found to be significantly overrepresented. This suggests that heritable sites that are associated with age may be involved in pathways related to immune response and tissue repair, which is congruent with rising evidence that immune responses, such as inflammation, are an increasing part of the adult aging process, especially in African-American populations. These results indicate that the epigenome may be a very important layer of heritability in which to explore the impact of cellular aging on chronic disease processes due to both genetic and environmental influences. The findings of the heritability analyses within the context of Aim 2 indicate that direct genetic studies of the epigenome (e.g. genome-wide association studies, or quantitative trait loci studies) may be able to provide new insight into the actual mechanisms of aging or chronic diseases in which the epigenome is a mediator of environmental and genetic effects.

Lastly, in Aim 3, we examined the predictive capability of novel epigenetic and inflammatory biomarkers within the scope of healthy and unhealthy aging by (1) utilizing risk factors of common chronic diseases (obesity, dyslipidemia, hypertension, diabetes, and chronic kidney disease) to estimate a biological age, (2) evaluating healthy and unhealthy aging (ζ), represented as the change across time points of the differences between chronological age and biological age (Δ), and (3) considering inflammatory and epigenetic biomarkers as predictors of healthy and unhealthy aging (ζ) using a longitudinal study of African American sibships from the GENOA study (Phase 1: N_1=1854, Phase 2: N_2=972, Phase 3: N_3=484). First, we estimated biological age by modeling chronological age as predicted by measures of obesity (BMI), hypertension (diastolic BP, pulse pressure), dyslipidemia (HDL, triglycerides), diabetes (glucose, insulin), and chronic kidney disease (serum creatinine) within Phase 1 data (LR R^2 =

51.5%), and applying this model to data from Phases II and III. Though these risk factors underlie many of the chronic diseases that have the highest morbidity and mortality rates in the United States (e.g. cardiovascular disease),(*3, 5, 178*) future studies using similar estimation techniques to achieve a measure of biological age may wish to consider additional measures of physical and cognitive function, hormonal regulation, psychosocial or lifestyle factors, and other suggested indicators of healthy aging,(*133, 134*) if available, in order to attain a more complete prediction of biological age that is representative of the wide range of morbidity and disability involved in the aging process. Despite this limitation, chronological and biological ages within the GENOA sample were highly correlated, with R^2 values of 0.80, 0.58, and 0.56 for Phases I, II, and III, respectively.

In our study, healthy and unhealthy aging (ζ) was represented as the change over two periods of time, of the differences between chronological age and biological age (Δ). We examined the associations between ζ, and inflammatory and epigenetic biomarkers, measured at Phase II. Regression relationships indicated significant associations (α=0.10) of ζ_1 with fibrinogen, and TNFR1, which are thought to play a central role in atherosclerotic plaque progression, and homocysteine, which may also be involved in atherosclerosis.(*137-139, 147, 148*) It also indicated significant associations of ζ_1 with 2 CpG which are located within *C12orf59* on chromosome 12, and *ABHD14A* on chromosome 3, which may indicate their involvement in disease-associated inflammatory processes, that include cancer development, atherosclerosis, and Alzheimer's disease.(*151-153*)

Regression relationships between inflammatory and methylation markers with aging from Phase II to Phase III (ζ_2) indicated significant associations (α=0.10) of monocyte chemotactic protein-1 (MCP1), 3 CpG sites, an interaction between cg05501357 and MCP1, and a principal component (estimated from 150 CpG sites) were significant (at $p<0.10$) in the prediction of healthy aging over this time period. The chemokine MCP1 is known for an array of inflammatory efforts, including effects within the development and progression of atherosclerotic lesions, CNS inflammatory processes, bone remodeling, and the genesis of kidney damage and renal dysfunction, especially within diabetic nephropathy.(*161*) The 3 CpG sites are found within the genes *HIPK3, BRD4,* and *LILRB3,* which all play roles within immune function and response.(*163-165*) Further, the 150 CpG sites from which principal components were estimated indicated gene enrichment within the biological role of neurogenesis, which may be a link to the relationship of DNA methylation and cognitive diseases, such as Alzheimer's disease, within the scope of the aging process. Since these associations are found in peripheral blood cells, these sites may be indicators of the impending damage and dysfunction of multiple organ systems (e.g. heart, brain, kidneys) that are associated with disease processes within aging through inflammatory mechanisms. By beginning to illuminate the interplay of epigenetics and inflammation as predictors of healthy aging processes, as implied in this Aim 3 analysis, insight can be gained into the potential targets for therapeutic interventions to increase health within the aging process. With the projected increase in size of the senior citizen population, and the shift of chronic disease patterns to earlier ages within the United States, it is important to understand preclinical symptoms that are predictive of healthy and unhealthy aging, in order to allow for better

and earlier medical interventions, and potentially new therapeutics for those exhibiting signs of unhealthy aging.

Limitations

Within the scope of human DNA methylation microarray studies, the GENOA study consists of a relatively large sample size (N=972) compared to many other epidemiological studies. This allows us better power to detect associations with meaningful effect sizes within the course of Aims 1 and 2 where the entire methylation sample size is utilized compared to other groups. However, Aim 3 becomes limited in power when Phase III data is utilized since there are only 752 total African American study participants, wherein 484 had epigenetic measurements, and only 345 had measurements across the risk factors for chronic disease that are appraised within the study. This decrease in sample size may limit the efficacy of the inferences made within the scope of assessment of healthy aging between Phases II and III (ζ_2). At the same time, the decrease in sample size and the older age of participants in Phase III may be indicating a survival bias of participants. Thus, it is possible that the results from predicting aging with inflammation and DNA methylation between Phases II and III are not exactly evaluating the effect of chronic disease processes occurring with aging, since the participants are still healthy enough to participate in the GENOA sample. The results between Phases II and III may be indicating completely different processes than the results of analysis of aging between Phases I and II.

Further, the GENOA study is a cross-sectional study that has collected data at three time points. One of the main limitations of cross-sectional studies is that causality

is difficult to determine since the order of exposure induction and disease initiation is often indistinguishable.(*179*) Since Aims 1 and 2 of this dissertation scrutinize the relationship between DNA methylation and chronological age at a single time point, this limitation will only become problematic when inferences of biological age are made through the proxy variable of chronological age, since it is unclear if the biological processes are causative of methylation levels, or if methylation levels are affecting biological processes. However, though DNA methylation may affect biological aging, chronological age is a fixed variable without molecular roots and, thus, isn't affected by molecular processes. Further, since we have available chronic disease risk factor measurements across three time points, with inflammation and DNA methylation available at the midpoint, we are able to assess healthy aging with respect to inflammation and DNA methylation measurements in order to measure the phenotype before and after aging assessments. We, however, are still limited within the assessment of biological age and healthy aging due to the cross-sectional nature of the study. Along these same lines, due to the cross-sectional nature of this study, we are unable to assess whether medication usage by individuals may be affecting DNA methylation. Future analysis may be needed to examine the effects of medications, such as statins. While few studies have assessed pharmaceutical use on DNA methylation levels, it is thought that some function through epigenetic effects.(*180*) For example, it is hypothesized that statins inhibit the enzyme HMG-CoA reductase in the liver, thus stimulating the production of LDL receptors, through an unknown epigenetic mechanism.(*181*) Since the effects of medications are unknown, we did not adjust for them in our models throughout

this dissertation. Thus, associations presented in this paper may be confounded by medication usage of study participants.

Further, given the tissue-specific nature of DNA methylation patterns,(*182, 183*) the selection of the appropriate cell population for study is an important consideration. In most large epidemiological studies, biological specimens from which DNA is extracted are typically limited to peripheral blood, saliva, and hair, since collection of other tissue types is significantly more invasive. While peripheral blood samples are convenient, they may also be an important cell population to assay because they are the orchestrators of the immune response and inflammatory pathways across multiple organ systems.(*184, 185*)

Peripheral blood cells are a heterogeneous population of immune cells, made up mainly of lymphocytes (~6-48%), neutrophils (~40-75%), monocytes (~4-14%), and eosinophils (~<7%), with a small proportion of basophils (<2%).(*91, 186*) The makeup of cells within peripheral blood may differ in methylation status, thus confounding studies of DNA methylation that utilize this peripheral blood for analysis.(*186*) To avoid confounding by proportion of cell type when using peripheral blood cells within studies of DNA methylation, an adjustment for cellular makeup should be employed during normalization. Peripheral blood cellular makeup is not available within the GENOA sample due to cost efficiency in such a large study, thus this correction cannot be applied. However, and of important note, when comparing a sub-sample of peripheral blood and transformed B-lymphocytes within the GENOA population, the average correlation of methylation measurement was 0.91 within African American men and 0.89 within

African American women,(*73*) indicating that the methylation measurements within the peripheral blood are similar to B-lymphocytes

Inflammatory markers within peripheral blood play an important role in predicting disease and mortality risks. For instance, in a group of 870 high-functioning elderly individuals, those with at least 3 elevated measured inflammatory markers exhibited 6.6 times the risk of 3-year mortality, and 3.2 times the risk of 7-year mortality, than individuals with no abnormal inflammatory marker values.(*187*) Further, individual inflammatory markers measured within the bloodstream, such as CRP, IL-6, IL-18, and MPO, have been found to be predictive of cardiovascular outcomes(*188-197*) and the development of diabetes.(*198, 199*). Thus, examining peripheral blood cells, despite their heterogeneous cellular makeup, is essential for understanding the role of DNA methylation within inflammatory processes.

Future directions.

The current set of studies presented in this dissertation indicates that the relationship between age and epigenome-wide DNA methylation levels is intricately intertwined, rallying involvement from across the genome during the aging process. Previous studies have indicated that epigenetics acts as a mediator between the accrued effects of age and gene expression(*40*) that may indicate higher risk for adverse disease outcomes, such as cancer(*37, 89*) or cardiovascular disease.(*99*) Since our sample is mostly a hypertensive population, we can add to this body of knowledge by hypothesizing that chronic disease processes, such as inflammatory pathways that lead to atherosclerotic processes, work synchronically to highlight these associations between

age and DNA methylation patterns. However, further studies are needed to decipher the cooperation among inflammation, DNA methylation, and aging processes, and the impact of this cooperation on chronic disease initiation and progression. Although our study advances the knowledge in this area by identifying different and specific sets of inflammation and DNA methylation biomarkers related to healthy and unhealthy aging prospectively and retrospectively, questions about the causal mechanisms involved in aging processes remain and need further scrutiny. Many questions remain, such whether inflammation leads to DNA methylation or vice versa, or what biological interactions are at play among DNA methylation, inflammation, chronic disease processes and aging. Exploration into the questions of whether DNA methylation within certain genomic regions may be more influenced by aging, inflammation, or chronic disease events, while methylation within other genomic regions may be more influential of aging and chronic disease processes is also still needed. Larger population-based studies with multiple longitudinal measurements of DNA methylation are necessitated to begin to shed light on the collaboration across cell types and throughout cascades of cellular events.

The GENOA African-American cohort is a unique cohort that evaluates disease outcomes on hypertensive sibships from Jackson, MS in which at least 2 siblings were initially diagnosed as hypertensive before age 60 by Phase I, results concerning healthy aging and DNA methylation are very specific to this population. This population may not mirror other population groups in the United States because it is comprised of participants that are predominantly female with high rates of hypertension (Phase I consisted of 69.1% females, of which 51.1% were hypertensive, compared to 41.9% of African American females within the general population in 1999-2002(*33*)). Because of

its selection process of hypertensive individuals, the GENOA cohort likely has earlier onset and/or further advanced stages of chronic diseases due to recruitment based on hypertension status, the relationships of DNA methylation and inflammation that are inferred within the scope of age and aging are population-specific to GENOA, and may be generalizable only to similar populations of African Americans. Again, this indicates the need for larger and more generalizable samples to explore the associations of methylation within age and aging processes to which we can compare biomarkers within GENOA and other special population groups. By exploring special populations separately within the context of aging research, we can examine the differential mechanisms of inflammation and methylation within the whole system of aging, which can provide novel insight within the field of healthy aging. Further, understanding how these special populations differ from a more generalizable population will further unravel how these inflammatory and methylation mechanisms promote or restrict disease processes.

Additionally, while we defined healthy and unhealthy aging with respect to chronic disease processes, others may argue that this is not a complete picture of disability and morbidity within aging processes, and that many other biochemical, physical, cognitive, and psychosocial factors must be scrutinized in order to represent aging more accurately.(*133, 134*) However, our study sought to elucidate the role of inflammation and methylation within aging processes. Throughout the aging process, the immune system changes in cellular makeup and reduces in functionality, and the composition of inflammatory markers becomes altered which is thought to play a large role in the decrease of physical function, and initiation or exacerbated states of

inflammatory-related chronic diseases such as cardiovascular disease and type 2 diabetes.(*45, 46, 48*) Since DNA methylation was measured in this study in peripheral blood cells, and this cell population directs immune response across multiple organ systems,(*184, 185*) we attempted to measure the relationship between DNA methylation and inflammation within the scope of inflammatory-related aging processes, thus defined by known inflammatory-related chronic diseases. The methodology and models used in our study performed well within the scope of the data set, with distinct CpG sites and inflammatory measures able to predict ~10% of the variation in healthy aging. However, of course, developing more thorough models to represent the aging processes by allowing for the inclusion of other variables that account for other aspects of the aging process may allow researchers to gain a more comprehensive insight into the molecular dialogues between DNA methylation and cellular processes.

Also, since aging and chronic diseases are multidimensional, discovering the true nature of the biological mechanisms underlying their processes is important and challenging. Integrating comprehensive information from genomic, epigenomic, transcriptomic, and proteomic data will be needed to further clarify the precise mechanisms acting within these pathophysiologic processes. This is a logical research progression for this field, as computational efficiency is improving. But integration of such multidimensional variables on the large scale necessary to explain the origins of chronic disease processes is still in its infancy.

Ultimately, the goal within the field of aging research is to understand the factors within the aging process that can be acted upon to help people age in a healthier manner, with less disability and morbidity. DNA methylation is a good target for intervention

strategies because it is a modifiable association between environment and gene expression.(23) DNA methylation profiles may lead to the discovery of novel targets for preclinical diagnosis or therapeutic interventions for chronic disease events, such as cardiovascular disease or chronic kidney disease. This is a promising avenue, as epigenetic targets have been implicated in certain cancers and a few pharmaceutical agents are in clinical trial phases.(*200, 201*) As more knowledge about epigenetic changes involved in the aging process are coupled with strategies to mitigate such changes, and comprehensive information from genomic, epigenomic, transcriptomic, and proteomic data are incorporated into the clinic, the result may be more personalized therapeutics to target specific mechanisms and pathways acting within an individual that lead to chronic diseases and unhealthy aging, and provide mitigating interventions.

Conclusion.

Age and aging processes are complex and multidimensional, involving changes in cellular composition and inflammatory makeup that are controlled by both genetics and the environment.(*46, 47, 202, 203*) The work presented in this dissertation is an explorative effort to illuminate the role of epigenetics as a mediator between inflammatory and aging processes. Since epigenetics links the environment to genetic processes,(*57-59*) it is an excellent source of information to interrogate when exploring intricate processes such as healthy and unhealthy aging. Significant findings within the scope of this dissertation have begun to elucidate the important associations between DNA methylation and age, and, simultaneously, the effect of methylation in the age spectrum is partly moderated through age-related genetic factors. Further, this dissertation has indicated the importance of specific DNA methylation markers as

indicators of healthy and unhealthy aging processes that may mediate cellular aging through inflammation within the GENOA sample. These results are an important contribution to public health by allowing important molecular insights into aging processes that may pave the way for novel clinical applications of DNA methylation (e.g. preclinical predictive models of health over time, therapeutic targets).

Appendices.

Appendix 1. Background on common risk factors for chronic diseases

Common Risk Factors for Chronic Disease

Hypertension: Hypertension, or high blood pressure, is defined as having systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg, taking antihypertensive medicine, or having at least twice been told by a physician or other health professional that one has high blood pressure.(*178*) Hypertension affects 1 in 3 American adults, with the highest rates among African Americans (41.4%).(*178*) Mortality rates for essential hypertension increase considerably as age increases, with young and middle aged U.S. adults having mortality rates fewer than 7.3 per 100,000, and senior citizens having morality rates up to 195.6 per 100,000.(*3*) In fact, among senior citizens, 66.5% of 65-74 year olds are hypertensive, and 75.0% of those aged 75+ are hypertensive.(*204*) Nearly everyone becomes hypertensive eventually because blood pressure naturally rises with age as a physiological mechanism to sustain blood flow through a stiffening arterial bed.(*205*) Hypertension is further associated with end-stage renal disease, cardiovascular events, and stroke, which are also common causes of morbidity and mortality in the United States.(*3, 5*)

Systolic and diastolic blood pressures represent extremes of arterial pressure fluctuation, where systolic is the maximum and diastolic is the minimum. Pulse pressure represents the difference between systolic and diastolic blood pressures. Systolic blood pressure tends to rise throughout life, while diastolic rises only until age 50-60, after which it declines due to stiffening of the arteries.(*206*) The Framingham Heart Study found the three measures of hypertension to be differentially predictive of target organ damage effects throughout different phases of the lifetime. At younger ages (<50 years), diastolic blood pressure is the best predictor of coronary heart disease (CHD); in patients between 50 and 59 years, all three measurements are comparative when predicting CHD; and at >60 years of age, pulse pressure is the best predictor of CHD.(*207*) Many groups currently regard pulse pressure as the best predictor of target organ damage.(*208-210*)

Diabetes. Diabetes is a grouping of metabolic diseases that result in defective insulin secretion, insulin action, or both, causing hyperglycemia. Chronic hyperglycemia is associated with damage and dysfunction of multiple organ systems, including the heart and arteries, brain, kidneys, and eyes.(*211*) The current standard of diabetes diagnosis as recommended in 2009 by an International Expert Committee and adopted by the American Diabetes Association, is having glycated hemoglobin (A1C), representing average plasma glucose concentration, >6.5%,(*212*) though the World Health Organization defines diabetes as fasting glucose ≥126 mg/dL (7.0 mmol/L) or a two-hour post glucose challenge value ≥200 mg/dL (11.1 mmol/L).(*213*) The present dissertation will use fasting plasma glucose and fasting plasma insulin as measures to represent diabetes.

Overall, the age-adjusted prevalence of diabetes among United States adults is 10.9%, with large racial disparities (African Americans: 18.7% prevalence, whites: 10.2% prevalence). Diabetes prevalence increases with age, and U.S. senior citizens have rates of 26.9%, which is over 7 times that of young adults (age 20-44), and nearly twice that of middle-aged adults (age 45-64).(*11*, *204*) Approximately, 10.9 million senior citizens in the United States are estimated to be diabetic.(*204*) Not only is diabetes the seventh leading cause of death in the United States,(*3*) but it is also a major cause of chronic disease, like heart disease and stroke. It is also the leading cause of kidney failure, non-traumatic lower-limb amputations, and new cases of blindness among United States adults.(*11*) Further, ~75% of diabetics are also hypertensive, creating an even greater downstream disease risk of macrovascular and microvascular complications, such as coronary artery disease, myocardial infarction, congestive heart failure, stroke, and peripheral vascular disease.(*214*) Diabetes is a major cause of morbidity in the United States, with elderly diabetics 2-3 times more likely to report inability to perform tasks like walking short distances, climbing stairs and doing housework, than non-diabetics of the same age group.(*11*)

Obesity. Obesity is an excess accumulation of fat mass, measured most accurately by magnetic resonance imaging or computer-assisted tomography scanning. Due to the expense of these tests, anthropometric measurements, like body mass index (BMI), waist and hip circumferences, and waist-hip ratio, are often used as a proxy for fat mass.(*215*) BMI is most commonly found in the literature to represent body fat mass, and thus will be used to represent a measure of obesity in this dissertation. BMI is measured in body weight per height-squared (kg/m^2). The World Health Organization

divides BMI into four major categories: underweight (<18.50 kg/m^2), normal weight (18.50-24.99 kg/m^2), overweight (≥25.00 kg/m^2), and obese (≥ 30.00 kg/m^2). However, obesity is further split into classes I, II, and III (BMI 30.00-34.99, 35.00-39.99, ≥ 40.00 kg/m^2, respectively) to better classify the shifting severity in obesity, while simultaneously identifying groups with increased risks of morbidity and mortality.(*216*)

Obesity is associated with 111,909 excess deaths in the United States in the year 2000, with the majority of excess deaths (82,066 deaths) occurring in individuals with a BMI of 35.0 or higher.(*217*) In the United States, obesity has a U-shaped relationship with age, wherein 55-64 year olds tend to have the highest rates of obesity >40.4%, though even the young adult (age 20-34) and the oldest adults (age 75+) have rates exceeding 25% of their population.(*204*) Obesity is highly associated with increased risks of many chronic diseases and chronic disease risk factors, such as hypertension, dyslipidemia, type 2 diabetes, heart disease, stroke, and osteoarthritis.(*9, 218-220*) This dissertation will focus on BMI as an indicator of obesity, since it is widely used in chronic disease studies and has been found to have similar associations with chronic diseases like diabetes, hypertension, and dyslipidemia that measures of waist-hip ratio and waist circumference.(*221, 222*)

Dyslipidemia. Dyslipidemia is commonly characterized by an unbalanced lipid triad of elevated triglycerides (>150 mg/dL) and small low density lipoproteins (LDL-C) particles (>100 mg/dL), and reduced high density lipoprotein (HDL-C) particles (<40 mg/dL).(*8*) The prevalence of dyslipidemia shows an increasing trend with age, and senior citizens are shown to have more than twice the prevalence (39.1%) of lipid imbalance than middle aged adults (age 45-55) (15.5%) within 6,704 participants the

Multi-Ethnic Study of Atherosclerosis.(*223*) Elevated levels of LDL cholesterol and reduced levels of HDL are associated with the high lifetime risk of developing coronary heart disease (CHD) in the United States, risks which are 49% for men and 32% for women.(*224*) Triglycerides are also implicated in the onset of CHD, though this risk is often implicated at least in part to their correlation with abnormal levels of HDL and LDL.(*8, 225*) In this dissertation, I use measurements of HDL-C, LDL-C, triglycerides, and total cholesterol to represent the lipid triad commonly affected in dyslipidemia.

Chronic Kidney Disease (CKD). CKD is characterized by the presence of kidney damage or decreased kidney function for a duration of at least three months. While glomerular filtration rate is considered to be the best marker of kidney function, wherein declines indicate progressive kidney disease,(*226*), estimation equations of glomerular filtration rate (eGFR) based on the Modification of Diet in Renal Disease (MDRD) Study take into effect age, sex, and racial differences of glomerular filtration of serum creatinine,(*227*) and, thus, their use as a predictor within this study are confounded by the comparison with outcome variables, such as age and a defined variable representing healthy aging. Thus, serum creatinine will be used as a proxy for kidney function. According to NHANES data, CKD has increased in prevalence over time in the United States.(*33*) Potential reasons for increasing prevalence of CKD include shifting demographics, predominantly in age, race, and sex, which are all risk factors for CKD. In 2005, more than 35% of adults over age 60 had clinical characteristics of CKD, and ≥65.5% of adults over 80 had clinical CKD, wherein only 9.1% of adults aged 41 to 59 had CKD. Further, 15.8% of U.S. women were diagnosed with CKD, as compared to 12.1% of men; and blacks also had a higher rate of CKD than did whites, with 16.0%

versus 14.3%, respectively.(*228*) CKD patients have an increased risk of cardiovascular diseases, end-stage renal disease, and mortality than those without chronic kidney disease. This is due, in part, to the increase in cardiovascular risk factors, such as hypertension and diabetes, but CKD is a risk factor for cardiovascular disease and mortality on its own.(*229*)

Appendix 2. Description of 12 inflammatory markers used within this dissertation

C-Reactive Protein (CRP).

CRP is primarily synthesized by hepatocytes in response to acute and chronic inflammation.(*230*) CRP activates elements of the innate immune system, such as the classical complement system and phagocytosis, and it can bind to immunoglobulin receptors.(*231*) It up-regulates the expression of adhesion molecules in endothelial cells, inhibits nitric-oxide synthase expression, and increases the release of cytokines IL-1, IL-6, IL-18, and TNF-α.(*232*) Normal levels of CRP in populations without acute illness are <2 mg/L, but levels can increase to 300 mg/L in populations exhibiting acute illness.(*233, 234*) Increases of CRP levels are found in response to acute and chronic stimuli, such as infection, burns, surgery, major trauma, and other inflammatory conditions.(*235*) Specifically, elevated levels of CRP are associated with risk factors for cardiovascular disease and metabolic syndrome, such as obesity, diabetes, hypertension, low HDL-C levels, and female sex.(*236, 237*) CRP has been successfully employed in predictive modeling for risk of cardiovascular endpoints, such as myocardial infarction, stroke, and CHD death.(*189-192*) It has been found to improve prediction of cardiovascular events when modeled with other inflammatory markers, such as serum amyloid A and intercellular adhesion molecule.(*193*)

Fibrinogen.

Fibrinogen is primarily synthesized by hepatocytes under the control of the IL-6 family of cytokines,(*238*) and circulates in plasma at a normal concentration of 200-400 mg/dL.(*239*) Because fibrinogen biosynthesis is mediated by IL-6, its production increases with stress and inflammation.(*240, 241*) The main purpose of fibrinogen is within the process of hemostatic balance. After conversion of fibrinogen to fibrin with thrombin, the product serves as the substrate for fibrin clot formation, binding to platelets to support platelet aggregation, wound healing, and serving as a template for thrombin binding and the fibrinolytic system.(*239, 242*) High fibrinogen levels are found to be associated with increased risk of cardiovascular disease, stroke, and nonvascular mortality.(*243-245*) In a population of 150,000 middle-aged and elderly patients, hazard ratios for the association between increased fibrinogen and CHD, stroke, and nonvascular mortality ranged from 1.8-2.42 after adjustment for age, sex, and C-Reactive Protein.(*245*)

Homocysteine (Hycs).

Homocysteine is an intermediary amino acid formed during the conversion of methionine to cytosine. It is metabolized by the processes of transsulfuration, which necessitates vitamin B-6 as a precursor, and remethylation, of which vitamin B12 is a cofactor. Deficiencies in vitamins B6 and 12, as well as folate, can increase blood levels of homocysteine.(*142-145*) A stable baseline level of homocysteine is reached when folate intake exceeds 400 µg/day.(*246*) The association between homocysteine levels and cardiovascular disease has been somewhat inconsistent,(*135-140*) with conclusions that homocysteine may be a marker of unhealthy lifestyle and poor dietary choices more

so than an independent risk factor for cardiovascular disease.(*138*) However, investigators within the Framingham Heart Study found that adding homocysteine to models with traditional risk factors to improved risk prediction of cardiovascular events.(*247*)

Intercellular adhesion molecule-1 (ICAM-1).

ICAM-1 is expressed at low-levels on vascular endothelial cells, lymphocytes, and monocytes,(*248*) and supports leukocyte-leukocyte, leukocyte-endothelial cell, and leukocyte-epithelial cell adhesion.(*249*) ICAM-1 also functions in the development of the nervous system,(*250*) and influences the immune system by binding and recruiting circulating leukocytes to the vascular endothelium,(*251*) binding T-lymphocytes with antigen presenting cells like macrophages,(*252*) and assisting in the transendothelial migration of inflammatory cells from the capillary bed into a target tissue.(*253*) Overexpression of ICAM-1 is a common characteristic of inflammatory and immune response,(*253, 254*) and increased expression of ICAM-1 across cell types has been found to occur with increased inflammatory cytokines, such as IL-1, TNF-α, and IFN-γ.(*248, 255*) ICAM-1 is found to be associated with risk factors for cardiovascular disease, such as ever smoking, diabetes, and female sex.(*256*) However, models for identifying individuals at higher risk of cardiac events have indicated mixed results in the predictive power of ICAM.(*257-259*)

Interleukin 18 (IL-18).

IL-18 is a member of the IL-1 superfamily of cytokines. Active IL-18 predominantly acts on macrophages and dendritic cells of the innate immune system,

though it is also expressed throughout the body in epithelial cells as well. It acts with other interleukins (12 and 5) in order to drive the helper T-cell and NK-cell response through induction of IFN-γ within disease processes, and, consequently, suppress immunoglobulin E (IgE) synthesis.(*260*) IL-18 is an important regulator of innate and acquired immune responses since it enhances T-cell and NK-cell maturation, thus it is a potent proinflammatory cytokine.(*260, 261*)

IL-18 is associated with insulin resistance and type 2 diabetes,(*262-266*) obesity,(*262, 267, 268*) hypertension,(*269*) and dyslipidemia.(*267*) IL-18 is highly expressed in atherosclerotic plaque macrophages,(*270*) and is thought to result in vulnerable, rupture-prone plaques via induction of IFN-γ which may inhibit fibrous cap formation.(*271, 272*) IL-18 is associated with metabolic syndrome and its components, and levels of IL-18 increase as the number of components of metabolic syndrome rises.(*267, 273*) Further, levels of IL-18 have been found to be predictive of cardiovascular events and mortality in populations with metabolic syndrome(*195, 196*) and coronary artery disease.(*197*)

Interleukin 6 (IL-6).

IL-6 is produced by a wide range of cell types, including fibroblasts,(*274*) endothelial cells,(*275*) keratinocytes,(*276*) T-cell lines,(*277, 278*) mast cells,(*279, 280*) and many tumor cell lines,(*281*) however lymphocytes represent a significant source of IL-6 production.(*278, 282, 283*) IL-6 is not usually produced by healthy cells, but its expression is induced by viral infections,(*284-287*) lipopolysaccharide,(*288*) and many inflammatory cytokines, such as IL-1,(*289*) TNF-α,(*290*) IFN-y,(*289, 291*) and platelet-

derived growth factor.(*292*) IL-6 is stimulates immune cell response via its activation of T-cells and control of their proliferation,(*293-295*) and the growth of B cells.(*296*)

Circulating IL-6 has been found to be associated with age and a number of chronic disease risk factors. Plasma levels of IL-6 have been validated in elderly cohorts, and age-related increases of circulating IL-6 have also been discovered in middle-aged subjects, who have IL-6 levels 1.8 times greater than those in younger controls.(*194, 297, 298*) Because of these strong age associations with IL-6 and the link between IL-6 and TNF-α, it is thought that IL-6 may be a good biomarker for total inflammation.(*299*) Further, because adipose tissue secretes IL-6, obesity is associated with higher circulating levels of IL-6.(*300-305*) More specifically, it is hypothesized that visceral obesity is more detrimental to health than subcutaneous obesity because omental adipose tissue produces higher levels of IL-6 than does subcutaneous adipose tissue.(*300*) Elevated IL-6 levels are also found to be associated with smoking, and are predictive of both myocardial infarction(*194*) and the development of diabetes.(*198, 199*)

Monocyte chemotactic protein-1 (MCP1).

MCP1, also known as CCL2, regulates migration and infiltration of monocytes and macrophages.(*306*) It is a chemokine produced by numerous cell types, including endothelial, epithelial, smooth muscle, mesangial, fibroblasts, astrocytes, and microglial cells,(*307-310*) but is most commonly produced by monocytes.(*311, 312*) MCP1 is the most studied member of the chemokine family, and is important within antiviral immune response within peripheral blood circulation and within tissues.(*306*) With the help of prostaglandin E2, MCP1 recruits monocytes to sites of active inflammation, and stimulates their maturity into macrophages.(*155, 156*)

Elevated levels of MCP1 has been implicated in the role of many disease processes throughout various life stages, including multiple sclerosis,(*313*) complications with HIV infection,(*314, 315*) and rheumatoid arthritis.(*316*) It has been implicated within insulin resistance and obesity since insulin appears to induce MCP1 expression in adipocytes, which may alter adipocyte function and metabolism.(*157, 158*) Further, MCP1 may play a fundamental role in the initiation and progression of atherosclerosis due to its role in recruitment of macrophages to vascular lesions.(*159, 160*) Polymorphisms within the gene encoding MCP1 have been found to be associated with increased risk of individuals suffering coronary artery disease.(*162*) Lastly, increased plasma levels of MCP1 following coronary artery balloon angioplasty are predictive of restenosis, indicating limited success of coronary interventions and higher morbidity within these patients.(*317*)

Myeloperoxidase (MPO).

MPO is a protein secreted by neutrophils and monocytes during innate inflammatory response. MPO has bactericidal properties(*318*) and is excreted predominantly by neutrophils during phagocytosis.(*319*) MPO oxidizes a number of substrates into free radical intermediates.(*320*) The MPO-derived oxidants are known to damage cells and tissues,(*319, 321-323*) which are implicated in the development of a number of inflammatory diseases, such as ischemia-reperfusion injury, respiratory distress syndrome, glomerulonephritis, arthritis, peptic ulcer formation, and gastric cancer.(*319, 324-326*) During atherosclerotic processes, MPO secreted from monocytes is thought to be the source responsible for the oxidation of LDL-C that is taken up by macrophage receptors and, consequently, forms foam cells.(*29, 327, 328*) Because MPO

is thought to be active throughout atherosclerotic processes, it may have utility in predicting the risk of cardiovascular disease development.(*188*)

Resistin.

Resistin is an adipocyte-derived polypeptide that is secreted by adipocytes and macrophages, and mainly targets the liver but also acts upon skeletal muscle and adipose tissue.(*329-331*) Resistin has more recently been implicated in inflammatory processes, wherein it is demonstrated to stimulate the synthesis and secretion of pro-inflammatory cytokines, such as TNF-α, IL-12, IL-1b, and IL-6 via the NF-κB signaling pathway.(*332, 333*) Resistin creates downstream hepatic insulin resistance,(*330, 334-336*) and engages overproduction of adipose tissue.(*337-342*) Thus, rises in resistin levels are correlated with obesity.(*343*) Resistin is involved in hepatic and skeletal muscle glucose metabolism through reduction of insulin action within the tissues.(*334-336, 344*) Resistin has also been implicated as the link between obesity and diabetes, since insulin-stimulated glucose uptake by adipocytes is enhanced by neutralization of resistin and reduced by treatment with resistin.(*345*) Furthermore, resistin may play a role in atherosclerosis since it has been shown to up-regulate both MCP1 and ICAM.(*346-348*)

Serum Amyloid A (SAA).

SAA is a conserved acute-phase protein, composed of three isotopes within plasma.(*349*) Upon stimulation by cytokines, such as IL-6 and TNF-α, SAA is synthesized primarily by hepatocytes.(*350-352*) SAA inhibits antibody production,(*353*) impedes platelet agglutination,(*354*) and induces chemotaxis for neutrophils and monocytes.(*355*) Thus, SAA appears to have an anti-inflammatory role in preventing

progression of cellular injury via tissue repair mechanisms.(*356*) During acute inflammatory responses, SAA levels can increase up to 1000 times that of baseline within just a few hours,(*357*) while SAA becomes the predominant apolipoprotein on HDL, displacing apoA1.(*358*) Chronic inflammatory states, however, change the milieu of SAA. Plasma SAA levels are chronically increased in obese individuals when compared to lean individuals, and the protein becomes expressed more predominantly by adipocytes than hepatocytes.(*359*) Further, during chronic inflammation, SAA appears to activate pro-inflammatory cytokines, such as IL-6 and MCP1, and SAA may alter fat metabolism, and impact insulin resistance and atherosclerosis through systemic changes in cytokine production.(*359*)

Tumor Necrosis Factor Receptors 1 and 2 (TNFR-1, TNFR-2).

TNF-α is a cytokine produced mainly by macrophages that is involved in acute phase inflammation, and also acts to protect the body against infectious agents and tumors, and aids in sleep regulation.(*360, 361*) TNF-α activates neutrophils and platelets, and enhances apoptosis and cell necrosis.(*362*) TNF-α binds to cells via receptors, TNFR-1 and TNFR-2. Most cell types express both TNFR-1 and TNFR-2, though TNFR-2 is preferentially expressed on cells of hematopoietic origin.(*363, 364*) The most important functions of the receptors is to mediate activation-induced cell death of T-cells(*365*) and to promote inflammatory response.(*366*) Most cytotoxic effects are mediated by TNFR-1,(*146*) while TNFR-2 is thought to mediate signals that promote tissue repair and angiogenesis.(*367*) NF-κB binds to the TNFR-associated death domain protein indirectly to TNFR-1, or directly to TNFR-2 in order to transduce intercellular signals and promote inflammatory response.(*368, 369*) Further, it is thought that the

receptors may act as mechanisms to protect from excessive TNF-α activity during inflammatory states.(*370*)

TNF-α is associated with many cardiovascular conditions, such as atherosclerosis, myocardial infarction, heart failure, and vascular endothelial response. Myocytes and macrophages within myocardial tissue are known to produce TNF-α.(*371, 372*) The cytokine increases within the myocardium after ischemia and reperfusion, burn trauma, myocardial infarction, and cardiopulmonary bypass.(*373-377*) Further, TNF-α is associated with obesity, adipocyte cell volume, and inhibition of glucose uptake in adipocytes.(*378-381*) TNF-α may mediate insulin resistance via obesity,(*305, 382, 383*) partly through increased IL-18 expression in myocytes.(*384*)

Appendix 3.The GENOA Study Population and Design

A. GENOA Study Population

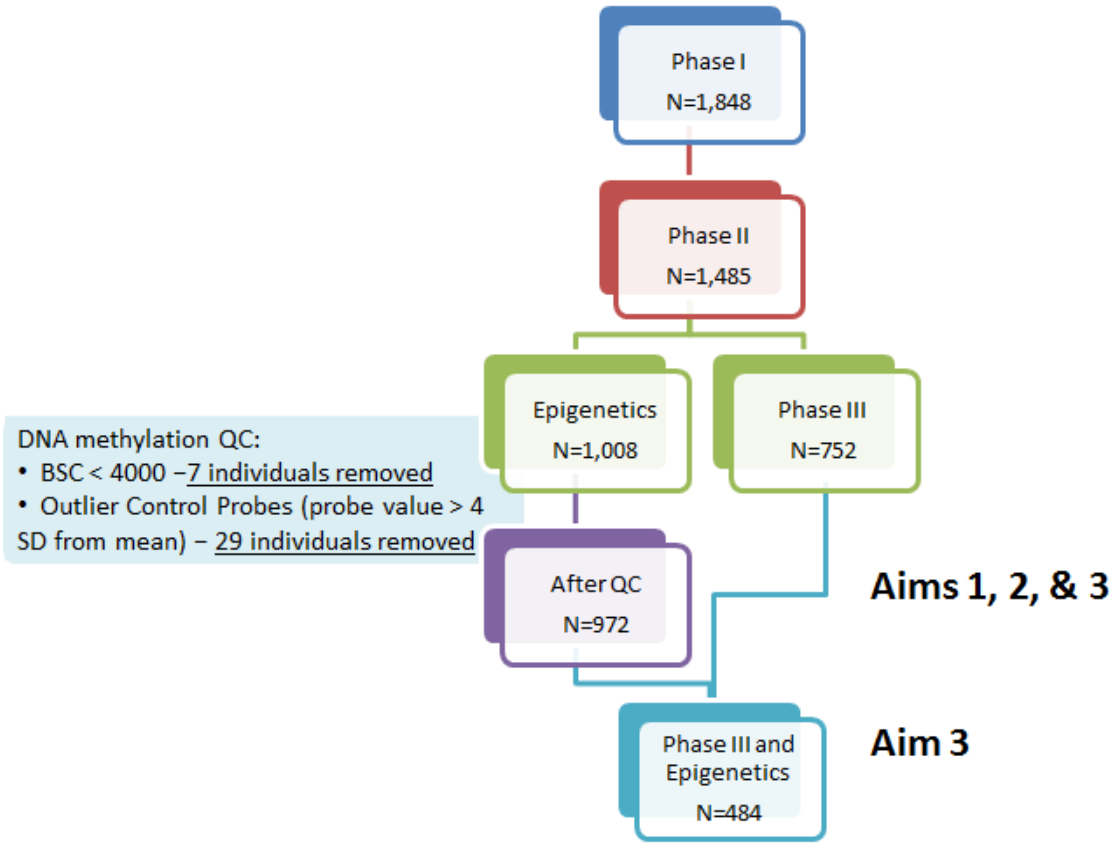
The Genetic Epidemiology Network of Arteriopathy (GENOA) study is a community-based study investigating the genetics of hypertension and its arteriosclerotic complications in non-Hispanic whites from Rochester, MN and African-Americans from Jackson, MS.(*72*) The present study will investigate the relationship between DNA methylation and aging processes in GENOA African-Americans. African-American sibships in which ≥2 siblings were diagnosed with primary hypertension before the age of 60 years (N=1,854) were recruited for an initial examination (Phase I: 1996-1999) that included standardized interviews concerning prescription drug usage, cigarette smoking, physical activity, history of hypertension, diabetes, and cardiovascular disease events; physical examination for blood pressure, height, weight, and waist circumference; and fasting blood samples for creatinine, total cholesterol, high-density lipoprotein (HDL)-cholesterol, triglycerides, glucose, and insulin. The second examination (Phase II: 2000-2004) included 1,482 Phase I participants, and included re-measurement of interview, physical examination, and blood characteristics from Phase I, as well as additional measurements of arteriosclerotic target organ damage of the heart (echocardiogram), kidney (estimated glomerular filtration rate), and peripheral arteries (ankle-brachial index). The third examination (Phase III: 2008-2011) included 752 African-American Phase II participants, and included re-measurement of interview, physical examination, and fasting blood characteristics from Phases I and II, as well as measurements of cognition and physical functioning, and coronary artery calcification (computed

tomography). Written informed consent was obtained from all subjects and approval was granted by participating Institutional Review Boards.

Due to financial limitations, DNA methylation was quantified on 1,008 of the 1,482 (68.0%) Phase II African-American participants. Methylation was assessed on DNA extracted from stored peripheral blood samples collected during the second examination. Epigenetics study participants were chosen based on the order of their Phase II exam, a method which is more likely to keep siblings in the same sibship within the study. By following this method, family-based analyses, including estimates of heritability, are executable. After quality control standards were employed (as described later in this section), 972 GENOA African-Americans had measurements of DNA methylation across 27,578 CpG sites (Figure 8).

Descriptive statistics of Phase II baseline variables, calculated using SAS software (SAS v.9.3, Cary, NC), comparing the 1,008 Phase II epigenetics study participants with the 474 Phase II participants who did not participate in the epigenetics study are presented in Supplementary Table 15. In brief, the epigenetics participants were similar in gender makeup to non-epigenetics participants (70.93% vs. 70.68% female), and epigenetics participants had higher prevalences of hypertension (82.64% vs. 70.68%), diabetes (30.56% vs. 27.00%), and having ever smoked ≥20 cigarettes within their lifetime (41.77% vs. 36.71%). Epigenetics study participants were also significantly older (mean: 66.34 vs. 56.07), less obese (mean: 31.1 kg/m^2 vs. 32.6 kg/m^2), and had significantly higher total cholesterol and HDL-C (mean: 203.98 vs. 197.93 and 58.08 vs. 55.42, respectively) than non-epigenetics study participants.

Figure 8. GENOA African American sample size.



B. Covariate Definition within GENOA

Traditional Risk Factors for Chronic Disease

Phenotypic measurements to be employed within the context of Aims 1 and 2 of this research, which analyze the associations between age and DNA methylation, and the heritability of DNA methylation markers, were collected during the Phase II examination within the African-American field center of the GENOA study. Aim 3 incorporates chronic disease risk factor data from Phases I, II, and III of the GENOA African

American field center, and inflammatory marker measurement from Phase II, along with epigenetics measurements from Phase II. Phenotype information was measured using consistent methodologies at each exam point.

Medical history and medication use was obtained via a questionnaire recorded the day of the exam. Blood was drawn by venipuncture after an overnight fast. Total serum cholesterol, high-density lipoprotein cholesterol (HDL-C), fasting glucose, and fasting insulin were measured by standard enzymatic methods.(*127, 385, 386*) Serum creatinine (SCr) was measured by the rate-Jaffe spectrophotometric method, in Phase I with standardized to later measurement methods using the equation: $SCr_{Plus} = 0.807 * (0.813 * SCr - 0.167) + 0.174$; in Phase II by either spectrophotometric method with standardization equation: $SCr_{Plus} = 0.807 * SCr + 0.174$ or by a recently nationally standardized procedure of Isotope Dilution Mass Spectrometry (IDMS); and solely via IDMS in Phase III.(*387, 388*) If the subject was being treated with insulin or oral agents, or had a fasting glucose level ≥126 mg/dL, they were considered to have diabetes.(*386*) Low-density lipoprotein cholesterol (LDL-C) was calculated as $LDL\text{-}C = Total\ Cholesterol - (HDL\text{-}C + {triglycerides}/{5})$ for individuals with triglyceride measurements < 200 mg/dL, or $LDL\text{-}C = Total\ Cholesterol - (HDL\text{-}C + 30)$ for triglyceride measurements ≥ 200 mg/dL.

Blood pressure was measured by a trained technician using a random-zero sphygmomanometer (Hawksley and Sons, London, UK) on participants who were lying in supine position in a quiet room with controlled temperature. Blood pressure was measured three times per participant, with the average of the last two values used for

analyses. Pulse pressure was calculated as the difference between systolic and diastolic blood pressures.(*127*) Hypertension was assessed as the average of the last two blood pressure level measures at the study visit (>140/90 mm Hg), or a prior diagnosis of hypertension and current treatment with antihypertensive medications.(*386*) In order to more accurately approximate the epigenetic effects on blood pressure, constants of 10 mm Hg were added to systolic blood pressure and 5 mm Hg were added to diastolic pressure if a patient reported taking blood pressure lowering medications, since anti-hypertensive medication usage may obscure the epigenetic association with hypertension. Adjusted pulse pressure was calculated as the difference between adjusted systolic and adjusted diastolic blood pressures. Brief information about all chronic disease risk factors can be seen in Table 17.

Table 17. Traditional chronic disease measurements used in the course of this dissertation.

Variable	Disease Measurement	Units
Body Mass Index (BMI)	Obesity	kg/m^2
Systolic Blood Pressure (BP)	Hypertension	mm Hg
Diastolic BP	Hypertension	mm Hg
Pulse Pressure	Hypertension	mm Hg
Adj. Systolic BP	Hypertension	mm Hg
Adj. Diastolic BP	Hypertension	mm Hg
Adj. Pulse Pressure	Hypertension	mm Hg
Total Cholesterol	Dyslipidemia	mg/dL
Triglycerides (trig)	Dyslipidemia	mg/dL
High density lipoproteins (HDL-C)	Dyslipidemia	mg/dL
Low density lipoproteins (LDL-C)	Dyslipidemia	mg/dL
Glucose	Diabetes	mg/dL
Insulin	Diabetes	mU/mL
Serum Creatinine	Chronic Kidney Disease	mg/dL

Novel Inflammatory Biomarkers of Chronic Disease.

Protein markers of vascular disease were measured in plasma (fibrinogen, serum amyloid A, interleukins 6 and 18, tumor necrosis factors 1 and 2, intercellular adhesion molecule, monocyte chemotactic protein-1, and resistin) or serum (C-reactive protein, myeloperoxidase, homocysteine) from Phase II blood samples, using commercially available solid-phase immunoassays and immunoturbidometric assays (Table 18).(*127, 128*) Additional information on biomarker assays, including precision, accuracy, stability, and methods of quality control have been previously described.(*128*)

Table 18. Measurement of inflammatory niomarkers

Biomarker	Sample Matrix	Measurement	Units
C-reactive protein (CRP)	Serum	Immunoturbidometric[1]	mg/L
Fibrinogen	Sodium-citrate plasma	Clot-based[2]	mg/dL
Homocysteine (Hycs)	Plasma	High-pressure liquid chromatography[3]	µmol/L
Intercellular adhesion molecule (ICAM)	EDTA Plasma	ELISA[4]	ng/mL
Interleukin 18 (IL-18)	EDTA Plasma	ELISA 6-plex[5]	pg/mL
Interleukin 6 (IL-6)	EDTA Plasma	ELISA 6-plex[5]	pg/mL
Monocyte chemotactic protein-1 (MCP1)	EDTA Plasma	ELISA 9-plex[5]	pg/mL
Myeloperoxidase (MPO)	Serum	ELISA[6]	ng/mL
Resistin	EDTA Plasma	ELISA[7]	ng/mL
Serum amyloid A (SAA)	EDTA Plasma	ELISA[8]	µg/mL
Tumor necrosis factor receptor-1 (TNFR-1)	EDTA Plasma	ELISA 6-plex[5]	pg/mL
Tumor necrosis factor receptor-2 (TNFR-2)	EDTA Plasma	ELISA 9-plex[5]	pg/mL

(1) Diasorin, Inc., Stillwater, MN; (2) Diagnostica STAGO, Asnieres, France; (3) Magera, Lacey, Casetta and Rinaldo (*389*); (4) R&D Systems, Minneapolis, MN; (5) SearchLight™, Pierce, Boston, MA; (6) ALPCO Diagnostics, Salem, NH; (7) BioVendor, Modrice, Czech Republic; (8) BioSource International, Camarillo, CA

Phenotypic variable transformation.

Before their use in statistical analysis, each continuous variable was assessed for its approximate normality by examining histograms and estimating skewness and kurtosis. If the variable distribution is heavily skewed, natural log transformation of the variable plus a constant is made, as ln(variable+1). For a list of variables and their transformations, see Supplementary Table 14. Outliers, as assessed as values of a variable that is more than its mean ± 4*standard deviation, was excluded from analysis due to their extremeness. By limiting the transformation to a log-transformation, the variables retain their ability to be interpreted in a biological context, and are comparable to previous studies using this data.

Supplementary Table 13. Comparison of baseline characteristics of Phase II Non-Epigenetics Participants with Phase II Epigenetics Participants. A) Continuous variables, B) Categorical variables.

A) Continuous clinical variables

	Phase II Non-Epigenetics Participants				Phase II Epigenetics Participants				
Variable	**N**	**% Missing**	**Range**	**Mean (SD)**	**N**	**% Missing**	**Range**	**Mean (SD)**	**Pr > |t|**
Age, years	474	0.00%	26.41-81.52	56.07 (9.21)	1008	0.00%	39.26-94.74	66.34 (7.6)	6.6E-49
BMI, kg/m²	472	0.42%	18.1-57.9	32.6 (7.4)	1001	0.69%	16.41-55.09	31.1 (6.09)	0.013
Systolic BP, mm Hg	473	0.21%	96-208	134.8 (19.3)	1006	0.20%	79-221	139.89 (21.12)	2.2E-06
Diastolic BP, mm Hg	474	0.00%	53-122	81.57 (10.2)	1008	0.00%	45-121	78.29 (11.01)	2.3E-05
Pulse Pressure, mm Hg	474	0.00%	19-127	53.41 (15.11)	1006	0.20%	26-127	61.64 (17.72)	6.0E-17
Adj. Systolic BP, mm Hg (1)	473	0.21%	97-218	142.05 (21.47)	1006	0.20%	89-231	148.15 (22.74)	5.5E-08
Adj. Diastolic BP, mm Hg (1)	474	0.00%	58-127	85.2 (10.95)	1008	0.00%	50-126	82.42 (11.43)	4.6E-04
Adj. Pulse Pressure, mm Hg (1)	473	0.21%	24-132	56.89 (15.72)	1004	0.40%	28-132	65.65 (18.28)	2.2E-18
Total Cholesterol, mg/dL	459	3.16%	72-348.5	197.93 (40.02)	1008	0.00%	73.5-354.5	203.98 (41.95)	7.8E-05
Triglycerides, mg/dL	454	4.22%	28.5-419.5	111.95 (58.94)	1004	0.40%	37-402.5	117.9 (56.68)	9.8E-04
ln (Triglycerides+1)	457	3.59%	3.38-6.45	4.62 (0.49)	1007	0.10%	3.64-6.27	4.69 (0.43)	0.69
HDL-C, mg/dL	457	3.59%	23.8-125.8	55.42 (16.66)	1005	0.30%	21.7-130.35	58.08 (17.48)	7.0E-03
ln (HDL+1)	460	2.95%	3.21-5.17	4 (0.3)	1008	0.00%	3.12-5.05	4.04 (0.29)	0.75
LDL-C, mg/dL	459	3.16%	23.6-253.75	121.1 (36.88)	1008	0.00%	24.85-272.1	123.84 (39.58)	0.064
Glucose, mg/dL	457	3.59%	43.5-296	108.28 (38.68)	998	0.99%	49.5-290	110.56 (34.31)	0.11
ln (Glucose+1)	457	3.59%	3.8-5.69	4.65 (0.28)	1004	0.40%	3.92-5.82	4.69 (0.27)	0.77
Insulin, mU/mL	163	65.61%	1.14-52.46	9.34 (8.53)	1005	0.30%	0.22-115.76	10.39 (12.45)	0.30
ln (Insulin+1)	163	65.61%	0.76-3.98	2.12 (0.62)	1006	0.20%	0.2-5.02	2.12 (0.74)	1.00
Menopause Age, years (2)	270	43.04%	23-59	42.6 (8.26)	699	30.65%	25-62	43.72 (7.85)	0.14

1 If a participant is listed as taking hypertensive medications, 10 mm Hg is added to systolic blood pressure, and 5 mm Hg is added to diastolic blood pressure, else recorded blood pressure is used in this variable.

2 Menopause age is represented for females only.

B) Measures of inflammation

	Phase II Non-Epigenetics Participants				Phase II Epigenetics Participants				
Variable	**N**	**% Missing**	**Range**	**Mean (SD)**	**N**	**% Missing**	**Range**	**Mean (SD)**	**Pr > |t|**
Serum Creatinine, mg/dL	460	2.95%	0.44-3.64	0.89 (0.29)	1008	0.00%	0.42-2.98	0.94 (0.3)	0.69
CRP, mg/L	347	26.79%	0.21-29.9	5.48 (5.8)	971	3.67%	0.21-29.9	6.05 (6.77)	0.37
ln (CRP+1)	347	26.79%	0.08-1.49	0.68 (0.33)	971	3.67%	0.08-1.49	0.7 (0.34)	0.87
Homocysteine, µmol/L	430	9.28%	4.5-25.2	9.38 (2.97)	1002	0.60%	4.7-31	10.57 (3.59)	7.1E-03
ln (Hycs+1)	430	9.28%	1.7-3.27	2.3 (0.26)	1002	0.60%	1.74-3.47	2.41 (0.28)	0.41
Fibrinogen, mg/dL	440	7.17%	123-684	362.44 (88.37)	1007	0.10%	120-680	369.35 (81)	1.6E-03
ICAM, ng/mL	199	58.02%	101-584	296.97 (83.09)	578	42.66%	17-625	275.38 (80.82)	3.2E-16
IL-6, pg/mL	241	49.16%	1.92-29.78	8.71 (5.06)	856	15.08%	1.42-36.45	9.09 (5.31)	0.54
ln (IL-6+1)	245	48.31%	0.46-1.58	0.95 (0.22)	878	12.90%	0.38-1.58	0.97 (0.22)	0.86
IL-18, pg/mL	240	49.37%	6.14-235.97	72.17 (38.66)	863	14.38%	1.86-252.42	69.92 (39.97)	0.19
ln (IL-18+1)	241	49.16%	0.85-2.49	1.81 (0.23)	872	13.49%	0.84-2.49	1.8 (0.25)	0.92
MPO, ng/mL	193	59.28%	11.74-141.47	47.43 (27.67)	954	5.36%	4-160.58	40.26 (25.04)	5.1E-07
ln (MPO+1)	196	58.65%	1.1-2.22	1.63 (0.23)	968	3.97%	0.7-2.22	1.56 (0.24)	0.60
MCP1, pg/mL	205	56.75%	247.44-2124.82	970.81 (319.89)	835	17.16%	159.81-2704.78	1053.43 (369.78)	2.3E-49
ln (MCP+1)	205	56.75%	2.4-3.33	2.96 (0.15)	849	15.77%	2.46-3.49	3.01 (0.16)	0.69
Resistin, ng/mL	312	34.18%	1-12.13	3.77 (1.96)	895	11.21%	1-12.32	4.07 (2.03)	0.40
ln (Resistin+1)	315	33.54%	0.3-1.2	0.65 (0.18)	902	10.52%	0.3-1.2	0.68 (0.17)	0.798
SAA, µg/mL	297	37.34%	2.35-146	26.75 (28.98)	731	27.48%	2.35-146	32.94 (30.88)	2.0E-05
ln (SAA+1)	297	37.34%	0.53-2.17	1.27 (0.39)	731	27.48%	0.53-2.17	1.38 (0.38)	0.51
TNFR-1, pg/mL	219	53.80%	102.78-3310.18	1113.11 (562.23)	872	13.49%	19.89-3844.07	1185.29 (546.83)	1.3E-27
ln (TNFR1+1)	217	54.22%	2.15-3.52	3 (0.21)	882	12.50%	2.17-3.68	3.04 (0.22)	0.76
TNFR-2, pg/mL	240	49.37%	388.3-4998.98	1751.03 (649.43)	849	15.77%	286.83-5333.39	1950.04 (831.08)	4.8E-117
ln (TNFR2+1)	241	49.16%	2.59-3.8	3.22 (0.16)	854	15.28%	2.59-3.8	3.26 (0.17)	0.70

C) Categorical variables

	Non-Epigenetics Participants			Epigenetics Participan		
	Count	**Total**	**% of Total**	**Count**	**Total**	**% of Total**
Female	335	474	70.68%	715	1,008	70.93%
Hypertensive	344	474	72.57%	833	1,008	82.64%
Has Diabetes	128	474	27.00%	308	1,008	30.56%
Ever Smoker	174	474	36.71%	421	1,008	41.77%

Supplementary Table 14. List of variables and transformations

Variable	Disease Measurement	Units	Description/Transformation
Age	Age	years	none
Body Mass Index (BMI)	Obesity	kg/m^2	$BMI = weight\ (kg)\ \frac{weight\ (kg)}{height^2\ (m^2)}$
Systolic Blood Pressure (BP)	Hypertension	mm Hg	none
Diastolic BP	Hypertension	mm Hg	none
Pulse Pressure	Hypertension	mm Hg	none
Adj. Systolic BP	Hypertension	mm Hg	If Hypertension = Yes, Adj Sys BP = Systolic BP + 10 mm Hg **Else if Hypertension = N then Adj Sys BP = Systolic BP**
Adj. Diastolic BP	Hypertension	mm Hg	If Hypertension = Yes, Adj Dia BP = Diastolic BP + 5 mm Hg Else if Hypertension = N, then Adj Dia BP = Diastolic BP
Adj. Pulse Pressure	Hypertension	mm Hg	Adj PP = Adj Sys BP – Adj Dia BP
Total Cholesterol	Dyslipidemia	mg/dL	Total molar mass of lipoproteins in blood
Triglycerides (trig)	Dyslipidemia	mg/dL	Trig is represented as: ln(Trig + 1)
High density lipoproteins (HDL-C)	Dyslipidemia	mg/dL	HDL is represented as: ln(HDL + 1)
Low density lipoproteins (LDL-C)	Dyslipidemia	mg/dL	If triglycerides < 200 mg/dL, LDL-C = $Total\ Cholesterol - (HDL\text{-}C + {}^{triglycerides}/_{5})$ Else if triglycerides ≥ 200 mg/dL, $LDL\text{-}C = Total\ Cholesterol - (HDL\text{-}C + 30)$
Glucose	Diabetes	mg/dL	Glucose is represented as: ln(Glucose + 1)
Insulin	Diabetes	mU/mL	Insulin is represented as: ln(Insulin + 1)
Serum Creatinine (SCr)	Chronic Kidney Disease	mg/dL	Lab value if measured by Isotope Dilution Mass Spectrometry (most of phase 2 values, and all of phase 3 values). Else: Phase 1: SCr=(0.807*(0.8134*SCr_Ph1-0.167)+0.1738); Phase 2: SCr =(0.807*SCr_Ph2)+0.1738.
C-Reactive Protein (CRP)	Inflammation	mg/L	CRP is represented as: ln(CRP+1)
Fibrinogen	Inflammation	mg/dL	None
Homocysteine	Inflammation	µmol/L	Hycs is represented as: ln(Hycs+1)
Intercellular Adhesion Molecule (ICAM)	Inflammation	ng/mL	None
Interleukin-6 (IL-6)	Inflammation	pg/mL	IL-6 is represented as: ln(IL-6+1)
Interleukin-18 (IL-18)	Inflammation	pg/mL	IL-18 is represented as: ln(IL-18+1)

Monocyte Chemotactic Protein-1 (MCP1)	Inflammation	pg/mL	MCP is represented as: ln(MCP1+1)
Myeloperoxidase (MPO)	Inflammation	ng/mL	MPO is represented as: ln(MPO+1)
Resistin	Inflammation	ng/mL	Resistin is represented as: ln(Resistin+1)
Serum Amyloid A (SAA)	Inflammation	µg/mL	SAA is represented as: ln(SAA+1)
Tumor Necrosis Factor Receptor-1 (TNFR-1)	Inflammation	pg/mL	TNFR1 is represented as: ln(TNFR-1+1)
Tumor Necrosis Factor Receptor-2 (TNFR-2)	Inflammation	pg/mL	TNFR2 is represented as: ln(TNFR-2+1

Appendix 4. Measurement of DNA methylation using the Illumina HumanMethylation27K microarray.

Illumina HumanMethylation27K

The Illumina HumanMethylation27 BeadChip (Illumina, San Diego, CA, USA) performs high-throughput genome-wide DNA methylation analysis, requiring only a small amount of bisulfite converted genomic DNA (1 µg) to assess methylation measurements across 27,578 CpG dinucleotides.(*390*) Each CpG site measured by Illumina has a primer length of 60 base pairs on each side of the CpG site, thus each CpG locus has a unique 122-base pair sequence identifying it.(*391*) The methylation sites measured by the Illumina Infinium 27K microarray span 14,495 genes, including 12,833 well-annotated genes described in NCBI National Center for Biotechnology Information (CCDS) database9 (Genome Build 36), 982 cancer-related targets, 144 methylation hotspots in cancer genes, and 110 micro RNA promoter regions.(*392*) On average, there are approximately two CpG sites measured for each gene represented on the Infinium 27K microarray, with an overrepresentation methylation hotspots in cancer genes, averaging 7.6 CpG sites per gene.(*392*)(Table 19)

Table 19. CpG sites and genes represented by Illumina Infinium HumanMethylation27K

Chr	Infinium 27K CpG Sites(*393*)	Infinium 27K Genes(*393*)	# CpG sites per Gene	Confirmed protein-encoding genes in human genome(*394*)	% of protein-encoding genes covered by Infinium 27K
1	2,904	1,538	1.89	2,012	76.44%
2	1,712	922	1.86	1,203	76.64%
3	1,523	795	1.92	1,040	76.44%
4	1,028	561	1.83	718	78.13%
5	1,159	622	1.86	849	73.26%
6	1,490	807	1.85	1,002	80.54%
7	1,260	632	1.99	866	72.98%
8	942	501	1.88	659	76.02%
9	1,076	566	1.90	785	72.10%
10	1,045	557	1.88	745	74.77%
11	1,735	845	2.05	1,258	67.17%
12	1,529	788	1.94	1,003	78.56%
13	493	258	1.91	318	81.13%
14	833	441	1.89	601	73.38%
15	829	416	1.99	562	74.02%
16	1,198	619	1.94	805	76.89%
17	1,586	832	1.91	1,158	71.85%
18	395	197	2.01	268	73.51%
19	1,905	992	1.92	1,399	70.91%
20	884	459	1.93	533	86.12%
21	315	180	1.75	225	80.00%
22	645	343	1.88	431	79.58%
X	1,085	599	1.81	815	73.50%
Y	7	5	1.40	45	11.11%
Total	**27,578**	**14,475**	**1.91**	**19,300**	**75.00%**

Though the Illumina Infinium 27K microarray measures CpG sites within genes spanning ~75% of the genome, it is by no means an exhaustive quantification of whole-genome DNA methylation. There are approximately 28 million CpG sites within the human genome,(*62*) a frequency of which is far less than expected based on genome GC content,(*395*) but the Infinium 27K chip is only able to critically assess a very small percentage of the total CpG sites. However, CpG methylation occurs differentially within CpG Islands – regions, approximately a few hundred base pairs in length, with elevated frequency of CpG sites.(*395*) The Infinium 27K chip contains 20,006 (72.5%) sites

within CpG Islands, methylation levels which may approximate those of methylation in surrounding CpG sites due to the correlation in DNA methylation in sites within CpG Islands. Lastly, since the Infinium 27K array more highly overlays CpG sites within cancer hotspots, and does not cover every gene in the genome, the results may be biased within this study of aging and chronic disease risk factors since crucial gene regions may be absent from the microarray.

DNA extraction and Methylation assay.

DNA was isolated from peripheral blood leukocytes from GENOA Phase II stored samples, using the PureGene DNA Isolation Kit from Gentra Systems (Minneapolis MN), and all DNA samples were run on agarose gels to verify quality. Next, 1 µg of each sample was bisulfite-converted utilizing the EZ-DNA Methylation Kit (Zymo Research, Irvine, CA). In the presence of bisulfite, unmethylated cytosines are chemically deaminated to uracil, while methylated cytosines are unaffected. Following bisulfite conversion, a thermocycling process with a short denaturation process (16 cycles of 95º, followed by 50º for one hour) occurs to improve bisulfite-conversion efficiency. Control samples were run in parallel to test the efficiency of the bisulfite conversion.

DNA samples were then denatured, neutralized, and whole-genome amplified overnight, which increases the amount of DNA sample by several thousand-fold without introducing large amounts of amplification bias, using reagents supplied by Illumina. (*396*) Samples were then enzymatically fragmented using end-point fragmentation to avoid overfragmentation, then purified by isopropanol precipitation. Samples were then hybrized to Illumina BeadChips, where the amplified, fragmented DNA samples anneal to locus-specific 50mers covering 27,578 CpG sites and a set of 56 control probes during

hybridization. Unhybridized and non-specifically hybridized DNA were then washed away. Two bead types are present for every CpG site interrogated: one corresponding to the methylated (cytosine or C) state and the other corresponding to the unmethylated (thymine or T) state. DNP- and biotin-labeled ddNTPs are used to extend the primer by a single base following allele-specific primer annealing. The array was fluorescently stained, scanned using the Illumina BeadXpress reader, and assessed for fluorescence intensities across the methylated and unmethylated bead types at 27,578 CpG sites. The scanner excites the fluorophore of the single-base extension product on the beads with a laser, then records high-resolution images of the light emitted from the fluorophores.(*396*) Twelve samples were assessed in parallel upon each BeadChip, with up to eight BeadChips analyzed in one batch within the BeadXpress Reader.(*390*)

Illumina Control Probes and Quality Control

The HumanMethylation27 BeadChip reads 56 control probes for each sample processed. These controls are split into sample-independent and sample-dependent controls. Sample-independent controls include staining, extension, target removal, and hybridization controls which evaluate the quality of each sample throughout each step in the process flow. The sample-dependent controls include bisulfite conversion, specificity, negative, and non-polymorphic controls, which evaluate performance across samples. Illumina recommends excluding samples that have bisulfite conversion control intensities of less than 4,000. Seven of our samples met the bisulfite conversion control exclusion criteria. We also established a criterion of excluding samples that had extreme outlying control probe values, which was calculated as a control probe value greater than 4 standard deviations from its mean. The expectation is that >99.99% of samples are

contained within ±4 standard deviations from the mean in a normal distribution. Twenty-nine samples were excluded due to having extreme outlier values of their control probes. (Figure 8.)

Data processing and methylation quantitation.

At each CpG site, fluorescent signals were measured from the site-specific **M** (methlyated) and **U** (unmethylated) bead types. The raw fluorescence data from the scanner was processed in Illumina BeadStudio software. To reduce batch and chip effects, the correlation structure among all 56 control probes was evaluated within channel to identify the most parsimonious subset of probes that explained the maximum amount of batch and chip variation across samples (5 probes in the red channel and 8 probes in the green channel; Table 20). Normalization was conducted by linearly regressing the 13 selected probes onto the intensity signals from the methylated and unmethylated bead types separately across each CpG site.

Table 20. List of Probes used to Standardize Methylated and Unmethylated Signals

Red Channel Probes	Green Channel Probes
EXTENSION 1190050	BISULFITE CONVERSION 4670278
EXTENSION 360446	BISULFITE CONVERSION 4670484
NON-POLYMORPHIC 1740025	BISULFITE CONVERSION 5290048
STAINING 4200736	EXTENSION 1190050
STAINING 4570020	EXTENSION 360446
	EXTENSION 520537
	NON-POLYMORPHIC 1740025
	STAINING 5340168

Two measures of methylation were calculated for each individual at every CpG site from the normalized methylated and unmethylated bead type signal intensities: the Beta Value and the M-Value. Beta Values are continuous variables ranging from 0 to 1

that are proportional to the percent methylation at each particular CpG site within each individual. The Beta Value(*104*) for an individual, *i*, at a single site, *k*, is calculated as:

$$Beta\ Value_{ik} = \frac{\max(\boldsymbol{M}_{ik},0)}{\max(\boldsymbol{U}_{ik},0)+\max(\boldsymbol{M}_{ik},0)+100}.$$

The constant added to the denominator of the Beta Value formula is a compensation for any negative value of signals that may arise from global background subtraction. Beta Values between 0 and 0.2 are generally considered to be unmethylated, and those between 0.8 and 1 are considered methylated, while values between 0.2 and 0.8 are considered semi-methylated.(*62*) The M-Value is also a commonly used measurement in microarray analysis that was more recently adapted for use in DNA methylation array data due to its ability to equalize the variance across the epigenetic sites.(*74*, *75*) The M-Value is calculated as:

$$M - Value_{ik} = log_2 \frac{\max(\boldsymbol{M}_{ik},0)+1}{\max(\boldsymbol{U}_{ik},0)+1},$$

where a constant is added to prevent large-scale changes caused by small intensity estimation errors.(*75*) The relationship between Beta and M-Values represents a logit transformation, where:

$$M - Value_{ik} = log_2 \frac{Beta\ Value_{ik}}{1 - Beta\ Value_{ik}}.$$

Unmethylated M-Values are considered to be < -2.0, methylated M-Values are > +2.0, and semi-methylated M-Values are between -2 and +2. In the results presented below, we discern the M-Value results more frequently than the Beta Value results because the statistical distributions of the M-Values conform to modeling assumptions more often than do those of the Beta Values.

DNA methylation sample and CpG site exclusion criteria.

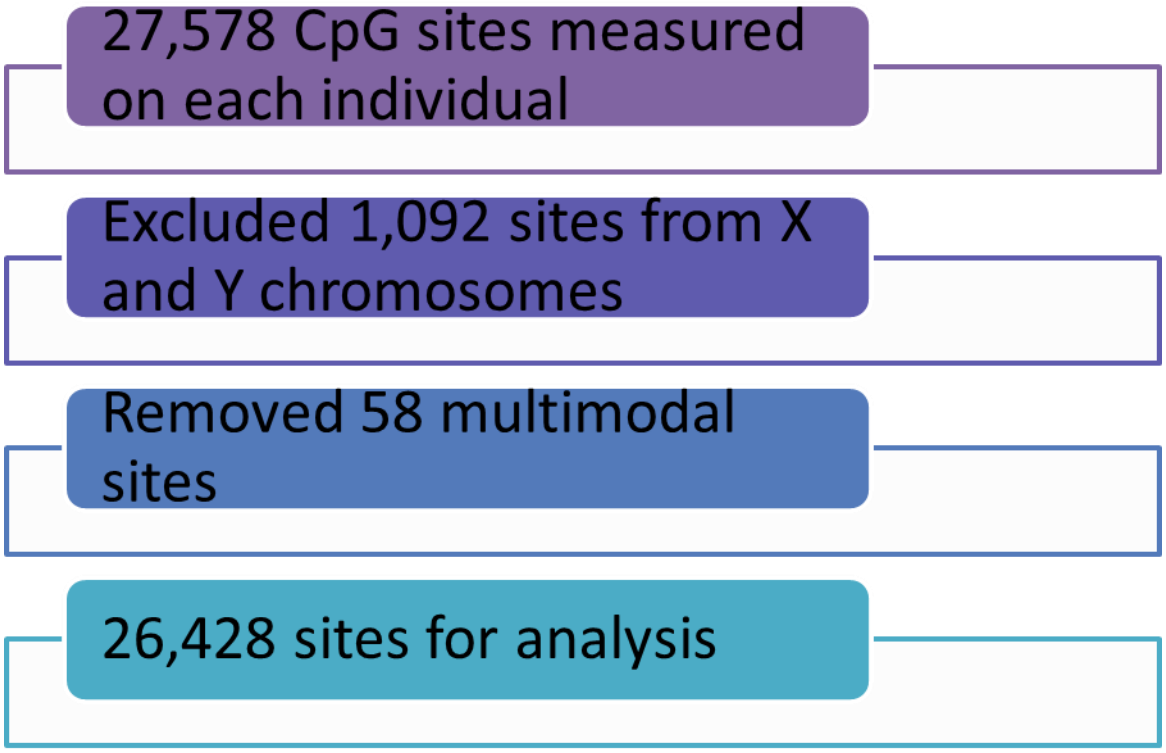
Epigenetic mechanisms are known to control X-chromosome inactivation, which occurs in all female mammals in order to achieve dosage compensation between the sexes. Methylation pattern differences between active and inactive X-chromosomes are vastly different, with inactive X-chromosomes displaying hypermethylation in gene-rich regions, and hypomethylation in gene-poor regions, and the active X-chromosome exhibiting the opposite.(*397*) Since each individual's sample is a heterogeneous collection of cells, these cells may exhibit immensely different levels of DNA methylation on the X-chromosome due to the population of cells expressing potentially different X-chromosomes within it. Thus, statistical associations between DNA methylation levels of CpG sites on the X-chromosome and phenotypic data may be weakened due to differing levels of methylation being read within the same cell population.

Further, the Illumina Infinium HumanMethylation27K microarray has been found to contain CpG site probes that are non-specific and cross-reactive to regions of the genome beyond their intended target. Within the Infinium 27K's sample of 1,085 CpG sites on the X-chromosome, 173 (15.9%) are found to have cross-reactive probes that may bind to non-target sequences around the genome, and are between 80-100% identical to non-target sequences. These probes were found after initially leading investigators to false discovery of sex-specific DNA methylation, and, thus, interpretation of these sites must be treated carefully.(*77*) In fact, Chen, Choufani, Ferreira, Grafodatskaya, Butcher and Weksberg (*77*) published a list of 2,984 CpG sites found to have probes that are 80-100% identical to non-target sequences elsewhere in the genome, which represents

10.82% of the total CpG sites on the Infinium 27K microarray. Chen, Choufani, Ferreira, Grafodatskaya, Butcher and Weksberg (*77*) also discovered a set of 876 (3.18%) polymorphic probes that overlap with at least one SNP, as recorded in the database dbSNP. Methylation levels of some CpG sites are thought to be entirely dependent on the genotype of nearby SNPs.(*77*)

Due to the considerations discussed above, autosomal CpG sites were analyzed within the scope of this dissertation. CpG sites excluded from the analysis included 1,085 X-chromosome CpG sites and 7 CpG sites on the Y-chromosome. A total of 58 autosomal CpG sites were removed from the analysis because they were found to be multimodal based on the Dip Test(*76*) using a cut-off of $p<0.001$ on the signal intensities of the methylated and/or unmethylated bead types. These multimodal sites clearly violated the statistical modeling assumptions of normality. To note, sex chromosomes had 46% of CpG sites identified as multimodal. This resulted in a total number of 26,428 CpG sites included in our analysis (Figure 9). Finally, 2,984 non-specific binding probes and 908 polymorphic probes overlapping with single nucleotide polymorphisms (SNPs)(*77*) were identified and denoted in the result tables. Though these sites are not removed from the analysis, their results should be interpreted with caution.

Figure 9. CpG sites used in analysis



Appendix 5. Description of methylation data within GENOA.

Summary of methylation data in the GENOA study

After exclusion criteria were employed, this study contained phenotype and methylation data for 972 African Americans across 26,428 CpG sites. The sample was predominantly female (70.7%) and hypertensive (82.5%), with mean age of 66.3 years and mean body mass index of 31.1 kg/m^2. Further descriptive statistics are presented in Table 21, and descriptive statistics before inclusion criteria were employed are available in Supplementary Table 15. This study population consisted of 197 singletons, and 296 sibships ranging in size from 2 to 10 siblings, with a mean of 2.6 siblings per sibship (Supplementary Table 16).

The majority of the 26,428 CpG sites were unmethylated in this population (15,227 (57.6%) sites had a mean Beta Value of <0.2; 15,217 (57.6%) had a mean M-Value of <-2.0). Across all CpG sites, the mean fluorescence intensities across the methylated bead types ranged from 482 to 39,693 (mean= 2,829), while mean fluorescence intensities on the unmethylated bead types ranged from 497 to 37,269 (mean= 6,866). Beta Value means ranged from 0.025 to 0.97, with an average mean Beta Value of 0.31, and M-Value means ranged from -5.37 to 5.07 with an average mean M-Value of -1.58 (Figure 10).

Table 21. Baseline characteristics of GENOA Phase II epigenetics participants after outlier removal

(a) Continuous Variables (Traditional Risk Factors for Common Chronic Diseases)

	N	Count of Outliers	Count of Missing Values	Range	Mean (SD)
Age, years	972	0	0	39.26-94.74	66.28 (7.6)
BMI, kg/m2	965	0	7	16.41-55.09	31.06 (6.09)
Systolic BP, mm Hg	970	0	2	79-221	139.75 (21.01)
Diastolic BP, mm Hg	972	0	0	45-121	78.33 (11)
Pulse Pressure, mm Hg	971	0	1	26-127	61.54 (17.77)
Systolic BP (Adj), mm Hg**	970	0	2	89-231	148 (22.66)
Diastolic BP (Adj), mm Hg**	972	0	0	50-126	82.46 (11.44)
Pulse Pressure (Adj), mm Hg**	969	0	3	28-132	65.54 (18.32)
Total cholesterol, mg/dL	972	0	0	73.5-354.5	203.73 (42.08)
Triglycerides, mg/dL	963	5	4	37-345	116.63 (53.77)
HDL cholesterol, mg/dL	967	2	3	21.7-122.25	57.91 (17.07)
LDL cholesterol, mg/dL	972	0	0	24.85-272.1	123.59 (39.73)
Glucose, mg/dL	951	11	10	49.5-245	108.6 (29.64)
Insulin, mU/mL	953	16	3	0.22-58.29	9.23 (8.25)
Serum creatinine, , mg/dL	961	11	0	0.42-2.16	0.92 (0.25)

** If a participant is listed as taking hypertensive medications, 10 mm Hg is added to systolic blood pressure, and 5 mm Hg is added to diastolic blood pressure, else recorded blood pressure is used in this variable.

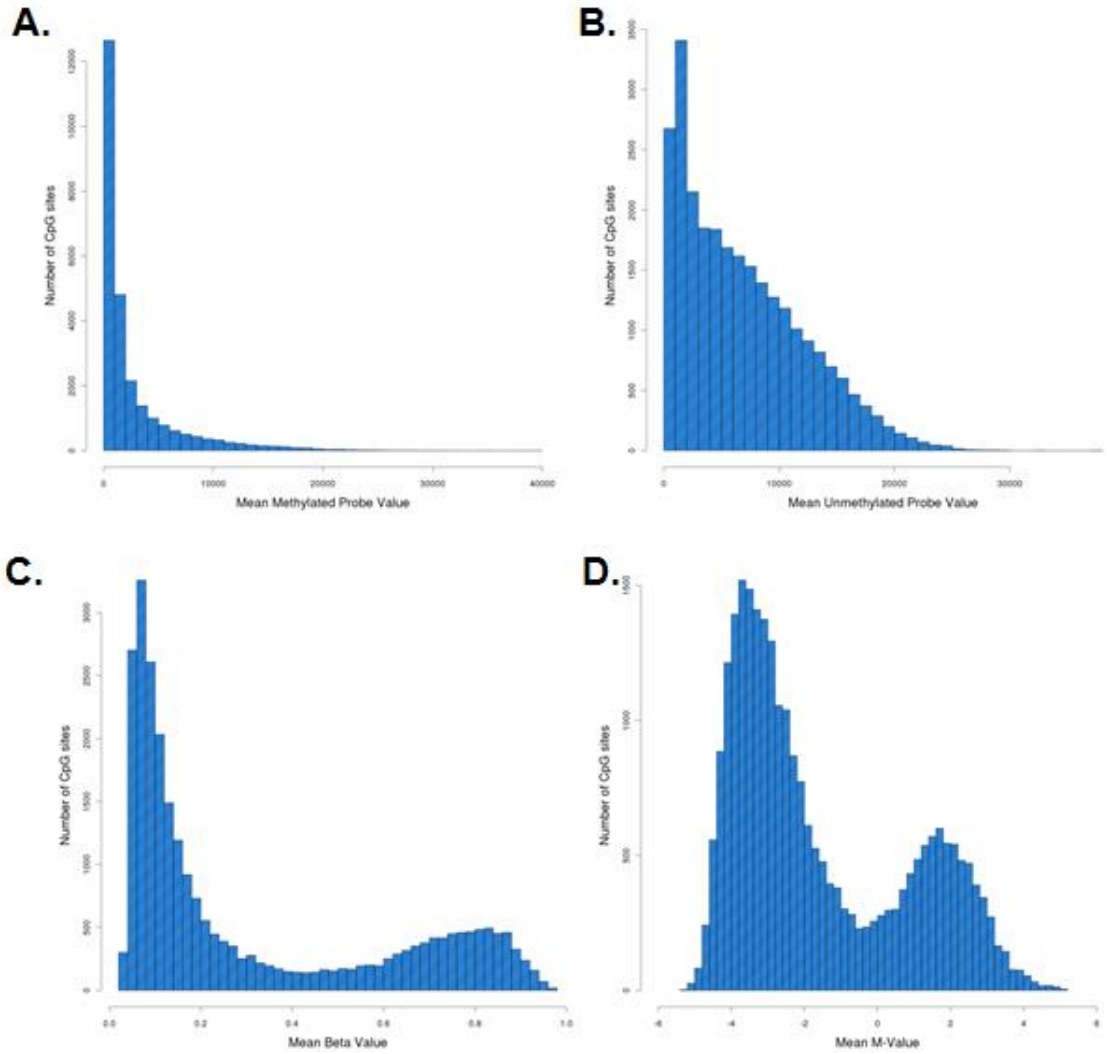
(b) Continuous Variables (Novel Inflammatory Markers)

	N	Count of Outliers	Count of Missing Values	Range	Mean (SD)
CRP, mg/L	936	0	36	0.21-29.9	6.03 (6.78)
Hycs, µmol/L	964	8	0	4.7-28.7	10.53 (3.49)
Fibrinogen, mg/dL	971	1	0	120-680	368.95 (80.98)
ICAM, ng/mL	556	3	413	17-543	274.2 (77.03)
IL6, pg/mL	818	8	146	1.42-29.35	8.83 (4.68)
IL18, pg/mL	829	4	139	1.86-214.85	69.09 (38.22)
MCP1, pg/mL	804	2	166	159.81-2,514.69	1,048.22 (359)
MPO, ng/mL	911	8	53	4-140.53	39.6 (23.05)
Resistin, ng/mL	859	1	112	1-12.16	4.07 (2.03)
SAA, µg/mL	704	0	268	2.35-146	32.94 (30.88)
TNFRI, pg/mL	840	3	129	112.01-3,108.14	1,167.39 (513.09)
TNFRII, pg/mL	818	1	153	286.83-5,170.48	1,945 (822.08)

(c) Categorical Variables

	Count	Total	Percent
Female sex	687	972	70.68%
Ever smoker	266	909	29.26%
Hypertensive	802	972	82.51%
Diabetic	298	972	30.66%

Figure 10. Distribution of means across 26,428 markers of methylation



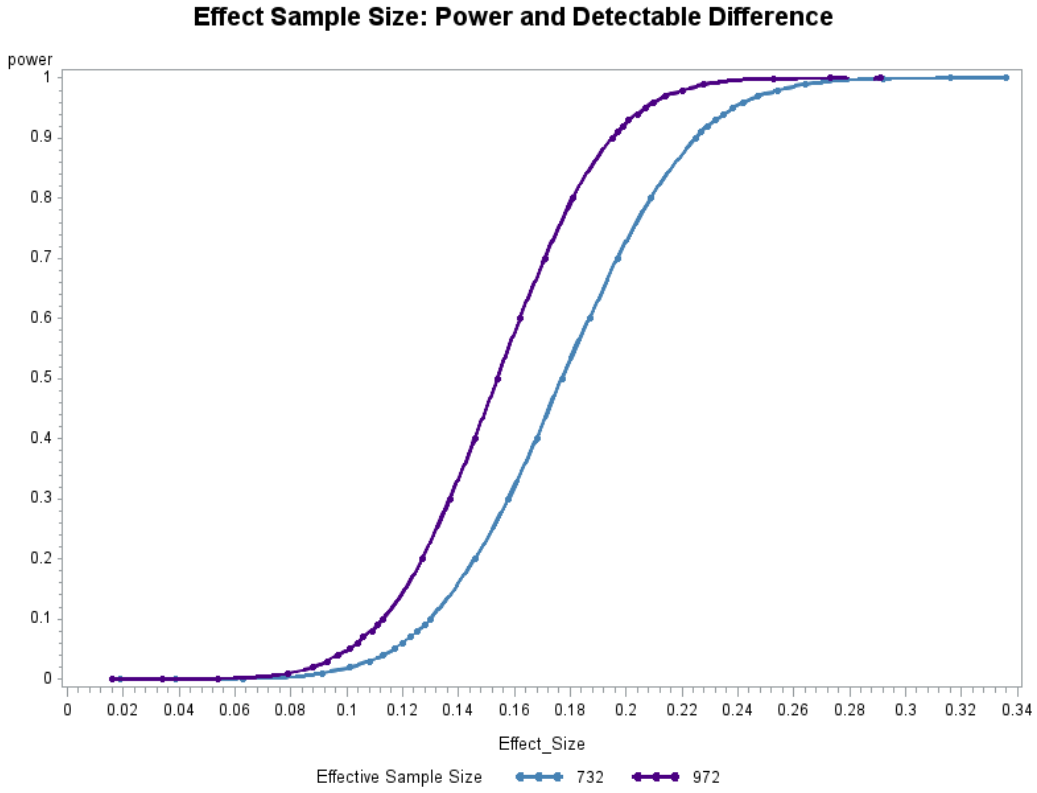
a) Methylated probe signal intensities (Range: 482 to 39,810, mean = 2,825),
b) Unmethylated probe signal intensities (Range: 497 to 37,310, mean = 6,865),
c) Beta (epigenetic) Values (Range: 0.03 to 0.97, mean =0.31)
d) M-Value (Range: -5.37 to 5.07, mean =-1.58)

Power Calculations

Power is calculated based on an effective sample size of 732 as calculated using a conservative estimate of the correlation amongst members of each sibship,(*398*) and compared to the power of the original sample sized of 972, which is the sample size if data is assumed uncorrelated. For quantitative traits, the effect sizes were estimated

using the program PS Power and Sample Size Calculations Version 3.0(*399*) and are indicated in Figure 11. In brief, at $\alpha=1.89x10^{-6}$, which is the Bonferroni-corrected α for this sample, when the effective sample size is 732, there will be >80% power to identify associations between DNA methylation markers with a standardized regression beta coefficient (effect size) >0.209, and for a power of 90% the effect size is >0.225. Using the full sample size of 972, at $\alpha=1.89x10^{-6}$, there will be >80% power to identify associations between DNA methylation markers with a standardized regression beta coefficient (effect size) >0.181, and for a power of 90% the effect size is >0.195.

Figure 11. Power and effect size.



Supplementary Table 15. Comparison of baseline characteristics of Phase II Non-Epigenetics Participants with Phase II Epigenetics Participants. A) Continuous clinical variables, B) Continuous biomarkers, C) Categorical descriptive variables.

A) Continuous clinical variables

	Phase II Non-Epigenetics Participants				Phase II Epigenetics Participants				
Variable	**N**	**% Missing**	**Range**	**Mean (SD)**	**N**	**% Missing**	**Range**	**Mean (SD)**	**Pr > |t|**
Age, years	474	0.00%	26.41-81.52	56.07 (9.21)	1008	0.00%	39.26-94.74	66.34 (7.6)	6.6E-49
BMI, kg/m²	472	0.42%	18.1-57.9	32.6 (7.4)	1001	0.69%	16.41-55.09	31.1 (6.09)	0.013
Systolic BP, mm Hg	473	0.21%	96-208	134.8 (19.3)	1006	0.20%	79-221	139.89 (21.12)	2.2E-06
Diastolic BP, mm Hg	474	0.00%	53-122	81.57 (10.2)	1008	0.00%	45-121	78.29 (11.01)	2.3E-05
Pulse Pressure, mm Hg	474	0.00%	19-127	53.41 (15.11)	1006	0.20%	26-127	61.64 (17.72)	6.0E-17
Adj. Systolic BP, mm Hg (1)	473	0.21%	97-218	142.05 (21.47)	1006	0.20%	89-231	148.15 (22.74)	5.5E-08
Adj. Diastolic BP, mm Hg (1)	474	0.00%	58-127	85.2 (10.95)	1008	0.00%	50-126	82.42 (11.43)	4.6E-04
Adj. Pulse Pressure, mm Hg (1)	473	0.21%	24-132	56.89 (15.72)	1004	0.40%	28-132	65.65 (18.28)	2.2E-18
Total Cholesterol, mg/dL	459	3.16%	72-348.5	197.93 (40.02)	1008	0.00%	73.5-354.5	203.98 (41.95)	7.8E-05
Triglycerides, mg/dL	454	4.22%	28.5-419.5	111.95 (58.94)	1004	0.40%	37-402.5	117.9 (56.68)	9.8E-04
ln (Triglycerides+1)	457	3.59%	3.38-6.45	4.62 (0.49)	1007	0.10%	3.64-6.27	4.69 (0.43)	0.69
HDL-C, mg/dL	457	3.59%	23.8-125.8	55.42 (16.66)	1005	0.30%	21.7-130.35	58.08 (17.48)	7.0E-03
ln (HDL+1)	460	2.95%	3.21-5.17	4 (0.3)	1008	0.00%	3.12-5.05	4.04 (0.29)	0.75
LDL-C, mg/dL	459	3.16%	23.6-253.75	121.1 (36.88)	1008	0.00%	24.85-272.1	123.84 (39.58)	0.064
Glucose, mg/dL	457	3.59%	43.5-296	108.28 (38.68)	998	0.99%	49.5-290	110.56 (34.31)	0.11
ln (Glucose+1)	457	3.59%	3.8-5.69	4.65 (0.28)	1004	0.40%	3.92-5.82	4.69 (0.27)	0.77
Insulin, mU/mL	163	65.61%	1.14-52.46	9.34 (8.53)	1005	0.30%	0.22-115.76	10.39 (12.45)	0.30
ln (Insulin+1)	163	65.61%	0.76-3.98	2.12 (0.62)	1006	0.20%	0.2-5.02	2.12 (0.74)	1.00
Menopause Age, years (2)	270	43.04%	23-59	42.6 (8.26)	699	30.65%	25-62	43.72 (7.85)	0.14

1 If a participant is listed as taking hypertensive medications, 10 mm Hg is added to systolic blood pressure, and 5 mm Hg is added to diastolic blood pressure, else recorded blood pressure is used in this variable.

2 Menopause age is represented for females only.

B) Measures of inflammation

	Phase II Non-Epigenetics Participants				Phase II Epigenetics Participants				
Variable	**N**	**% Missing**	**Range**	**Mean (SD)**	**N**	**% Missing**	**Range**	**Mean (SD)**	**Pr > |t|**
Serum Creatinine, mg/dL	460	2.95%	0.44-3.64	0.89 (0.29)	1008	0.00%	0.42-2.98	0.94 (0.3)	0.69
CRP, mg/L	347	26.79%	0.21-29.9	5.48 (5.8)	971	3.67%	0.21-29.9	6.05 (6.77)	0.37
ln (CRP+1)	347	26.79%	0.08-1.49	0.68 (0.33)	971	3.67%	0.08-1.49	0.7 (0.34)	0.87
Homocysteine, µmol/L	430	9.28%	4.5-25.2	9.38 (2.97)	1002	0.60%	4.7-31	10.57 (3.59)	7.1E-03
ln (Hycs+1)	430	9.28%	1.7-3.27	2.3 (0.26)	1002	0.60%	1.74-3.47	2.41 (0.28)	0.41
Fibrinogen, mg/dL	440	7.17%	123-684	362.44 (88.37)	1007	0.10%	120-680	369.35 (81)	1.6E-03
ICAM, ng/mL	199	58.02%	101-584	296.97 (83.09)	578	42.66%	17-625	275.38 (80.82)	3.2E-16
IL-6, pg/mL	241	49.16%	1.92-29.78	8.71 (5.06)	856	15.08%	1.42-36.45	9.09 (5.31)	0.54
ln (IL-6+1)	245	48.31%	0.46-1.58	0.95 (0.22)	878	12.90%	0.38-1.58	0.97 (0.22)	0.86
IL-18, pg/mL	240	49.37%	6.14-235.97	72.17 (38.66)	863	14.38%	1.86-252.42	69.92 (39.97)	0.19
ln (IL-18+1)	241	49.16%	0.85-2.49	1.81 (0.23)	872	13.49%	0.84-2.49	1.8 (0.25)	0.92
MPO, ng/mL	193	59.28%	11.74-141.47	47.43 (27.67)	954	5.36%	4-160.58	40.26 (25.04)	5.1E-07
ln (MPO+1)	196	58.65%	1.1-2.22	1.63 (0.23)	968	3.97%	0.7-2.22	1.56 (0.24)	0.60
MCP1, pg/mL	205	56.75%	247.44-2124.82	970.81 (319.89)	835	17.16%	159.81-2704.78	1053.43 (369.78)	2.3E-49
ln (MCP+1)	205	56.75%	2.4-3.33	2.96 (0.15)	849	15.77%	2.46-3.49	3.01 (0.16)	0.69
Resistin, ng/mL	312	34.18%	1-12.13	3.77 (1.96)	895	11.21%	1-12.32	4.07 (2.03)	0.40
ln (Resistin+1)	315	33.54%	0.3-1.2	0.65 (0.18)	902	10.52%	0.3-1.2	0.68 (0.17)	0.798
SAA, µg/mL	297	37.34%	2.35-146	26.75 (28.98)	731	27.48%	2.35-146	32.94 (30.88)	2.0E-05
ln (SAA+1)	297	37.34%	0.53-2.17	1.27 (0.39)	731	27.48%	0.53-2.17	1.38 (0.38)	0.51
TNFR-1, pg/mL	219	53.80%	102.78-3310.18	1113.11 (562.23)	872	13.49%	19.89-3844.07	1185.29 (546.83)	1.3E-27
ln (TNFR1+1)	217	54.22%	2.15-3.52	3 (0.21)	882	12.50%	2.17-3.68	3.04 (0.22)	0.76
TNFR-2, pg/mL	240	49.37%	388.3-4998.98	1751.03 (649.43)	849	15.77%	286.83-5333.39	1950.04 (831.08)	4.8E-117
ln (TNFR2+1)	241	49.16%	2.59-3.8	3.22 (0.16)	854	15.28%	2.59-3.8	3.26 (0.17)	0.70

C) Categorical variables

	Non-Epigenetics Participants			Epigenetics Participan		
	Count	**Total**	**% of Total**	**Count**	**Total**	**% of Total**
Female	335	474	70.68%	715	1,008	70.93%
Hypertensive	344	474	72.57%	833	1,008	82.64%
Has Diabetes	128	474	27.00%	308	1,008	30.56%
Ever Smoker	174	474	36.71%	421	1,008	41.77%

Supplementary Table 16. Sibship sizes in GENOA Epigenetic population.

Sibship Size	Sibship Count	Number of Participants
10	1	10
9	0	0
8	0	0
7	1	7
6	4	24
5	7	35
4	31	124
3	71	213
2	181	362
1	197	197
TOTAL:	**296 sibships 197 singletons**	**972**

Average sibship size: 2.6

Works Cited.

1. J. M. Wiener, J. Tilly, Population ageing in the United States of America: implications for public programmes. *International journal of epidemiology* **31**, 776 (2002).
2. O. H. Franco *et al.*, Changing course in ageing research: The healthy ageing phenotype. *Maturitas* **63**, 13 (2009).
3. A. M. Minino, S. L. Murphy, J. Xu, K. D. Kochanek, Deaths: final data for 2008. *National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System* **59**, 1 (2011).
4. A. S. Go *et al.*, Heart disease and stroke statistics--2013 update: a report from the American Heart Association. *Circulation* **127**, e6 (Jan 1, 2013).
5. National High Blood Pressure Education Program, (2004).
6. N. Bergvall, S. Cnattingius, Familial (shared environmental and genetic) factors and the foetal origins of cardiovascular diseases and type 2 diabetes: a review of the literature. *Journal of internal medicine* **264**, 205 (2008).
7. J. Kunes, J. Zicha, The interaction of genetic and environmental factors in the etiology of hypertension. *Physiological Research / Academia Scientiarum Bohemoslovaca* **58 Suppl 2**, S33 (2009).
8. E. T. o. H. B. C. i. A. National Cholesterol Education Program Expert Panel on Detection, Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* **106**, 3143 (2002).
9. National Task Force on the Prevention Treatment of Obesity, Overweight, obesity, and health risk. *Archives of Internal Medicine* **160**, 898 (2000).
10. T. Temelkova-Kurktschiev, T. Stefanov, Lifestyle and genetics in obesity and type 2 diabetes. *Experimental and clinical endocrinology & diabetes : official journal, German Society of Endocrinology [and] German Diabetes Association* **120**, 1 (2012).
11. Centers for Disease Control and Prevention, National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. (2011).
12. E. Crimmins, S. Vasunilashorn, J. K. Kim, D. Alley, Biomarkers related to aging in human populations. *Advances in Clinical Chemistry* **46**, 161 (2008).
13. L. Ferrucci, F. Giallauria, J. M. Guralnik, Epidemiology of aging. *Radiologic clinics of North America* **46**, 643 (2008).
14. J. Hebebrand, A. L. Volckmar, N. Knoll, A. Hinney, Chipping away the 'missing heritability': GIANT steps forward in the molecular elucidation of obesity - but still lots to go. *Obesity facts* **3**, 294 (2010).
15. A. M. Herskind *et al.*, The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Human genetics* **97**, 319 (1996).

16. C. M. Povel, J. M. Boer, E. J. Feskens, Shared genetic variance between the features of the metabolic syndrome: heritability studies. *Molecular genetics and metabolism* **104**, 666 (2011).

17. T. Reed, D. M. Dick, Heritability and validity of healthy physical aging (wellness) in elderly male twins. *Twin research : the official journal of the International Society for Twin Studies* **6**, 227 (2003).

18. T. A. Manolio, Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* **363**, 166 (2010).

19. E. R. Fox *et al.*, Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study. *Human molecular genetics* **20**, 2273 (2011).

20. C. L. Wassel *et al.*, Genetic determinants of the ankle-brachial index: a meta-analysis of a cardiovascular candidate gene 50K SNP panel in the candidate gene association resource (CARe) consortium. *Atherosclerosis* **222**, 138 (2012).

21. J. Yang *et al.*, Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature genetics* **44**, 369 (2012).

22. M. J. Fazzari, J. M. Greally, Introduction to epigenomics and epigenome-wide analysis. *Methods in molecular biology (Clifton, N.J.)* **620**, 243 (2010).

23. R. Jaenisch, A. Bird, Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics* **33 Suppl**, 245 (2003).

24. M. J. Fazzari, J. M. Greally, Epigenomics: beyond CpG islands. *Nature reviews.Genetics* **5**, 446 (2004).

25. H. Slomko, H. J. Heo, F. H. Einstein, Minireview: Epigenetics of obesity and diabetes in humans. *Endocrinology* **153**, 1025 (2012).

26. Administration on Aging, U.S. Department of Health and Human . Challenges of Global Aging. **April 2009**, (2009).

27. C. A. Depp, D. V. Jeste, Definitions and predictors of successful aging: a comprehensive review of larger quantitative studies. *The American Journal of Geriatric Psychiatry : Official Journal of the American Association for Geriatric Psychiatry* **14**, 6 (2006).

28. N. M. Peel, R. J. McClure, H. P. Bartlett, Behavioral determinants of healthy aging. *American Journal of Preventive Medicine* **28**, 298 (2005).

29. R. Ross, Atherosclerosis--an inflammatory disease. *The New England journal of medicine* **340**, 115 (1999).

30. W. C. Cushman, The burden of uncontrolled hypertension: morbidity and mortality associated with disease progression. *Journal of clinical hypertension (Greenwich, Conn.)* **5**, 14 (2003).

31. G. Anderson, J. Horvath, The growing burden of chronic disease in America. *Public health reports (Washington, D.C.: 1974)* **119**, 263 (2004).

32. V. Bollati *et al.*, Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mechanisms of ageing and development* **130**, 234 (2009).

33. National Center for Health Statistics. (National Center for Health Statistics (US), Hyattsvill, MD, 2011), vol. Feb.

34. National Task Force on the Prevention Treatment of Obesity, Overweight, obesity, and health risk. National Task Force on the Prevention and Treatment of Obesity. *Archives of Internal Medicine* **160**, 898 (2000).

35. P. Dandona, A. Aljada, A. Bandyopadhyay, Inflammation: the link between insulin resistance, obesity and diabetes. *Trends in immunology* **25**, 4 (Jan, 2004).

36. T. A. Pearson *et al.*, Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation* **107**, 499 (Jan 28, 2003).

37. P. W. Laird, R. Jaenisch, DNA methylation and cancer. *Human molecular genetics* **3 Spec No**, 1487 (1994).

38. S. Maegawa *et al.*, Widespread and tissue specific age-related DNA methylation changes in mice. *Genome research* **20**, 332 (2010).

39. R. F. Thompson *et al.*, Tissue-specific dysregulation of DNA methylation in aging. *Aging cell* **9**, 506 (2010).

40. M. F. Fraga *et al.*, Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10604 (2005).

41. J. Madrigano *et al.*, Aging and epigenetics: Longitudinal changes in gene-specific DNA methylation. *Epigenetics : official journal of the DNA Methylation Society* **7**, (2012).

42. L. Backdahl, A. Bushell, S. Beck, Inflammatory signalling as mediator of epigenetic modulation in tissue-specific chronic inflammation. *The international journal of biochemistry & cell biology* **41**, 176 (2009).

43. S. Ghosh *et al.*, Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. *Epigenetics : official journal of the DNA Methylation Society* **5**, 527 (2010).

44. L. Ferrero-Miliani, O. H. Nielsen, P. S. Andersen, S. E. Girardin, Chronic inflammation: importance of NOD2 and NALP3 in interleukin-1beta generation. *Clinical and experimental immunology* **147**, 227 (2007).

45. A. Larbi *et al.*, Aging of the immune system as a prognostic factor for human longevity. *Physiology (Bethesda, Md.)* **23**, 64 (2008).

46. S. Vasto *et al.*, Biomarkes of aging. *Frontiers in bioscience (Scholar edition)* **2**, 392 (2010).

47. K. L. Gibson *et al.*, B-cell diversity decreases in old age and is correlated with poor health status. *Aging Cell* **8**, 18 (Feb, 2009).

48. G. Candore, C. Caruso, G. Colonna-Romano, Inflammation, genetic background and longevity. *Biogerontology* **11**, 565 (2010).

49. D. F. Terry, M. Wilcox, M. A. McCormick, E. Lawler, T. T. Perls, Cardiovascular advantages among the offspring of centenarians. *The journals of gerontology. Series A, Biological sciences and medical sciences* **58**, M425 (May, 2003).

50. C. Franceschi *et al.*, Genes involved in immune response/inflammation, IGF1/insulin pathway and response to oxidative stress play a major role in the genetics of human longevity: the lesson of centenarians. *Mech Ageing Dev* **126**, 351 (Feb, 2005).

51. M. Capri *et al.*, The genetics of human longevity. *Annals of the New York Academy of Sciences* **1067**, 252 (May, 2006).

52. E. Cevenini *et al.*, Age-related inflammation: the contribution of different organs, tissues and systems. How to face it for therapeutic approaches. *Current pharmaceutical design* **16**, 609 (2010).

53. R. S. Illingworth, A. P. Bird, CpG islands--'a rough guide'. . *FEBS Lett.* **583**, 1713 (2009).

54. G. Kelsey, Genomic imprinting--roles and regulation in development. *Endocrine development* **12**, 99 (2007).

55. I. M. Mansuy, S. Mohanna, Epigenetics and the Human Brain: Where Nurture Meets Nature. *Cerebrum* **May**, 1 (2011).

56. J. P. Hamilton, Epigenetics: principles and practice. *Digestive diseases (Basel, Switzerland)* **29**, 130 (2011).

57. F. A. Champagne, Epigenetic influence of social experiences across the lifespan. *Developmental psychobiology* **52**, 299 (May, 2010).

58. J. P. Issa, CpG-island methylation in aging and cancer. *Current topics in microbiology and immunology* **249**, 101 (2000).

59. T. H. Bestor, Cytosine methylation and the unequal developmental potentials of the oocyte and sperm genomes. *American Journal of Human Genetics* **62**, 1269 (1998).

60. D. L. Foley *et al.*, Prospects for epigenetic epidemiology. *American journal of epidemiology* **169**, 389 (Feb 15, 2009).

61. R. A. Marrie, Demographic, genetic, and environmental factors that modify disease course. *Neurologic clinics* **29**, 323 (May, 2011).

62. M. Bibikova *et al.*, Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics* **1**, 177 (Oct, 2009).

63. I. R. Gowers *et al.*, Age-related loss of CpG methylation in the tumour necrosis factor promoter. *Cytokine* **56**, 792 (2011).

64. S. B. Baylin, J. G. Herman, Promoter hypermethylation--can this change alone ever designate true tumor suppressor gene function? *Journal of the National Cancer Institute* **93**, 664 (2001).

65. S. B. Baylin, J. G. Herman, DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends in genetics : TIG* **16**, 168 (2000).

66. S. P. Barros, S. Offenbacher, Epigenetics: connecting environment and genotype to phenotype and disease. *Journal of dental research* **88**, 400 (2009).

67. J. G. Herman, S. B. Baylin, Gene silencing in cancer in association with promoter hypermethylation. *The New England journal of medicine* **349**, 2042 (2003).

68. A. Hunter *et al.*, DNA methylation is associated with altered gene expression in AMD. *Investigative ophthalmology & visual science* **53**, 2089 (2012).

69. M. M. Suzuki, A. Bird, DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews.Genetics* **9**, 465 (2008).

70. C. Yao *et al.*, Reproducibility and concordance of differential DNA methylation and gene expression in cancer. *PloS one* **7**, e29686 (2012).

71. T. M. Gill, E. A. Gahbauer, T. E. Murphy, L. Han, H. G. Allore, Risk factors and precipitants of long-term disability in community mobility: a cohort study of older persons. *Annals of Internal Medicine* **156**, 131 (2012).

72. P. R. Daniels *et al.*, Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *The American Journal of Medicine* **116**, 676 (2004).

73. Y. V. Sun *et al.*, Comparison of the DNA methylation profiles of human peripheral blood cells and transformed B-lymphocytes. *Human genetics* **127**, 651 (2010).

74. R. A. Irizarry *et al.*, Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome research* **18**, 780 (2008).

75. P. Du *et al.*, Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC bioinformatics* **11**, 587 (2010).

76. J. A. Hartigan, P. M. Hartigan, The Dip Test of Unimodality. *The Annals of Statistics* **13**, 70 (1985).

77. Y. A. Chen *et al.*, Sequence overlap between autosomal and sex-linked probes on the Illumina HumanMethylation27 microarray. *Genomics* **97**, 214 (2011).

78. J. Pinheiro, D. Bates, D. S. DebRoy, R. C. D. Team, nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-102. (2011).

79. A. Maj, lmmfit: Goodness-of-fit-measures for linear mixed models with one-level-grouping. R package version 1.0., (2011).

80. A. Agrawal, J. Tay, G. E. Yang, S. Agrawal, S. Gupta, Age-associated epigenetic modifications in human DNA increase its immunogenicity. *Aging* **2**, 93 (2010).

81. E. G. Hoal-van Helden, P. D. van Helden, Age-related methylation changes in DNA may reflect the proliferative potential of organs. *Mutation research* **219**, 263 (1989).

82. V. L. Wilson, R. A. Smith, S. Ma, R. G. Cutler, Genomic 5-methyldeoxycytidine decreases with age. *The Journal of biological chemistry* **262**, 9948 (1987).

83. W. Barbot, A. Dupressoir, V. Lazar, T. Heidmann, Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction. *Nucleic acids research* **30**, 2365 (2002).

84. L. Menendez, B. B. Benigno, J. F. McDonald, L1 and HERV-W retrotransposons are hypomethylated in human ovarian carcinomas. *Molecular cancer* **3**, 12 (2004).

85. P. A. Jones, D. Takai, The role of DNA methylation in mammalian epigenetics. *Science (New York, N.Y.)* **293**, 1068 (2001).

86. S. Bocklandt *et al.*, Epigenetic predictor of age. *PLoS One* **6**, e14821 (2011).

87. R. S. Alisch *et al.*, Age-associated DNA methylation in pediatric populations. *Genome Res* **22**, 623 (Apr, 2012).

88. S. Numata *et al.*, DNA methylation signatures in development and aging of the human prefrontal cortex. *Am J Hum Genet* **90**, 260 (Feb 10, 2012).

89. A. E. Teschendorff *et al.*, Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* **20**, 440 (Apr, 2010).

90. R. P. Talens *et al.*, Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell* **11**, 694 (Aug, 2012).

91. J. Van den Bossche *et al.*, Reference intervals for a complete blood count determined on different automated haematology analysers: Abx Pentra 120 Retic, Coulter Gen-S, Sysmex SE 9500, Abbott Cell Dyn 4000 and Bayer Advia 120. *Clinical chemistry and laboratory medicine : CCLM / FESCC* **40**, 69 (Jan, 2002).

92. P. M. Visscher, W. G. Hill, N. R. Wray, Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* **9**, 255 (Apr, 2008).

93. D. S. Falconer, T. F. C. Mackay, *Introduction to Quantitative Genetics, Fourth Edition*. (Longman, Essex, England, 1996).

94. T. Niwa *et al.*, Inflammatory processes triggered by Helicobacter pylori infection cause aberrant DNA methylation in gastric epithelial cells. *Cancer research* **70**, 1430 (Feb 15, 2010).

95. B. T. Heijmans *et al.*, Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A* **105**, 17046 (Nov 4, 2008).

96. J. Madrigano *et al.*, Prolonged exposure to particulate pollution, genes associated with glutathione pathways, and DNA methylation in a cohort of older men. *Environmental health perspectives* **119**, 977 (Jul, 2011).

97. L. Tarantini *et al.*, Effects of particulate matter on genomic DNA methylation content and iNOS promoter methylation. *Environmental health perspectives* **117**, 217 (Feb, 2009).

98. B. Richardson, R. Yung, Role of DNA methylation in the regulation of cell function. *J Lab Clin Med* **134**, 333 (Oct, 1999).

99. A. Baccarelli, S. Ghosh, Environmental exposures, epigenetics and cardiovascular disease. *Current opinion in clinical nutrition and metabolic care* **15**, 323 (Jul, 2012).

100. D. Thomas, Gene--environment-wide association studies: emerging approaches. *Nat Rev Genet* **11**, 259 (Apr, 2010).

101. Y. C. Lee, C. Q. Lai, J. M. Ordovas, L. D. Parnell, A Database of Gene-Environment Interactions Pertaining to Blood Lipid Traits, Cardiovascular Disease and Type 2 Diabetes. *Journal of data mining in genomics & proteomics* **2**, (Jan 1, 2011).

102. C. H. Andreasen, G. Andersen, Gene-environment interactions and obesity--further aspects of genomewide association studies. *Nutrition (Burbank, Los Angeles County, Calif.)* **25**, 998 (Oct, 2009).

103. A. L. Lazarus *et al.*, Epigenomic Indicators of Age in African Americans. *Manuscript in preparation.*

104. M. Bibikova *et al.*, High-throughput DNA methylation profiling using universal bead arrays. *Genome research* **16**, 383 (2006).

105. L. Almasy, J. Blangero, Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics* **62**, 1198 (1998).

106. P. Haggarty, Nutrition and the epigenome. *Progress in molecular biology and translational science* **108**, 427 (2012).

107. L. Gong, Y. X. Pan, H. Chen, Gestational low protein diet in the rat mediates Igf2 gene expression in male offspring via altered hepatic DNA methylation. *Epigenetics* **5**, 619 (Oct 1, 2010).

108. S. Altmann *et al.*, Somatic cytochrome c (CYCS) gene expression and promoter-specific DNA methylation in a porcine model of prenatal exposure to maternal dietary protein excess and restriction. *The British journal of nutrition* **107**, 791 (Mar, 2012).

109. C. Jousse *et al.*, Perinatal undernutrition affects the methylation and expression of the leptin gene in adults: implication for the understanding of metabolic syndrome. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **25**, 3271 (Sep, 2011).

110. T. F. Oberlander *et al.*, Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. *Epigenetics : official journal of the DNA Methylation Society* **3**, 97 (2008).

111. I. C. Weaver *et al.*, Reversal of maternal programming of stress responses in adult offspring through methyl supplementation: altering epigenetic marking later in life. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **25**, 11045 (Nov 23, 2005).

112. I. C. Weaver, M. J. Meaney, M. Szyf, Maternal care effects on the hippocampal transcriptome and anxiety-mediated behaviors in the offspring that are reversible in adulthood. *Proc Natl Acad Sci U S A* **103**, 3480 (Feb 28, 2006).

113. F. P. Perera *et al.*, Prenatal airborne polycyclic aromatic hydrocarbon exposure and child IQ at age 5 years. *Pediatrics* **124**, e195 (Aug, 2009).

114. F. P. Perera *et al.*, Effects of transplacental exposure to environmental pollutants on birth outcomes in a multiethnic population. *Environmental health perspectives* **111**, 201 (Feb, 2003).

115. F. P. Perera *et al.*, Effect of prenatal exposure to airborne polycyclic aromatic hydrocarbons on neurodevelopment in the first 3 years of life among inner-city children. *Environmental health perspectives* **114**, 1287 (Aug, 2006).

116. K. Gervin *et al.*, Extensive variation and low heritability of DNA methylation identified in a twin study. *Genome research* **21**, 1813 (2011).

117. J. T. Bell *et al.*, Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS genetics* **8**, e1002629 (2012).

118. K. Gervin *et al.*, Extensive variation and low heritability of DNA methylation identified in a twin study. *Genome research* **21**, 1813 (2011).

119. D. A. Hosack, G. Dennis, Jr., B. T. Sherman, H. C. Lane, R. A. Lempicki, Identifying biological themes within lists of genes with EASE. *Genome biology* **4**, R70 (2003).

120. J. Folkman, Role of angiogenesis in tumor growth and metastasis. *Seminars in oncology* **29**, 15 (Dec, 2002).

121. J. A. Ware, M. Simons, Angiogenesis in ischemic heart disease. *Nature medicine* **3**, 158 (Feb, 1997).

122. M. Esteller, Epigenetics in cancer. *The New England journal of medicine* **358**, 1148 (2008).

123. A. P. Feinberg, Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**, 433 (2007).

124. J. J. Fuster *et al.*, Control of cell proliferation in atherosclerosis: insights from animal models and human studies. *Cardiovascular research* **86**, 254 (May 1, 2010).

125. T. Yates, M. J. Davies, K. Khunti, Obesity and chronic disease in younger people: an unfolding crisis. *The British journal of general practice : the journal of the Royal College of General Practitioners* **62**, 4 (Jan, 2012).

126. L. P. Fried, J. M. Guralnik, Disability in older adults: evidence regarding significance, etiology, and risk. *Journal of the American Geriatrics Society* **45**, 92 (Jan, 1997).

127. T. Coutinho, S. T. Turner, T. H. Mosley, I. J. Kullo, Biomarkers associated with pulse pressure in African-Americans and non-Hispanic whites. *American journal of hypertension* **25**, 145 (2012).

128. C. X. Kim *et al.*, Sex and ethnic differences in 47 candidate proteomic markers of cardiovascular disease: the Mayo Clinic proteomic markers of arteriosclerosis study. *PloS one* **5**, e9065 (2010).

129. J. Kuha, AIC and BIC : Comparisons of Assumptions and Performance. *Sociological Methods & Research* **33**, 188 (2004).

130. S. J. McLaughlin, A. M. Jette, C. M. Connell, An examination of healthy aging across a conceptual continuum: prevalence estimates, demographic patterns, and validity. *The journals of gerontology. Series A, Biological sciences and medical sciences* **67**, 783 (Jun, 2012).

131. H. Jee *et al.*, Development and application of biological age prediction models with physical fitness and physiological components in Korean adults. *Gerontology* **58**, 344 (2012).

132. C. Y. Bae *et al.*, Development of models for predicting biological age (BA) with physical, biochemical, and hormonal parameters. *Archives of gerontology and geriatrics* **47**, 253 (Sep-Oct, 2008).

133. J. W. Rowe, R. L. Kahn, Human aging: usual and successful. *Science* **237**, 143 (Jul 10, 1987).

134. C. A. Depp, S. J. Glatt, D. V. Jeste, Recent advances in research on successful or healthy aging. *Current psychiatry reports* **9**, 7 (Feb, 2007).

135. C. J. Boushey, S. A. Beresford, G. S. Omenn, A. G. Motulsky, A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. Probable benefits of increasing folic acid intakes. *JAMA* **274**, 1049 (Oct 4, 1995).

136. W. G. Christen, U. A. Ajani, R. J. Glynn, C. H. Hennekens, Blood levels of homocysteine and increased risks of cardiovascular disease: causal or casual? *Arch Intern Med* **160**, 422 (Feb 28, 2000).

137. R. Clarke *et al.*, Hyperhomocysteinemia: an independent risk factor for vascular disease. *N Engl J Med* **324**, 1149 (Apr 25, 1991).

138. T. J. Cleophas, N. Hornstra, B. van Hoogstraten, J. van der Meulen, Homocysteine, a risk factor for coronary artery disease or not? A meta-analysis. *Am J Cardiol* **86**, 1005 (Nov 1, 2000).

139. U. B. Fallon, Y. Ben-Shlomo, P. Elwood, J. B. Ubbink, G. D. Smith, Homocysteine and coronary heart disease in the Caerphilly cohort: a 10 year follow up. *Heart (British Cardiac Society)* **85**, 153 (Feb, 2001).

140. P. Knekt *et al.*, Hyperhomocystinemia: a risk factor or a consequence of coronary heart disease? *Arch Intern Med* **161**, 1589 (Jul 9, 2001).

141. C. L. Ulrey, L. Liu, L. G. Andrews, T. O. Tollefsbol, The impact of metabolism on DNA methylation. *Hum Mol Genet* **14 Spec No 1**, R139 (Apr 15, 2005).

142. E. B. Rimm *et al.*, Folate and vitamin B6 from diet and supplements in relation to risk of coronary heart disease among women. *JAMA* **279**, 359 (Feb 4, 1998).

143. K. Robinson *et al.*, Low circulating folate and vitamin B6 concentrations: risk factors for stroke, peripheral vascular disease, and coronary artery disease. European COMAC Group. *Circulation* **97**, 437 (Feb 10, 1998).

144. E. G. Vermeulen *et al.*, Effect of homocysteine-lowering treatment with folic acid plus vitamin B6 on progression of subclinical atherosclerosis: a randomised, placebo-controlled trial. *Lancet* **355**, 517 (Feb 12, 2000).

145. S. Voutilainen, T. H. Rissanen, J. Virtanen, T. A. Lakka, J. T. Salonen, Low dietary folate intake is associated with an excess incidence of acute coronary events: The Kuopio Ischemic Heart Disease Risk Factor Study. *Circulation* **103**, 2674 (Jun 5, 2001).

146. H. Hsu, J. Xiong, D. V. Goeddel, The TNF receptor 1-associated protein TRADD signals cell death and NF-kappa B activation. *Cell* **81**, 495 (May 19, 1995).

147. G. E. McKellar, D. W. McCarey, N. Sattar, I. B. McInnes, Role for TNF in atherosclerosis? Lessons from autoimmune disease. *Nature reviews. Cardiology* **6**, 410 (Jun, 2009).

148. J. Levenson, P. Giral, M. Razavian, J. Gariepy, A. Simon, Fibrinogen and silent atherosclerosis in subjects with cardiovascular risk factors. *Arteriosclerosis, thrombosis, and vascular biology* **15**, 1263 (Sep, 1995).

149. S. M. Razavian, M. Del Pino, A. Simon, J. Levenson, Increase in erythrocyte disaggregation shear stress in hypertension. *Hypertension* **20**, 247 (Aug, 1992).

150. T. N. C. f. B. I. (NCBI), in *NCBI Gene*. (2013), vol. 2013.

151. S. Srikantan, M. Gorospe, HuR function in disease. *Front Biosci* **17**, 189 (2012).

152. J. Hoshino, J. Aruga, A. Ishiguro, K. Mikoshiba, Dorz1, a novel gene expressed in differentiating cerebellar granule neurons, is down-regulated in Zic1-deficient mouse. *Brain research. Molecular brain research* **120**, 57 (Dec 12, 2003).

153. E. Salero, R. Perez-Sen, J. Aruga, C. Gimenez, F. Zafra, Transcription factors Zic1 and Zic2 bind and transactivate the apolipoprotein E gene promoter. *J Biol Chem* **276**, 1881 (Jan 19, 2001).

154. G. D. Schellenberg, T. J. Montine, The genetics and neuropathology of Alzheimer's disease. *Acta neuropathologica* **124**, 305 (Sep, 2012).

155. I. Kurth *et al.*, Monocyte selectivity and tissue localization suggests a role for breast and kidney-expressed chemokine (BRAK) in macrophage development. *The Journal of experimental medicine* **194**, 855 (Sep 17, 2001).

156. M. N. Ajuebor *et al.*, Endogenous monocyte chemoattractant protein-1 recruits monocytes in the zymosan peritonitis model. *Journal of leukocyte biology* **63**, 108 (Jan, 1998).

157. N. Kamei *et al.*, Overexpression of monocyte chemoattractant protein-1 in adipose tissues causes macrophage recruitment and insulin resistance. *J Biol Chem* **281**, 26602 (Sep 8, 2006).

158. P. Sartipy, D. J. Loskutoff, Monocyte chemoattractant protein 1 in obesity and insulin resistance. *Proc Natl Acad Sci U S A* **100**, 7265 (Jun 10, 2003).

159. M. Namiki *et al.*, Local overexpression of monocyte chemoattractant protein-1 at vessel wall induces infiltration of macrophages and formation of atherosclerotic lesion: synergism with hypercholesterolemia. *Arteriosclerosis, thrombosis, and vascular biology* **22**, 115 (Jan, 2002).

160. K. F. Kusano *et al.*, Significance of the level of monocyte chemoattractant protein-1 in human atherosclerosis. *Circ J* **68**, 671 (Jul, 2004).

161. A. Yadav, V. Saini, S. Arora, MCP-1: chemoattractant with a role beyond immunity: a review. *Clinica chimica acta; international journal of clinical chemistry* **411**, 1570 (Nov 11, 2010).

162. C. Szalai *et al.*, Involvement of polymorphisms in the chemokine system in the susceptibility for coronary artery disease (CAD). Coincidence of elevated Lp(a) and MCP-1 -2518 G/G genotype in CAD patients. *Atherosclerosis* **158**, 233 (Sep, 2001).

163. J. F. Curtin, T. G. Cotter, Live and let die: regulatory mechanisms in Fas-mediated apoptosis. *Cellular signalling* **15**, 983 (Nov, 2003).

164. A. C. Belkina, G. V. Denis, BET domain co-regulators in obesity, inflammation and cancer. *Nature reviews. Cancer* **12**, 465 (Jul, 2012).

165. L. Borges, M. L. Hsu, N. Fanger, M. Kubin, D. Cosman, A family of human lymphoid and myeloid Ig-like receptors, some of which bind to MHC class I molecules. *Journal of immunology (Baltimore, Md. : 1950)* **159**, 5192 (Dec 1, 1997).

166. E. H. Corder *et al.*, Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921 (Aug 13, 1993).

167. M. W. Logue *et al.*, A comprehensive genetic association study of Alzheimer disease in African Americans. *Archives of neurology* **68**, 1569 (Dec, 2011).

168. S. C. Wang, B. Oelze, A. Schumacher, Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLoS One* **3**, e2698 (2008).

169. M. Kivipelto *et al.*, Midlife vascular risk factors and Alzheimer's disease in later life: longitudinal, population based study. *BMJ (Clinical research ed.)* **322**, 1447 (Jun 16, 2001).

170. L. J. Launer *et al.*, Midlife blood pressure and dementia: the Honolulu-Asia aging study. *Neurobiology of aging* **21**, 49 (Jan-Feb, 2000).

171. J. A. Luchsinger *et al.*, Aggregation of vascular risk factors and risk of incident Alzheimer disease. *Neurology* **65**, 545 (Aug 23, 2005).

172. I. Skoog *et al.*, 15-year longitudinal study of blood pressure and dementia. *Lancet* **347**, 1141 (Apr 27, 1996).

173. I. Skoog, D. Gustafson, Update on hypertension and Alzheimer's disease. *Neurological research* **28**, 605 (Sep, 2006).

174. G. Hannum *et al.*, Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Molecular cell*, (Nov 20, 2012).

175. I. W. Webster, A. R. Logie, A relationship between functional age and health status in female subjects. *Journal of gerontology* **31**, 546 (Sep, 1976).

176. T. L. Dubina, V. A. Dyundikova, E. V. Zhuk, Biological age and its estimation. II. Assessment of biological age of albino rats by multiple regression analysis. *Experimental gerontology* **18**, 5 (1983).

177. R. Hochschild, Improving the precision of biological age determinations. Part 1: A new approach to calculating biological age. *Experimental gerontology* **24**, 289 (1989).

178. V. L. Roger *et al.*, Heart disease and stroke statistics--2012 update: a report from the American Heart Association. *Circulation* **125**, e2 (2012).

179. W. D. Flanders, L. Lin, J. L. Pirkle, S. P. Caudill, Assessing the direction of causality in cross-sectional studies. *American journal of epidemiology* **135**, 926 (Apr 15, 1992).

180. A. B. Csoka, M. Szyf, Epigenetic side-effects of common pharmaceuticals: a potential new field in medicine and pharmacology. *Medical hypotheses* **73**, 770 (Nov, 2009).

181. C. R. Sirtori, Pharmacology and mechanism of action of the new HMG-CoA reductase inhibitors. *Pharmacological research : the official journal of the Italian Pharmacological Society* **22**, 555 (Sep-Oct, 1990).

182. J. Igarashi *et al.*, Quantitative analysis of human tissue-specific differences in methylation. *Biochemical and biophysical research communications* **376**, 658 (2008).

183. E. Schilling, M. Rehli, Global, comparative analysis of tissue-specific promoter CpG methylation. *Genomics* **90**, 314 (2007).

184. G. J. Blake, P. M. Ridker, Inflammatory mechanisms in atherosclerosis: from laboratory evidence to clinical application. *Italian Heart Journal : official journal of the Italian Federation of Cardiology* **2**, 796 (2001).

185. R. S. Schwartz *et al.*, Detecting vulnerable plaque using peripheral blood: inflammatory and cellular markers. *Journal of interventional cardiology* **16**, 231 (2003).

186. S. Moverare-Skrtic, D. Mellstrom, L. Vandenput, M. Ehrich, C. Ohlsson, Peripheral blood leukocyte distribution and body mass index are associated with the methylation pattern of the androgen receptor promoter. *Endocrine* **35**, 204 (Apr, 2009).

187. D. B. Reuben *et al.*, Peripheral blood markers of inflammation predict mortality and functional decline in high-functioning community-dwelling older persons. *Journal of the American Geriatrics Society* **50**, 638 (Apr, 2002).

188. M. L. Brennan, S. L. Hazen, Emerging role of myeloperoxidase and oxidant stress markers in cardiovascular risk assessment. *Current opinion in lipidology* **14**, 353 (Aug, 2003).

189. S. de Ferranti, N. Rifai, C-reactive protein and cardiovascular disease: a review of risk prediction and interventions. *Clinica chimica acta; international journal of clinical chemistry* **317**, 1 (Mar, 2002).

190. J. Danesh, R. Collins, P. Appleby, R. Peto, Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies. *JAMA* **279**, 1477 (May 13, 1998).

191. L. H. Kuller, R. P. Tracy, J. Shaten, E. N. Meilahn, Relation of C-reactive protein and coronary heart disease in the MRFIT nested case-control study. Multiple Risk Factor Intervention Trial. *American journal of epidemiology* **144**, 537 (Sep 15, 1996).

192. P. M. Ridker, M. Cushman, M. J. Stampfer, R. P. Tracy, C. H. Hennekens, Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *N Engl J Med* **336**, 973 (Apr 3, 1997).

193. P. M. Ridker, C. H. Hennekens, J. E. Buring, N. Rifai, C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N Engl J Med* **342**, 836 (Mar 23, 2000).

194. H. J. Cohen, C. F. Pieper, T. Harris, K. M. Rao, M. S. Currie, The association of plasma IL-6 levels with functional disability in community-dwelling elderly. *The journals of gerontology. Series A, Biological sciences and medical sciences* **52**, M201 (Jul, 1997).

195. C. Espinola-Klein *et al.*, Impact of inflammatory markers on cardiovascular mortality in patients with metabolic syndrome. *European journal of cardiovascular prevention and rehabilitation : official journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology* **15**, 278 (Jun, 2008).

196. M. Troseid, I. Seljeflot, E. M. Hjerkinn, H. Arnesen, Interleukin-18 is a strong predictor of cardiovascular events in elderly men with the metabolic syndrome: synergistic effect of inflammation and hyperglycemia. *Diabetes Care* **32**, 486 (Mar, 2009).

197. S. Blankenberg *et al.*, Interleukin-18 is a strong predictor of cardiovascular death in stable and unstable angina. *Circulation* **106**, 24 (Jul 2, 2002).

198. M. I. Schmidt *et al.*, Markers of inflammation and prediction of diabetes mellitus in adults (Atherosclerosis Risk in Communities study): a cohort study. *Lancet* **353**, 1649 (May 15, 1999).

199. A. D. Pradhan, J. E. Manson, N. Rifai, J. E. Buring, P. M. Ridker, C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. *JAMA* **286**, 327 (Jul 18, 2001).

200. H. Heyn, M. Esteller, DNA methylation profiling in the clinic: applications and challenges. *Nat Rev Genet* **13**, 679 (Oct, 2012).

201. C. B. Yoo, P. A. Jones, Epigenetic therapy of cancer: past, present and future. *Nature reviews. Drug discovery* **5**, 37 (Jan, 2006).

202. E. Cevenini *et al.*, Age-related inflammation: the contribution of different organs, tissues and systems. How to face it for therapeutic approaches. *Curr Pharm Des* **16**, 609 (2010).

203. S. Vasto *et al.*, Inflammatory networks in ageing, age-related diseases and longevity. *Mech Ageing Dev* **128**, 83 (Jan, 2007).

204. National Center for Health Statistics, Health, United States, 2010: With Special Feature on Death and Dying. (2011).

205. M. F. O'Rourke, A. Adji, M. Namasivayam, J. Mok, Arterial aging: a review of the pathophysiology and potential for pharmacological intervention. *Drugs & aging* **28**, 779 (2011).

206. H. Smulyan, M. E. Safar, Blood pressure measurement: retrospective and prospective views. *American journal of hypertension* **24**, 628 (2011).

207. S. S. Franklin *et al.*, Does the relation of blood pressure to coronary heart disease risk change with aging? The Framingham Heart Study. *Circulation* **103**, 1245 (2001).

208. A. Benetos *et al.*, Pulse pressure: a predictor of long-term cardiovascular mortality in a French male population. *Hypertension* **30**, 1410 (1997).

209. J. A. Millar, A. F. Lever, V. Burke, Pulse pressure as a risk factor for cardiovascular events in the MRC Mild Hypertension Trial. *Journal of hypertension* **17**, 1065 (1999).

210. P. Verdecchia *et al.*, Ambulatory pulse pressure: a potent predictor of total cardiovascular risk in hypertension. *Hypertension* **32**, 983 (1998).

211. American Diabetes Association, Diagnosis and classification of diabetes mellitus. *Diabetes care* **34 Suppl 1**, S62 (2011).

212. M. J. Gillett, International Expert Committee report on the role of the A1c assay in the diagnosis of diabetes: Diabetes Care 2009; 32(7): 1327-1334. *The Clinical biochemist.Reviews / Australian Association of Clinical Biochemists* **30**, 197 (2009).

213. W. H. Organization, Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia. Report of a WHO/IDF ConsultatIon. (2006).

214. A. N. Long, S. Dagogo-Jack, Comorbidities of diabetes and hypertension: mechanisms and approach to target organ protection. *Journal of clinical hypertension (Greenwich, Conn.)* **13**, 244 (2011).

215. F. X. Pi-Sunyer, Obesity: criteria and classification. *The Proceedings of the Nutrition Society* **59**, 505 (2000).

216. W. H. Organization, Obesity: preventing and managing the global epidemic. Report of a WHO consultation. *World Health Organization technical report series* **894**, i (2000).

217. K. M. Flegal, B. I. Graubard, D. F. Williamson, M. H. Gail, Excess deaths associated with underweight, overweight, and obesity. *JAMA : the journal of the American Medical Association* **293**, 1861 (2005).

218. D. E. Alley, V. W. Chang, J. Doshi, The shape of things to come: obesity, aging, and disability. *LDI issue brief* **13**, 1 (2008).

219. E. W. Gregg, J. M. Guralnik, Is disability obesity's price of longevity? *JAMA : the journal of the American Medical Association* **298**, 2066 (2007).

220. L. B. National Heart, Institute. National Institute of, Diabetes Digestive, Kidney, Diseases,, Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults--The Evidence Report. National Institutes of Health. National Institutes of Health. *Obesity research* **6 Suppl 2**, 51S (1998).

221. R. Huxley, S. Mendis, E. Zheleznyakov, S. Reddy, J. Chan, Body mass index, waist circumference and waist:hip ratio as predictors of cardiovascular risk--a review of the literature. *European journal of clinical nutrition* **64**, 16 (2010).

222. G. Vazquez, S. Duval, D. R. Jacobs, Jr., K. Silventoinen, Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. *Epidemiologic reviews* **29**, 115 (2007).

223. D. C. Goff, Jr. *et al.*, Dyslipidemia prevalence, treatment, and control in the Multi-Ethnic Study of Atherosclerosis (MESA): gender, ethnicity, and coronary artery calcium. *Circulation* **113**, 647 (2006).

224. D. Lloyd-Jones *et al.*, Heart disease and stroke statistics--2010 update: a report from the American Heart Association. *Circulation* **121**, e46 (2010).

225. D. Preiss, N. Sattar, Lipids, lipid modifying agents and cardiovascular risk: a review of the evidence. *Clinical endocrinology* **70**, 815 (2009).

226. T. H. Hostetter, J. L. Olson, H. G. Rennke, M. A. Venkatachalam, B. M. Brenner, Hyperfiltration in remnant nephrons: a potentially adverse response to renal ablation. *The American journal of physiology* **241**, F85 (Jul, 1981).

227. A. S. Levey *et al.*, A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med* **130**, 461 (Mar 16, 1999).

228. US Renal Data System, N. I. o. D. a. D. a. K. D. National Institutes of Health, Ed. (Bethesda, MD, 2012).

229. J. J. Snyder, A. J. Collins, Association of preventive health care with atherosclerotic heart disease and mortality in CKD. *Journal of the American Society of Nephrology : JASN* **20**, 1614 (Jul, 2009).

230. M. B. Pepys, C-reactive protein fifty years on. *Lancet* **1**, 653 (Mar 21, 1981).

231. S. Black, I. Kushner, D. Samols, C-reactive Protein. *J Biol Chem* **279**, 48487 (Nov 19, 2004).

232. S. P. Ballou, G. Lozanski, Induction of inflammatory cytokine release from cultured human monocytes by C-reactive protein. *Cytokine* **4**, 361 (Sep, 1992).

233. C. Gabay, I. Kushner, Acute-phase proteins and other systemic responses to inflammation. *N Engl J Med* **340**, 448 (Feb 11, 1999).

234. E. M. Macy, T. E. Hayes, R. P. Tracy, Variability in the measurement of C-reactive protein in healthy subjects: implications for reference intervals and epidemiological applications. *Clin Chem* **43**, 52 (Jan, 1997).

235. B. Young, M. Gleeson, A. W. Cripps, C-reactive protein: a critical review. *Pathology* **23**, 118 (Apr, 1991).

236. M. Miller, M. Zhan, S. Havas, High attributable risk of elevated C-reactive protein level to conventional coronary heart disease risk factors: the Third National Health and Nutrition Examination Survey. *Arch Intern Med* **165**, 2063 (Oct 10, 2005).

237. I. Lemieux *et al.*, Elevated C-reactive protein: another component of the atherothrombotic profile of abdominal obesity. *Arteriosclerosis, thrombosis, and vascular biology* **21**, 961 (Jun, 2001).

238. G. A. Tennent *et al.*, Human plasma fibrinogen is synthesized in the liver. *Blood* **109**, 1971 (Mar 1, 2007).

239. D. Collen, G. N. Tytgat, H. Claeys, R. Piessens, Metabolism and distribution of fibrinogen. I. Fibrinogen turnover in physiological conditions in humans. *British journal of haematology* **22**, 681 (Jun, 1972).

240. F. Green, S. Humphries, Control of plasma fibrinogen levels. *Bailliere's clinical haematology* **2**, 945 (Oct, 1989).

241. A. Woods, D. J. Brull, S. E. Humphries, H. E. Montgomery, Genetics of inflammation and risk of coronary artery disease: the central role of interleukin-6. *European heart journal* **21**, 1574 (Oct, 2000).

242. R. F. Doolittle, Fibrinogen and fibrin. *Scientific American* **245**, 126 (Dec, 1981).

243. A. R. Folsom, K. K. Wu, W. D. Rosamond, A. R. Sharrett, L. E. Chambless, Prospective study of hemostatic factors and incidence of coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) Study. *Circulation* **96**, 1102 (Aug 19, 1997).

244. J. J. Stec *et al.*, Association of fibrinogen with cardiovascular risk factors and cardiovascular disease in the Framingham Offspring Population. *Circulation* **102**, 1634 (Oct 3, 2000).

245. J. Danesh *et al.*, Plasma fibrinogen level and the risk of major cardiovascular diseases and nonvascular mortality: an individual participant meta-analysis. *JAMA* **294**, 1799 (Oct 12, 2005).

246. J. Selhub, P. F. Jacques, P. W. Wilson, D. Rush, I. H. Rosenberg, Vitamin status and intake as primary determinants of homocysteinemia in an elderly population. *JAMA* **270**, 2693 (Dec 8, 1993).

247. V. Veeranna *et al.*, Homocysteine and reclassification of cardiovascular disease risk. *Journal of the American College of Cardiology* **58**, 1025 (Aug 30, 2011).

248. R. Rothlein, M. L. Dustin, S. D. Marlin, T. A. Springer, A human intercellular adhesion molecule (ICAM-1) distinct from LFA-1. *Journal of immunology (Baltimore, Md. : 1950)* **137**, 1270 (Aug 15, 1986).

249. C. W. M. Smith, S.D.; Rothlein, R.; Toman, C. ; Anderson, D.C., Cooperative interactions of LFA-1 and MAC-1 with intercellular adhesion molecule-1 in facilitating adherence and transendothelial migration of human neutrophils in vitro. *J. Clin. Invest.* **83**, 2008 (1989).

250. M. Tessier-Lavigne, C. S. Goodman, The molecular biology of axon guidance. *Science* **274**, 1123 (Nov 15, 1996).

251. S. M. Albelda, C. W. Smith, P. A. Ward, Adhesion molecules and inflammatory injury. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **8**, 504 (May, 1994).

252. A. Grakoui *et al.*, The immunological synapse: a molecular machine controlling T cell activation. *Science* **285**, 221 (Jul 9, 1999).

253. T. A. Springer, Adhesion receptors of the immune system. *Nature* **346**, 425 (Aug 2, 1990).

254. J. S. Pober, R. S. Cotran, The role of endothelial cells in inflammation. *Transplantation* **50**, 537 (Oct, 1990).

255. J. S. Pober *et al.*, Two distinct monokines, interleukin 1 and tumor necrosis factor, each independently induce biosynthesis and transient expression of the same antigen on the surface of cultured human vascular endothelial cells. *Journal of immunology (Baltimore, Md. : 1950)* **136**, 1680 (Mar 1, 1986).

256. S. J. Hwang *et al.*, Circulating adhesion molecules VCAM-1, ICAM-1, and E-selectin in carotid atherosclerosis and incident coronary heart disease cases: the Atherosclerosis Risk In Communities (ARIC) study. *Circulation* **96**, 4219 (Dec 16, 1997).

257. Y. C. Doo *et al.*, Associations between C-reactive protein and circulating cell adhesion molecules in patients with unstable angina undergoing coronary intervention and their clinical implication. *Clinical cardiology* **28**, 47 (Jan, 2005).

258. G. S. Hillis *et al.*, Elevated soluble P-selectin levels are associated with an increased risk of early adverse events in patients with presumed myocardial ischemia. *Am Heart J* **143**, 235 (Feb, 2002).

259. G. S. Hillis *et al.*, Soluble intercellular adhesion molecule-1 as a predictor of early adverse events in patients with chest pain compatible with myocardial ischemia. *Annals of emergency medicine* **38**, 223 (Sep, 2001).

260. C. A. Dinarello, Interleukin 1 and interleukin 18 as mediators of inflammation and the aging process. *Am J Clin Nutr* **83**, 447S (Feb, 2006).

261. J. A. Gracie, S. E. Robertson, I. B. McInnes, Interleukin-18. *Journal of leukocyte biology* **73**, 213 (Feb, 2003).

262. J. M. Bruun, B. Stallknecht, J. W. Helge, B. Richelsen, Interleukin-18 in plasma and adipose tissue: effects of obesity, insulin resistance, and weight loss. *European journal of endocrinology / European Federation of Endocrine Societies* **157**, 465 (Oct, 2007).

263. M. Bosch *et al.*, Circulating IL-18 concentration is associated with insulin sensitivity and glucose tolerance through increased fat-free mass. *Diabetologia* **48**, 1841 (Sep, 2005).

264. M. Straczkowski *et al.*, Increased serum interleukin-18 concentration is associated with hypoadiponectinemia in obesity, independently of insulin resistance. *International journal of obesity (2005)* **31**, 221 (Feb, 2007).

265. K. Esposito *et al.*, Cytokine milieu tends toward inflammation in type 2 diabetes. *Diabetes Care* **26**, 1647 (May, 2003).

266. C. P. Fischer, L. B. Perstrup, A. Berntsen, P. Eskildsen, B. K. Pedersen, Elevated plasma interleukin-18 is a marker of insulin-resistance in type 2 diabetic and non-diabetic humans. *Clinical immunology (Orlando, Fla.)* **117**, 152 (Nov, 2005).

267. J. Hung, B. M. McQuillan, C. M. Chapman, P. L. Thompson, J. P. Beilby, Elevated interleukin-18 levels are associated with the metabolic syndrome independent of obesity and insulin resistance. *Arteriosclerosis, thrombosis, and vascular biology* **25**, 1268 (Jun, 2005).

268. A. Zirlik *et al.*, Interleukin-18, the metabolic syndrome, and subclinical atherosclerosis: results from the Dallas Heart Study. *Arteriosclerosis, thrombosis, and vascular biology* **27**, 2043 (Sep, 2007).

269. S. W. Rabkin, The role of interleukin 18 in the pathogenesis of hypertension-induced vascular disease. *Nature clinical practice. Cardiovascular medicine* **6**, 192 (Mar, 2009).

270. Z. Mallat *et al.*, Expression of interleukin-18 in human atherosclerotic plaques and relation to plaque instability. *Circulation* **104**, 1598 (Oct 2, 2001).

271. A. K. Robertson, G. K. Hansson, T cells in atherogenesis: for better or for worse? *Arteriosclerosis, thrombosis, and vascular biology* **26**, 2421 (Nov, 2006).

272. M. L. Leon, S. H. Zuckerman, Gamma interferon: a central mediator in atherosclerosis. *Inflammation research : official journal of the European Histamine Research Society ... [et al.]* **54**, 395 (Oct, 2005).

273. G. P. Van Guilder, G. L. Hoetzer, J. J. Greiner, B. L. Stauffer, C. A. Desouza, Influence of metabolic syndrome on biomarkers of oxidative stress and inflammation in obese adults. *Obesity (Silver Spring, Md.)* **14**, 2127 (Dec, 2006).

274. J. Weissenbach *et al.*, Two interferon mRNAs in human fibroblasts: in vitro translation and Escherichia coli cloning studies. *Proc Natl Acad Sci U S A* **77**, 7152 (Dec, 1980).

275. C. Corbel, F. Melchers, The synergism of accessory cells and of soluble alpha-factors derived from them in the activation of B cells to proliferation. *Immunological reviews* **78**, 51 (Apr, 1984).

276. H. Baumann, G. P. Jahreis, D. N. Sauder, A. Koj, Human keratinocytes and monocytes release factors which regulate the synthesis of major acute phase plasma proteins in hepatic cells from man, rat, and mouse. *J Biol Chem* **259**, 7331 (Jun 10, 1984).

277. T. Hirano *et al.*, Purification to homogeneity and characterization of human B-cell differentiation factor (BCDF or BSFp-2). *Proc Natl Acad Sci U S A* **82**, 5490 (Aug, 1985).

278. J. Van Snick *et al.*, Purification and NH2-terminal amino acid sequence of a T-cell-derived lymphokine with growth factor activity for B-cell hybridomas. *Proc Natl Acad Sci U S A* **83**, 9679 (Dec, 1986).

279. M. Plaut *et al.*, Mast cell lines produce lymphokines in response to cross-linkage of Fc epsilon RI or to calcium ionophores. *Nature* **339**, 64 (May 4, 1989).

280. L. Hultner *et al.*, Mouse bone marrow-derived IL-3-dependent mast cells and autonomous sublines produce IL-6. *Immunology* **67**, 408 (Jul, 1989).

281. T. Hirano *et al.*, Complementary DNA for a novel human interleukin (BSF-2) that induces B lymphocytes to produce immunoglobulin. *Nature* **324**, 73 (Nov 6-12, 1986).

282. T. Hirano *et al.*, Excessive production of interleukin 6/B cell stimulatory factor-2 in rheumatoid arthritis. *European journal of immunology* **18**, 1797 (Nov, 1988).

283. L. A. Aarden, E. R. De Groot, O. L. Schaap, P. M. Lansdorp, Production of hybridoma growth factor by human monocytes. *European journal of immunology* **17**, 1411 (Oct, 1987).

284. S. Cayphas *et al.*, Identification of an interleukin HP1-like plasmacytoma growth factor produced by L cells in response to viral infection. *Journal of immunology (Baltimore, Md. : 1950)* **139**, 2965 (Nov 1, 1987).

285. K. Nakajima *et al.*, Induction of IL-6 (B cell stimulatory factor-2/IFN-beta 2) production by HIV. *Journal of immunology (Baltimore, Md. : 1950)* **142**, 531 (Jan 15, 1989).

286. K. Frei *et al.*, On the cellular source and function of interleukin 6 produced in the central nervous system in viral diseases. *European journal of immunology* **19**, 689 (Apr, 1989).

287. P. B. Sehgal *et al.*, Regulation of the acute phase and immune responses in viral disease. Enhanced expression of the beta 2-interferon/hepatocyte-stimulating factor/interleukin 6 gene in virus-infected human fibroblasts. *The Journal of experimental medicine* **167**, 1951 (Jun 1, 1988).

288. R. P. Nordan, M. Potter, A macrophage-derived factor required by plasmacytomas for survival and proliferation in vitro. *Science* **233**, 566 (Aug 1, 1986).

289. M. R. Shalaby, A. Waage, T. Espevik, Cytokine regulation of interleukin 6 production by human endothelial cells. *Cellular immunology* **121**, 372 (Jul, 1989).

290. J. Van Damme, S. Cayphas, G. Opdenakker, A. Billiau, J. Van Snick, Interleukin 1 and poly(rI).poly(rC) induce production of a hybridoma growth factor by human fibroblasts. *European journal of immunology* **17**, 1 (Jan, 1987).

291. J. Sanceau, F. Beranger, C. Gaudelet, J. Wietzerbin, IFN-gamma is an essential cosignal for triggering IFN-beta 2/BSF-2/IL-6 gene expression in human monocytic cell lines. *Annals of the New York Academy of Sciences* **557**, 130 (1989).

292. M. Kohase, L. T. May, I. Tamm, J. Vilcek, P. B. Sehgal, A cytokine network in human diploid fibroblasts: interactions of beta-interferons, tumor necrosis factor, platelet-derived growth factor, and interleukin-1. *Molecular and cellular biology* **7**, 273 (Jan, 1987).

293. M. Lotz *et al.*, B cell stimulating factor 2/interleukin 6 is a costimulant for human thymocytes and T lymphocytes. *The Journal of experimental medicine* **167**, 1253 (Mar 1, 1988).

294. C. Uyttenhove, P. G. Coulie, J. Van Snick, T cell growth and differentiation induced by interleukin-HP1/IL-6, the murine hybridoma/plasmacytoma growth factor. *The Journal of experimental medicine* **167**, 1417 (Apr 1, 1988).

295. R. D. Garman, K. A. Jacobs, S. C. Clark, D. H. Raulet, B-cell-stimulatory factor 2 (beta 2 interferon) functions as a second signal for interleukin 2 production by mature murine T cells. *Proc Natl Acad Sci U S A* **84**, 7629 (Nov, 1987).

296. T. Kishimoto, T. Hirano, Molecular regulation of B lymphocyte response. *Annual review of immunology* **6**, 485 (1988).

297. A. Catania *et al.*, Cytokine antagonists in aged subjects and their relation with cellular immunity. *The journals of gerontology. Series A, Biological sciences and medical sciences* **52**, B93 (Mar, 1997).

298. R. H. Straub *et al.*, Serum dehydroepiandrosterone (DHEA) and DHEA sulfate are negatively correlated with serum interleukin-6 (IL-6), and DHEA inhibits IL-6 secretion from mononuclear cells in man in vitro: possible link between endocrinosenescence and immunosenescence. *The Journal of clinical endocrinology and metabolism* **83**, 2012 (Jun, 1998).

299. H. Bruunsgaard, B. K. Pedersen, Age-related inflammatory cytokines and disease. *Immunology and allergy clinics of North America* **23**, 15 (Feb, 2003).

300. S. K. Fried, D. A. Bunkin, A. S. Greenberg, Omental and subcutaneous adipose tissues of obese subjects release interleukin-6: depot difference and regulation by glucocorticoid. *The Journal of clinical endocrinology and metabolism* **83**, 847 (Mar, 1998).

301. P. A. Kern, S. Ranganathan, C. Li, L. Wood, G. Ranganathan, Adipose tissue tumor necrosis factor and interleukin-6 expression in human obesity and insulin resistance. *American journal of physiology. Endocrinology and metabolism* **280**, E745 (May, 2001).

302. V. Mohamed-Ali *et al.*, Subcutaneous adipose tissue releases interleukin-6, but not tumor necrosis factor-alpha, in vivo. *The Journal of clinical endocrinology and metabolism* **82**, 4196 (Dec, 1997).

303. M. Garaulet, F. Perex-Llamas, T. Fuente, S. Zamora, F. J. Tebar, Anthropometric, computed tomography and fat cell data in an obese population: relationship with insulin, leptin, tumor necrosis factor-alpha, sex hormone-binding globulin and sex hormones. *European journal of endocrinology / European Federation of Endocrine Societies* **143**, 657 (Nov, 2000).

304. M. Diamant *et al.*, The association between abdominal visceral fat and carotid stiffness is mediated by circulating inflammatory markers in uncomplicated type 2 diabetes. *The Journal of clinical endocrinology and metabolism* **90**, 1495 (Mar, 2005).

305. G. S. Hotamisligil, N. S. Shargill, B. M. Spiegelman, Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. *Science* **259**, 87 (Jan 1, 1993).

306. S. L. Deshmane, S. Kremlev, S. Amini, B. E. Sawaya, Monocyte chemoattractant protein-1 (MCP-1): an overview. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research* **29**, 313 (Jun, 2009).

307. S. D. Cushing *et al.*, Minimally modified low density lipoprotein induces monocyte chemotactic protein 1 in human endothelial cells and smooth muscle cells. *Proc Natl Acad Sci U S A* **87**, 5134 (Jul, 1990).

308. T. J. Standiford, S. L. Kunkel, S. H. Phan, B. J. Rollins, R. M. Strieter, Alveolar macrophage-derived cytokines induce monocyte chemoattractant protein-1 expression from human pulmonary type II-like epithelial cells. *J Biol Chem* **266**, 9912 (May 25, 1991).

309. Z. Brown *et al.*, IL-1 receptor antagonist inhibits monocyte chemotactic peptide 1 generation by human mesangial cells. *Kidney international* **42**, 95 (Jul, 1992).

310. B. P. Barna *et al.*, Regulation of monocyte chemoattractant protein-1 expression in adult human non-neoplastic astrocytes is sensitive to tumor necrosis factor (TNF) or antibody to the 55-kDa TNF receptor. *Journal of neuroimmunology* **50**, 101 (Feb, 1994).

311. T. Yoshimura, E. A. Robinson, S. Tanaka, E. Appella, E. J. Leonard, Purification and amino acid analysis of two human monocyte chemoattractants produced by phytohemagglutinin-stimulated human blood mononuclear leukocytes. *Journal of immunology (Baltimore, Md. : 1950)* **142**, 1956 (Mar 15, 1989).

312. T. Yoshimura *et al.*, Human monocyte chemoattractant protein-1 (MCP-1). Full-length cDNA cloning, expression in mitogen-stimulated blood mononuclear leukocytes, and sequence similarity to mouse competence gene JE. *FEBS letters* **244**, 487 (Feb 27, 1989).

313. N. Tanuma, H. Sakuma, A. Sasaki, Y. Matsumoto, Chemokine expression by astrocytes plays a role in microglia/macrophage activation and subsequent neurodegeneration in secondary progressive multiple sclerosis. *Acta neuropathologica* **112**, 195 (Aug, 2006).

314. G. Flora *et al.*, Proinflammatory synergism of ethanol and HIV-1 Tat protein in brain tissue. *Experimental neurology* **191**, 2 (Jan, 2005).

315. A. B. Ragin *et al.*, Monocyte chemoattractant protein-1 correlates with subcortical brain injury in HIV infection. *Neurology* **66**, 1255 (Apr 25, 2006).

316. S. Rantapaa-Dahlqvist, K. Boman, A. Tarkowski, G. Hallmans, Up regulation of monocyte chemoattractant protein-1 expression in anti-citrulline antibody and immunoglobulin M rheumatoid factor positive subjects precedes onset of inflammatory response and development of overt rheumatoid arthritis. *Annals of the rheumatic diseases* **66**, 121 (Jan, 2007).

317. F. Cipollone *et al.*, Elevated circulating levels of monocyte chemoattractant protein-1 in patients with restenosis after coronary angioplasty. *Arteriosclerosis, thrombosis, and vascular biology* **21**, 327 (Mar, 2001).

318. S. J. Klebanoff, A peroxidase-mediated antimicrobial system in leukocytes. *J Clin Invest* **46**, (1967, 1967).

319. S. Klebanoff, *Myeloperoxidase: occurrence and biological function.* E. K. Everse J, Grisham MB, Ed., Peroxidases in Chemistry and Biology (CRC Press, Boca Raton, 1991).

320. L. A. Marquez, H. B. Dunford, Kinetics of oxidation of tyrosine and dityrosine by myeloperoxidase compounds I and II. Implications for lipoprotein peroxidation studies. *J Biol Chem* **270**, 30434 (Dec 22, 1995).

321. S. J. McKenzie, M. S. Baker, G. D. Buffinton, W. F. Doe, Evidence of oxidant-induced injury to epithelial cells during inflammatory bowel disease. *J Clin Invest* **98**, 136 (Jul 1, 1996).

322. T. Tatsumi, H. Fliss, Hypochlorous acid and chloramines increase endothelial permeability: possible involvement of cellular zinc. *The American journal of physiology* **267**, H1597 (Oct, 1994).

323. M. C. Vissers, A. C. Carr, A. L. Chapman, Comparison of human red cell lysis by hypochlorous and hypobromous acids: insights into the mechanism of lysis. *The Biochemical journal* **330 (Pt 1)**, 131 (Feb 15, 1998).

324. C. C. Winterbourn, M. C. Vissers, A. J. Kettle, Myeloperoxidase. *Current opinion in hematology* **7**, 53 (Jan, 2000).

325. R. H. Demling, The modern version of adult respiratory distress syndrome. *Annual review of medicine* **46**, 193 (1995).

326. W. G. Couser, Pathogenesis of glomerulonephritis. *Kidney international. Supplement* **42**, S19 (Jul, 1993).

327. J. W. Heinecke, Mechanisms of oxidative damage by myeloperoxidase in atherosclerosis and other inflammatory disorders. *J Lab Clin Med* **133**, 321 (Apr, 1999).

328. A. Daugherty, J. L. Dunn, D. L. Rateri, J. W. Heinecke, Myeloperoxidase, a catalyst for lipoprotein oxidation, is expressed in human atherosclerotic lesions. *J Clin Invest* **94**, 437 (Jul, 1994).

329. P. G. McTernan, C. M. Kusminski, S. Kumar, Resistin. *Current opinion in lipidology* **17**, 170 (Apr, 2006).

330. M. Pravenec *et al.*, Transgenic and recombinant resistin impair skeletal muscle glucose metabolism in the spontaneously hypertensive rat. *J Biol Chem* **278**, 45209 (Nov 14, 2003).

331. H. Satoh *et al.*, Adenovirus-mediated chronic "hyper-resistinemia" leads to in vivo insulin resistance in normal rats. *J Clin Invest* **114**, 224 (Jul, 2004).

332. N. Silswal *et al.*, Human resistin stimulates the pro-inflammatory cytokines TNF-alpha and IL-12 in macrophages by NF-kappaB-dependent pathway. *Biochem Biophys Res Commun* **334**, 1092 (Sep 9, 2005).

333. M. Bokarewa, I. Nagaev, L. Dahlberg, U. Smith, A. Tarkowski, Resistin, an adipokine with potent proinflammatory properties. *Journal of immunology (Baltimore, Md. : 1950)* **174**, 5789 (May 1, 2005).

334. M. W. Rajala, S. Obici, P. E. Scherer, L. Rossetti, Adipose-derived resistin and gut-derived resistin-like molecule-beta selectively impair insulin action on glucose production. *J Clin Invest* **111**, 225 (Jan, 2003).

335. E. D. Muse *et al.*, Role of resistin in diet-induced hepatic insulin resistance. *J Clin Invest* **114**, 232 (Jul, 2004).

336. R. R. Banerjee *et al.*, Regulation of fasted blood glucose by resistin. *Science* **303**, 1195 (Feb 20, 2004).

337. P. G. McTernan *et al.*, Resistin and type 2 diabetes: regulation of resistin expression by insulin and rosiglitazone and the effects of recombinant resistin on lipid and glucose metabolism in human differentiated adipocytes. *The Journal of clinical endocrinology and metabolism* **88**, 6098 (Dec, 2003).

338. J. N. Fain, P. S. Cheema, S. W. Bahouth, M. Lloyd Hiler, Resistin release by human adipose tissue explants in primary culture. *Biochem Biophys Res Commun* **300**, 674 (Jan 17, 2003).

339. D. B. Savage *et al.*, Resistin / Fizz3 expression in relation to obesity and peroxisome proliferator-activated receptor-gamma action in humans. *Diabetes* **50**, 2199 (Oct, 2001).

340. J. Janke, S. Engeli, K. Gorzelniak, F. C. Luft, A. M. Sharma, Resistin gene expression in human adipocytes is not related to insulin resistance. *Obes Res* **10**, 1 (Jan, 2002).

341. S. Kaser *et al.*, Resistin messenger-RNA expression is increased by proinflammatory cytokines in vitro. *Biochem Biophys Res Commun* **309**, 286 (Sep 19, 2003).

342. L. Patel *et al.*, Resistin is expressed in human macrophages and directly regulated by PPAR gamma activators. *Biochem Biophys Res Commun* **300**, 472 (Jan 10, 2003).

343. C. Pagano *et al.*, Increased serum resistin in adults with prader-willi syndrome is related to obesity and not to insulin resistance. *The Journal of clinical endocrinology and metabolism* **90**, 4335 (Jul, 2005).

344. S. M. Rangwala *et al.*, Abnormal glucose homeostasis due to chronic hyperresistinemia. *Diabetes* **53**, 1937 (Aug, 2004).

345. C. M. Steppan *et al.*, The hormone resistin links obesity to diabetes. *Nature* **409**, 307 (Jan 18, 2001).

346. M. S. Burnett *et al.*, The potential role of resistin in atherogenesis. *Atherosclerosis* **182**, 241 (Oct, 2005).

347. W. Y. Hsu *et al.*, Resistin induces monocyte-endothelial cell adhesion by increasing ICAM-1 and VCAM-1 expression in endothelial cells via p38MAPK-dependent pathway. *Journal of cellular physiology* **226**, 2181 (Aug, 2011).

348. S. Verma *et al.*, Resistin promotes endothelial cell activation: further evidence of adipokine-endothelial interaction. *Circulation* **108**, 736 (Aug 12, 2003).

349. B. Kluve-Beckerman, M. L. Drumm, M. D. Benson, Nonexpression of the human serum amyloid A three (SAA3) gene. *DNA and cell biology* **10**, 651 (Nov, 1991).

350. I. Kushner, The acute phase response: an overview. *Methods in enzymology* **163**, 373 (1988).

351. M. R. Edbrooke *et al.*, Constitutive and NF-kappa B-like proteins in the regulation of the serum amyloid A gene by interleukin 1. *Cytokine* **3**, 380 (Sep, 1991).

352. H. Baumann, J. Gauldie, The acute phase response. *Immunology today* **15**, 74 (Feb, 1994).

353. M. A. Aldo-Benson, M. D. Benson, SAA suppression of immune response in vitro: evidence for an effect on T cell-macrophage interaction. *Journal of immunology (Baltimore, Md. : 1950)* **128**, 2390 (Jun, 1982).

354. S. Zimlichman, A. Danon, I. Nathan, G. Mozes, R. Shainkin-Kestenbaum, Serum amyloid A, an acute phase protein, inhibits platelet activation. *J Lab Clin Med* **116**, 180 (Aug, 1990).

355. R. Badolato *et al.*, Serum amyloid A is a chemoattractant: induction of migration, adhesion, and tissue infiltration of monocytes and polymorphonuclear leukocytes. *The Journal of experimental medicine* **180**, 203 (Jul 1, 1994).

356. T. Yamada, Serum amyloid A (SAA): a concise review of biology, assay methods and clinical usefulness. *Clinical chemistry and laboratory medicine : CCLM / FESCC* **37**, 381 (Apr, 1999).

357. A. S. Whitehead *et al.*, Identification of novel members of the serum amyloid A protein superfamily as constitutive apolipoproteins of high density lipoprotein. *J Biol Chem* **267**, 3862 (Feb 25, 1992).

358. G. A. Coetzee *et al.*, Serum amyloid A-containing human high density lipoprotein 3. Density, size, and apolipoprotein composition. *J Biol Chem* **261**, 9644 (Jul 25, 1986).

359. R. Z. Yang *et al.*, Acute-phase serum amyloid A: an inflammatory adipokine and potential link between obesity and its metabolic complications. *PLoS medicine* **3**, e287 (Jun, 2006).

360. J. Vilcek, T. H. Lee, Tumor necrosis factor. New insights into the molecular mechanisms of its multiple actions. *J Biol Chem* **266**, 7313 (Apr 25, 1991).

361. J. M. Krueger *et al.*, Sleep. A physiologic role for IL-1 beta and TNF-alpha. *Annals of the New York Academy of Sciences* **856**, 148 (Sep 29, 1998).

362. W. Fiers, Tumor necrosis factor. Characterization at the molecular, cellular and in vivo level. *FEBS letters* **285**, 199 (Jul 22, 1991).

363. R. S. Al-Lamki *et al.*, Expression of tumor necrosis factor receptors in normal kidney and rejecting renal transplants. *Laboratory investigation; a journal of technical methods and pathology* **81**, 1503 (Nov, 2001).

364. H. T. Idriss, J. H. Naismith, TNF alpha and the TNF receptor superfamily: structure-function relationship(s). *Microscopy research and technique* **50**, 184 (Aug 1, 2000).

365. K. F. Chan, M. R. Siegel, J. M. Lenardo, Signaling by the TNF receptor superfamily and T cell homeostasis. *Immunity* **13**, 419 (Oct, 2000).

366. T. Zhou, J. D. Mountz, R. P. Kimberly, Immunobiology of tumor necrosis factor receptor superfamily. *Immunologic research* **26**, 323 (2002).

367. J. R. Bradley, TNF-mediated inflammatory disease. *The Journal of pathology* **214**, 149 (Jan, 2008).

368. H. Hsu, J. Huang, H. B. Shu, V. Baichwal, D. V. Goeddel, TNF-dependent recruitment of the protein kinase RIP to the TNF receptor-1 signaling complex. *Immunity* **4**, 387 (Apr, 1996).

369. N. L. Malinin, M. P. Boldin, A. V. Kovalenko, D. Wallach, MAP3K-related kinase involved in NF-kappaB induction by TNF, CD95 and IL-1. *Nature* **385**, 540 (Feb 6, 1997).

370. K. J. Van Zee *et al.*, Tumor necrosis factor soluble receptors circulate during experimental and clinical inflammation and can protect against excessive tumor necrosis factor alpha in vitro and in vivo. *Proc Natl Acad Sci U S A* **89**, 4845 (Jun 1, 1992).

371. D. R. Meldrum, Tumor necrosis factor in the heart. *The American journal of physiology* **274**, R577 (Mar, 1998).

372. F. J. Neumann *et al.*, Cardiac release of cytokines and inflammatory responses in acute myocardial infarction. *Circulation* **92**, 748 (Aug 15, 1995).

373. R. Latini *et al.*, Cytokines in acute myocardial infarction: selective increase in circulating tumor necrosis factor, its soluble receptor, and interleukin-1 receptor antagonist. *Journal of cardiovascular pharmacology* **23**, 1 (Jan, 1994).

374. A. Herskowitz, S. Choi, A. A. Ansari, S. Wesselingh, Cytokine mRNA expression in postischemic/reperfused myocardium. *The American journal of pathology* **146**, 419 (Feb, 1995).

375. J. W. Horton, Cellular basis for burn-mediated cardiac dysfunction in adult rabbits. *The American journal of physiology* **271**, H2615 (Dec, 1996).

376. D. Cameron, Initiation of white cell activation during cardiopulmonary bypass: cytokines and receptors. *Journal of cardiovascular pharmacology* **27 Suppl 1**, S1 (1996).

377. B. G. Hattler, A. Zeevi, C. V. Oddis, M. S. Finkel, Cytokine induction during cardiac surgery: analysis of TNF-alpha expression pre- and postcardiopulmonary bypass. *Journal of cardiac surgery* **10**, 418 (Jul, 1995).

378. H. S. Park, J. Y. Park, R. Yu, Relationship of obesity and visceral adiposity with serum concentrations of CRP, TNF-alpha and IL-6. *Diabetes research and clinical practice* **69**, 29 (Jul, 2005).

379. M. Olszanecka-Glinianowicz, B. Zahorska-Markiewicz, J. Janowska, A. Zurakowski, Serum concentrations of nitric oxide, tumor necrosis factor (TNF)-alpha and TNF soluble receptors in women with overweight and obesity. *Metabolism: clinical and experimental* **53**, 1268 (Oct, 2004).

380. M. Good *et al.*, TNF and TNF receptor expression and insulin sensitivity in human omental and subcutaneous adipose tissue--influence of BMI and adipose distribution. *Diabetes & vascular disease research : official journal of the International Society of Diabetes and Vascular Disease* **3**, 26 (May, 2006).

381. G. Winkler *et al.*, Expression of tumor necrosis factor (TNF)-alpha protein in the subcutaneous and visceral adipose tissue in correlation with adipocyte cell volume, serum TNF-alpha, soluble serum TNF-receptor-2 concentrations and C-peptide level. *European journal of endocrinology / European Federation of Endocrine Societies* **149**, 129 (Aug, 2003).

382. A. T. Cheung *et al.*, An in vivo model for elucidation of the mechanism of tumor necrosis factor-alpha (TNF-alpha)-induced insulin resistance: evidence for differential regulation of insulin signaling by TNF-alpha. *Endocrinology* **139**, 4928 (Dec, 1998).

383. I. Wybranska, M. Malczewska-Malec, S. Niedbal, J. W. Naskalski, A. Dembinska-Kiec, The TNF-alpha gene NcoI polymorphism at position -308 of the promoter influences insulin resistance, and increases serum triglycerides after postprandial lipaemia in familiar obesity. *Clinical chemistry and laboratory medicine : CCLM / FESCC* **41**, 501 (Apr, 2003).

384. R. Krogh-Madsen, P. Plomgaard, K. Moller, B. Mittendorfer, B. K. Pedersen, Influence of TNF-alpha and IL-6 infusions on insulin sensitivity and expression of IL-18 in humans. *American journal of physiology. Endocrinology and metabolism* **291**, E108 (Jul, 2006).

385. A. A. Ellington *et al.*, Association of plasma resistin with glomerular filtration rate and albuminuria in hypertensive adults. *Hypertension* **50**, 708 (2007).

386. I. J. Kullo *et al.*, Ethnic differences in peripheral arterial disease in the NHLBI Genetic Epidemiology Network of Arteriopathy (GENOA) study. *Vascular medicine (London, England)* **8**, 237 (2003).

387. A. D. Rule *et al.*, A comparison of serum creatinine-based methods for identifying chronic kidney disease in hypertensive individuals and their siblings. *Am J Hypertens* **19**, 608 (Jun, 2006).

388. P. Komenda, M. Beaulieu, D. Seccombe, A. Levin, Regional implementation of creatinine measurement standardization. *Journal of the American Society of Nephrology : JASN* **19**, 164 (Jan, 2008).

389. M. J. Magera, J. M. Lacey, B. Casetta, P. Rinaldo, Method for the determination of total homocysteine in plasma and urine by stable isotope dilution and electrospray tandem mass spectrometry. *Clin Chem* **45**, 1517 (Sep, 1999).

390. D. J. Weisenberger, D. Van Den Berg, B. P. Berman, P. W. Laird, Comprehensive DNA Methylation Analysis on the Illumina® Infinium® Assay Platform. *Application Note: Illumina Epigenetic Analysis*, (2008).

391. Illumina Inc., in *Technical Note: Epigenetics*. (San Diego, CA, 2010).

392. Illumina Inc., in *Data Sheet: Epigenetics*. (San Diego, CA, 2012).

393. Illumina Inc., Illumina_HumanMethylation27-v1.2_Annotation2.csv, Ed. (Illumina Inc, 2012).

394. P. Flicek *et al.*, Ensembl 2012. *Nucleic acids research* **40**, (2012).

395. S. Saxonov, P. Berg, D. L. Brutlag, A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**, 1412 (Jan 31, 2006).

396. Illumina Inc. (San Diego, CA, 2010), vol. November.

397. A. Wutz, Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat Rev Genet* **12**, 542 (Aug, 2011).

398. A. J. Scott, D. Holt, The Effect of Two-Stage Sampling on Ordinary Least Squares Methods. *Journal of the American Statistical Association* **77**, 848 (Dec., 1982, 1982).

399. W. D. Dupont, W. D. Plummer. (2009).