Epidemiologic approaches

to understanding mechanisms of cardiovascular diseases:

Genes, environment, and DNA methylation

by

Min A Jhun

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in the University of Michigan
2015

Doctoral Committee:

Assistant Professor Sung Kyun Park, Co-Chair
Professor Patricia A. Peyser, Co-Chair
Emeritus Professor John D. Kalbfleisch
Professor Sharon L.R. Kardia
Assistant Research Scientist Jennifer A. Smith

Acknowledgements

Heartfelt appreciation to my advisors, Dr. Patricia Peyser and Dr. Sung Kyun Park, friends, and family

Table of Contents

List of figures

List of tables

List of abbreviations

ABI: ankle brachial index

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1

AHRR: aryl-hydrocarbon receptor repressor

AKT3:v-akt murine thymoma viral oncogene homolog 3

ARIC: atherosclerosis risk in communities

ASE: American Society of Echocardiography

AZA: 5'deoxy-Azacytidine

BMI: body mass index

CAC: coronary artery calcification

CBFB: core-binding factor beta subunit

CEU: Utah residents with Northern and Western European ancestry from the CEPH collection

CI: confidence interval

CVD: cardiovascular disease

CHD: coronary heart disease

DBP: diastolic blood pressure

DNA: deoxyribonucleic acid

EWAS: epigenome-wide association studies

F2RL3: coagulation factor II (thrombin) receptor-like 3

F7: factor VII

GENOA: genetic epidemiology network of arteriopathy

GPR15: G protein-coupled receptor 15

GPR25: G protein-coupled receptor 25

GWAS: genome-wide association studies

HDL: high density lipoprotein

HNRPUL1: heterogeneous nuclear ribonucleoprotein U-like 1

HWE: hardy-Weinberg equilibrium

IQR: inter quartile range

KXRF: K-x-ray fluorescence instrument

LD: linkage disequilibrium

LDL: low density lipoprotein

LFEF: left ventricular ejection fraction

LIM2: lens intrinsic membrane protein 2 (19kDa)

LRRC32: leucine rich repeat containing 32

LVH: left ventricular hypertrophy

LVMI: left ventricular mass index

Me-CpG: methylation of a CpG site

MeQTL: methylation quantitative trait loci

MI: myocardial infarction

MR: Mendelian randomization

MTHRF: methylenetetrahydrofolate reductase

NAS: normative aging study

NCBP1: nuclear cap binding protein subunit 1 (80kDa)

NHANES: National Health and Nutrition Examination Survey

OR2B6: olfactory receptor, family 2, subfamily B, member 6

PAR-4: protease-activated receptor 4

POU3F1: POU class 3 homeobox 1

PP: pulse pressure

SBP: systolic blood pressure

SEM: structural equation models

SNP: single nucleotide polymorphisms

SPATA12: spermatogenesis associated 12

VA: Veterans administration

VDR: vitamin D receptor

VDRE: vitamin D receptor responsive element

YRI: Yoruba in Ibadan, Nigeria

ZG16B: zymogen granule protein 16B

ZNF384: zinc finger protein 384

Abstract

Cardiovascular disease (CVD) is the leading cause of death in the United States. Arterial stiffness, left ventricular dysfunction, and left ventricular hypertrophy are risk factors for CVD and are measured by subclinical markers such as pulse pressure, left ventricular ejection fraction, and left ventricular mass index, respectively. Complex diseases, such as CVD, are caused by combination and interaction of various factors, such as genetic, epigenetic, and environmental factors. Although disease predisposition at the DNA level is rarely modifiable, the same genetic variant can be associated with various phenotypes in different environments. For example, epigenetic modification of DNA occurs naturally as a response to changes in environmental exposures. Thus, epigenetic modification may be considered as an intermediate phenotype which is a result of new kind of gene by environment interaction where the gene transcription is dependent on environmental factors.

The goal of this dissertation is to provide insight into the mechanisms of CVD by combining information of genetic, epigenetic, and environmental factors. Specifically this dissertation investigates effect modification by candidate genetic polymorphisms and explores mediation analysis using epigenetic modification. First, this dissertation focuses on the effect modification by vitamin D receptor (*VDR*) genetic polymorphisms in the association between bone lead levels and pulse pressure in older unrelated non-Hispanic White men (Normative Aging Study) (Aim 1). Next, the focus is on the mediative effect of DNA methylation in the association between cigarette smoking and inflammation in older African American siblings (Genetic Epidemiology Network of Arteriopathy (GENOA)) using two-step epigenetic

Mendelian randomization (MR) approaches (Aim 2). The last part of this dissertation is dedicated to investigating the role of DNA methylation in cigarette smoking, inflammation, and subclinical markers of CVD in older African American siblings (GENOA) using network MR approaches and structural equation models (Aim 3).

In Aim 1, we found that subjects with the minor frequency alleles of *VDR Bsm1 or Taq1* may be more susceptible to cumulative lead exposure-related elevated pulse pressure. Adjusting for potential confounders, pulse pressure was 2.5 mmHg (95% CI: 0.4-4.7) and 1.9 mmHg (95% CI: 0.1-3.8) greater per interquartile range increase in tibia lead (15µg/g) and patella lead (20 µg/g), respectively, in those with at least one minor frequency allele in *Bsm1* compared with those with major frequency allele homozygotes. In Aim 2, we identified that the odds ratio for current smoking was 1.96 for a one-unit increase in rs4887071 which results in 0.44 decrease in M-values of cg03636183 on *F2RL3* gene, and a 5% increase in interleukin-18 (IL-18) levels (pg/ml), and results in 0.09 decrease in M-values of cg13500388 on *CBFB* gene and a 3% increase in C-reactive protein (CRP) levels (mg/L). In Aim 3, we found that direct effects of smoking were significant for log transformed CRP, log transformed interleukin-6 (IL-6), and fibrinogen. The direct effect of smoking was not significant for log(IL-18) levels. However, the indirect effect from smoking to log(IL-18) levels mediated by cg03636183 on *F2RL3* gene, was significant. The indirect path had a significant effect on left ventricular mass index (LVMI).

In summary, results suggest that *VDR*-related calcium metabolism pathways and DNA methylation play a role in adverse effects of cumulative lead exposure and cigarette smoking, respectively, on CVD. These findings together expand our understanding of mechanisms behind CVD development and can inform future studies.

Chapter 1

Introduction

This dissertation focuses on gene by environment interaction in risk factors of cardiovascular diseases (CVD). The introduction is composed of eight subsections to provide background on CVD in terms of its public health importance; risk factors of CVD focusing on smoking and inflammation; relevant subclinical markers including pulse pressure, left ventricular ejection fraction, and left ventricular mass index; genetic variation at the DNA level and epigenetic modification; relation to vitamin D and the vitamin D receptor (*VDR*) gene; relation to lead exposure; gene by environment interaction; and the Mendelian randomization (MR) approach. In the last section, the need to consider gene by environment interactions, through direct measures of environmental exposures or through indirect measures such as DNA methylation in response to environmental exposure to better understand variation in subclinical markers of CVD is highlighted. The overarching goal and hypotheses for each aim are presented last.

Importance of Cardiovascular Disease

CVD is the leading cause of death in the United States.[1] In 2010, based on the National Vital Statistics Reports[1], there were 780,213 deaths (age-adjusted death rate: 234 per 100,000) from CVD which accounted for 31.6% of total deaths. Age-adjusted death rates from CVD were 1.4 times higher in men than in women (282 vs. 196 per 100,000) and were 1.3 times higher in African Americans than in non-Hispanic Whites (304 vs. 230 per 100,000).

In 2012, based on data from the National Health Interview Survey, the number of non-institutionalized adults aged 18 and over who had been diagnosed with heart disease - coronary heart disease (CHD), angina, myocardial infarction (MI), or any other heart condition or disease - reached 26.6 million which equates to a prevalence of 11.3% (age adjusted: 10.8%).[2] The prevalence went up to 24.4% among 65-74 year-olds, and 36.9% among 75 year-olds and over. The overall prevalence was higher in men (12.1%) than in women (9.7%). Six percent of adults aged 18 and over had been diagnosed with CHD, angina, or had a MI. The percentage was higher in men (7.6%) than in women (4.8%). Among adults aged 18 and over, 25.5% had been told they have hypertension or high blood pressure on two or more visits (age adjusted: 23.9%). The percentage went up to 52.3% among 65-74 year-olds, and 59.2% among 75 year-olds and over. Men (24.7%) were more likely to have hypertension than women (23.2%). African Americans (32.9%) were more likely to have hypertension than non-Hispanic Whites (22.9%). Therefore, it is important to study risk factors and biological/genetic mechanisms of the CVD among older individuals in various race/ethnic populations.

Inflammation and Cardiovascular Disease

Chronic systemic inflammation plays a key role in atherosclerosis development.[3, 4] Smoking[5] and alcohol consumption[6] are two major environmental factors that can induce inflammatory responses. The levels of inflammatory biomarkers in the blood stream help to quantify the degree of inflammation and to predict future CVD events.[7] Several inflammatory biomarkers that have been found to be associated with CVD events include C-reactive protein (CRP)[8, 9], interleukin-6 (IL-6)[10], interleukin-18 (IL-18)[11, 12], and fibrinogen.[13, 14]

C-reactive Protein

CRP is produced as a non-specific acute response to inflammation, infection, and tissue damage.[15] CRP is produced by hepatocytes under control of IL-6 and binds to lipids and lipoproteins;[16] phospholipids on damaged cell membranes[17] and apoptotic cells;[18] and microorganisms.[19] CRP activates generation and release of proinflammatory cytokines. CRP cannot recognize antigenic epitopes, however, it recognizes altered self-molecules and certain pathogens and provides early defense.[20] In 468 healthy adult volunteers, the median of the serum CRP level was 0.8 mg/L with a range between 0.07 to 29 mg/L.[21] In the study, over 90% of the study participants have serum CRP levels < 3mg/L. The distribution of serum CRP level is significantly different by gender and race/ethnicity: CRP was higher in African Americans than in European Americans, and it was higher in women than in men in the Dallas Heart Study.[22] Serum CRP level tends to increase with aging[23] and it is useful in risk prediction of CVD events.[8, 9, 24-26]

Interleukin-6

IL-6 is produced by lymphocytes, monocytes, macrophages, adipocytes, endothelial cells, and vascular smooth muscle cells.[27-29] This cytokine stimulates differentiation of B cells and T cells, and activation of NK cells and macrophages.[30-32] IL-6 binds to and activates the IL-6 receptor, and the activated complex binds to the signal-transducing component of the membrane-bound receptor (gp130) and stimulates inflammatory responses.[33] IL-6 has both pro-inflammatory and anti-inflammatory properties.[34] IL-6 has anti-inflammatory effects on cells that express both IL-6 receptor and gp130, while it has pro-inflammatory effects on cells that only express gp130. Chronic high IL-6 levels are largely determined by adipose tissue.[35] Production of IL-6 is stimulated by CRP,[36] and CRP expression in liver is stimulated by IL-6.[37] Serum IL-6 levels are

associated with left ventricular dysfunction[38-40] and cardiac hypertrophy.[41, 42] It is a strong predictor of CVD events including mortality.[43-45]

Interkeukin-18

IL-18 is found at the sites of chronic inflammation and have been found to be expressed in various cell types including macrophages, dendritic cells,[46] osteoblasts,[47] intestinal mucosal cells,[48] and synovial fibroblasts.[49] Expression of the *IL18* gene produces the inactive precursor of IL-18 which is activated by the endoprotease IL-1β (ICE)[50] and protease 3.[51] Active IL-18 can combine with IL-18 receptor, which is expressed in diverse cell types including macrophages[52], neutrophils[53], natural killer cells,[54] and vascular endothelial cells.[52] IL-18 is involved in regulation of innate and adaptive immunity.[55, 56] It promotes production and release of cytokines,[57] maturation of T cells and natural killer cells,[58] and neutrophil activation.[53] IL-18 binding protein binds to IL-18 and inhibit the function of IL-18.[59] Therefore, a careful consideration of the levels of IL-18, IL-18 receptor, and IL-18 binding protein is needed to estimate the activity of IL-18. IL-18 levels have been found to be associated with not only inflammatory diseases including rheumatoid arthritis,[60] insulin-dependent diabetes mellitus (a.k.a. type 1 diabetes),[61] multiple sclerosis,[62] and inflammatory bowel disease,[63, 64] but also other chronic diseases such as cancer,[65] and CVD.[11, 12]

Fibrinogen

Fibrinogen is a coagulation factor that plays an important role in blood clotting.[66] Fibrinogen is produced by hepatocytes in liver and is circulated through the blood stream.[67] Damaged endothelial cells of blood vessels go through a coagulation cascade involving cleavage of fibrinogen by thrombin to form fibrin.[68] Fibrinogen not only activates platelets and fibrin

formation, and increases serum viscosity, but also is involved in cytokine production and macrophage adhesion.[69] Fibrinogen plays an important role in plaque formation and atherosclerosis[70] and its level has been suggested as a predictor of CVD events.[13, 14, 71]

Selected Subclinical Markers of Cardiovascular Disease

Pulse Pressure and Arterial Stiffness

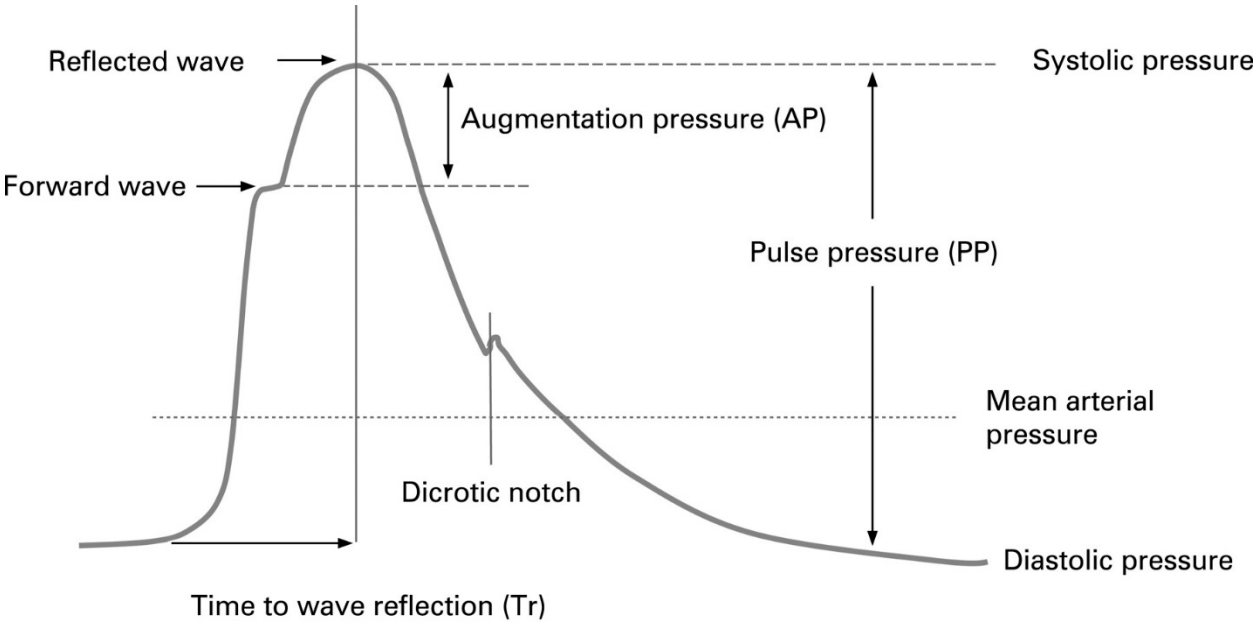

Figure 1.1 Blood pressure measurements and pulse pressure. Adapted from *Thorax* 2008; 63:306-311.

Pulse pressure, the difference between peak systolic blood pressure (SBP) and end diastolic blood pressure (DBP), is a surrogate measure of arterial stiffness (**Figure 1.1**). Arterial stiffness is a loss of elasticity and compliance of arteries. Pulse pressure is an independent risk factor for CVD such as CHD, stroke, and heart failure.[72-74] Pulse pressure has been demonstrated as an independent predictor of long-term cardiovascular mortality. Among the elderly, a 10 mmHg increase in pulse pressure has been associated with 10% to 25% increased risks for CVD-related

mortality including CHD, congestive heart failure, and cerebrovascular disease.[75-77] In risk prediction of CHD for the middle-aged and elderly, pulse pressure performed better than SBP and DBP.[78] In the Framingham Heart Study, as the study population became older, pulse pressure became more important in coronary risk prediction than SBP or DBP.[79] Pulse pressure is the outcome in Aim 1.

Left Ventricular Ejection Fraction and Systolic Function

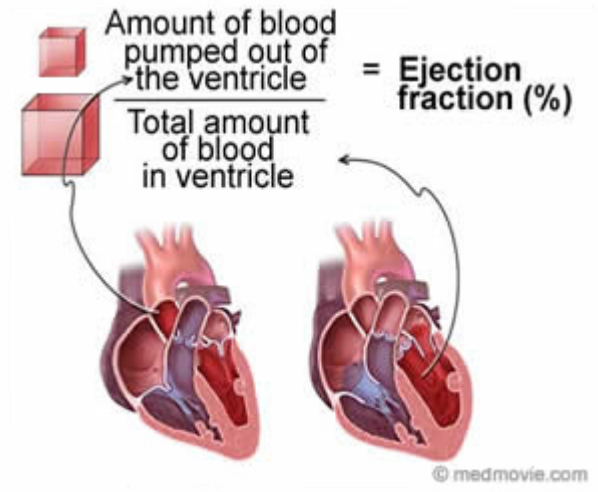

Figure 1.2 Left ventricular ejection fraction

Left ventricular ejection fraction (LVEF), a measure of left ventricular systolic function, is calculated as a percentage of the amount of blood pumped out of the left ventricle compared to the total amount of blood in ventricle (**Figure 1.2**). LVEF ranges from 0% to 100%. LVEF between 55% and 70% is considered normal. A LVEF of 50% or lower is considered reduced and is associated with increased risk for heart failure, the most rapidly growing form of heart disease in the U.S, and sudden cardiac arrest, which claims 300,000 lives a year. As many as 20 million Americans may have a reduced ejection fraction without any symptoms such a shortness of breath or fatigue. Causes of lower ejection fraction includes dilated cardiomyopathy, heart failure, problems with the heart's valves, and long-standing uncontrolled high blood pressure.

In a recent study of 850 coronary angiography patients, the Gensini score, a measure of the severity of coronary atherosclerosis, was associated with LVEF (β= -0.194, P<0.0001) independent of other risk factors including age, body mass index (BMI), blood pressure, fasting blood glucose level, blood lipids, and leucocyte count.[80] In a study of 7,599 chronic heart failure patients, lower LVEF was associated with a greater risk of MI.[81] Each 10% reduction in LVEF (below 45%) was associated with a significant increase in risk of all-cause death, cardiovascular death, sudden death, death due to heart failure, death due to MI, and other cardiovascular related deaths. LVEF is one of the outcomes in Aim 3.

Left Ventricular Mass Index and Hypertrophy

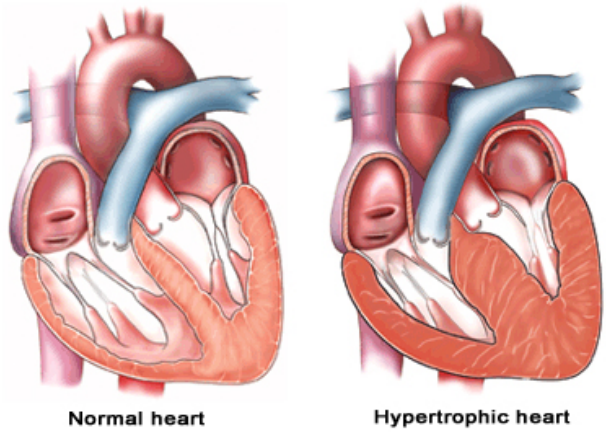

Figure 1.3 Cardiac hypertrophy. Adapted from http://www.columbiasurgery.org/hcm/diagnosis_hcm.html

Left ventricular mass index (LVMI) is a marker of left ventricular hypertrophy (LVH) (**Figure 1.3**).[82] LVMI is calculated as left ventricular mass divided by body surface area[83] or height$^{2.7}$.[84] In a study of 40-59 years old men, the prevalence of LVH ranged from 2.7-3.2% in the normative group without hypertension while it ranged from 11.8-14.5% in the hypertensive group.[85] In another study of an essential hypertensive population, the prevalence of LVH ranged from 59% to 73% depending on different criteria.[86] Systemic inflammation and endothelial

damage in the blood vessel wall were associated with LVH.[87] CRP level, a marker of inflammation, was positively associated with LVMI.[88] LVMI has been suggested as a predictive marker for CVD events in the hypertensive population.[89] LVMI is an outcome in Aim 3.

Genetics of Cardiovascular Disease

SNPs in Cardiovascular Disease

Numerous studies have identified genetic polymorphisms that are associated with CVD and its risk factors. The very early large-scale genome-wide association study (GWAS) by the Wellcome Trust Consortium identified single nucleotide polymorphisms (SNPs) in the 9p21.3 region associated with CHD in approximately 2,000 cases and 3,000 controls.[90] For the most significant SNP, rs1333049, the reported odds ratio was 1.47 for heterozygotes and 1.9 for risk allele homozygotes. The discovery of the 9p21 region was confirmed in several other GWASs.[91-94] The 9p21 region has also been associated with other traits related to CVD such as ankle brachial index (ABI)[95] and coronary artery calcification (CAC).[96] A recent large-scale GWAS identified several new SNPs that are associated with CHD.[97] According to the paper, the combined set of previously identified SNPs and newly identified SNPs explains 10% of genetic variance of CHD. The most significant SNPs in each region are more frequently found in regulatory regions such as 5'-untranslated regions than in inter-genic regions.[98]

Based on the GWAS Catalog (http://www.genome.gov/gwastudies) by the National Human Genome research Institute (NHGRI), there were 174 SNPs (p-value $\leq$ 5E-8) found to be associated with cardiovascular disease by the end of May 2014.

DNA Methylation and Cardiovascular Disease Risks

Epigenetic modification, including DNA methylation, histone acetylation and methylation, and RNA-associated silencing, can lead to changes in gene expression and phenotype without making changes to the DNA sequence. It is an important mechanism explaining the variation in phenotype observed among individuals with the same genotype. Some epigenetic patterns are inherited from parent to child but can be changed in response to environmental exposure. Epigenetic modification can explain the difference in phenotype between monozygotic twins.[99] It also can explain changes in gene expression level and phenotype of the same individual over time.[100]

One example of epigenetic modification, DNA methylation, plays an important role in the development of atherosclerosis. In a study of knockout mice with deficiency in Methylenetetrahydrofolate reductase (MTHFR), global DNA hypomethylation was followed by the formation of aortic fatty streaks.[101] Another mouse study also suggested that changes in global DNA methylation profiles are early markers of atherosclerosis and may play a causative role in atherosclerosis.[102] The same study also showed that atherogenic lipoproteins promote global DNA hypermethylation in a human monocyte cell line. A change in methylation pattern is observed in the promoter region of genes such as extracellular superoxide dismutase, estrogen receptor-alpha, endothelial nitric oxide synthase and 15-lipoxygenase.[103-106]

A recent study investigated genetic-epigenetic interaction in CHD.[107] The study was conducted in 253 unrelated subjects from the Verona Heart Project. The authors studied SNPs and DNA methylation levels in Factor VII (*F7*) gene, which is involved in the coagulation pathway and atherothrombosis. The methylation sites of the *F7* gene were located at CpG islands and two known functional polymorphisms: a SNP (402 base-pair upstream of the gene, G>A)

and a 10 base-pair insertion (323 base-pair upstream of the gene). This study found that methylation of the SNP at the *F7* gene promoter region reduced serum FVIIa concentration only among those with GG genotype (Pearson correlation coefficient between DNA methylation levels and serum FVIIa concentration = -0.201, p = 0.012) but not in other genotypes (p = 0.30). The CHD-free group showed higher methylation associated with lower FVIIa concentrations compared to the CHD group. The study suggests the possible role of genetic-epigenetic interaction in CHD development. In this dissertation, measures of DNA methylation are included in Aims 2 and 3.

Vitamin D, Vitamin D receptor, and Cardiovascular Disease

Vitamin D and VDR

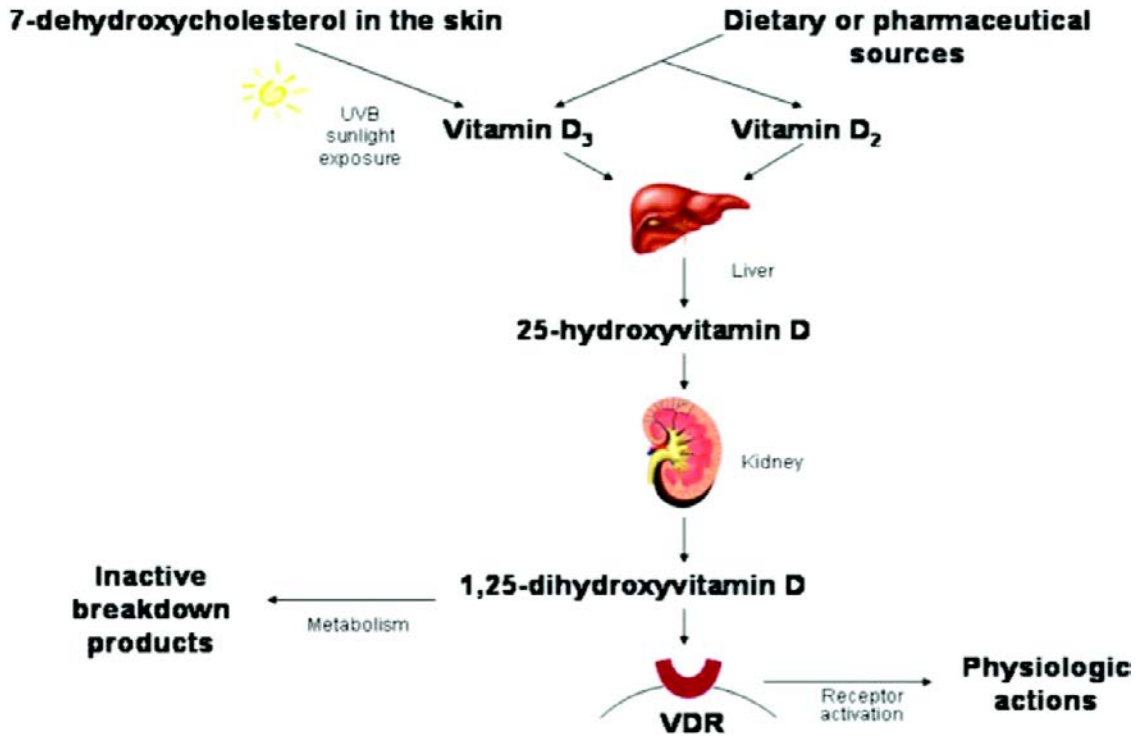

Figure 1.4 Vitamin D metabolism and activation to the steroidal hormone calcitriol. Adapted from Simpton R U *Circulation.* 2011; 124:1808-1810.

Vitamin D_2 and Vitamin D_3 are two forms of vitamin D in the human body. Vitamin Ds are hydroxylated to 25-hydroxyvitamin D in the liver. An additional hydroxylation forms 1,25-dihydroxyvitamin D (**Figure 1.4**). Vitamin D plays an important role in calcium metabolism.[108] Calcitriol (1,25-dihydroxyvitamin D), the active hormonal form of vitamin D, stimulates increased calcium absorption from the gut during calcium deficiency. It also stimulates the expression of genes engaged in absorption of calcium in the intestine. The effect of calcitriol is mediated by its interaction with the vitamin D receptor (VDR). The complex of calcitriol and VDR acts as a transcription factor regulating the gene expression of calcium-binding receptors.[109]

VDR and Cardiovascular Disease

Importantly, vitamin D and the VDR are involved in not only calcium metabolism but also regulation of the cardiovascular system and development of CVD. VDRs are distributed in most of the body cells including cardiomyocytes, vascular smooth muscles, and endothelium of the blood vessels.[110-112] In a rat study, expression of the VDR and enzymes involved in vitamin D synthesis and metabolism, 1-alpha-hydrozylase and 24-hydroxylase, increased in the hypertrophic heart.[113] Cardiac myocytes isolated from VDR knockout mice show accelerated rates of contraction and relaxation.[114] Deletion of the *VDR* gene in cardiomyocytes results in cardiac hypertrophy.[115] Vitamin D has been demonstrated to regulate endothelial nitric oxide synthase and arterial stiffness in a mouse study.[116] In apolipoprotein E knockout mice, vitamin D deficiency was associated with atherosclerosis.[117]

The effect of VDR on immune response and inflammation has been related to atherosclerosis.[118] Vitamin D levels are involved in arterial stiffness and arterial aging.[119, 120] Lower serum vitamin D_3 level is known to be associated with hypertension.[121] VDR is also involved in the renin-angiotensin system, cell proliferation and differentiation, anti-inflammation, and anti-fibrosis.[122] Circulating vitamin D levels are found to be associated with left ventricular geometry and function, including ejection fraction, end-systolic diameter, and fractional shortening.[123] Vitamin D supplementation is known to decrease the circulating level of pro-brain natriuretic peptide, a biomarker of CVD, and high-sensitivity CRP levels, a marker of inflammation.[124] Vitamin D deficiency from either diet or genetic polymorphisms in the *VDR* gene has been associated with CVD.[125-128] Based on a study of the National Health and Nutrition Examination Survey (NHANES) 2001 to 2004 data, prevalence of hypovitaminosis D (<30ng/ml) was associated with CHD (OR=1.48, 95%CI=(1.14, 1.91)).[129]

Genetics of the VDR

The human *VDR* gene is 63 kilo base pairs long with 9 exons.[130] Exon1 has 6 isoforms and alternative splicing of the *VDR* gene results in at least 14 transcript variants encoding different proteins.[131] The *VDR* genetic polymorphisms are differentially distributed among ethnic populations with different linkage disequilibrium (LD) patterns.[132] In the HapMap Phase II samples of 60 unrelated individuals from northern and western European origin (CEU) and 60 unrelated Yoruba individuals (YRI) from Nigeria, the difference in allele frequency of SNPs in the *VDR* gene between the two ethnic populations ranges from 0.03 to 0.78 (**Table 1.1**). The distribution and size of haplotype blocks are also different among ethnic groups. The sizes of haplotype blocks are smaller in YRI than in CEU, as expected. The differences in allele frequencies could potentially explain, in part, some of the differences in prevalence of CVD and its risk factors between different ethnic groups. SNPs in the *VDR* gene are included in Aim 1.

Lead

Importance and measurement of Lead

Lead exposure is a major public health issue and there are multiple sources of exposure including inhalation and ingestion of lead from indoor dust or outdoor soil, occupational exposures, hand-to-mouth contact with lead-containing products for children, dust or chips of peeling lead-containing paint, and ingestion of lead in drinking water conveyed through lead pipes.[133] Recent reductions in the use of lead in gasoline, paint, and drinking water resulted in reduction in blood lead levels.[134] However, even though blood lead levels have been gradually reduced since the phase-out of leaded gasoline in 1970s (blood lead levels in US population in 1976-1980 (NHANES-II) and in NHANES 2007-2008 were 12.8µg/dL and 1.3 µg/dL, respectively),[135]

cumulative lead levels are still substantial in the elderly and associated with diverse diseases including cognitive function decline and CVD in the United States.[136, 137]

Lead is a toxic heavy metal and lead toxicity is found in almost every system in the body. Over 90% of the total body lead burden in adults is accumulated in the bone and only about 1% is found in the blood.[138] Blood lead is the most common biomarker of lead exposure levels in epidemiologic studies. While blood lead measurement reflects current or recent lead exposure levels, bone lead measurement is used as an index of cumulative lead exposure levels. Therefore, bone lead measurements serve as a better biomarker to study chronic disease.[139]

Lead and Cardiovascular Disease

Lead is known to increase the risk of CVD. Several studies have identified a positive association between lead exposure and hypertension, which is a major risk factor for CVD. Martin et al. found an association between blood lead levels and increases in blood pressure, as well as an association between tibia bone lead levels and hypertension in the Baltimore Memory Study.[137] Cheng et al. reported an increased risk of incident hypertension with higher levels of lead in tibia and patella bones using data from the Normative Aging Study (NAS).[140] Another study, which used the NAS data, found an association between high lead levels in bone and blood and hypertension among subjects with low dietary calcium intake.[141] Blood lead levels have also been associated with increased risk of pregnancy-induced hypertension.[142] The association between lead exposure and the risk of hypertension in pregnant women was also demonstrated even in those with blood lead levels less than 2 μg/dL.[143]

One of the biological processes that likely explains the association between lead exposure and the increased risk of CVD is arterial stiffness, which is associated with aging and is accelerated by hypertension.[144] Some studies have identified a positive association of lead

exposure with increases in pulse pressure, a measure of arterial stiffness, and with clinical CVD events. The study by Perlstein et al. suggests that lead accumulation may contribute to the increase in pulse pressure in the NAS.[145] Zhang et al. also reported a deleterious impact of cumulative lead on pulse pressure with effect modification by hemochromatosis (*HFE* H63D) genetic polymorphism in the NAS.[146]

Lead and VDR

VDR is involved in lead absorption and accumulation. Vitamin D plays an important role in calcium metabolism, which is shared by lead.[108] From its divalent cation characteristic, lead competes to bind to calcium-binding receptors.[147] When calcium levels are low, the calcium-binding protein may bind lead instead of calcium, resulting in elevated absorption of lead.[148] The *VDR* genetic variants have been identified as potential genetic factors that can influence the absorption, retention and accumulation of lead in the human body.[149] Schwartz et al. examined former organolead manufacturing workers and found that the *VDR* gene variant *Bsm1* modifies the association between age and tibia bone lead levels.[150] In this study, the cumulative level of lead in bone and the rate of reabsorption and excretion of lead over time were higher for those with the *VDR Bsm1* variant. This study suggests that *VDR* genetic variant may play a role in susceptibility to lead accumulation.

There have been studies examining effect modification of lead by *VDR* gene in relation to several diseases. Krieg et al. found that the association between lead and cognitive function was modified by *VDR* genetic variants, and the relationship between the serum homocysteine levels and blood lead concentrations varied by *VDR* genotypes.[151] Weaver et al. reported that *VDR* genotypes modified the association between lead and serum creatinine level and the association

between lead and renal outcomes in lead workers.[152] SNPs in the *VDR* gene are included in Aim 1.

Mendelian Randomization

Observational epidemiologic studies have often suffered from confounding and reverse causation. The MR approach is a type of instrumental variable analysis,[153] in which a genetic variant is used as an instrument variable to avoid biases from confounding and reverse causation.[154] Genetic variants in DNA sequence, such as SNPs, may serve as an appropriate instrument because the meiotic assortment of genetic variants during gamete formation is random. As we are born with the randomly assigned genetic variants, common confounders such as age, gender, and socioeconomic status do not have effects on genetic variants. Therefore, using genetic variants as instruments for the exposure of interest helps to avoid confounding when we assess the relationship between an environmentally modifiable exposure and an outcome. Another advantage of MR is that we can make an inference on the direction of the association between two variables using a genetic variant as an instrument. The problem of reverse causation often exists in epidemiologic studies. For example, patients stop smoking or drinking alcohol after they are diagnosed for a disease. However, by assumption, the instrument has an effect on the outcome only through the exposure. Therefore, we can evaluate the direction of the association from one variable to another variable by randomizing one variable with genetic variants. Using an appropriate instrument enables us to make an inference that is comparable to randomized clinical trials while using observational epidemiologic studies.

The use of the MR approach can be limited by the availability of an appropriate instrument.[155] A genetic variant that satisfies the following conditions can serve as an appropriate instrument: 1) the instrument has a causal effect (here the causal effect means directed effect

from instrument to exposure) on the (environmental) exposure, 2) the association between the instrument and the (disease) outcome is only through the exposure, and 3) the instrument does not share common causes (confounders) with the (disease) outcome. If there are no appropriate genetic variants that satisfy the conditions, the MR approach may not be able to provide stable unbiased estimates. Another limitation of this approach is a confounding that may be introduced by pleiotropic effects of genetic variants. Genetic variants often have pleiotropic effects, meaning that one genetic variant may have effects on several phenotypes. The pleiotropic effect may open other paths to the outcome resulting in introducing a bias in estimation. The MR approach is often less powerful than conventional regression approaches, hence, it may require a larger sample size.[156]

In spite of the limitations of the approach, the MR approach has become more widely used recently.[157] Advances in genotyping and sequencing techniques have enhanced the availability of genetic variants as an instrumental variable. In addition, several extensions of the MR approach are becoming available and provide enhanced flexibility of the application of this approach.[157]

Two-Step Epigenetic Mendelian Randomization

Two-step epigenetic MR was proposed as an extension of the MR approach to integrate epigenetic modifications.[158] DNA methylation is a kind of epigenetic modification. It is responsive to environmental exposures and plays an important role in regulation of gene expression.[159] Therefore, DNA methylation can be considered as an intermediate phenotype that mediates between environmental exposure and disease outcome. The two-step epigenetic MR approach is composed of two steps of MR. The first step assesses the relationship between an environmental exposure and DNA methylation, and the second step assesses the relationship

between DNA methylation and an outcome. This approach requires instrumental variables for each step. Recent findings from methylation quantitative trait loci (MeQTL) studies support the idea of a genetic proxy for DNA methylation.[160] In addition to that, recent epigenome-wide association studies (EWAS) identified differently methylated CpG sites by various environmental exposures supporting the mediating role of DNA methylation.[161-164]

The two-step epigenetic MR approach is limited by the availability of an appropriate instrument, confounding from pleiotropic effects, and reduced statistical power. However, this approach enables us to identify DNA methylation sites that mediate an environmental exposure and an outcome. The CpG sites of differential methylation can shed light on the biological/genetic mechanisms behind the elevated disease risk induced by an environmental exposure. The two-step MR approach is used in Aim 2.

Network Mendelian Randomization

The network MR approach is another extension of MR to investigate more complex paths of variables of interest.[165] This approach enables us to estimate both direct and indirect effects of an exposure on an outcome. The method can be implemented using either a sequence of regressions or structural equation models (SEM).

This approach requires some additional assumptions to estimate direct and indirect effects[166] using instrumental variables: 1) no interaction effect of exposure and mediator on outcome, 2) no bidirectional associations, and 3) sequential ignorability[167] meaning that the exposure is statistically independent of the mediators and outcomes, and the mediator is statistically independent of the outcome given exposure and confounders. When the assumptions are satisfied, the network MR approach that is implemented using software for SEM provides extensive flexibility in applying the MR approach. Network MR is used in Aim 3.

Gene by Environment Interaction in Cardiovascular Disease

Complex diseases, such as CVD, are influenced by a combination and interaction of various factors. Risk of CVD is affected by many factors including genetic variants, epigenetic modification, age, gender, race/ethnicity, smoking, alcohol consumption, diet, obesity, diabetes mellitus, physical activity, socioeconomic status, hypertension, dyslipidemia, arterial stiffness, left ventricular dysfunction, and environmental exposures like lead.[168-170] Most studies focus on genetic variants ignore other factors or include them as confounders. In contrast, many studies that focus on environmental exposures or behaviors tend to ignore genetic variability among individuals. While a major assumption of many studies is that there is no gene-environment interaction, evidence exists for such interactions.[171]

Evidence for gene by environment interaction implies that the phenotype for a specific genotype depends on the environmental exposures. The environment may be measured directly, for example by cumulative lead exposure or self-reported smoking, or more indirectly by epigenetic modification in response to environmental exposures. Epigenetic modification occurs as a response to changes in environmental and lifestyle factors.[172] Thus, epigenetic modification is known to explain, in part, the interaction between genes and environmental factors in development of various diseases including CVD.[173-179] Epigenetic modification may be seen as an intermediate phenotype which is a result of the interaction between genes and environmental factors.

Overarching goal

The goal of this dissertation is to understand biological mechanisms of CVD using gene-environment interaction. Specifically this dissertation examines effect modification by candidate genetic polymorphisms and explores mediation analysis using epigenetic modification on

measures of inflammation as well as CVD risk markers. Inflammatory biomarkers including CRP, IL-6, IL-18, and fibrinogen were considered. Pulse pressure, a measure of arterial stiffness, left ventricular ejection fraction (LVEF), a measure of systolic function, and left ventricular mass index (LVMI), a measure of left ventricular hypertrophy, are the subclinical markers of CVD risk and are the outcomes of interest. Aim 1 of this dissertation examines whether there is evidence for an interaction between *VDR* genetic polymorphisms and cumulative bone lead levels on pulse pressure in unrelated non-Hispanic White men. Aim 2 examines the mediating effect of DNA methylation on the relationship between smoking and inflammatory markers in African American siblings. Aim 3 investigates the causal paths of smoking, DNA methylation, and inflammatory markers extended to include subclinical markers of CVD in African American siblings.

Hypothesis for Aim 1

VDR genetic polymorphisms modify the association between cumulative bone lead exposure levels and longitudinal measures of pulse pressure.

Hypothesis for Aim 2

DNA methylation mediates the association between cigarette smoking and inflammatory markers.

Hypothesis for Aim 3

DNA methylation mediates the path from cigarette smoking to inflammation to either LVEF or LVMI.

Studies on gene-environment interaction may shed light on identifying biological mechanisms underlying CVD development. The studies of gene-environment interaction in this dissertation,

using both directly measured environmental exposures, such as lead measurements, as well as indirectly measured environmental exposures, such as DNA methylation levels, together will further our understanding of the factors affecting CVD risk markers in diverse populations.

Tables

Table 1.1 Allele frequencies of VDR genetic polymorphisms in the HapMap Phase II CEU and YRI populations. Partly adapted from Lins et al. *Genet and Mol Biol.* 2011; 34(3):377-385.

SNP	Allele	CEU	YRI	Difference in allele frequency
rs4077869	A	NA	0.40	NA
rs11568820 (*Cdx2*)	C	0.79	0.02	0.78
rs4516035	A	0.58	0.99	0.41
rs10783219	A	0.33	0.00	0.33
rs7302235	A	0.74	0.43	0.31
rs3890734	C	0.68	0.88	0.20
rs2853559	A	0.42	0.17	0.25
rs2853564	A	0.58	0.91	0.33
rs2254210	C	0.63	0.66	0.03
rs10735810 (*FokI*)	C	0.53	0.83	0.31
rs886441	A	0.81	0.58	0.23
rs2239179	C	0.42	0.29	0.13
rs2248098	A	0.43	0.38	0.04
rs2239185	A	NA	0.54	NA
rs1544410 (*BsmI*)	C	0.53	0.71	0.19
rs7975232 (*ApaI*)	G	0.42	0.38	0.05
rs731236 (*TaqI*)	A	0.53	0.75	0.22
rs9729	G	0.41	0.34	0.08
rs7968585	A	0.59	0.63	0.04
rs11608702	A	0.36	0.18	0.18
rs2544040	C	1.00	0.89	0.12

SNP: Single Nucleotide Polymorphism; CEU: Northern and western European; YRI: Yoruba in Ibadan, Nigeria; Cdx2: caudal type homeobox 2; NA: Not available

References

1. Murphy SL XJ, Kochanek KD. Deaths: Final data for 2010. *Natl Vital Stat Rep.* 2013;61
2. Summary health statistics for u.S. Adults: National health interview survey, 2012 *Data from the National Health Interview Survey*. 2014
3. Malik IS. Inflammation in cardiovascular disease. *Journal of the Royal College of Physicians of London*. 2000;34:205-207
4. Kaplan RC, Frishman WH. Systemic inflammation as a cardiovascular disease risk factor and as a potential target for drug therapy. *Heart disease*. 2001;3:326-332
5. Wannamethee SG, Lowe GD, Shaper AG, Rumley A, Lennon L, Whincup PH. Associations between cigarette smoking, pipe/cigar smoking, and smoking cessation, and haemostatic and inflammatory markers for cardiovascular disease. *European heart journal*. 2005;26:1765-1773
6. Imhof A, Froehlich M, Brenner H, Boeing H, Pepys MB, Koenig W. Effect of alcohol consumption on systemic markers of inflammation. *Lancet*. 2001;357:763-767
7. Van Lente F. Markers of inflammation as predictors in cardiovascular disease. *Clinica chimica acta; international journal of clinical chemistry*. 2000;293:31-52
8. Koenig W, Sund M, Frohlich M, Fischer HG, Lowel H, Doring A, Hutchinson WL, Pepys MB. C-reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: Results from the monica (monitoring trends and determinants in cardiovascular disease) augsburg cohort study, 1984 to 1992. *Circulation*. 1999;99:237-242
9. Ridker PM, Hennekens CH, Buring JE, Rifai N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *The New England journal of medicine*. 2000;342:836-843
10. Ridker PM, Rifai N, Stampfer MJ, Hennekens CH. Plasma concentration of interleukin-6 and the risk of future myocardial infarction among apparently healthy men. *Circulation*. 2000;101:1767-1772
11. Blankenberg S, Tiret L, Bickel C, Peetz D, Cambien F, Meyer J, Rupprecht HJ, AtheroGene I. Interleukin-18 is a strong predictor of cardiovascular death in stable and unstable angina. *Circulation*. 2002;106:24-30
12. Blankenberg S, Luc G, Ducimetiere P, Arveiler D, Ferrieres J, Amouyel P, Evans A, Cambien F, Tiret L, Group PS. Interleukin-18 and the risk of coronary heart disease in european men: The prospective epidemiological study of myocardial infarction (prime). *Circulation*. 2003;108:2453-2459
13. Kannel WB, Wolf PA, Castelli WP, D'Agostino RB. Fibrinogen and risk of cardiovascular disease. The framingham study. *JAMA*. 1987;258:1183-1186
14. Stec JJ, Silbershatz H, Tofler GH, Matheney TH, Sutherland P, Lipinska I, Massaro JM, Wilson PF, Muller JE, D'Agostino RB, Sr. Association of fibrinogen with cardiovascular risk factors and cardiovascular disease in the framingham offspring population. *Circulation*. 2000;102:1634-1638
15. Pepys MB, Baltz ML. Acute phase proteins with special reference to c-reactive protein and related proteins (pentaxins) and serum amyloid a protein. *Advances in immunology*. 1983;34:141-212
16. Pepys MB, Rowe IF, Baltz ML. C-reactive protein: Binding to lipids and lipoproteins. *International review of experimental pathology*. 1985;27:83-111

17. Li YP, Mold C, Du Clos TW. Sublytic complement attack exposes c-reactive protein binding sites on cell membranes. *Journal of immunology*. 1994;152:2995-3005
18. Chang MK, Binder CJ, Torzewski M, Witztum JL. C-reactive protein binds to both oxidized ldl and apoptotic cells through recognition of a common ligand: Phosphorylcholine of oxidized phospholipids. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99:13043-13048
19. Steel DM, Whitehead AS. The major acute phase reactants: C-reactive protein, serum amyloid p component and serum amyloid a protein. *Immunology today*. 1994;15:81-88
20. Du Clos TW. Function of c-reactive protein. *Annals of medicine*. 2000;32:274-278
21. Shine B, de Beer FC, Pepys MB. Solid phase radioimmunoassays for human c-reactive protein. *Clinica chimica acta; international journal of clinical chemistry*. 1981;117:13-23
22. Khera A, McGuire DK, Murphy SA, Stanek HG, Das SR, Vongpatanasin W, Wians FH, Jr., Grundy SM, de Lemos JA. Race and gender differences in c-reactive protein levels. *Journal of the American College of Cardiology*. 2005;46:464-469
23. Hutchinson WL, Koenig W, Frohlich M, Sund M, Lowe GD, Pepys MB. Immunoradiometric assay of circulating c-reactive protein: Age-related values in the adult general population. *Clinical chemistry*. 2000;46:934-938
24. Tracy RP, Lemaitre RN, Psaty BM, Ives DG, Evans RW, Cushman M, Meilahn EN, Kuller LH. Relationship of c-reactive protein to risk of cardiovascular disease in the elderly. Results from the cardiovascular health study and the rural health promotion project. *Arteriosclerosis, thrombosis, and vascular biology*. 1997;17:1121-1127
25. Ridker PM, Stampfer MJ, Rifai N. Novel risk factors for systemic atherosclerosis: A comparison of c-reactive protein, fibrinogen, homocysteine, lipoprotein(a), and standard cholesterol screening as predictors of peripheral arterial disease. *JAMA*. 2001;285:2481-2485
26. Albert CM, Ma J, Rifai N, Stampfer MJ, Ridker PM. Prospective study of c-reactive protein, homocysteine, and plasma lipid levels as predictors of sudden cardiac death. *Circulation*. 2002;105:2595-2599
27. Loppnow H, Libby P. Proliferating or interleukin 1-activated human vascular smooth muscle cells secrete copious interleukin 6. *The Journal of clinical investigation*. 1990;85:731-738
28. Kaneko K, Kanda T, Yokoyama T, Nakazato Y, Iwasaki T, Kobayashi I, Nagai R. Expression of interleukin-6 in the ventricles and coronary arteries of patients with myocardial infarction. *Research communications in molecular pathology and pharmacology*. 1997;97:3-12
29. Gwechenberger M, Mendoza LH, Youker KA, Frangogiannis NG, Smith CW, Michael LH, Entman ML. Cardiac myocytes produce interleukin-6 in culture and in viable border zone of reperfused infarctions. *Circulation*. 1999;99:546-551
30. Lotz M, Jirik F, Kabouridis P, Tsoukas C, Hirano T, Kishimoto T, Carson DA. B cell stimulating factor 2/interleukin 6 is a costimulant for human thymocytes and t lymphocytes. *The Journal of experimental medicine*. 1988;167:1253-1258
31. Luger TA, Krutmann J, Kirnbauer R, Urbanski A, Schwarz T, Klappacher G, Kock A, Micksche M, Malejczyk J, Schauer E, et al. Ifn-beta 2/il-6 augments the activity of human natural killer cells. *Journal of immunology*. 1989;143:1206-1209
32. Somers W, Stahl M, Seehra JS. 1.9 a crystal structure of interleukin 6: Implications for a novel mode of receptor dimerization and signaling. *The EMBO journal*. 1997;16:989-997

33. Romano M, Sironi M, Toniatti C, Polentarutti N, Fruscella P, Ghezzi P, Faggioni R, Luini W, van Hinsbergh V, Sozzani S, Bussolino F, Poli V, Ciliberto G, Mantovani A. Role of il-6 and its soluble receptor in induction of chemokines and leukocyte recruitment. *Immunity*. 1997;6:315-325
34. Scheller J, Chalaris A, Schmidt-Arras D, Rose-John S. The pro- and anti-inflammatory properties of the cytokine interleukin-6. *Biochimica et biophysica acta*. 2011;1813:878-888
35. Yudkin JS, Stehouwer CD, Emeis JJ, Coppack SW. C-reactive protein in healthy subjects: Associations with obesity, insulin resistance, and endothelial dysfunction: A potential role for cytokines originating from adipose tissue? *Arteriosclerosis, thrombosis, and vascular biology*. 1999;19:972-978
36. Verma S, Li SH, Badiwala MV, Weisel RD, Fedak PW, Li RK, Dhillon B, Mickle DA. Endothelin antagonism and interleukin-6 inhibition attenuate the proatherogenic effects of c-reactive protein. *Circulation*. 2002;105:1890-1896
37. Modur V, Li Y, Zimmerman GA, Prescott SM, McIntyre TM. Retrograde inflammatory signaling from neutrophils to endothelial cells by soluble interleukin-6 receptor alpha. *The Journal of clinical investigation*. 1997;100:2752-2756
38. Raymond RJ, Dehmer GJ, Theoharides TC, Deliargyris EN. Elevated interleukin-6 levels in patients with asymptomatic left ventricular systolic dysfunction. *American heart journal*. 2001;141:435-438
39. Kosmala W, Derzhko R, Przewlocka-Kosmala M, Orda A, Mazurek W. Plasma levels of tnf-alpha, il-6, and il-10 and their relationship with left ventricular diastolic function in patients with stable angina pectoris and preserved left ventricular systolic performance. *Coronary artery disease*. 2008;19:375-382
40. Yan AT, Yan RT, Cushman M, Redheuil A, Tracy RP, Arnett DK, Rosen BD, McClelland RL, Bluemke DA, Lima JA. Relationship of interleukin-6 with regional and global left-ventricular function in asymptomatic individuals without clinical cardiovascular disease: Insights from the multi-ethnic study of atherosclerosis. *European heart journal*. 2010;31:875-882
41. Hirota H, Yoshida K, Kishimoto T, Taga T. Continuous activation of gp130, a signal-transducing receptor component for interleukin 6-related cytokines, causes myocardial hypertrophy in mice. *Proceedings of the National Academy of Sciences of the United States of America*. 1995;92:4862-4866
42. Melendez GC, McLarty JL, Levick SP, Du Y, Janicki JS, Brower GL. Interleukin 6 mediates myocardial fibrosis, concentric hypertrophy, and diastolic dysfunction in rats. *Hypertension*. 2010;56:225-231
43. Tsutamoto T, Hisanaga T, Wada A, Maeda K, Ohnishi M, Fukai D, Mabuchi N, Sawaki M, Kinoshita M. Interleukin-6 spillover in the peripheral circulation increases with the severity of heart failure, and the high plasma level of interleukin-6 is an important prognostic predictor in patients with congestive heart failure. *Journal of the American College of Cardiology*. 1998;31:391-398
44. Biasucci LM, Liuzzo G, Fantuzzi G, Caligiuri G, Rebuzzi AG, Ginnetti F, Dinarello CA, Maseri A. Increasing levels of interleukin (il)-1ra and il-6 during the first 2 days of hospitalization in unstable angina are associated with increased risk of in-hospital coronary events. *Circulation*. 1999;99:2079-2084

45. Wollert KC, Drexler H. The role of interleukin-6 in the failing heart. *Heart failure reviews*. 2001;6:95-103
46. Stoll S, Jonuleit H, Schmitt E, Muller G, Yamauchi H, Kurimoto M, Knop J, Enk AH. Production of functional il-18 by different subtypes of murine and human dendritic cells (dc): Dc-derived il-18 enhances il-12-dependent th1 development. *European journal of immunology*. 1998;28:3231-3239
47. Udagawa N, Horwood NJ, Elliott J, Mackay A, Owens J, Okamura H, Kurimoto M, Chambers TJ, Martin TJ, Gillespie MT. Interleukin-18 (interferon-gamma-inducing factor) is produced by osteoblasts and acts via granulocyte/macrophage colony-stimulating factor and not via interferon-gamma to inhibit osteoclast formation. *The Journal of experimental medicine*. 1997;185:1005-1012
48. Pizarro TT, Michie MH, Bentz M, Woraratanadharm J, Smith MF, Jr., Foley E, Moskaluk CA, Bickston SJ, Cominelli F. Il-18, a novel immunoregulatory cytokine, is up-regulated in crohn's disease: Expression and localization in intestinal mucosal cells. *Journal of immunology*. 1999;162:6829-6835
49. Gracie JA, Forsey RJ, Chan WL, Gilmour A, Leung BP, Greer MR, Kennedy K, Carter R, Wei XQ, Xu D, Field M, Foulis A, Liew FY, McInnes IB. A proinflammatory role for il-18 in rheumatoid arthritis. *The Journal of clinical investigation*. 1999;104:1393-1401
50. Gu Y, Kuida K, Tsutsui H, Ku G, Hsiao K, Fleming MA, Hayashi N, Higashino K, Okamura H, Nakanishi K, Kurimoto M, Tanimoto T, Flavell RA, Sato V, Harding MW, Livingston DJ, Su MS. Activation of interferon-gamma inducing factor mediated by interleukin-1beta converting enzyme. *Science*. 1997;275:206-209
51. Sugawara S, Uehara A, Nochi T, Yamaguchi T, Ueda H, Sugiyama A, Hanzawa K, Kumagai K, Okamura H, Takada H. Neutrophil proteinase 3-mediated induction of bioactive il-18 secretion by human oral epithelial cells. *Journal of immunology*. 2001;167:6568-6575
52. Gerdes N, Sukhova GK, Libby P, Reynolds RS, Young JL, Schonbeck U. Expression of interleukin (il)-18 and functional il-18 receptor on human vascular endothelial cells, smooth muscle cells, and macrophages: Implications for atherogenesis. *The Journal of experimental medicine*. 2002;195:245-257
53. Leung BP, Culshaw S, Gracie JA, Hunter D, Canetti CA, Campbell C, Cunha F, Liew FY, McInnes IB. A role for il-18 in neutrophil activation. *Journal of immunology*. 2001;167:2879-2886
54. Hyodo Y, Matsui K, Hayashi N, Tsutsui H, Kashiwamura S, Yamauchi H, Hiroishi K, Takeda K, Tagawa Y, Iwakura Y, Kayagaki N, Kurimoto M, Okamura H, Hada T, Yagita H, Akira S, Nakanishi K, Higashino K. Il-18 up-regulates perforin-mediated nk activity without increasing perforin messenger rna expression by binding to constitutively expressed il-18 receptor. *Journal of immunology*. 1999;162:1662-1668
55. Okamura H, Tsutsui H, Kashiwamura S, Yoshimoto T, Nakanishi K. Interleukin-18: A novel cytokine that augments both innate and acquired immunity. *Advances in immunology*. 1998;70:281-312
56. Gracie JA, Robertson SE, McInnes IB. Interleukin-18. *Journal of leukocyte biology*. 2003;73:213-224
57. Okamura H, Tsutsi H, Komatsu T, Yutsudo M, Hakura A, Tanimoto T, Torigoe K, Okura T, Nukada Y, Hattori K, et al. Cloning of a new cytokine that induces ifn-gamma production by t cells. *Nature*. 1995;378:88-91

58. Micallef MJ, Ohtsuki T, Kohno K, Tanabe F, Ushio S, Namba M, Tanimoto T, Torigoe K, Fujii M, Ikeda M, Fukuda S, Kurimoto M. Interferon-gamma-inducing factor enhances t helper 1 cytokine production by stimulated human t cells: Synergism with interleukin-12 for interferon-gamma production. *European journal of immunology*. 1996;26:1647-1651
59. Kim SH, Eisenstein M, Reznikov L, Fantuzzi G, Novick D, Rubinstein M, Dinarello CA. Structural requirements of six naturally occurring isoforms of the il-18 binding protein to inhibit il-18. *Proceedings of the National Academy of Sciences of the United States of America*. 2000;97:1190-1195
60. Yamamura M, Kawashima M, Taniai M, Yamauchi H, Tanimoto T, Kurimoto M, Morita Y, Ohmoto Y, Makino H. Interferon-gamma-inducing activity of interleukin-18 in the joint with rheumatoid arthritis. *Arthritis and rheumatism*. 2001;44:275-285
61. Nicoletti F, Conget I, Di Marco R, Speciale AM, Morinigo R, Bendtzen K, Gomis R. Serum levels of the interferon-gamma-inducing cytokine interleukin-18 are increased in individuals at high risk of developing type i diabetes. *Diabetologia*. 2001;44:309-311
62. Nicoletti F, Di Marco R, Mangano K, Patti F, Reggio E, Nicoletti A, Bendtzen K, Reggio A. Increased serum levels of interleukin-18 in patients with multiple sclerosis. *Neurology*. 2001;57:342-344
63. Monteleone G, Trapasso F, Parrello T, Biancone L, Stella A, Iuliano R, Luzza F, Fusco A, Pallone F. Bioactive il-18 expression is up-regulated in crohn's disease. *Journal of immunology*. 1999;163:143-147
64. Furuya D, Yagihashi A, Komatsu M, Masashi N, Tsuji N, Kobayashi D, Watanabe N. Serum interleukin-18 concentrations in patients with inflammatory bowel disease. *Journal of immunotherapy*. 2002;25 Suppl 1:S65-67
65. Vidal-Vanaclocha F, Mendoza L, Telleria N, Salado C, Valcarcel M, Gallot N, Carrascal T, Egilegor E, Beaskoetxea J, Dinarello CA. Clinical and experimental approaches to the pathophysiology of interleukin-18 in cancer progression. *Cancer metastasis reviews*. 2006;25:417-434
66. Fuss C, Palmaz JC, Sprague EA. Fibrinogen: Structure, function, and surface interactions. *Journal of vascular and interventional radiology : JVIR*. 2001;12:677-682
67. Herrick S, Blanc-Brude O, Gray A, Laurent G. Fibrinogen. *The international journal of biochemistry & cell biology*. 1999;31:741-746
68. Lisman T, Weeterings C, de Groot PG. Platelet aggregation: Involvement of thrombin and fibrin(ogen). *Frontiers in bioscience : a journal and virtual library*. 2005;10:2504-2517
69. Szaba FM, Smiley ST. Roles for thrombin and fibrin(ogen) in cytokine/chemokine production and macrophage adhesion in vivo. *Blood*. 2002;99:1053-1059
70. Koenig W. Fibrin(ogen) in cardiovascular disease: An update. *Thrombosis and haemostasis*. 2003;89:601-609
71. Kannel WB, D'Agostino RB, Belanger AJ. Update on fibrinogen as a cardiovascular risk factor. *Annals of epidemiology*. 1992;2:457-466
72. Okada K, Iso H, Cui R, Inoue M, Tsugane S. Pulse pressure is an independent risk factor for stroke among middle-aged japanese with normal systolic blood pressure: The jphc study. *Journal of hypertension*. 2011;29:319-324
73. Glasser SP, Halberg DL, Sands C, Gamboa CM, Muntner P, Safford M. Is pulse pressure an independent risk factor for incident acute coronary heart disease events? The regards study. *American journal of hypertension*. 2014;27:555-563

74. Vaccarino V, Holford TR, Krumholz HM. Pulse pressure and risk for myocardial infarction and heart failure in the elderly. *Journal of the American College of Cardiology*. 2000;36:130-138
75. Domanski M, Norman J, Wolz M, Mitchell G, Pfeffer M. Cardiovascular risk assessment using pulse pressure in the first national health and nutrition examination survey (nhanes i). *Hypertension*. 2001;38:793-797
76. Assmann G, Cullen P, Evers T, Petzinna D, Schulte H. Importance of arterial pulse pressure as a predictor of coronary heart disease risk in procam. *European heart journal*. 2005;26:2120-2126
77. Benetos A, Safar M, Rudnichi A, Smulyan H, Richard JL, Ducimetieere P, Guize L. Pulse pressure: A predictor of long-term cardiovascular mortality in a french male population. *Hypertension*. 1997;30:1410-1415
78. Franklin SS, Khan SA, Wong ND, Larson MG, Levy D. Is pulse pressure useful in predicting risk for coronary heart disease? The framingham heart study. *Circulation*. 1999;100:354-360
79. Franklin SS, Larson MG, Khan SA, Wong ND, Leip EP, Kannel WB, Levy D. Does the relation of blood pressure to coronary heart disease risk change with aging? The framingham heart study. *Circulation*. 2001;103:1245-1249
80. Jia EZ, Xu ZX, Yang ZJ, Zhu TB, Wang LS, Cao KJ, Ma WZ. Severity of coronary atherosclerosis is an independent predictor of the left ventricular ejection fraction. *Clinical and experimental pharmacology & physiology*. 2011;38:109-112
81. Solomon SD, Anavekar N, Skali H, McMurray JJ, Swedberg K, Yusuf S, Granger CB, Michelson EL, Wang D, Pocock S, Pfeffer MA, Candesartan in Heart Failure Reduction in Mortality I. Influence of ejection fraction on cardiovascular outcomes in a broad spectrum of heart failure patients. *Circulation*. 2005;112:3738-3744
82. Foppa M, Duncan BB, Rohde LE. Echocardiography-based left ventricular mass estimation. How should we define hypertrophy? *Cardiovascular ultrasound*. 2005;3:17
83. Devereux RB, Lutas EM, Casale PN, Kligfield P, Eisenberg RR, Hammond IW, Miller DH, Reis G, Alderman MH, Laragh JH. Standardization of m-mode echocardiographic left ventricular anatomic measurements. *Journal of the American College of Cardiology*. 1984;4:1222-1230
84. de Simone G, Daniels SR, Devereux RB, Meyer RA, Roman MJ, de Divitiis O, Alderman MH. Left ventricular mass and body size in normotensive children and adults: Assessment of allometric relations and impact of overweight. *Journal of the American College of Cardiology*. 1992;20:1251-1260
85. Antoniucci D, Seccareccia F, Menotti A, Dovellini EV, Prati PL, Rovelli F, Fazzini PF. Prevalence and correlates of echocardiographic determined left ventricular hypertrophy in 2318 asymptomatic middle-aged men: The eccis project. Epidemiolgia e clinica della cardiopatia ischemica silente. *Giornale italiano di cardiologia*. 1997;27:363-369
86. Coca A, Gabriel R, de la Figuera M, Lopez-Sendon JL, Fernandez R, Sagastagoitia JD, Garcia JJ, Barajas R. The impact of different echocardiographic diagnostic criteria on the prevalence of left ventricular hypertrophy in essential hypertension: The vitae study. Ventriculo izquierdo tension arterial espana. *Journal of hypertension*. 1999;17:1471-1480
87. Salles GF, Fiszman R, Cardoso CR, Muxfeldt ES. Relation of left ventricular hypertrophy with systemic inflammation and endothelial damage in resistant hypertension. *Hypertension*. 2007;50:723-728

88. Iwashima Y, Horio T, Kamide K, Rakugi H, Ogihara T, Kawano Y. C-reactive protein, left ventricular mass index, and risk of cardiovascular disease in essential hypertension. *Hypertension research : official journal of the Japanese Society of Hypertension*. 2007;30:1177-1185
89. Casale PN, Devereux RB, Milner M, Zullo G, Harshfield GA, Pickering TG, Laragh JH. Value of echocardiographic measurement of left ventricular mass in predicting cardiovascular morbid events in hypertensive men. *Annals of internal medicine*. 1986;105:173-178
90. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661-678
91. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316:1488-1491
92. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491-1493
93. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, Konig IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H, Wtccc, the Cardiogenics C. Genomewide association analysis of coronary artery disease. *The New England journal of medicine*. 2007;357:443-453
94. Coronary Artery Disease C, Samani NJ, Deloukas P, Erdmann J, Hengstenberg C, Kuulasmaa K, McGinnis R, Schunkert H, Soranzo N, Thompson J, Tiret L, Ziegler A. Large scale association analysis of novel genetic loci for coronary artery disease. *Arteriosclerosis, thrombosis, and vascular biology*. 2009;29:774-780
95. Murabito JM, White CC, Kavousi M, Sun YV, Feitosa MF, Nambi V, Lamina C, Schillert A, Coassin S, Bis JC, Broer L, Crawford DC, Franceschini N, Frikke-Schmidt R, Haun M, Holewijn S, Huffman JE, Hwang SJ, Kiechl S, Kollerits B, Montasser ME, Nolte IM, Rudock ME, Senft A, Teumer A, van der Harst P, Vitart V, Waite LL, Wood AR, Wassel CL, Absher DM, Allison MA, Amin N, Arnold A, Asselbergs FW, Aulchenko Y, Bandinelli S, Barbalic M, Boban M, Brown-Gentry K, Couper DJ, Criqui MH, Dehghan A, den Heijer M, Dieplinger B, Ding J, Dorr M, Espinola-Klein C, Felix SB, Ferrucci L, Folsom AR, Fraedrich G, Gibson Q, Goodloe R, Gunjaca G, Haltmayer M, Heiss G, Hofman A, Kieback A, Kiemeney LA, Kolcic I, Kullo IJ, Kritchevsky SB, Lackner KJ, Li X, Lieb W, Lohman K, Meisinger C, Melzer D, Mohler ER, 3rd, Mudnic I, Mueller T, Navis G, Oberhollenzer F, Olin JW, O'Connell J, O'Donnell CJ, Palmas W, Penninx BW, Petersmann A, Polasek O, Psaty BM, Rantner B, Rice K, Rivadeneira F, Rotter JI, Seldenrijk A, Stadler M, Summerer M, Tanaka T, Tybjaerg-Hansen A, Uitterlinden AG, van Gilst WH, Vermeulen SH, Wild SH, Wild PS, Willeit J, Zeller T,

Zemunik T, Zgaga L, Assimes TL, Blankenberg S, Boerwinkle E, Campbell H, Cooke JP, de Graaf J, Herrington D, Kardia SL, Mitchell BD, Murray A, Munzel T, Newman AB, Oostra BA, Rudan I, Shuldiner AR, Snieder H, van Duijn CM, Volker U, Wright AF, Wichmann HE, Wilson JF, Witteman JC, Liu Y, Hayward C, Borecki IB, Ziegler A, North KE, Cupples LA, Kronenberg F. Association between chromosome 9p21 variants and the ankle-brachial index identified by a meta-analysis of 21 genome-wide association studies. *Circulation. Cardiovascular genetics*. 2012;5:100-112

96. Johnson AD, Hwang SJ, Voorman A, Morrison A, Peloso GM, Hsu YH, Thanassoulis G, Newton-Cheh C, Rogers IS, Hoffmann U, Freedman JE, Fox CS, Psaty BM, Boerwinkle E, Cupples LA, O'Donnell CJ. Resequencing and clinical associations of the 9p21.3 region: A comprehensive investigation in the framingham heart study. *Circulation*. 2013;127:799-810
97. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K, Anderson JL, Ardissino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, Buysschaert I, Cardiogenics, Carlquist JF, Chen L, Cichon S, Codd V, Davies RW, Dedoussis G, Dehghan A, Demissie S, Devaney JM, Diemert P, Do R, Doering A, Eifert S, Mokhtari NE, Ellis SG, Elosua R, Engert JC, Epstein SE, de Faire U, Fischer M, Folsom AR, Freyer J, Gigante B, Girelli D, Gretarsdottir S, Gudnason V, Gulcher JR, Halperin E, Hammond N, Hazen SL, Hofman A, Horne BD, Illig T, Iribarren C, Jones GT, Jukema JW, Kaiser MA, Kaplan LM, Kastelein JJ, Khaw KT, Knowles JW, Kolovou G, Kong A, Laaksonen R, Lambrechts D, Leander K, Lettre G, Li M, Lieb W, Loley C, Lotery AJ, Mannucci PM, Maouche S, Martinelli N, McKeown PP, Meisinger C, Meitinger T, Melander O, Merlini PA, Mooser V, Morgan T, Muhleisen TW, Muhlestein JB, Munzel T, Musunuru K, Nahrstaedt J, Nelson CP, Nothen MM, Olivieri O, Patel RS, Patterson CC, Peters A, Peyvandi F, Qu L, Quyyumi AA, Rader DJ, Rallidis LS, Rice C, Rosendaal FR, Rubin D, Salomaa V, Sampietro ML, Sandhu MS, Schadt E, Schafer A, Schillert A, Schreiber S, Schrezenmeir J, Schwartz SM, Siscovick DS, Sivananthan M, Sivapalaratnam S, Smith A, Smith TB, Snoep JD, Soranzo N, Spertus JA, Stark K, Stirrups K, Stoll M, Tang WH, Tennstedt S, Thorgeirsson G, Thorleifsson G, Tomaszewski M, Uitterlinden AG, van Rij AM, Voight BF, Wareham NJ, Wells GA, Wichmann HE, Wild PS, Willenborg C, Witteman JC, Wright BJ, Ye S, Zeller T, Ziegler A, Cambien F, Goodall AH, Cupples LA, Quertermous T, Marz W, Hengstenberg C, Blankenberg S, Ouwehand WH, Hall AS, Deloukas P, Thompson JR, Stefansson K, Roberts R, Thorsteinsdottir U, O'Donnell CJ, McPherson R, Erdmann J, Consortium CA, Samani NJ. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature genetics*. 2011;43:333-338
98. Schork AJ, Thompson WK, Pham P, Torkamani A, Roddey JC, Sullivan PF, Kelsoe JR, O'Donovan MC, Furberg H, Tobacco, Genetics C, Bipolar Disorder Psychiatric Genomics C, Schizophrenia Psychiatric Genomics C, Schork NJ, Andreassen OA, Dale AM. All snps are not created equal: Genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated snps. *PLoS genetics*. 2013;9:e1003449
99. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C,

Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M. Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102:10604-10609

100. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekstrom TJ, Harris TB, Launer LJ, Eiriksdottir G, Leppert MF, Sapienza C, Gudnason V, Feinberg AP. Intra-individual change over time in DNA methylation with familial clustering. *JAMA*. 2008;299:2877-2883
101. Chen Z, Karaplis AC, Ackerman SL, Pogribny IP, Melnyk S, Lussier-Cacan S, Chen MF, Pai A, John SW, Smith RS, Bottiglieri T, Bagley P, Selhub J, Rudnicki MA, James SJ, Rozen R. Mice deficient in methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and decreased methylation capacity, with neuropathology and aortic lipid deposition. *Human molecular genetics*. 2001;10:433-443
102. Lund G, Andersson L, Lauria M, Lindholm M, Fraga MF, Villar-Garea A, Ballestar E, Esteller M, Zaina S. DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein e. *The Journal of biological chemistry*. 2004;279:29147-29154
103. Laukkanen MO, Mannermaa S, Hiltunen MO, Aittomaki S, Airenne K, Janne J, Yla-Herttuala S. Local hypomethylation in atherosclerosis found in rabbit ec-sod gene. *Arteriosclerosis, thrombosis, and vascular biology*. 1999;19:2171-2178
104. Post WS, Goldschmidt-Clermont PJ, Wilhide CC, Heldman AW, Sussman MS, Ouyang P, Milliken EE, Issa JP. Methylation of the estrogen receptor gene is associated with aging and atherosclerosis in the cardiovascular system. *Cardiovascular research*. 1999;43:985-991
105. Chan Y, Fish JE, D'Abreo C, Lin S, Robb GB, Teichert AM, Karantzoulis-Fegaras F, Keightley A, Steer BM, Marsden PA. The cell-specific expression of endothelial nitric-oxide synthase: A role for DNA methylation. *The Journal of biological chemistry*. 2004;279:35087-35100
106. Liu C, Xu D, Sjoberg J, Forsell P, Bjorkholm M, Claesson HE. Transcriptional regulation of 15-lipoxygenase expression by promoter methylation. *Experimental cell research*. 2004;297:61-67
107. Friso S, Lotto V, Choi SW, Girelli D, Pinotti M, Guarini P, Udali S, Pattini P, Pizzolo F, Martinelli N, Corrocher R, Bernardi F, Olivieri O. Promoter methylation in coagulation f7 gene influences plasma fvii concentrations and relates to coronary artery disease. *Journal of medical genetics*. 2012;49:192-199
108. Feldman D, Pike JW, Glorieux FH. *Vitamin d*. London, San Diego: Elsevier, Academic Press; 2005.
109. Shao A, Wood RJ, Fleet JC. Increased vitamin d receptor level enhances 1,25-dihydroxyvitamin d3-mediated gene expression and calcium transport in caco-2 cells. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. 2001;16:615-624
110. Nibbelink KA, Tishkoff DX, Hershey SD, Rahman A, Simpson RU. 1,25(oh)2-vitamin d3 actions on cell proliferation, size, gene expression, and receptor localization, in the hl-1 cardiac myocyte. *The Journal of steroid biochemistry and molecular biology*. 2007;103:533-537

111. Wu-Wong JR, Nakane M, Ma J, Ruan X, Kroeger PE. Effects of vitamin d analogs on gene expression profiling in human coronary artery smooth muscle cells. *Atherosclerosis*. 2006;186:20-28
112. Merke J, Milde P, Lewicka S, Hugel U, Klaus G, Mangelsdorf DJ, Haussler MR, Rauterberg EW, Ritz E. Identification and regulation of 1,25-dihydroxyvitamin d3 receptor activity and biosynthesis of 1,25-dihydroxyvitamin d3. Studies in cultured bovine aortic endothelial cells and human dermal capillaries. *The Journal of clinical investigation*. 1989;83:1903-1915
113. Chen S, Glenn DJ, Ni W, Grigsby CL, Olsen K, Nishimoto M, Law CS, Gardner DG. Expression of the vitamin d receptor is increased in the hypertrophic heart. *Hypertension*. 2008;52:1106-1112
114. Tishkoff DX, Nibbelink KA, Holmberg KH, Dandu L, Simpson RU. Functional vitamin d receptor (vdr) in the t-tubules of cardiac myocytes: Vdr knockout cardiomyocyte contractility. *Endocrinology*. 2008;149:558-564
115. Chen S, Law CS, Grigsby CL, Olsen K, Hong TT, Zhang Y, Yeghiazarians Y, Gardner DG. Cardiomyocyte-specific deletion of the vitamin d receptor gene results in cardiac hypertrophy. *Circulation*. 2011;124:1838-1847
116. Andrukhova O, Slavic S, Zeitz U, Riesen SC, Heppelmann MS, Ambrisko TD, Markovic M, Kuebler WM, Erben RG. Vitamin d is a regulator of endothelial nitric oxide synthase and arterial stiffness in mice. *Molecular Endocrinology*. 2014;28:53-64
117. Ellam T, Hameed A, ul Haque R, Muthana M, Wilkie M, Francis SE, Chico TJ. Vitamin d deficiency and exogenous vitamin d excess similarly increase diffuse atherosclerotic calcification in apolipoprotein e knockout mice. *PloS one*. 2014;9:e88767
118. Falk E. Pathogenesis of atherosclerosis. *Detection of Vulnerable Plaques*. 2006;47:C7-C12
119. Giallauria F, Milaneschi Y, Tanaka T, Maggio M, Canepa M, Elango P, Vigorito C, Lakatta EG, Ferrucci L, Strait J. Arterial stiffness and vitamin d levels: The baltimore longitudinal study of aging. *The Journal of Clinical Endocrinology & Metabolism*. 2012;97:3717-3723
120. Cozzolino M, Stucchi A, Rizzo MA, Soldati L, Cusi D, Ciceri P, Brenna I, Elli F, Gallieni M. Reprint of: Vitamin d receptor activation and prevention of arterial ageing. *Proceedings and abstracts from the 8th International Symposium on Amino Acid and Protein Metabolism in Health and Disease*. 2013;23, Supplement 1:S31-S36
121. Goel RK, Lal H. Role of vitamin d supplementation in hypertension. *Indian journal of clinical biochemistry : IJCB*. 2011;26:88-90
122. Yang L, Ma J, Zhang X, Fan Y, Wang L. Protective role of the vitamin d receptor. *Cellular immunology*. 2012;279:160-166
123. Fall T, Shiue I, Bergea af Geijerstam P, Sundstrom J, Arnlov J, Larsson A, Melhus H, Lind L, Ingelsson E. Relations of circulating vitamin d concentrations with left ventricular geometry and function. *European journal of heart failure*. 2012;14:985-991
124. Amin A, Minaee S, Chitsazan M, Naderi N, Taghavi S, Ardeshiri M. Can vitamin d supplementation improve the severity of congestive heart failure? *Congestive heart failure*. 2013;19:E22-28
125. Wang TJ, Pencina MJ, Booth SL, Jacques PF, Ingelsson E, Lanier K, Benjamin EJ, D'Agostino RB, Wolf M, Vasan RS. Vitamin d deficiency and risk of cardiovascular disease. *Circulation*. 2008;117:503-511

126. Judd SE, Tangpricha V. Vitamin d deficiency and risk for cardiovascular disease. *The American journal of the medical sciences*. 2009;338:40-44
127. Siadat ZD, Kiani K, Sadeghi M, Shariat AS, Farajzadegan Z, Kheirmand M. Association of vitamin d deficiency and coronary artery disease with cardiovascular risk factors. *Journal of research in medical sciences : the official journal of Isfahan University of Medical Sciences*. 2012;17:1052-1055
128. Motiwala SR, Wang TJ. Vitamin d and cardiovascular risk. *Current hypertension reports*. 2012;14:209-218
129. Kim DH, Sabour S, Sagar UN, Adams S, Whellan DJ. Prevalence of hypovitaminosis d in cardiovascular diseases (from the national health and nutrition examination survey 2001 to 2004). *The American journal of cardiology*. 2008;102:1540-1544
130. Miyamoto K, Kesterson RA, Yamamoto H, Taketani Y, Nishiwaki E, Tatsumi S, Inoue Y, Morita K, Takeda E, Pike JW. Structural organization of the human vitamin d receptor chromosomal gene and its promoter. *Mol Endocrinol*. 1997;11:1165-1179
131. Crofts LA, Hancock MS, Morrison NA, Eisman JA. Multiple promoters direct the tissue-specific expression of novel n-terminal variant human vitamin d receptor gene transcripts. *Proceedings of the National Academy of Sciences of the United States of America*. 1998;95:10529-10534
132. Lins TC, Vieira RG, Grattapaglia D, Pereira RW. Population analysis of vitamin d receptor polymorphisms and the role of genetic ancestry in an admixed population. *Genetics and molecular biology*. 2011;34:377-385
133. Integrated science assessment for lead. 2012
134. Pirkle JL, Brody DJ, Gunter EW, Kramer RA, Paschal DC, Flegal KM, Matte TD. The decline in blood lead levels in the united states. The national health and nutrition examination surveys (nhanes). *JAMA*. 1994;272:284-291
135. Mahaffey KR, Annest JL, Roberts J, Murphy RS. National estimates of blood lead levels: United states, 1976-1980: Association with selected demographic and socioeconomic factors. *The New England journal of medicine*. 1982;307:573-579
136. Bandeen-Roche K, Glass TA, Bolla KI, Todd AC, Schwartz BS. Cumulative lead dose and cognitive function in older adults. *Epidemiology*. 2009;20:831-839
137. Martin D, Glass TA, Bandeen-Roche K, Todd AC, Shi W, Schwartz BS. Association of blood lead and tibia lead with blood pressure and hypertension in a community sample of older adults. *American Journal of Epidemiology*. 2006;163:467-478
138. Barry PS. A comparison of concentrations of lead in human tissues. *British journal of industrial medicine*. 1975;32:119-139
139. Hu H, Shih R, Rothenberg S, Schwartz BS. The epidemiology of lead toxicity in adults: Measuring dose and consideration of other methodologic issues. *Environ Health Perspect*. 2007;115:455-462
140. Cheng Y, Schwartz J, Sparrow D, Aro A, Weiss ST, Hu H. Bone lead and blood lead levels in relation to baseline blood pressure and the prospective development of hypertension the normative aging study. *American Journal of Epidemiology*. 2001;153:164-171
141. Elmarsafawy SF, Jain NB, Schwartz J, Sparrow D, Nie H, Hu H. Dietary calcium as a potential modifier of the relationship of lead burden to blood pressure. *Epidemiology*. 2006;17:531-537

142. Yazbeck C, Thiebaugeorges O, Moreau T, Goua V, Debotte G, Sahuquillo J, Forhan A, Foliguet B, Magnin G, Slama R, Charles MA, Huel G. Maternal blood lead levels and the risk of pregnancy-induced hypertension: The eden cohort study. *Environmental health perspectives*. 2009;117:1526-1530
143. Wells EM, Navas-Acien A, Herbstman JB, Apelberg BJ, Silbergeld EK, Caldwell KL, Jones RL, Halden RU, Witter FR, Goldman LR. Low-level lead exposure and elevations in blood pressure during pregnancy. *Environmental health perspectives*. 2011;119:664-669
144. Mitchell GF, Hwang S-J, Vasan RS, Larson MG, Pencina MJ, Hamburg NM, Vita JA, Levy D, Benjamin EJ. Arterial stiffness and cardiovascular events: The framingham heart study. *Circulation*. 2010;121:505-511
145. Perlstein T, Weuve J, Schwartz J, Sparrow D, Wright R, Litonjua A, Nie H, Hu H. Cumulative community-level lead exposure and pulse pressure: The normative aging study. *Environmental health perspectives*. 2007;115:1696-1700
146. Zhang A, Park SK, Wright RO, Weisskopf MG, Mukherjee B, Nie H, Sparrow D, Hu H. Hfe h63d polymorphism as a modifier of the effect of cumulative lead exposure on pulse pressure: The normative aging study. *Environmental health perspectives*. 2010;118:1261-1266
147. Rosen JF, Pounds JG. Quantitative interactions between pb2+ and ca2+ homeostasis in cultured osteoclastic bone cells. *Toxicology and applied pharmacology*. 1989;98:530-543
148. Pounds JG. Effect of lead intoxication on calcium homeostasis and calcium-mediated cell function: A review. *Neurotoxicology*. 1984;5:295-331
149. Onalaja AO, Claudio L. Genetic susceptibility to lead poisoning. *Environmental health perspectives*. 2000;108 Suppl 1:23-28
150. Schwartz BS, Stewart WF, Kelsey KT, Simon D, Park S, Links JM, Todd AC. Associations of tibial lead levels with bsmi polymorphisms in the vitamin d receptor in former organolead manufacturing workers. *Environmental health perspectives*. 2000;108:199-203
151. Krieg Jr EF, Butler MA, Chang M-h, Liu T, Yesupriya A, Dowling N, Lindegren ML. Lead and cognitive function in vdr genotypes in the third national health and nutrition examination survey. *Neurotoxicology and teratology*. 2010;32:262-272
152. Weaver VM, Lee B-K, Todd AC, Ahn K-D, Shi W, Jaar BG, Kelsey KT, Lustberg ME, Silbergeld EK, Parsons PJ, Wen J, Schwartz BS. Effect modification by δ-aminolevulinic acid dehydratase, vitamin d receptor, and nitric oxide synthase gene polymorphisms on associations between patella lead and renal function in lead workers. *Environmental research*. 2006;102:61-69
153. Greenland S. An introduction to instrumental variables for epidemiologists. *International journal of epidemiology*. 2000;29:722-729
154. Smith GD, Ebrahim S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology*. 2003;32:1-22
155. Smith GD, Ebrahim S. Mendelian randomization: Prospects, potentials, and limitations. *International journal of epidemiology*. 2004;33:30-42
156. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for mendelian randomization studies using multiple genetic variants. *International journal of epidemiology*. 2011;40:740-752

157. Davey Smith G, Hemani G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Human molecular genetics*. 2014;23:R89-98
158. Relton CL, Davey Smith G. Two-step epigenetic mendelian randomization: A strategy for establishing the causal role of epigenetic processes in pathways to disease. *International journal of epidemiology*. 2012;41:161-176
159. Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature genetics*. 2003;33 Suppl:245-254
160. Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, Tylavsky FA, Conneely KN. Methylation quantitative trait loci (meqtls) are consistently detected across ancestry, developmental stage, and tissue type. *BMC genomics*. 2014;15:145
161. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27k discovery and replication. *American journal of human genetics*. 2011;88:450-457
162. Philibert RA, Plume JM, Gibbons FX, Brody GH, Beach SR. The impact of recent alcohol use on genome wide DNA methylation signatures. *Frontiers in genetics*. 2012;3:54
163. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, Belvisi MG, Brown R, Vineis P, Flanagan JM. Epigenome-wide association study in the european prospective investigation into cancer and nutrition (epic-turin) identifies novel genetic loci associated with smoking. *Human molecular genetics*. 2013;22:843-851
164. Zhao R, Zhang R, Li W, Liao Y, Tang J, Miao Q, Hao W. Genome-wide DNA methylation patterns in discordant sib pairs with alcohol dependence. *Asia-Pacific psychiatry : official journal of the Pacific Rim College of Psychiatrists*. 2013;5:39-50
165. Burgess S, Daniel RM, Butterworth AS, Thompson SG, the E-IC. Network mendelian randomization: Using genetic variants as instrumental variables to investigate mediation in causal pathways. *International journal of epidemiology*. 2014
166. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992;3:143-155
167. Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Stat Sci*. 2010;25:51-71
168. Mozaffarian D, Wilson PWF, Kannel WB. Beyond established and novel risk factors: Lifestyle risk factors for cardiovascular disease. *Circulation*. 2008;117:3031-3038
169. Safar ME, Levy BI, Struijker-Boudier H. Current perspectives on arterial stiffness and pulse pressure in hypertension and cardiovascular diseases. *Circulation*. 2003;107:2864-2869
170. Nucifora G, Bax JJ, van Werkhoven JM, Boogers MJ, Schuijf JD. Coronary artery calcium scoring in cardiovascular risk assessment. *Cardiovascular therapeutics*. 2011;29:e43-53
171. Hutter CM, Mechanic LE, Chatterjee N, Kraft P, Gillanders EM, Tank NCIG-ET. Gene-environment interactions in cancer epidemiology: A national cancer institute think tank report. *Genetic epidemiology*. 2013;37:643-657
172. Tammen SA, Friso S, Choi SW. Epigenetics: The link between nature and nurture. *Molecular aspects of medicine*. 2013;34:753-764
173. van Vliet J, Oates NA, Whitelaw E. Epigenetic mechanisms in the context of complex diseases. *Cellular and molecular life sciences : CMLS*. 2007;64:1531-1538

174. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature*. 2007;447:433-440
175. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nature reviews. Genetics*. 2007;8:253-262
176. Abdolmaleky HM, Smith CL, Faraone SV, Shafa R, Stone W, Glatt SJ, Tsuang MT. Methylomics in psychiatry: Modulation of gene-environment interactions may be through DNA methylation. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 2004;127B:51-59
177. Lu Q, Qiu X, Hu N, Wen H, Su Y, Richardson BC. Epigenetics, disease, and therapeutic interventions. *Ageing research reviews*. 2006;5:449-467
178. Foley DL, Craig JM, Morley R, Olsson CA, Dwyer T, Smith K, Saffery R. Prospects for epigenetic epidemiology. *Am J Epidemiol*. 2009;169:389-400
179. Udali S, Guarini P, Moruzzi S, Choi SW, Friso S. Cardiovascular epigenetics: From DNA methylation to microrna s. *Molecular aspects of medicine*. 2013;34:883-901

Chapter 2

Effect modification by Vitamin D receptor genetic polymorphisms in the association between cumulative lead exposure and pulse pressure: a longitudinal study

Introduction

Lead (Pb) is known to increase the risk of cardiovascular diseases (CVD). A recent expert review by the National Toxicology Program (NTP) concluded that there is sufficient evidence that long-term cumulative exposure, even at low level exposure, which can be assessed by bone lead levels used in our study, is associated with elevated blood pressure.[1] Several studies have identified a positive association between lead exposure and high blood pressure, which is a major risk factor for CVD. Martin et al[2] found an association between blood lead levels and increases in blood pressure, as well as an association between tibia bone lead levels and hypertension in the Baltimore Memory Study. Cheng et al[3] reported an increased risk of incident hypertension with higher levels of lead in tibia and patella bones in the Normative Aging Study (NAS). Another study, which used the NAS data, found an association between high lead levels in bone and blood and hypertension among subjects with low dietary calcium intake.[4] Blood lead levels have also been associated with increased risk of pregnancy-induced hypertension.[5] The association between lead exposure and the risk of hypertension in pregnant women was also demonstrated even in those with blood lead levels less than 2 μg/dL.[6] Lead exposure may also induce endothelial injury and atherosclerosis.[7-9] The study by Perlstein et al[10] suggests that lead

accumulation may contribute to the increase in pulse pressure, a measure of arterial stiffness, and with clinical cardiovascular events. Zhang et al[11] also reported a deleterious impact of cumulative lead on pulse pressure with effect modification by hemochromatosis genetic polymorphism in the NAS.

Vitamin D receptor (VDR) is involved in lead absorption and accumulation.[12] Vitamin D plays an important role in calcium metabolism, which is shared by lead.[13] Calcitriol, an active hormonal form of vitamin D, stimulates increased calcium absorption from the gut during calcium deficiency. Calcitriol also stimulates the expression of genes engaged in absorption of calcium in the intestine. The effects of vitamin D and calcitriol are mediated by their interaction with the VDR. The complex of calcitriol and VDR acts as a transcription factor regulating the gene expression of calcium-binding receptors. Because lead is a divalent cation, lead competes with calcium to bind to calcium-binding receptors.[14] When calcium levels are low, the calcium-binding protein may bind lead instead of calcium, resulting in elevated absorption of lead.[15] The *VDR* genetic variants have been identified as potential genetic factors that can influence the absorption, retention and accumulation of lead in the human body.[12] Schwartz et al. examined former organolead manufacturing workers and found that *Bsm1* variant on the *VDR* gene modifies the association between age and tibia bone lead levels.[16] In this study, the cumulative level of lead in bone and the rate of reabsorption and excretion of lead over time were higher for those with the *VDR Bsm1* variant. This study suggests that the *VDR* genetic variant may play a role in susceptibility to lead accumulation.

Vitamin D and VDR are involved in arterial stiffness and arterial aging.[17, 18] Vitamin D has been demonstrated to regulate endothelial nitric oxide synthase and arterial stiffness in a mouse study.[19] Lower serum vitamin D_3 levels are known to be associated with hypertension.[20]

VDR is also involved in the renin-angiotensin system, cell proliferation and differentiation, anti-inflammation, and anti-fibrosis.[21] The effect of VDR on immune response and inflammation has been related to atherosclerosis.[22]

There have been studies examining effect modification of lead by *VDR* in relation to several diseases.[23, 24] However, effect modification by *VDR* genotype in the association between lead and subclinical CVD measures, including arterial stiffness in longitudinal settings, has not yet been examined. In this study, we investigate effect modification by the *VDR* gene in the association between cumulative lead exposure measured by bone lead levels and pulse pressure, a marker of arterial stiffness.

Methods

Study population

The Normative Aging Study (NAS) is a longitudinal study of the aging process established by the Veterans Administration in 1963 at the VA Outpatient Clinic in Boston, Massachusetts. The participants were 2,280 mostly White men aged 21 to 80 years with no past or present known chronic conditions (heart disease, cancer, recurrent asthma, sinusitis, bronchitis, diabetes, gout, peptic ulcer, or hypertension) [25]. The NAS followed up on the participants every 3 to 5 years.

The NAS participants were invited to obtain bone lead measurements between 1991 and 1999 at the Ambulatory Clinical Research Center of the Brigham and Women's Hospital in Boston, Massachusetts (N=866). Of the participants with bone lead measurements, nine participants were excluded due to unreliable bone lead measurements (detailed description in the lead exposure section). Of 857 participants, 727 participants were successfully genotyped for at least one single nucleotide polymorphisms (SNP) on the *VDR* gene including *Bsm1* (rs1544410),

Taq1 (rs731236), *Apa1* (rs7975232), and *Fok1* (rs10735810). The present analysis includes pulse pressure measured at the time of bone lead measurement (baseline, 1991-1999) and follow-up data through June 22, 2011. A total of 3,100 observations (727 participants) were used in this study. Each participating institute's institutional review board approved this study and written informed consent was collected from each participant.

Blood pressure

The participants visited the study center in the morning. The participants were asked not to smoke or drink for at least 12 hours before a visit. Seating systolic blood pressure (SBP) and fifth-phase diastolic blood pressure (DBP) were measured to the nearest 2 mmHg. Blood pressures were measured in the left arm and then in the right arm with a standard mercury sphygmomanometer with a 14-cm cuff. The mean measurements of the left and right arms were used in this study. Pulse pressure was calculated as the difference between SBP and DBP.

Lead exposure

Bone lead measurement is used as an index of cumulative lead exposure levels. Tibia (the mid-shaft of the left tibia, cortical) and patella (the left patella, trabecular) bone lead levels were measured using a K-x-ray fluorescence instrument (KXRF) (ABIOMED, Danvers, MA).[26] Participants with higher than 10 μg/g or 15 μg/g uncertainty of tibia or patella bone lead levels, respectively, (reflecting precision of the estimates) were excluded (n=9). More details were published in a previous study.[27]

Genotyping

Multiplex polymerase chain reaction assays were designed with Sequenom Spectro DESIGNER software (Sequenon, Inc, San Diego, CA) by inputting sequences containing the SNP site and

100 base pairs of flanking sequence on either side of the SNP. In the *VDR* gene, four single nucleotide polymorphisms (SNP) including *Bsm1* (rs1544410), *Taq1* (rs731236), *Apa1* (rs7975232), and *Fok1* (rs10735810) were genotyped. More details on genotyping were provided in a previous study.[27]

Statistical analysis

When the data were investigated without assuming any inheritance models for the SNPs, we observed that a dominant inheritance model (in terms of minor frequency allele) fits the data best (data not shown). For dominant model, genotype was coded as 0 if a subject had no minor frequency allele (ancestral type), or was coded as 1 if a subject had one or two minor frequency alleles (variant type). The study population was partitioned based on the genotypes and compared with regard to baseline characteristics. For each of the four SNPs, the Hardy-Weinberg Equilibrium was checked for an evidence of inbreeding, population stratification, and problems in genotyping.

To account for the repeated measurements on pulse pressure and covariates, linear mixed effects models with random intercepts were implemented. We decomposed age at examination into age-at-baseline and time-since-baseline to capture the baseline age effect as well as the longitudinal aging effect.[28] The main effects of bone lead levels and *VDR* genotype, and their interaction term were fitted adjusting for time-since-baseline and an interaction term between time-since-baseline and bone lead levels in addition to the following covariates: age at baseline, race (White or not), body mass index (BMI), smoking (pack-years), alcohol intake (two or more drinks/day; Yes/No), calcium intake from food (calcium (mg) /day), diabetes status (Yes/No), antihypertensive medication status (Yes/No), family history of hypertension (Yes/No), education (less than high school, high school, some college, or four year college or more), and age at

baseline by *VDR* genotype interaction. The interaction between time-since-baseline and bone lead would capture different trajectories of pulse pressure over time in relation to bone lead levels. A three-way interaction among time-since-baseline, bone lead and *VDR* genetic polymorphism was initially considered but not included in our final model because it was almost null, suggesting that the lead by *VDR* genotype interaction did not change over time and that the lead by time-since-baseline interaction did not differ between the *VDR* genotypes. The mixed model we used is described as follows:

$Y_{ij} \sim \beta_0 + \beta_1 \cdot lead + \beta_2 \cdot SNP + \beta_3 \cdot lead * SNP + \beta_4 \cdot time + \beta_5 \cdot lead * time + covariates + u_i + \epsilon_{ij}$ $(Equation\ 1)$,

where Y_{ij} is pulse pressure of subject *i* at time *j*, β_0 is a fixed intercept, $\beta_1 - \beta_5$ are fixed coefficients representing the estimated effects of each following term, u_i is the random intercept that reflects unexplained subject to subject heterogeneity that induces correlation among observations from the same subject, and ϵ_{ij} is a random error.

To interpret longitudinal associations of pulse pressure with tibia levels, we computed the predicted values of pulse pressure from the model with the continuous tibia lead variable using the tibia bone lead values at the 25th percentile and the 75th percentile of the distribution at the baseline (time=0) and after 10 years of follow-up (time=10) for the ancestral type and variant type, with all covariates held constant at the mean (continuous variables) or zero (categorical variables).

Results

All the four SNPs on the *VDR* gene, *Bsm1, Taq1, Apa1,* and *Fok1*, are common SNPs with a minor allele frequency range of 0.37 to 0.45 (**Table 2.1**). All SNPs were in Hardy-Weinberg equilibrium. The *Bsm1* is in a strong linkage disequilibrium ($r^2 = 0.92$) with *Taq1* in this study

population but not with *Apa1* ($r^2 = 0.54$) or *Fok1* ($r^2 = 0.001$). Genotypes missing rates were low (range: 1 - 6%). Among 727 participants, 442 participants (61%) had at least one copy of the *Bsm1* minor frequency allele (bb or Bb, variant type) and 238 participants (33%) were homozygous for the major frequency allele (BB, ancestral type). Forty seven subjects (6%) had a missing *Bsm1* genotype.

The mean age at baseline of the study population was 66 years (range: 48-93 years). The participants were followed for up to 20 years. The median follow-up period was longer for subjects who had at least one copy of the *Bsm1* minor frequency allele than for subjects who did not have the minor frequency allele (12 years vs. 9 years, **Table 2.2**). The number of follow-up examinations ranged from 1 to 8 with a median of 4. More than half of the participants were examined at least 5 times over 10 years.

The subjects with a missing *Bsm1* genotype were not substantially different from the remaining subjects for the baseline characteristics including age, race, BMI, SBP, DBP, pulse pressure, and blood pressure control medication status (**Table 2.2**). Tibia and patella bone lead levels, and the proportion of subjects with a family history of hypertension were slightly lower for those with the missing genotype. In contrast, calcium, sodium, potassium, and alcohol intake were slightly higher among subjects with the missing *Bsm1* genotype. The baseline characteristics were similar between participants with ancestral type and those with variant type. Participants with ancestral type smoked more, had a higher DBP, and had slightly lower prevalence of whites.

In longitudinal analyses, the association between bone lead levels and pulse pressure changed over time whereas the beta for the three-way interaction term of *VDR* genotype, bone lead and time-since-baseline was close to zero. This suggests that the association between bone

lead and pulse pressure changed over time but the difference in the association between bone lead and pulse pressure by *VDR* genotype was constant during the follow-up.

Table 2.2 shows the estimated regression coefficients and 95% confidence intervals for the association between bone lead and pulse pressure by *VDR* genotype at baseline (i.e., time is fixed at zero). With an interquartile range (IQR) increase in tibia lead (15µg/g), pulse pressure was 2.5 mmHg (95% CI: 0.4-4.7) greater for the participants with variant type on *Bsm1* compared with the participants with ancestral type (**Table 2.3**). With an IQR increase in patella lead (20µg/g), pulse pressure was 1.9 mmHg (95% CI: 0.1-3.8) greater for the participants with at least one copy of the minor frequency allele in *Bsm1* compared with the participants without the minor frequency allele (**Table 2.4**). Similar results were found for *Taq1*. The interaction effect was relatively smaller for *Apa1* and *Fok1*.

In Figure 2.1, we show the predicted values of pulse pressure from the model with the continuous tibia lead variable by *Bsm1* genotype at baseline (time=0) and after ten years (time=10). At baseline, as tibia bone lead level increased from 13µg/g (25^{th} percentile) to 28 µg/g (75% percentile), pulse pressure increased by 0.07 mmHg (an IQR increase from 50.68 mmHg to 50.75 mmHg) among subjects with *Bsm1* ancestral type, while pulse pressure increased by 2.6 mmHg (from 49.68 mmHg to 52.28 mmHg) among subjects with *Bsm1* variant type. After ten years, the marginal association (when the two lines were combined ignoring the *Bsm1* genotype) between tibia bone lead levels and pulse pressure became weaker. After 10 years of follow-up, as tibia bone lead level increased from 13µg/g (25^{th} percentile) to 28 µg/g (75% percentile), pulse pressure decreased by 1.9 mmHg (from 54.09 mmHg to 52.19 mmHg) among *Bsm1* ancestral type while pulse pressure increased by 0.63 mmHg (from 53.09 mmHg to 53.72 mmHg) among *Bsm1* variant type for the IQR change in tibia lead levels. In spite of the changing

association between bone lead levels and pulse pressure over time, the effect modifications by *VDR Bsm1* and *Taq1* genotypes on the association between bone lead levels and pulse pressure persists over time.

We also conducted cross-sectional analyses using baseline data as a sensitivity analysis. The same covariates except time-since-baseline and its interaction with lead exposure levels were examined. Additional sensitivity analyses were done in order to investigate other possible confounders. Verifying analytical consistency, the interaction effect of bone lead levels and the *VDR* genotype on pulse pressure were examined by (i) adding the square of baseline age, (ii) adding sodium (Na) and potassium (K) intake, (iii) adding heart rate (sitting, beats/min), (iv) adding triglyceride level, (v) adding high-density lipoprotein (HDL) and total cholesterol-to-HDL ratio, (vi) adding total calories (kcal/day) and physical activity (expended (fast walk adjusted), kcal/week), (vii) replacing BMI with height and waist circumference, (viii) replacing two alcohol drinks per day with grams per day of alcohol, (ix) replacing smoking in pack-years with current smoker or not, (x) ignoring calcium intake (obtained from a food frequency questionnaire), (xi) ignoring observations of subjects who are taking blood pressure control medication, and (xii) separating the blood pressure control medication variable into two variables: a calcium channel blocker and others.

We found slightly larger beta values in the cross-sectional baseline analyses (**Table 2.5**). The magnitudes of the beta estimates slightly varied by different covariate sets but the conclusions were consistent (data not shown).

Discussion

Using longitudinal observations, we found a stronger association between cumulative bone lead levels and pulse pressure in participants with at least one minor frequency allele on *Bsm1* or

Taq1. The results were consistent for tibia and patella bone lead levels. Our results were robust to cardiovascular risk related confounding factors in the sensitivity analyses. To our knowledge, this is the first study showing the interplay of *VDR* genetic polymorphisms and cumulative lead exposure levels on a CVD subclinical measure in the longitudinal setting.

Pulse pressure is a marker of arterial stiffness, and elevated pulse pressure is a risk factor of CVD. Pulse pressure has been demonstrated as an independent predictor of long-term cardiovascular mortality. A 10 mmHg increase in pulse pressure has been associated with a 10 to 25% increase in risk for CVD related mortality among older adults including coronary heart disease, congestive heart failure, and cerebrovascular disease.[29-31] In the present study, we found that individuals with at least one *VDR Bsm1* variant had a 2.5 mmHg greater pulse pressure in relation to every 15 μg/g increase in cumulative (tibia) lead exposure. We interpret this suggesting that individuals with *VDR Bsm1* at least one variant may have 2.5% to 6% greater risks for CVD mortality for every 15 μg/g increase in cumulative lead exposure.

The two genetic polymorphisms, that we found to interact with the cumulative lead exposure, are common SNPs with minor allele frequency over 0.4 in American Caucasians (**Table 2.1**). Over 65% of European descendants have at least one copy of these genetic polymorphisms, based on HapMap Phase 3 European ancestry samples,[32] suggesting that over 65% of European descendants may be susceptible to cumulative lead exposure related elevation in pulse pressure. Thus, the lead by *VDR* interaction found in this study may explain some portion of CVD events and mortality in the elderly.

Underlying biological mechanisms, by which the *VDR* genetic polymorphisms may modify the effect of lead on cardiovascular disease, are not well understood. Individuals with *VDR* genetic variants may have higher body burdens of lead, suggesting that the *VDR* gene may

modify the toxicokinetics of lead.[12, 24, 33, 34] In a study of Korean lead workers, blood lead levels and tibia bone lead levels were significantly higher in participants with the *VDR Bsm1* minor frequency allele.[35] Another study of former organolead manufacturing workers in the eastern United States suggested that the *VDR Bsm1* variant influences lead uptake and subsequent release of lead from bone.[16] The interaction between the VDR and lead can be explained by calcium metabolism. The VDR plays a crucial role in calcium absorption and metabolism, which is shared by lead from its divalent cation characteristic. Calcium deficiency was demonstrated to increase lead absorption in the gastrointestinal tract in chicks.[36, 37] Ingestion of lead inhibits the effect of vitamin D and its metabolites on intestinal calcium transport in rats.[38] In addition, associations between high bone and blood lead levels and hypertension were also found among subjects with low dietary calcium intake in the NAS.[4]

An interaction between the *VDR* gene and lead was also found to be involved in diverse disease development processes. In a study of the US general population, adults aged 60 years and older with the *VDR* rs2239185–rs731236 (*Taq1*) CC haplotype showed a negative association between blood lead and serum homocysteine, a risk factor of CVD and neurodegenerative disease, while those with the *VDR* rs2239185–rs731236 CT or TT showed a positive association.[23] In the same study, adults aged 20 to 59 years of age who had *VDR* rs2239185–rs731236 CC or TT haplotypes, showed significant decline in cognitive function with increased blood lead concentration while those CT haplotype did not show significant decline. In a study of lead and creatinine among Korean lead workers, the *VDR Bsm1* genotype was also found to modify the association between tibia bone lead and renal function as assessed by serum creatinine level and creatinine clearance.[24] The lead workers with at least one minor frequency allele on *VDR Bsm1* showed worse renal function with higher lead exposure

levels. Another study of the Korean lead workers demonstrated an effect modification of the *VDR Bsm1* genotype on the association between lead exposure and blood pressure.[39] Among lead workers with the *VDR Bsm1* variant, SBP were 2.7-3.7 mmHg higher and prevalence of hypertension was higher (OR =2.1). Thus, our results and those of others support the concept of interaction between the *VDR* genotypes and lead.

We observed a significant negative interaction between time-since-baseline and tibia bone lead levels in longitudinal analyses. The negative interaction can be interpreted in two different ways: the interaction effect between tibia bone lead and pulse pressure may decreases during the follow-up or the interaction effect between time-since-baseline and pulse pressure may decrease with higher lead exposure levels. A possible explanation for the decreasing effect of cumulative lead levels on pulse pressure during the follow-up is that people who had developed health related problems or diseases were more likely to drop out during follow-up.[40] Hence, the participants who stayed longer in the study may be healthier than those who dropped out. Another possible explanation is that other atherosclerotic risk factors became more pronounced as the study participants get older. As a result, the association between lead exposure and pulse pressure may seem attenuated over time. Alternatively, we have already reported that bone lead concentrations are falling in these participants, and more quickly for patella lead.[41] This suggests that the weaker associations between the baseline bone lead levels and the follow-up pulse pressures could reflect the lower exposure at later follow-up visits. In spite of the changing association between bone lead levels and pulse pressure over time, our main interest, the effect modifications by *VDR Bsm1* and *Taq1* genotypes on the association between bone lead levels and pulse pressure were consistent over time (**Figure 2.1**).

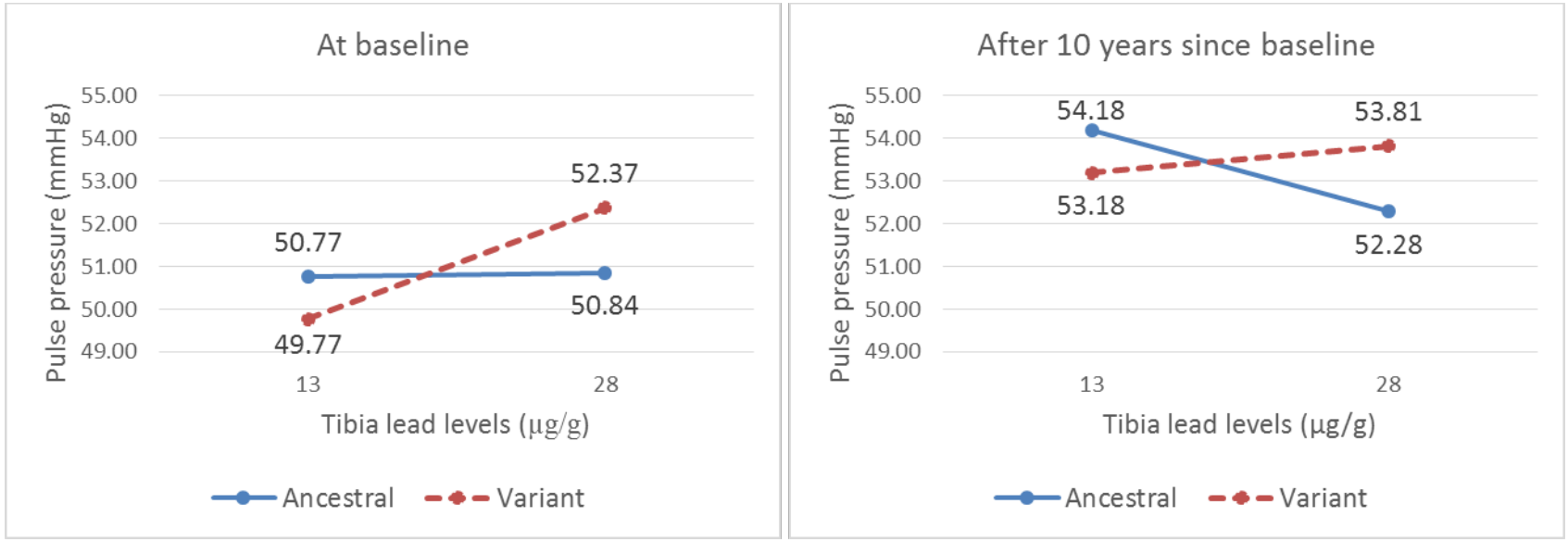

Figure 2.1 The predicted values of pulse pressure from the linear mixed model with the continuous tibia lead variable using the tibia bone lead values at the 25th percentile and the 75th percentile of the distribution at baseline (left) and after 10 years (right) for the ancestral type (solid line) and variant (dashed line), with all other covariates held constant at the mean (continuous variables: age at baseline (65yrs), BMI (28kg/m^2), and calcium intake from food (800mg/day)) or zero (categorical variables: race (white), smoking (non-smoker), alcohol intake (less than two drinks/day), diabetes status (no), family history of hypertension (no), and education (completed high school)).

We see consistent results with regard to *Bsm1* and *Taq1*. These SNPs are closely located to each other in the *VDR* gene (distance between *Bsm1* and *Taq1* = 1 kilo base pairs, size of the *VDR* gene = 63 kilo base pairs). The two SNPs are in high linkage disequilibrium ($r^2 = 0.92$). Among participants with at least one minor frequency allele on *Bsm1*, 98% had at least one minor frequency allele on *Taq1*. This explains why *Bsm1* and *Taq1* show similar effect modification signals. And these consistent results support that our finding is less likely to be a false positive from genotyping errors on *Bsm1* or *Taq1*. In the *VDR* gene, *Bsm1* is located in intron8, and *Taq1* is located in exon9. Genetic polymorphisms in intron regions, where splice enhances or silencers bind, can have an effect on alternative splicing.[42] More than 60% of the alternatively spliced variants in humans results in changes in the protein structure which may result in conformational changes.[43] However, whether these SNPs, *Bsm1* and *Taq1*, are the functional polymorphism with an effect on the structure of vitamin D receptor or on the affinity of the receptor, or whether they are in high linkage disequilibrium with some other functional SNPs is unclear. It should be further investigated in animal studies or *in vitro* studies.

The strengths of the current study include reliable bone lead measurements and extensive follow-ups up to 20 years (median follow-up of 9 years and median number of follow-up examinations of 4). The repeated measurements in pulse pressure and covariates increase statistical power to detect the gene by environment interactions. However, the NAS is an older cohort of predominantly white male participants. Hence, the findings may not be generalizable to women, younger individuals, and other ethnicities.

Conclusion

Lead toxicity is found in almost every system in the body. Over 90% of the total body lead burden in adults is accumulated in the bone and only about 1% is found in the blood.[44] Even

though blood lead levels have been gradually reduced since the phase-out of leaded gasoline in the 1970s, cumulative lead levels are still substantial in the elderly and associated with diverse diseases including cognitive function decline and hypertension.[2, 45]. Our finding suggests that subjects with the minor frequency alleles of *VDR Bsm1* or *Taq1* may be more susceptible to cumulative lead exposure-related elevated pulse pressure. If the *VDR* gene is involved in the association between lead and pulse pressure, calcium metabolism may play an important role in lead toxicity in CVD. The observed interaction between cumulative lead levels and the *VDR Bsm1* or *Taq1* persists over time during the follow-up. This implies that the elderly experience adverse effects from their early lead exposure, even though current ambient lead levels are low. Our findings suggest the importance of restricting early exposure to lead to avoid its persistent adverse health effects in the subjects' later life.

Tables

Table 2.1 Characteristics of genetic markers in *VDR* gene.

SNP	Missing	Number of minor frequency allele		MAF	HWE Statistics	HWE P-value
		0	1 or 2			
Bsm1	47 (6%)	238 (33%)	442 (61%)	0.42	1.22	0.27
Taq1	4 (1%)	242 (33%)	481 (66%)	0.41	0.76	0.38
Apa1	7 (1%)	211 (29%)	509 (70%)	0.45	0.64	0.42
Fok1	22 (3%)	274 (38%)	432 (59%)	0.37	0.25	0.62

SNP: Single Nucleotide Polymorphism; MAF: Minor Allele Frequency; HWE: Hardy-Weinberg Equilibrium.

Table 2.2 Characteristics of study population by the number of genetic variants at baseline.

Characteristics		**Missing**	**No. of minor frequency allele on *Bsm1***		
	All	***Bsm1***	**0**	**1**	**2**
Number of subjects	727	47 (6%)	238 (33%)	316 (43%)	126 (17%)
Follow-up (years, mean±SD)	10.6±5.5	9.6±6.2	9.9±5.8	11.1±5.2	11.3±5.1
No. of follow-up exams (median (Q1-Q3))	4 (3-6)	4 (2-6)	4 (2-6)	4.5 (3-6)	4.5 (3-6)
Continuous variables (mean±SD)					
Age at baseline (years)	66.4±7.2	66.8±8.9	66.7±7.1	66.1±7.1	66.6±6.9
Height (m)	1.7±0.1	1.7±0.1	1.7±0.1	1.7±0.1	1.7±0.1
Waist circumference (cm)	984.3±94.1	97.6±9.2	98.9±9.4	985.1±92.8	976.4±98.3
Body Mass Index (kg/m^2)	27.9±3.7	27.4±3.5	28.2±3.9	27.9±3.5	27.6±3.7
High-density lipoprotein (HDL) cholesterol (mg/dL)	47.8±12.4	47.4±13.2	48.7±12.3	47.5±12.7	47.3±11.5
Total cholesterol-to-HDL ratio	5.1±1.5	5.3±1.9	4.9±1.3	5.1±1.6	5.0±1.3
Smoking (pack-years)	21.0±25.1	21.5±24.9	24.5±28.2	19.6±23.5	17.7±22.2
Calcium intake (mg/day)	806±404	916±425	780±388	817±401	789±427
Sodium intake (mg/day)	3855±1841	4022±1673	3770±1700	3970±2093	3658±1407
Potassium intake (mg/day)	3363±1386	3895±1413	3212±1232	3467±1568	3191±1067
Total calories intake (kcal/day)	1992±637	2208±770	1936±651	2015±615	1955±599
Physical activity (kcal/week)	2002±1788	1963±1743	1879±1755	2131±1833	1921±1752
Systolic blood pressure (mmHg)	136.0±17.3	135.5±18.2	137.6±17.9	135.2±17.8	135.3±14.4

Diastolic blood pressure (mmHg)	81.7±9.6	80.7±10.6	83.2±9.6	81.3±9.3	80.3±9.4
Pulse pressure (mmHg)	54.3±14.7	54.8±15.9	54.5±14.5	53.8±15.3	55.0±12.8
Tibia lead level (μg/g)	21.2±13.2	20.4±14.5	22.2±13.3	20.6±13.2	21.2±12.8
Patella lead level (μg/g)	30.5±19.3	27.3±15.9	31.9±21.4	29.3±17.8	32.0±19.9
Blood lead level (μg/dL)	5.9±3.9	6.0±4.4	6.5±4.1	5.7±3.7	5.6±3.7
Categorical variables (n(%))					
Race (white or not)	703 (97%)	45 (96%)	227 (95%)	310 (96%)	121 (96%)
Alcohol (two or more drinks/day)	148 (20%)	7 (15%)	42 (18%)	76 (24%)	23 (18%)
Diabetes (diagnosed or taking medication)	94 (13%)	7 (15%)	28 (12%)	39 (12%)	20 (16%)
Antihypertensive medication	128 (18%)	8 (17%)	44 (18%)	54 (17%)	22 (17%)
Family history of hypertension	438 (60%)	21 (45%)	141 (59%)	201 (64%)	75 (60%)
Education: Less than high school	70 (10%)	1 (2%)	31 (13%)	19 (6%)	19 (15%)
Complete high school	248 (34%)	14 (30%)	87 (37%)	106 (34%)	41 (33%)
Some college	179 (25%)	15 (32%)	53 (22%)	81 (26%)	30 (24%)
College or more	204 (28%)	16 (34%)	62 (26%)	97 (31%)	29 (23%)

Table 2.3 Adjusted changes in pulse pressure (mmHg) with an IQR (15µg/g) increase in tibia lead levels.

SNP	N	Ancestral vs. Variant β (95% CI)*	Interaction term β (95% CI)*	P
Bsm1	816	Ancestral 0.1 (-1.8, 1.9)	2.5 (0.4, 4.7)	0.02
	1626	Variant 2.6 (1.2, 4.0)		
Taq1	827	Ancestral 0.4 (-1.4, 2.2)	2.0 (-0.1, 4.1)	0.06
	1760	Variant 2.4 (1.1, 3.8)		
Apa1	795	Ancestral 1.5 (-0.4, 3.4)	0.3 (-1.9, 2.4)	0.81
	1787	Variant 1.8 (0.4, 3.1)		
Fok1	974	Ancestral 2.1 (0.5, 3.7)	-0.02 (-2.0, 2.0)	0.99
	1548	Variant 2.1 (0.6, 3.6)		

SNP: Single Nucleotide Polymorphism; N: Number of observations; IQR: Inter-quartile range; P: p-value of the interaction term; Ancestral: major frequency allele homozygotes; Variant: minor frequency allele homozygotes and heterozygotes.

*To compute effect estimates from longitudinal models, the time term was fixed at zero.

Table 2.4 Adjusted changes in pulse pressure (mmHg) with an IQR (20µg/g) increase in patella lead levels.

SNP	N	Ancestral vs. Variant β (95% CI)*	Interaction term β (95% CI)*	P
Bsm1	811	Ancestral 0.0 (-1.6, 1.5)	1.9 (0.1, 3.8)	0.04
	1619	Variant 1.9 (0.5, 3.2)		
Taq1	822	Ancestral -0.1 (-1.6, 1.4)	2.0 (0.2, 3.8)	0.03
	1753	Variant 2.0 (0.7, 3.3)		
Apa1	789	Ancestral -0.1 (-1.8, 1.7)	1.7 (-0.3, 3.6)	0.09
	1781	Variant 1.6 (0.4, 2.8)		
Fok1	965	Ancestral 1.4 (-0.1, 2.9)	0.3 (-1.5, 2.2)	0.72
	1545	Variant 1.7 (0.3, 3.1)		

SNP: Single Nucleotide Polymorphism; N: Number of observations; IQR: Inter-quartile range; P: p-value of the interaction term; Ancestral: major frequency allele homozygotes; Variant: minor frequency allele homozygotes and heterozygotes.

*To compute effect estimates from longitudinal models, the time term was fixed at zero.

Table 2.5 Adjusted changes in pulse pressure (mmHg) with at least one minor allele in VDR gene per IQR increase in bone lead marker using baseline data.

	Tibia lead IQR = 15µg/g				Patella lead IQR = 20µg/g			
		Wild vs. variant	Interaction term			Wild vs. variant	Interaction term	
SNP	**N**	**β (95% CI)**	**β (95% CI)**	**P**	**N**	**β (95% CI)**	**β (95% CI)**	**P**
Bsm1	612	Ancestral -0.5 (-2.7, 1.7) Variant 2.6 (0.9, 4.2)	3.0 (0.3, 5.8)	0.03	608	Ancestral 0.5 (-1.4, 2.4) Variant 1.5 (-0.1, 3.1)	2.0 (-0.4, 4.4)	0.10
Taq1	649	Ancestral -0.1 (-2.3, 2.1) Variant 2.3 (0.7, 3.9)	2.4 (-0.3, 5.1)	0.08	645	Ancestral -0.6 (-2.4, 1.2) Variant 1.6 (0.0, 3.1)	2.1 (-0.2, 4.5)	0.07
Apa1	647	Ancestral 0.9 (-1.5, 3.3) Variant 1.6 (0.1, 3.2)	0.7 (-2.0, 3.5)	0.60	603	Ancestral -0.6 (-2.8, 1.6) Variant 1.1 (-0.4, 2.5)	1.6 (-0.9, 4.2)	0.20
Fok1	634	Ancestral 2.1 (0.1, 4.1) Variant 1.2 (-0.5, 3.0)	-0.9 (-3.5, 1.7)	0.51	630	Ancestral 1.1 (-0.7, 2.9) Variant 0.6 (-1.0, 2.3)	-0.5 (-2.8, 1.9)	0.71

SNP: Single Nucleotide Polymorphism; N: Number of observations; IQR: Inter-quartile range; P: p-value of the interaction term; Ancestral: major frequency allele homozygotes; Variant: minor frequency allele homozygotes and heterozygotes.

References

1. National Toxicology P. Ntp monograph: Health effects of low-level lead. *NTP monograph*. 2012:i-148
2. Martin D, Glass TA, Bandeen-Roche K, Todd AC, Shi W, Schwartz BS. Association of blood lead and tibia lead with blood pressure and hypertension in a community sample of older adults. *American Journal of Epidemiology*. 2006;163:467-478
3. Cheng Y, Schwartz J, Sparrow D, Aro A, Weiss ST, Hu H. Bone lead and blood lead levels in relation to baseline blood pressure and the prospective development of hypertension the normative aging study. *American Journal of Epidemiology*. 2001;153:164-171
4. Elmarsafawy SF, Jain NB, Schwartz J, Sparrow D, Nie H, Hu H. Dietary calcium as a potential modifier of the relationship of lead burden to blood pressure. *Epidemiology*. 2006;17:531-537
5. Yazbeck C, Thiebaugeorges O, Moreau T, Goua V, Debotte G, Sahuquillo J, Forhan A, Foliguet B, Magnin G, Slama R, Charles MA, Huel G. Maternal blood lead levels and the risk of pregnancy-induced hypertension: The eden cohort study. *Environmental health perspectives*. 2009;117:1526-1530
6. Wells EM, Navas-Acien A, Herbstman JB, Apelberg BJ, Silbergeld EK, Caldwell KL, Jones RL, Halden RU, Witter FR, Goldman LR. Low-level lead exposure and elevations in blood pressure during pregnancy. *Environmental health perspectives*. 2011;119:664-669
7. Revis NW, Zinsmeister AR, Bull R. Atherosclerosis and hypertension induction by lead and cadmium ions: An effect prevented by calcium ion. *Proceedings of the National Academy of Sciences of the United States of America*. 1981;78:6494-6498
8. Kaji T, Suzuki M, Yamamoto C, Mishima A, Sakamoto M, Kozuka H. Severe damage of cultured vascular endothelial cell monolayer after simultaneous exposure to cadmium and lead. *Archives of environmental contamination and toxicology*. 1995;28:168-172
9. Vaziri ND. Mechanisms of lead-induced hypertension and cardiovascular disease. *American journal of physiology. Heart and circulatory physiology*. 2008;295:H454-465
10. Perlstein T, Weuve J, Schwartz J, Sparrow D, Wright R, Litonjua A, Nie H, Hu H. Cumulative community-level lead exposure and pulse pressure: The normative aging study. *Environmental health perspectives*. 2007;115:1696-1700
11. Zhang A, Park SK, Wright RO, Weisskopf MG, Mukherjee B, Nie H, Sparrow D, Hu H. Hfe h63d polymorphism as a modifier of the effect of cumulative lead exposure on pulse pressure: The normative aging study. *Environmental health perspectives*. 2010;118:1261-1266
12. Onalaja AO, Claudio L. Genetic susceptibility to lead poisoning. *Environmental health perspectives*. 2000;108 Suppl 1:23-28
13. Feldman D, Pike JW, Glorieux FH. *Vitamin d*. London, San Diego: Elsevier, Academic Press; 2005.
14. Rosen JF, Pounds JG. Quantitative interactions between pb2+ and ca2+ homeostasis in cultured osteoclastic bone cells. *Toxicology and applied pharmacology*. 1989;98:530-543
15. Pounds JG. Effect of lead intoxication on calcium homeostasis and calcium-mediated cell function: A review. *Neurotoxicology*. 1984;5:295-331

16. Schwartz BS, Stewart WF, Kelsey KT, Simon D, Park S, Links JM, Todd AC. Associations of tibial lead levels with bsmi polymorphisms in the vitamin d receptor in former organolead manufacturing workers. *Environmental health perspectives*. 2000;108:199-203
17. Cozzolino M, Stucchi A, Rizzo MA, Soldati L, Cusi D, Ciceri P, Brenna I, Elli F, Gallieni M. Reprint of: Vitamin d receptor activation and prevention of arterial ageing. *Proceedings and abstracts from the 8th International Symposium on Amino Acid and Protein Metabolism in Health and Disease*. 2013;23, Supplement 1:S31-S36
18. Giallauria F, Milaneschi Y, Tanaka T, Maggio M, Canepa M, Elango P, Vigorito C, Lakatta EG, Ferrucci L, Strait J. Arterial stiffness and vitamin d levels: The baltimore longitudinal study of aging. *The Journal of Clinical Endocrinology & Metabolism*. 2012;97:3717-3723
19. Andrukhova O, Slavic S, Zeitz U, Riesen SC, Heppelmann MS, Ambrisko TD, Markovic M, Kuebler WM, Erben RG. Vitamin d is a regulator of endothelial nitric oxide synthase and arterial stiffness in mice. *Molecular Endocrinology*. 2014;28:53-64
20. Goel RK, Lal H. Role of vitamin d supplementation in hypertension. *Indian journal of clinical biochemistry : IJCB*. 2011;26:88-90
21. Yang L, Ma J, Zhang X, Fan Y, Wang L. Protective role of the vitamin d receptor. *Cellular immunology*. 2012;279:160-166
22. Falk E. Pathogenesis of atherosclerosis. *Detection of Vulnerable Plaques*. 2006;47:C7-C12
23. Krieg Jr EF, Butler MA, Chang M-h, Liu T, Yesupriya A, Dowling N, Lindegren ML. Lead and cognitive function in vdr genotypes in the third national health and nutrition examination survey. *Neurotoxicology and teratology*. 2010;32:262-272
24. Weaver VM, Lee B-K, Todd AC, Ahn K-D, Shi W, Jaar BG, Kelsey KT, Lustberg ME, Silbergeld EK, Parsons PJ, Wen J, Schwartz BS. Effect modification by δ-aminolevulinic acid dehydratase, vitamin d receptor, and nitric oxide synthase gene polymorphisms on associations between patella lead and renal function in lead workers. *Environmental research*. 2006;102:61-69
25. Bell B, Rose C, Damon A. The normative aging study: An interdisciplinary and longitudinal study of health and aging. *Aging Human Develop.* 1972;3:5-17
26. Kim R, Aro A, Rotnitzky A, Amarasiriwardena C, Hu H. K x-ray fluorescence measurements of bone lead concentration: The analysis of low-level data. *Physics in Medicine and Biology*. 1995;40:1475-1485
27. Wright RO, Silverman EK, Schwartz J, Tsaih SW, Senter J, Sparrow D, Weiss ST, Aro A, Hu H. Association between hemochromatosis genotype and lead exposure among elderly men: The normative aging study. *Environmental health perspectives*. 2004;112:746-750
28. Morrell CH, Brant LJ, Ferrucci L. Model choice can obscure results in longitudinal studies. *The journals of gerontology. Series A, Biological sciences and medical sciences*. 2009;64:215-222
29. Assmann G, Cullen P, Evers T, Petzinna D, Schulte H. Importance of arterial pulse pressure as a predictor of coronary heart disease risk in procam. *European heart journal*. 2005;26:2120-2126
30. Baena-Diez JM, Bermudez-Chillida N, Garcia-Lareo M, Olivia Byram A, Vidal-Solsona M, Vilato-Garcia M, Gomez-Fernandez C, Vasquez-Lazo JE. [role of pulse pressure,

systolic blood pressure, and diastolic blood pressure in the prediction of cardiovascular risk. Cohort study]. *Medicina clinica*. 2008;130:361-365
31. Domanski M, Norman J, Wolz M, Mitchell G, Pfeffer M. Cardiovascular risk assessment using pulse pressure in the first national health and nutrition examination survey (nhanes i). *Hypertension*. 2001;38:793-797
32. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghori MJ, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467:52-58
33. Xi-Biao Y, Cui-E W, Hua F, Shui-Lian Y, Yi-Wen L, Wei-Min N. Associations of blood lead levels, kidney function, and blood pressure with δ-aminolevulinic acid dehydratase and vitamin d receptor gene polymorphisms. *Toxicology Mechanisms and Methods*. 2003;13:139-146
34. Rezende V, Barbosa F, Jr., Montenegro M, Sandrim V, Gerlach R, Tanus-Santos J. Haplotypes of vitamin d receptor modulate the circulating levels of lead in exposed subjects. *Archives of Toxicology*. 2008;82:29-36
35. Schwartz BS, Lee BK, Lee GS, Stewart WF, Simon D, Kelsey K, Todd AC. Associations of blood lead, dimercaptosuccinic acid-chelatable lead, and tibia lead with polymorphisms in the vitamin d receptor and [delta]-aminolevulinic acid dehydratase genes. *Environmental health perspectives*. 2000;108:949-954
36. Edelstein S, Fullmer CS, Wasserman RH. Gastrointestinal absorption of lead in chicks: Involvement of the cholecalciferol endocrine system. *The Journal of nutrition*. 1984;114:692-700
37. Fullmer CS. Intestinal lead and calcium absorption: Effect of 1,25-dihydroxycholecalciferol and lead status. *Proc Soc Exp Biol Med*. 1990;194:258-264
38. Smith CM, DeLuca HF, Tanaka Y, Mahaffey KR. Effect of lead ingestion on functions of vitamin d and its metabolites. *The Journal of nutrition*. 1981;111:1321-1329
39. Lee BK, Lee GS, Stewart WF, Ahn KD, Simon D, Kelsey KT, Todd AC, Schwartz BS. Associations of blood pressure and hypertension with lead dose measures and polymorphisms in the vitamin d receptor and delta-aminolevulinic acid dehydratase genes. *Environmental health perspectives*. 2001;109:383-389
40. Weuve J. Magnitude matters: Beyond detection in the presence of selection in research on socioeconomic inequalities in health. *Epidemiology (Cambridge, Mass.)*. 2013;24:10-13
41. Wilker E, Korrick S, Nie LH, Sparrow D, Vokonas P, Coull B, Wright RO, Schwartz J, Hu H. Longitudinal changes in bone lead levels: The va normative aging study. *Journal*

of occupational and environmental medicine / American College of Occupational and Environmental Medicine. 2011;53:850-855
42. Ward AJ, Cooper TA. The pathobiology of splicing. *The Journal of pathology*. 2010;220:152-163
43. Yura K, Shionyu M, Hagino K, Hijikata A, Hirashima Y, Nakahara T, Eguchi T, Shinoda K, Yamaguchi A, Takahashi K-i, Itoh T, Imanishi T, Gojobori T, Go M. Alternative splicing in human transcriptome: Functional and structural influence on proteins. *Gene*. 2006;380:63-71
44. Barry PS. A comparison of concentrations of lead in human tissues. *British journal of industrial medicine*. 1975;32:119-139
45. Bandeen-Roche K, Glass TA, Bolla KI, Todd AC, Schwartz BS. Cumulative lead dose and cognitive function in older adults. *Epidemiology*. 2009;20:831-839

Chapter 3

Modeling the causal role of DNA methylation in the association between smoking and inflammation in African Americans

Introduction

Cigarette smoking is a well-known risk factor for many diseases including cardiovascular disease (CVD).[1] Activation of inflammation has been suggested as a potential mechanism behind the smoking-induced cardiovascular risk.[2] Efforts to identify genetic components involved in inflammatory response, which is induced by cigarette smoking, have been extended to epigenome-wide association studies (EWAS).[3-7] Recently, epigenetic modification has gotten more attention in environmental epidemiology because of its plasticity and responsiveness to environmental changes.[8] Epigenetic modification is one mechanism for regulation of gene expression levels.[9] The most extensively studied epigenetic modification in human population is DNA methylation. DNA methylation is an addition of a methyl group to a Cytosine (C) base that is followed by a Guanine (G) base. C followed by G is denoted CpG where p indicates that C and G are connected by a phosphodiester bond. DNA methylation levels in gene promoter regions are often inversely associated with gene expression levels of the gene.[10] DNA methylation plays an important role in embryonic development,[11] cell differentiation,[12] and X chromosome inactivation.[13]

Several EWAS identified differentially methylated CpG sites related to cigarette smoking status.[3-7, 14] Methylation levels of the CpG sites in coagulation factor II (thrombin) receptor-like 3

(*F2RL3)*, also known as protease-activated receptor 4 (*PAR-4*),[3, 5, 14] and aryl-hydrocarbon receptor repressor (*AHRR)*[4-6] have been consistently associated with cigarette smoking. However, the role of DNA methylation levels as a mediator in the association between cigarette smoking and inflammatory responses has not been investigated yet. Because of its responsive characteristic, DNA methylation may not only have an effect on inflammatory response, but also may be affected by inflammation. Therefore, DNA methylation brings another complexity from the possibility of reverse causation into the investigation.

The Mendelian randomization (MR) approach is an instrumental variable analysis using a genetic marker as an instrument. This approach enables us to examine the causal relationship in observational studies while avoiding confounding and reverse causation.[15] Because of random assortment of genetic alleles during gamete formation, the assignment of a specific genetic allele is random. When a genetic marker is highly associated with the environmental exposure status, the assignment of environmental exposure using the genetic marker as an instrument can be considered as random. Therefore, the MR approach may be comparable to randomized clinical trials (RCT). Because we are born with the genetic allele, the genetic allele would not change by the environmental exposure status or other factors, therefore, the MR approach is less likely to be affected by confounding. In addition, the instrument has an effect on outcome only through the exposure by assumption,[16] hence, we can evaluate the direction of the association from one variable to another variable by using the MR approach.

In this study, we hypothesized that cigarette smoking is associated with DNA methylation levels, and DNA methylation levels are associated with inflammatory responses. To investigate the role of the DNA methylation as a mediator, we implemented the two-step epigenetic MR approach.[17]

Methods

Study population

The Genetic Epidemiology Network of Arteriopathy (GENOA) was established as a part of the Family Blood Pressure Program by the National Heart Lung and Blood Institute in 1996.[18] From its inception, GENOA's long-term objective has been to elucidate the genetics of hypertension and its arteriosclerotic target-organ damage, including both atherosclerotic (macrovascular) and arteriolosclerotic (microvascular) complications involving the heart, brain, kidneys, and peripheral arteries. GENOA is a community-based study of sibships that incudes individuals with a personal or family history of hypertension.

The African-American cohort of GENOA, which is the focus of Aims 2 and 3, is from Jackson, Mississippi.[19] In Jackson, the sibships were recruited through hypertensive probands who had participated in the Atherosclerosis Risk in Communities (ARIC) study, which is a probability sample of 45 to 64-year-old African American residents of the community.[20] At the enrollment stage of GENOA (Phase I: 1995-2000), 1,854 subjects who were members of 683 sibships with at least two individuals diagnosed with essential hypertension before age sixty were recruited into the study. All siblings in the sibship were invited to participate, both normotensive and hypertensive. Exclusion criteria of the GENOA study were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. Eighty percent (1,482 subjects) of the initial study population returned for the second examination (Phase II: 2000-2005). Study visits were made in the morning after an overnight fast of at least eight hours. Data on demographic information, medical history, clinical characteristics, lifestyle factors, and blood samples were collected in each phase. Written informed consent was obtained from all subjects and approval was granted by participating

institutional review boards. Because both DNA methylation levels and inflammatory markers were measured at the second examination, only data from Phase II were used in this study.

Genotypes

For 1,263 participants, genotype data was obtained using the Affymetrix® Genome-Wide Human SNP Array 6.0 platform for about one million SNPs. For 269 samples, genotype data was obtained using Illumina® Human 1M-Duo BeadChip for about one million SNPs. For 508 samples, genotype data was obtained using the Affymetrix® Genome-Wide Human SNP Array 6.0 platform for about one million SNPs, separately. Quality control was performed based on the following exclusion criteria in order: <95% SNP call rate, <1% minor allele frequency (MAF), and <95% sample call rate.

To get more dense markers, we imputed the genotype data using the 1,000 Genomes Project data.[21] The three sets of samples were pre-phased and imputed separately. The DNA strands were aligned to positive strand. A physical map was updated from NCBI build 36 (hg 18) to build 37 (hg 19) using LiftOver software (http://genome.sph.umich.edu/wiki/LiftOver). Pre-phasing was done using the Segmented HAPlotype Estimation and Imputation Tool (SHAPEIT) version 2.[22] Pre-phasing was done separately for each chromosome. Imputation was performed using IMPUTE version 2[23] and reference panels from the 1,000 Genomes project Phase I integrated variant set release (v3) in NCBI build 37 (hg19) coordinates (released on March 2012). For imputation, each chromosome was split into 5 mega base pair (Mbp) long segments, resulting in 565 segments across the genome. Outlier samples from principal component analyses (>6 standard deviations) were excluded. Monomorphic markers were also excluded. Finally, there were 30,022,375 imputed SNPs for 1599 samples.

DNA methylation

DNA methylation levels were measured from blood samples collected in Phase II. Peripheral blood leukocytes were isolated from blood samples and used to measure DNA methylation levels. The EZ DNA Methylation Gold Kit (Zymo Research, Orange CA) was used for bisulfite conversion. The methylation assay was performed at the Mayo Clinic Advanced Genomics Technology Center using Illumina® Infinium HumanMethylation27 BeadChips and Illumina BeadXpress reader. DNA methylation was measured from blood samples on a single array, in two different color channels (red and green). A total of 1,008 individuals were measured for 27,578 CpG sites.

As a quality control, seven samples were excluded from analysis due to poor bisulfite conversion efficiency (intensity < 4,000). An additional 28 samples were removed because of poor background signals. Finally, there were 973 samples left. The *lumi* package[24] in R software was used for background adjustment, color balance adjustment, and quantile normalization. After quality control and normalization, the distribution of methylation intensity became similar across the samples. The CpG site intensities were used to calculate beta-values and M-values, a ratio of methylated allele intensity and total intensity+100, and a log2 ratio of methylated and unmethylated probe intensities, respectively.[25] Cell proportions for CD8 T lymphocytes, CD4 T lymphocytes, natural killer cells, B cells, monocytes, granulocytes were estimated using Houseman's method.[26]

Smoking variable

Current smoking status was obtained during Phase II from questionnaire asking a question "Do you now smoke cigarettes?".

Inflammatory markers

Inflammatory markers were measured from blood samples obtained during Phase II. Circulating CRP levels were measured by a highly sensitive immunoturbidimetric assay.[27] Fibrinogen level was measured from citrated plasma using the clotting time based Clauss method.[28] Interleukin-6 (IL-6) and interleukin-18 (IL-18) were measured using a contracted service with SearchLight™ Technologies (Boston, MA).[29] CRP, IL-6, and IL-18 were log-transformed to reduce skewness.

Assumption checking

A. Assumptions for instrumental variable analysis[30]

   a. The instrument has a causal effect (here causal effect means directed effect from instrument to exposure) on the exposure: genetic variants are rarely changed by other factors (except somatic mutations in cancer cells[31]). Therefore, any genetic variants that are statistically associated with the exposure satisfy this assumption. We have two instruments: one for smoking status (SNP S) and the other for DNA methylation levels (SNP C). To find SNP S that satisfies this assumption, the association between SNP S and smoking status was evaluated using the logistic mixed effects model with an alpha level of 0.05 (Equation 3.1).

$$log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 * SNP\ S_{ij} + \beta_2 * age_{ij} + \beta_3 * gender_{ij} + u_j \quad (Equation\ 3.1)$$

   where *P* is a probability of being a current smoker, u_j is a random intercept for sibship *j*, and *i* indicates subject *i*.

   To check the assumption for SNP C, the association between the SNP C and DNA methylation levels was evaluated using the linear mixed effects model (Equation 3.2).

$$DNA\ methylation\ levels = \beta_0 + \beta_1 * SNP\ C_{ij} + \beta_2 * age_{ij} + \beta_3 * gender_{ij} + u_j + \varepsilon_{ij} \quad (Equation\ 3.2)$$

where u_j is a random intercept for sibship j, and ε_{ij} is a random error for an individual i in sibship j.

b. The association between the instrument and the outcome is only through the exposure. This assumption was checked by estimating a direct effect of the instrument on the outcome while the exposure is in the model. To check the assumption for SNP S, the direct effect of the SNP S on DNA methylation was assessed using the linear mixed effects model (with an alpha level of 0.05) including smoking status as a covariate (Equation 3.3). DNA methylation levels can be different depending on cell proportions. Hence, the model was adjusted for cell proportions.

$$DNA\ methylation\ levels = \beta_0 + \beta_1 * SNP\ S_{ij} + \beta_2 * current\ smoking\ status_{ij} + \beta_3 * age_{ij} + \beta_4 * gender_{ij} + \beta_5 * CD8T_{ij} + \beta_6 * CD4T_{ij} + \beta_7 * NK_{ij} + \beta_8 * Bcell_{ij} + \beta_9 * Mono_{ij} + \beta_{10} * Gran_{ij} + u_j + \varepsilon_{ij} \quad (Equation\ 3.3)$$

where u_j is a random intercept for sibship j, and ε_{ij} is a random error for an individual i in sibship j.

c. To check the assumption for SNP C, the direct effects of SNP C on inflammatory markers were assessed using the linear mixed effects models (with an alpha level of 0.05) including DNA methylation as a covariate (Equation 3.4). DNA methylation levels can be different depending on cell proportions. Hence, the model was adjusted for cell proportions.

$$Inflammatory\ marker\ levels$$

$$= \beta_0 + \beta_1 * SNP\ C_{ij} + \beta_2 * DNA\ methylation\ levels_{ij}$$

$$+ \beta_3 * age_{ij} + \beta_4 * gender_{ij} + \beta_5 * CD8T_{ij} + \beta_6 * CD4T_{ij} + \beta_7 * NK_{ij}$$

$$+ \beta_8 * Bcell_{ij} + \beta_9 * Mono_{ij} + \beta_{10} * Gran_{ij} + u_j + \varepsilon_{ij} \quad (Equation\ 3.4)$$

where u_j is a random intercept for sibship j, ε_{ij} is a random error for an individual i in sibship j, and inflammatory markers are log(CRP), log(IL-6), log(IL-18), and fibrinogen.

d. The instrument does not share common causes with the outcome. Because genetic variants are rarely changed by other factors, this assumption is satisfied for any SNP.

Two-step epigenetic Mendelian randomization

In this study, we investigated the causal relation from smoking to inflammatory response mediated by DNA methylation using a two-step epigenetic MR approach (**Figure 3.1**).[17] Previous, a study of GENOA African Americans identified 15 CpG sites including cg03636183, cg19859270, cg04983977, cg13668129, cg13500388, cg13633560, cg03340878, cg01500140, cg00353953, cg11314684, cg03330058, cg13745870, cg17791651, cg14223444, and cg26259865 that were significantly associated with current smoking status (a derived binary variable on whether a participant smoked within the past year) (Bonferroni corrected p-value < 0.05).[14] Out of those 15 CpG sites, 5 CpG sites including cg03636183, cg19859270, cg13668129, cg01500140, and cg11314684 were significantly associated with current smoking status in an independent replication study. In our study, those 15 CpG sites found in GENOA study were considered.

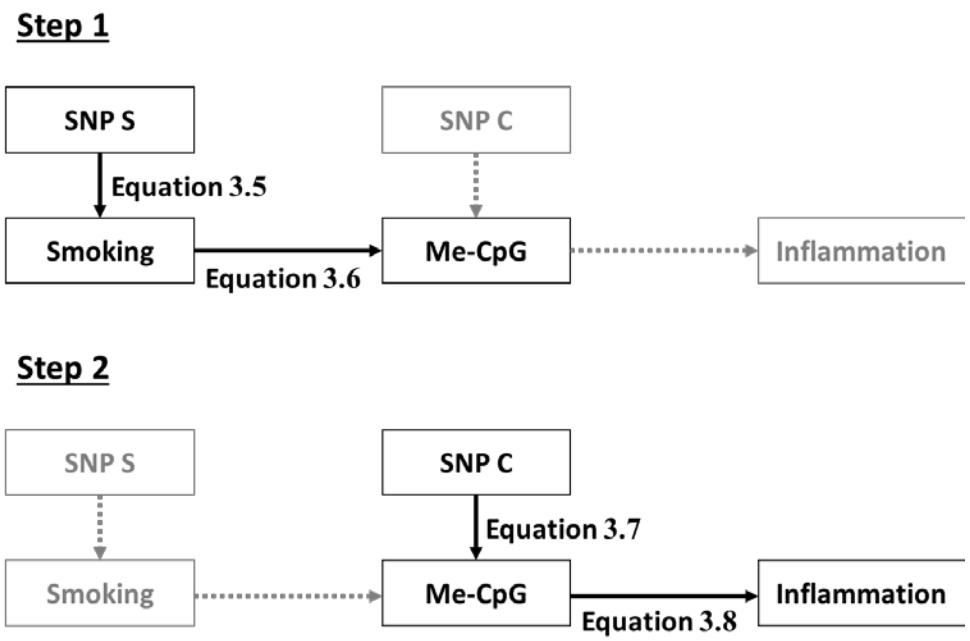

Figure 3.1. Two-step epigenetic Mendelian randomization. Rectangular dark boxes represent variables of interest at each step, and light boxes represent variables of interest at the other step. Solid arrows indicate the directed association of interest at each step, and dashed arrows indicate the directed association of interest at the other step. SNP S is an instrumental SNP for smoking, and SNP C is an instrumental SNP for DNA methylation levels of CpG sites. Equations 3.5-3.8 were used to obtain estimates at each step.

Step 1: The MR approach was applied to examine the causal relationship between smoking and differential DNA methylation at the 15 CpG sites using selected SNPs (SNP S in Figure 3.1) as instrumental variables. Nicotinic acetylcholine receptor alpha3, alpha5, and beta4 genes (*CHRNA3*, *CHRNA5*, and *CHRNB4*) have been demonstrated to be associated with onset of smoking[32], serum cotinine level [33], nicotine dependency [34], nicotine addiction [35], and smoking cessation success [36]. We investigated SNPs within 10kb from *CHRNA5*, *CHRNA3*, and *CHRNB4* genes to identify an appropriate instrument SNP for step 1. Using logistic mixed effects models,

the SNPs were tested for an association with current smoking status adjusting for age, gender, and a random intercept for sibship effect. The allele dosage from imputation was used for SNPs assuming an additive inheritance model. Using the most significantly associated SNP with a MAF larger than or equal to 5% as an instrument, the MR approach was implemented by running two-stage least squares. First, current smoking status was regressed on the instrument SNP S using a logistic mixed effects model adjusting for age, gender, and a random intercept for sibship effect (Equation 3.5). Second, DNA methylation levels were regressed on the predicted values from equation 5 adjusting for age, gender, cell proportions, and a random intercept for sibship effect (Equation 3.6).

$$log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 * SNP\ S + \beta_2 * age + \beta_3 * gender + u_j \quad (Equation\ 3.5)$$

where *P* is a probability of being a current smoker, u_j is a random intercept for sibship *j*, and *i* indicates subject *i*.

$$DNA\ methylation\ levels = \beta_0 + \beta_1 * \widehat{smoking} + \beta_2 * age + \beta_3 * gender + \beta_4 * CD8T + \beta_5 * CD4T + \beta_6 * NK + \beta_7 * Bcell + \beta_8 * Mono + \beta_9 * Gran + u_j + \varepsilon_{ij} \quad (Equation\ 3.6)$$

where u_j is a random intercept for sibship j, ε_{ij} is a random error for an individual i in sibship j, and $\widehat{smoking}$ is a predicted values from (Equation 3.5).

To correct for the multiple testing problem, a Bonferroni correction was applied to the results of Step 1. Since fifteen DNA methylation sites were tested, 0.003 (=0.05/15) was used as a criterion for a significant p-value.

Step 2 The MR approach was applied to examine the causal relationship between DNA methylation and inflammatory responses using SNPs as instrumental variables. We investigated SNPs within 10kb from the genes where the 15 CpG sites are located to identify appropriate instrument SNPs (SNP C in Figure 3.1) for step 2. Using linear mixed effects models, the SNPs were tested for an association with each of the CpG site methylation levels adjusting for age, gender, and a random intercept for sibship effect. Using the selected SNP as an instrument, the MR approach was implemented by running two-stage least squares. First, DNA methylation levels were regressed on the instrument SNP using linear mixed effects models adjusting for age, gender, and a random intercept for sibship effect (Equation 3.7). Second, inflammatory marker levels were regressed on the predicted values from equation 3.7 adjusting for age, gender, cell proportions, and a random intercept for sibship effect (Equation 3.8).

$$DNA\ methylation\ levels = \beta_0 + \beta_1 * SNP\ C + \beta_2 * age + \beta_3 * gender + u_j + \varepsilon_{ij} \quad (Equation\ 3.7)$$

where u_j is a random intercept for sibship j, and ε_{ij} is a random error for an individual i in sibship j.

$$Inflammatory\ marker = \beta_0 + \beta_1 * \widehat{methylation} + \beta_2 * age + \beta_3 * gender + \beta_4 * CD8T + \beta_5 * CD4T + \beta_6 * NK + \beta_7 * Bcell + \beta_8 * Mono + \beta_9 * Gran + u_j + \varepsilon_{ij} \quad (Equation\ 3.8)$$

where u_j is a random intercept for sibship j, ε_{ij} is a random error for an individual i in sibship j, $\widehat{methylation}$ is predicted values from (Equation 3.7), and inflammatory markers are log(CRP), log(IL-6), log(IL-18), fibrinogen.

For DNA methylation levels, beta-values have heteroscedasticity for highly methylated or unmethylated CpG sites. Thus, M-values were used as a primary measure due to their favorable statistical properties compared to beta-values.[25] To resolve the multiple testing problem, a Bonferroni correction was applied to the results of the step 2 MR. To run the analysis, the lme() function in the nlme R package and SAS® 9.4 were used.

Results

Study population descriptive

The total of 848 GENOA African-Americans with both measured genotypes and DNA methylation levels and were included in the analysis. The mean age was 67 years (standard deviation (SD) = 7 years), and 72% of the population were women. In this study population, 11% were current smokers, 30% were previous smokers, and 59% were never smokers. The majority of the study participants have been diagnosed for hypertension (77%), as expected (**Table 3.1**). Adjusting for age and gender, current cigarette smoking status was associated with inflammatory markers including log(CRP) ($p = 0.03$), log(IL-6) ($p = 0.01$), log(IL-18) ($p = 0.06$), and fibrinogen ($p < 0.01$).

Assumption checking

A.1. Assumption for the instrument SNP for smoking

a. The instrument SNP is associated with the exposure.

   Among the 335 SNPs in the *CHRNA5*, *CHRNA3*, and *CHRNB4* genes with MAF larger than or equal to 5%, rs4887071 was the most significantly associated with smoking status. The odds ratio (OR) for being a current smoker was 1.96 (95% CI: 1.12-3.42, p=0.02) for

every one copy of the minor allele of rs4887071 (minor allele = A, minor allele frequency = 7%) adjusting for age and gender in GENOA African Americans.

b. The instrument SNP affects the outcome only through the exposure.

The direct effect of rs4887071 on DNA methylation levels was not significant for methylation levels considered except for the CpG site in *LIM2* ($p<0.05$) (**Table 3.2***)*. Therefore, the next SNP significantly associated with smoking, rs7170068 (minor allele frequency = 23%), was used as an instrument for *LIM2*.

A.2. Assumption for instrument SNPs for DNA methylation

a. The instrument SNP has an effect on the DNA methylation. To avoid a bias from a weak instrument, it is suggested to use an instrument with the F-value larger than 10.[37] Unfortunately, we could not find any instrument SNPs with F-value larger than 10 for cg03340878 in *OR2B6,* cg13745870 in *SPATA12,* cg13668129 in *HNRPUL1*, cg04983977 in *GPR25,* and cg19859270 in *GPR15* (**Table 3.3**). We are aware of the possible bias from weak instruments for these CpG sites in the results.

b. The instrument SNPs of DNA methylation affect the inflammatory responses only through the DNA methylation. All the instrument SNPs satisfy this assumption except rs11076844, which is an instrument SNP for *LOC124220* and has a strong direct effect on log(CRP) levels (data not shown).

Step 1 Mendelian randomization: Smoking -> DNA methylation

In GENOA imputed genotype data using the 1000 genomes data as a reference panels, there were 997 SNPs within *CHRNA5, CHRNA3, and CHRNB4* genes. To avoid selecting a rare variant as an instrument variable, we only considered 294 SNPs with a minor allele frequency

greater than or equal to 0.05. We tested the association between the 294 SNPs and current smoking status using logistic mixed effect models. The most significant SNP, rs4887071, was selected as the instrument for step 1 MR (p=0.02). The SNP has a F-value smaller than 10 and it may introduce a bias from weak instrument. To check the bias, a sensitivity analysis was performed using a genetic risk score that is composed of three SNPs with F-value of 13. The estimates and p-values were consistent using the genetic risk score (data not shown). The SNP, rs4887071, which is SNP S in Figure 3.1, is located in the intron region of *CHRNB4* gene which makes a long-noncoding RNA (lncRNA). This lncRNA is a conserved target of microRNA (hsa-miR-4659a-3p).[38] Rs4887071 causes changes in the structure of the lncRNA which results in changes in the binding affinity between the lncRNA the microRNA (**Figure 3.2**).[39, 40] The OR for being a current smoker is 1.96 (=e $^{0.67}$, 95% CI: 1.12-3.42) for every one copy of the minor allele of rs4887071 adjusting for age and gender in the GENOA African Americans.

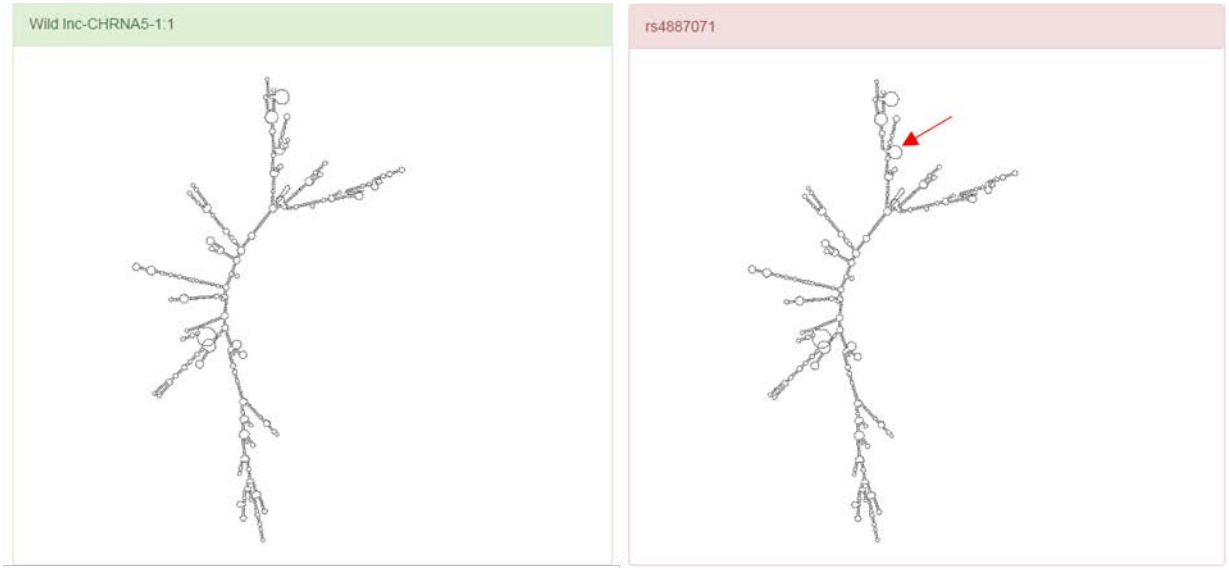

Figure 3.2. The effect of rs4887071 on the structure of lncRNA from *CHRNA5* gene. The red arrow points out the bulge caused by rs4887071 (http://bioinfo.life.hust.edu.cn/lncRNASNP/).

Current smoking status was found to be associated with differential DNA methylation on the 15 CpG sites in GENOA.[14] We reevaluated the association using the MR approach based on

Equations 3.5 and 3.6 (**Table 3.4)**. Current smoking status, which is genetically affected by rs4887071, makes changes on DNA methylation levels of cg03636183 in *F2RL3* gene, cg19859270 in *GPR15* gene, cg13500388 in *CBFB* gene, and cg04983977 in *GPR25* gene (criterion for p-value = 0.05/15 = 3.33E-03) (**Table 3.4**). With a minor allele of rs4887071, OR for current smoking was 1.96 which results in 0.44 (=log(1.96)*0.660) decrease in M-values of cg03636183 on *F2RL3* gene. M-values of cg03636183 range between 0.12 and 5.13 with a mean of 3.06 (**Table 3.5**). With a minor allele of rs4887071, OR for current smoking was 1.96 which results in 0.14 (=log(1.96)*0.203) decrease in M-values of cg19859270 on *GPR15* gene. M-values of cg19859270 ranges between -0.01 and 4.88 with a mean of 3.77. With a minor allele of rs4887071, OR for current smoking was 1.96 which results in 0.09 (=log(1.96)*0.138) decrease in M-values of cg13500388 on *CBFB* gene (M-value ranges between -1.47 and 1.94 with a mean of 0.32) and 0.06 (=log(1.96)*0.093) increase in M-values of cg04983977 on *GPR25* gene (M-values of ranges between -1.97 and 2.87 with a mean of 1.64). The results were mostly consistent between beta-values and M-values.

Step 2 Mendelian randomization: DNA methylation -> Inflammation

We further investigated whether the changes in DNA methylation levels induced by current smoking status can make changes in inflammatory responses. Adjusting for age and gender, the SNP that was the most significantly associated with the DNA methylation levels with a minor allele frequency larger than or equal to 5% was selected as an instrument for each CpG site based on Equation 2 (**Table 3.3**). Most of the SNPs, which were selected as instruments, were significantly associated with the methylation levels of each CpG sites. However, we could not find an appropriate instrument SNP for cg19859270 on *GPR15* gene (p-value > 0.05). Because

there was no SNP that satisfied the assumptions for an instrument, cg19859270 was excluded for step 2.

The three CpG sites, cg03636183 in *F2RL3* gene, cg13500388 in *CBFB* gene, and cg04983977 in *GPR25* gene that were found to be associated with smoking status in step 1 were considered in step 2 MR. We found that DNA methylation levels of a few CpG sites cause changes in inflammatory markers including IL-18 and CRP (**Table 3.6**).

With a minor allele of rs4887071, OR for current smoking was 1.96 which results in 0.44 decrease in M-values of cg03636183 on *F2RL3* gene and 0.09 decrease in M-values of cg13500388 on *CBFB* gene. The 0.44 decrease in M-values of cg03636183 on *F2RL3* gene results in a 5% ($e^{0.44*0.11}$=1.05) increase in IL-18 levels (pg/ml). The 0.09 decrease in M-values of cg13500388 on *CBFB* gene results in a 3% ($e^{0.09*0.30}$=1.03) increase in CRP levels (mg/L) (**Table 3.7**).

Discussion

The association between smoking and inflammation is well known.[41] However, the biological mechanisms behind the association are not fully understood. In particular, the role of DNA methylation is largely unknown. We investigated the causal role of DNA methylation in the association between smoking and inflammation using the two-step epigenetic MR approach. We identified CpG sites that respond to smoking exposure and mediate the effect of smoking on inflammation. We demonstrated that current smoking status causes changes in DNA methylation levels of CpG sites on *F2RL3*, *GPR15*, *CBFB*, and *GPR25* genes. Cg03636183 on the *F2RL3* gene has been found to be associated with smoking not only in GENOA and the Grady Trauma Project study but also in the Langzeiterfolge der KARdiOLogischen Anschlussheilbehandlung study and the Family and Community Health Study.[3, 6, 14] Methylation in the *F2RL3* gene has

been suggested as a biomarker of current and lifetime smoking exposure.[42-44] Methylation of *GPR15* was also found to be associated with current smoking, cumulative smoke exposure and time since smoking.[45]

The methylation level changes of CpG sites on *F2RL3* and *CBFB* genes are associated with changes in the levels of inflammatory markers IL-18 and CRP. The *F2RL3* gene has been found to be associated with platelet activation and perioperative myocardial injury,[46] post infectious irritable bowel syndrome,[47] and gastric cancer.[48] The *CBFB* gene makes a beta subunit of a core-binding transcription factor and it binds to Runt-related transcription factors (*RUNX*1, 2, and 3) that are involved in hematopoiesis and osteogenesis.[49] The *CBFB* gene has been found to be associated with breast cancer, prostate and ovarian cancer, and acute myeloid leukemia.[50-52]

Cigarette smoking has been associated with elevated levels of inflammatory markers. In a study of NHANES 1999-2002, current smokers had 36% (95% CI: 8-70%), 46% (95% CI: 10-93%), and 73% (95% CI: 30-130%) higher levels of serum CRP among men aged 20-39, 40-59, 60 and older, respectively, compared to non-smokers.[53] Among smokers, a dose-response relationship was found between pack-years and serum CRP levels.[54] High serum CRP and IL-18 levels are associated with coronary artery disease and type 2 diabetes.[55-57]

The two-step epigenetic MR approach enables us to identify the CpG sites that are involved in the association between an environmental exposure and an outcome. This approach can expand our understanding on the role of DNA methylation in disease development and provides information on biological mechanisms behind it. However, the MR has limitations. This approach relies on strong assumptions: every directed pathway from the instrument SNP to DNA methylation passes through smoking in step 1, and every directed pathway from the instrument SNP to inflammatory markers passes through DNA methylation. However, pleiotropic effects of

genes and SNPs that are found to be associated with human complex diseases and traits are abundant.[58] Additional paths from the instrument SNP for smoking status to DNA methylation, and paths from the instrument SNP for DNA methylation to inflammatory markers can potentially introduce biases to the estimates from the MR approach. The MR approach also relies on the assumption that the relationship between variables is linear. Because a non-linear relationship is commonly found in epidemiologic studies,[59] we might have missed a causal relationship which is not linear. Another limitation of this approach is the availability of an appropriate instrument. To avoid a bias from a weak instrument, it is suggested to use an instrument with the F-value larger than 10.[37] However, the instrument SNP for cigarette smoking, and the instrument SNPs for DNA methylation levels of cg03340878, cg13745870, cg13668129, cg04983977, and cg19859270 (**Table 3.3**) had F-values smaller than 10. Use of those SNPs might introduce bias in effect size estimation. Statistical power of the MR approach is influenced by the strength of the association between an instrument and the exposure, and the association between the exposure and the outcome. Therefore, use of those SNPs with small F-values might also deteriorate statistical power of the MR approach.

The other limitation of this study is that we did not consider the direct effect of smoking on inflammatory responses. To assess the magnitude and significance of the indirect effect of smoking that is mediated by DNA methylation levels, the MR approach should be applied in the framework of causal mediation analysis. In fact, a new method to study causal mediation with MR, named network MR, was recently published.[60] The approach implemented in Aim 3 is similar to the newly published approach because both direct and indirect effects are considered to evaluate the causal mediation effect of DNA methylation.

Compared to the general population, there are a relatively smaller number of current smokers in the elderly population. Smoking status is likely to be affected by infection or disease conditions. From the race/ethnic differences in prevalence of smoking and DNA methylation levels, the results may not be generalizable to other race/ethnic populations. The results may be better supported if the finding is confirmed in other independent study populations.

In conclusion, we found evidence for the mediation of DNA methylation levels in the association between smoking and inflammation that will inform our next study using MR.

Tables

Table 3.1 Descriptive characteristics of GENOA African American study population (N=848) at Phase 2 Examination.

Continuous variables	**N**	**Mean**	**SD**
Age at examination (years)	848	67	7
Height (cm)	844	168	9
Weight (kg)	844	88	18
Body Mass Index (kg/m^2)	844	31	6
Pack years	348	25	22
Age started smoking	348	20	6
Age quit smoking	252	48	12
Alcohol (number of drinks/week)	122	4	4
Education (years)	848	12	4
Systolic blood pressure (sitting, mmHg)	848	154	22
Diastolic blood pressure (sitting, mmHg)	848	92	12
Pulse pressure (sitting, mmHg)	848	62	18
C reactive protein (mg/L)	821	6	7
Fibrinogen (mg/dL)	823	369	81
Interleukin-6 (pg/ml)	741	10	7
Interleukin-18 (pg/ml)	735	72	45
High density lipoprotein (mg/dL)	848	59	18
Low density lipoprotein (mg/dL)	848	121	38
Lipid lp(a) (mg/dL)	823	63	51
Categorical variables		**N**	**%**
Female		610	72%
Current smoker		96	11%
Previous smoker		252	30%
Never smoker		500	59%
Currently taking anti-hypertensive medication		634	75%
Ever diagnosed with hypertension		661	78%
Ever diagnosed with heart attack or MI		43	5%

Ever diagnosed with diabetes	249	29%
Ever diagnosed with cancer	71	8%

Table 3.2 Estimates of direct effect of rs4887071 on DNA methylation levels (N=848).

CpG site	Gene	Estimate	SE	Pr > |t|
cg01500140	*LIM2*	0.08	0.04	0.03
cg19859270	*GPR15*	-0.07	0.04	0.07
cg13500388	*CBFB*	-0.07	0.05	0.17
cg03636183	*F2RL3*	-0.14	0.11	0.20
cg13745870	*SPATA12*	-0.03	0.03	0.31
cg04983977	*GPR25*	0.04	0.04	0.34
cg03340878	*OR2B6*	-0.03	0.03	0.34
cg03330058	*ABTB1*	-0.04	0.06	0.47
cg13668129	*HNRPUL1*	-0.02	0.03	0.49
cg17791651	*POU3F1*	0.05	0.07	0.49
cg11314684	*AKT3*	0.02	0.04	0.55
cg26259865	*ZG16B*	-0.02	0.04	0.62
cg14223444	*NCBP1*	-0.02	0.04	0.63
cg00353953	*ZNF384*	0.00	0.05	0.92
cg13633560	*LRRC32*	0.00	0.05	0.95

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 3.3 Instrument SNPs of 15 CpG sites to be used for step 2 Mendelian randomization (SNP C in Figure 3.1).

CpG site	Gene	Step 2 instrument SNP	Coded allele frequency	Estimate for SNP	SE for SNP	P-value for SNP	F-value
cg13500388	*CBFB*	rs8048014	0.15	0.59	0.02	8.69E-82	630
cg03636183	*F2RL3*	rs2227341	0.19	-1.15	0.05	4.11E-71	505
cg14223444	*NCBP1*	rs60028882	0.37	-0.20	0.01	1.24E-41	237
cg03330058	*ABTB1*	rs7627836	0.16	-0.39	0.03	5.75E-30	155
cg13633560	*LRRC32*	rs4945097	0.56	0.13	0.02	5.85E-10	40
cg11314684	*AKT3*	rs10157763	0.15	0.13	0.02	4.51E-08	31
cg26259865	*LOC124220*	rs11076844	0.83	-0.13	0.02	4.60E-08	31
cg17791651	*POU3F1*	rs34860389	0.22	-0.19	0.03	5.02E-08	31
cg00353953	*ZNF384*	rs35787939	0.06	0.20	0.04	3.53E-06	22
cg01500140	*LIM2*	rs10409027	0.13	0.11	0.03	2.74E-04	14
cg03340878	*OR2B6*	rs276364	0.07	0.08	0.02	2.50E-03	9
cg13745870	*SPATA12*	rs9857168	0.84	-0.05	0.02	5.68E-03	8
cg13668129	*HNRNPUL1*	rs12980267	0.14	-0.04	0.02	7.41E-03	7
cg04983977	*GPR25*	rs2292099	0.8	0.04	0.02	3.55E-02	4
cg19859270	*GPR15*	rs6790026	0.28	-0.03	0.02	1.08E-01	3

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 3.4 Results of the step 1 Mendelian randomization (N=848) based on Equation 5 and 6.

		Mendelian randomization (M-values)			Mendelian randomization (Beta-values)		
CpG site	Gene	β	SE	P	β	SE	P
cg03636183	*F2RL3*	-0.660	0.087	<.0001	-0.059	0.007	<.0001
cg19859270	*GPR15*	-0.203	0.029	<.0001	-0.010	0.002	<.0001
cg13500388	*CBFB*	-0.138	0.043	0.001	-0.023	0.007	0.001
cg04983977	*GPR25*	0.093	0.030	0.002	0.012	0.004	0.006
cg03340878	*OR2B6*	-0.070	0.025	0.005	-0.012	0.004	0.005
cg01500140	*LIM2*	0.079	0.028	0.005	0.008	0.003	0.023
cg03330058	*ABTB1*	-0.142	0.050	0.005	-0.022	0.008	0.005
cg13668129	*HNRPUL1*	-0.057	0.021	0.006	-0.005	0.002	0.018
cg13633560	*LRRC32*	-0.094	0.039	0.015	-0.012	0.005	0.012
cg17791651	*POU3F1*	-0.085	0.050	0.087	-0.013	0.008	0.079
cg26259865	*LOC124220*	-0.049	0.031	0.120	-0.007	0.004	0.085
cg00353953	*ZNF384*	-0.050	0.033	0.130	-0.007	0.005	0.174
cg13745870	*SPATA12*	-0.027	0.020	0.170	-0.005	0.004	0.173
cg14223444	*NCBP1*	-0.018	0.027	0.510	-0.003	0.005	0.522
cg11314684	*AKT3*	-0.010	0.030	0.750	-0.001	0.004	0.837

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 3.5 Distributions of M-values and Beta-values.

	M-value				Beta-value			
CpG	Mean	SD	Min	Max	Mean	SD	Min	Max
cg03636183	3.06	0.96	0.12	5.13	0.88	0.08	0.51	0.97
cg19859270	3.77	0.35	-0.01	4.88	0.93	0.02	0.45	0.97
cg04983977	1.64	0.34	-1.97	2.87	0.75	0.05	0.19	0.88
cg13668129	-2.19	0.24	-3.07	0.50	0.17	0.03	0.10	0.57
cg13500388	0.32	0.49	-1.47	1.94	0.54	0.08	0.26	0.79
cg13633560	-1.76	0.43	-3.71	-0.25	0.21	0.05	0.05	0.44
cg03340878	-0.29	0.32	-1.83	1.08	0.44	0.06	0.20	0.67
cg01500140	1.93	0.45	-1.35	3.15	0.79	0.05	0.27	0.90
cg00353953	-1.28	0.39	-2.52	0.16	0.27	0.06	0.12	0.51
cg11314684	-1.32	0.35	-2.85	0.36	0.28	0.05	0.11	0.56
cg03330058	-0.71	0.54	-3.80	1.11	0.38	0.09	0.06	0.68
cg13745870	0.07	0.27	-0.91	1.17	0.49	0.05	0.33	0.68
cg17791651	0.90	0.59	-1.24	3.06	0.64	0.09	0.29	0.89
cg14223444	-0.53	0.31	-1.49	1.87	0.40	0.05	0.25	0.77
cg26259865	1.45	0.36	-3.35	2.93	0.73	0.05	0.08	0.88

Table 3.6 Results of the step 2 Mendelian randomization.

CpG site	Gene	log (IL-6) N=741 Estimate	SE	P	log (IL-18) N=735 Estimate	SE	P	log (CRP) N=821 Estimate	SE	P	Fibrinogen N=823 Estimate	SE	P
cg03636183	*F2RL3*	-0.02	0.04	0.51	-0.11	0.04	0.003	-0.05	0.06	0.400	-2.87	4.66	0.54
cg13500388	*CBFB*	-0.15	0.06	0.01	-0.06	0.06	0.300	-0.30	0.10	0.003	-15.80	7.69	0.04
cg04983977	*GPR25*	0.44	0.17	0.01	0.27	0.18	0.140	0.51	0.30	0.090	46.10	22.82	0.04

Table 3.7 Effect of current smoking status on DNA methylation levels and inflammatory markers.

		Step 1 Mendelian randomization				Step 2 Mendelian randomization	
CpG site	Gene	Estimate	SE	P	Change in M-value	Inflammatory marker	Change
cg03636183	*F2RL3*	-0.660	0.087	<.0001	-0.44	IL-18	5% increase
cg19859270	*GPR15*	-0.203	0.029	<.0001	-0.14		
cg13500388	*CBFB*	-0.138	0.043	0.001	-0.09	CRP	3% increase
cg04983977	*GPR25*	0.093	0.030	0.002	0.06		

CBFB: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, IL-18: interleukin-18, CRP: C-reactive protein

References

1. Ockene IS, Miller NH. Cigarette smoking, cardiovascular disease, and stroke: A statement for healthcare professionals from the american heart association. American heart association task force on risk reduction. *Circulation*. 1997;96:3243-3247
2. Wannamethee SG, Lowe GD, Shaper AG, Rumley A, Lennon L, Whincup PH. Associations between cigarette smoking, pipe/cigar smoking, and smoking cessation, and haemostatic and inflammatory markers for cardiovascular disease. *European heart journal*. 2005;26:1765-1773
3. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27k discovery and replication. *American journal of human genetics*. 2011;88:450-457
4. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, Strauch K, Waldenberger M, Illig T. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PloS one*. 2013;8:e63812
5. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, Belvisi MG, Brown R, Vineis P, Flanagan JM. Epigenome-wide association study in the european prospective investigation into cancer and nutrition (epic-turin) identifies novel genetic loci associated with smoking. *Human molecular genetics*. 2013;22:843-851
6. Dogan MV, Shields B, Cutrona C, Gao L, Gibbons FX, Simons R, Monick M, Brody GH, Tan K, Beach SR, Philibert RA. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from african american women. *BMC genomics*. 2014;15:151
7. Harlid S, Xu Z, Panduri V, Sandler DP, Taylor JA. Cpg sites associated with cigarette smoking: Analysis of epigenome-wide data from the sister study. *Environmental health perspectives*. 2014;122:673-678
8. Feil R, Fraga MF. Epigenetics and the environment: Emerging patterns and implications. *Nature reviews. Genetics*. 2011;13:97-109
9. Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature genetics*. 2003;33 Suppl:245-254
10. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics*. 2007;39:457-466
11. Geiman TM, Muegge K. DNA methylation in early development. *Molecular reproduction and development*. 2010;77:105-113
12. Nagae G, Isagawa T, Shiraki N, Fujita T, Yamamoto S, Tsutsumi S, Nonaka A, Yoshiba S, Matsusaka K, Midorikawa Y, Ishikawa S, Soejima H, Fukayama M, Suemori H, Nakatsuji N, Kume S, Aburatani H. Tissue-specific demethylation in cpg-poor promoters during cellular differentiation. *Human molecular genetics*. 2011;20:2710-2721
13. Goto T, Monk M. Regulation of x-chromosome inactivation in development in mice and humans. *Microbiology and molecular biology reviews : MMBR*. 1998;62:362-378
14. Sun YV, Smith AK, Conneely KN, Chang Q, Li W, Lazarus A, Smith JA, Almli LM, Binder EB, Klengel T, Cross D, Turner ST, Ressler KJ, Kardia SL. Epigenomic association analysis identifies smoking-related DNA methylation sites in african americans. *Human genetics*. 2013;132:1027-1037

15. Smith GD, Ebrahim S. Mendelian randomization: Prospects, potentials, and limitations. *International journal of epidemiology*. 2004;33:30-42
16. Greenland S. An introduction to instrumental variables for epidemiologists. *International journal of epidemiology*. 2000;29:722-729
17. Relton CL, Davey Smith G. Two-step epigenetic mendelian randomization: A strategy for establishing the causal role of epigenetic processes in pathways to disease. *International journal of epidemiology*. 2012;41:161-176
18. Investigators F. Multi-center genetic study of hypertension: The family blood pressure program (fbpp). *Hypertension*. 2002;39:3-9
19. Daniels PR, Kardia SL, Hanis CL, Brown CA, Hutchinson R, Boerwinkle E, Turner ST, Genetic Epidemiology Network of Arteriopathy s. Familial aggregation of hypertension treatment and control in the genetic epidemiology network of arteriopathy (genoa) study. *The American journal of medicine*. 2004;116:676-681
20. investigators TA. The atherosclerosis risk in communities (aric) study: Design and objectives. *American journal of epidemiology*. 1989;129:687-702
21. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56-65
22. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nature methods*. 2013;10:5-6
23. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics*. 2012;44:955-959
24. Du P, Kibbe WA, Lin SM. Lumi: A pipeline for processing illumina microarray. *Bioinformatics*. 2008;24:1547-1548
25. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of beta-value and m-value methods for quantifying methylation levels by microarray analysis. *BMC bioinformatics*. 2010;11:587
26. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics*. 2012;13:86
27. Kullo IJ, Seward JB, Bailey KR, Bielak LF, Grossardt BR, Sheedy PF, 2nd, Peyser PA, Turner ST. C-reactive protein is related to arterial wave reflection and stiffness in asymptomatic subjects from the community. *American journal of hypertension*. 2005;18:1123-1129
28. Khaleghi M, Singletary LA, Kondragunta V, Bailey KR, Turner ST, Mosley TH, Jr., Kullo IJ. Haemostatic markers are associated with measures of vascular disease in adults with hypertension. *Journal of human hypertension*. 2009;23:530-537
29. Kim CX, Bailey KR, Klee GG, Ellington AA, Liu G, Mosley TH, Jr., Rehman H, Kullo IJ. Sex and ethnic differences in 47 candidate proteomic markers of cardiovascular disease: The mayo clinic proteomic markers of arteriosclerosis study. *PloS one*. 2010;5:e9065
30. Hernan MA, Robins JM. Instruments for causal inference: An epidemiologist's dream? *Epidemiology*. 2006;17:360-372
31. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T,

Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Cahill DP, Louis DN, Goldstraw P, Nicholson AG, Brasseur F, Looijenga L, Weber BL, Chiew YE, DeFazio A, Greaves MF, Green AR, Campbell P, Birney E, Easton DF, Chenevix-Trench G, Tan MH, Khoo SK, Teh BT, Yuen ST, Leung SY, Wooster R, Futreal PA, Stratton MR. Patterns of somatic mutation in human cancer genomes. *Nature*. 2007;446:153-158

32. Stephens SH, Hartz SM, Hoft NR, Saccone NL, Corley RC, Hewitt JK, Hopfer CJ, Breslau N, Coon H, Chen X, Ducci F, Dueker N, Franceschini N, Frank J, Han Y, Hansel NN, Jiang C, Korhonen T, Lind PA, Liu J, Lyytikainen LP, Michel M, Shaffer JR, Short SE, Sun J, Teumer A, Thompson JR, Vogelzangs N, Vink JM, Wenzlaff A, Wheeler W, Yang BZ, Aggen SH, Balmforth AJ, Baumeister SE, Beaty TH, Benjamin DJ, Bergen AW, Broms U, Cesarini D, Chatterjee N, Chen J, Cheng YC, Cichon S, Couper D, Cucca F, Dick D, Foroud T, Furberg H, Giegling I, Gillespie NA, Gu F, Hall AS, Hallfors J, Han S, Hartmann AM, Heikkila K, Hickie IB, Hottenga JJ, Jousilahti P, Kaakinen M, Kahonen M, Koellinger PD, Kittner S, Konte B, Landi MT, Laatikainen T, Leppert M, Levy SM, Mathias RA, McNeil DW, Medland SE, Montgomery GW, Murray T, Nauck M, North KE, Pare PD, Pergadia M, Ruczinski I, Salomaa V, Viikari J, Willemsen G, Barnes KC, Boerwinkle E, Boomsma DI, Caporaso N, Edenberg HJ, Francks C, Gelernter J, Grabe HJ, Hops H, Jarvelin MR, Johannesson M, Kendler KS, Lehtimaki T, Magnusson PK, Marazita ML, Marchini J, Mitchell BD, Nothen MM, Penninx BW, Raitakari O, Rietschel M, Rujescu D, Samani NJ, Schwartz AG, Shete S, Spitz M, Swan GE, Volzke H, Veijola J, Wei Q, Amos C, Cannon DS, Grucza R, Hatsukami D, Heath A, Johnson EO, Kaprio J, Madden P, Martin NG, Stevens VL, Weiss RB, Kraft P, Bierut LJ, Ehringer MA. Distinct loci in the chrna5/chrna3/chrnb4 gene cluster are associated with onset of regular smoking. *Genetic epidemiology*. 2013;37:846-859
33. Keskitalo K, Broms U, Heliovaara M, Ripatti S, Surakka I, Perola M, Pitkaniemi J, Peltonen L, Aromaa A, Kaprio J. Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (chrna3/chrna5/chrnb4) on chromosome 15. *Human molecular genetics*. 2009;18:4007-4012
34. Greenbaum L, Rigbi A, Teltsh O, Lerer B. Role of genetic variants in the chrna5-chrna3-chrnb4 cluster in nicotine dependence risk: Importance of gene-environment interplay. *Molecular psychiatry*. 2009;14:828-830
35. Berrettini WH, Doyle GA. The chrna5-a3-b4 gene cluster in nicotine addiction. *Molecular psychiatry*. 2012;17:856-866
36. Chen LS, Baker TB, Piper ME, Breslau N, Cannon DS, Doheny KF, Gogarten SM, Johnson EO, Saccone NL, Wang JC, Weiss RB, Goate AM, Bierut LJ. Interplay of genetic risk factors (chrna5-chrna3-chrnb4) and cessation treatments in smoking cessation success. *The American journal of psychiatry*. 2012;169:735-742
37. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in mendelian randomization studies. *International journal of epidemiology*. 2011;40:755-764
38. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grutzner F, Kaessmann H. The evolution of lncrna repertoires and expression patterns in tetrapods. *Nature*. 2014;505:635-640

39. Park C, Yu N, Choi I, Kim W, Lee S. Lncrnator: A comprehensive resource for functional investigation of long non-coding rnas. *Bioinformatics*. 2014;30:2480-2485
40. Gong J, Liu W, Zhang J, Miao X, Guo AY. Lncrnasnp: A database of snps in lncrnas and their potential functions in human and mouse. *Nucleic acids research*. 2015;43:D181-186
41. Shiels MS, Katki HA, Freedman ND, Purdue MP, Wentzensen N, Trabert B, Kitahara CM, Furr M, Li Y, Kemp TJ, Goedert JJ, Chang CM, Engels EA, Caporaso NE, Pinto LA, Hildesheim A, Chaturvedi AK. Cigarette smoking and variations in systemic immune and inflammation markers. *Journal of the National Cancer Institute*. 2014;106
42. Breitling LP, Salzmann K, Rothenbacher D, Burwinkel B, Brenner H. Smoking, f2rl3 methylation, and prognosis in stable coronary heart disease. *European heart journal*. 2012;33:2841-2848
43. Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. F2rl3 methylation as a biomarker of current and lifetime smoking exposures. *Environmental health perspectives*. 2014;122:131-137
44. Zhang Y, Yang R, Burwinkel B, Breitling LP, Holleczek B, Schottker B, Brenner H. F2rl3 methylation in blood DNA is a strong predictor of mortality. *International journal of epidemiology*. 2014;43:1215-1225
45. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, Agusti A, Anderson W, Lomas DA, Demeo DL. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Human molecular genetics*. 2012;21:3073-3082
46. Muehlschlegel JD, Perry TE, Liu KY, Fox AA, Smith SA, Lichtner P, Collard CD, Shernan SK, Hartwig JH, Body SC, Hoffmeister KM. Polymorphism in the protease-activated receptor-4 gene region associates with platelet activation and perioperative myocardial injury. *American journal of hematology*. 2012;87:161-166
47. Han W, Wang Z, Lu X, Guo C. Protease activated receptor 4 status of mast cells in post infectious irritable bowel syndrome. *Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society*. 2012;24:113-119, e182
48. Zhang Y, Yu G, Jiang P, Xiang Y, Li W, Lee W, Zhang Y. Decreased expression of protease-activated receptor 4 in human gastric cancer. *The international journal of biochemistry & cell biology*. 2011;43:1277-1283
49. Chuang LS, Ito K, Ito Y. Runx family: Regulation and diversification of roles through interacting proteins. *International journal of cancer. Journal international du cancer*. 2013;132:1260-1271
50. Mendoza-Villanueva D, Deng W, Lopez-Camacho C, Shore P. The runx transcriptional co-activator, cbfbeta, is essential for invasion of breast cancer cells. *Molecular cancer*. 2010;9:171
51. Davis JN, Rogers D, Adams L, Yong T, Jung JS, Cheng B, Fennell K, Borazanci E, Moustafa YW, Sun A, Shi R, Glass J, Mathis JM, Williams BJ, Meyers S. Association of core-binding factor beta with the malignant phenotype of prostate and ovarian cancer cells. *Journal of cellular physiology*. 2010;225:875-887
52. O'Reilly J, Chipper L, Springall F, Herrmann R. A unique structural abnormality of chromosome 16 resulting in a cbf beta-myh11 fusion transcript in a patient with acute myeloid leukemia, fab m4. *Cancer genetics and cytogenetics*. 2000;121:52-55
53. Dietrich T, Garcia RI, de Pablo P, Schulze PC, Hoffmann K. The effects of cigarette smoking on c-reactive protein concentrations in men and women and its modification by

exogenous oral hormones in women. *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*. 2007;14:694-700

54. Hastie CE, Haw S, Pell JP. Impact of smoking cessation and lifetime exposure on c-reactive protein. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco*. 2008;10:637-642
55. Koenig W, Sund M, Frohlich M, Fischer HG, Lowel H, Doring A, Hutchinson WL, Pepys MB. C-reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: Results from the monica (monitoring trends and determinants in cardiovascular disease) augsburg cohort study, 1984 to 1992. *Circulation*. 1999;99:237-242
56. Suchanek H, Mysliwska J, Siebert J, Wieckiewicz J, Hak L, Szyndler K, Kartanowicz D. High serum interleukin-18 concentrations in patients with coronary artery disease and type 2 diabetes mellitus. *European cytokine network*. 2005;16:177-185
57. Thorand B, Lowel H, Schneider A, Kolb H, Meisinger C, Frohlich M, Koenig W. C-reactive protein as a predictor for incident diabetes mellitus among middle-aged men: Results from the monica augsburg cohort study, 1984-1998. *Archives of internal medicine*. 2003;163:93-99
58. Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, Rudan I, McKeigue P, Wilson JF, Campbell H. Abundant pleiotropy in human complex diseases and traits. *American journal of human genetics*. 2011;89:607-618
59. May S, Bigelow C. Modeling nonlinear dose-response relationships in epidemiologic studies: Statistical approaches and practical challenges. *Dose-response : a publication of International Hormesis Society*. 2005;3:474-490
60. Burgess S, Daniel RM, Butterworth AS, Thompson SG, the E-IC. Network mendelian randomization: Using genetic variants as instrumental variables to investigate mediation in causal pathways. *International journal of epidemiology*. 2014

Chapter 4

Modeling causal direct and indirect effects of smoking mediated by DNA methylation on inflammation, and subclinical markers of cardiovascular disease

Introduction

Cigarette smoking induces inflammatory responses and is a risk factor for cardiovascular disease (CVD).[1, 2] Cigarette smoking has been found to be associated with several inflammatory markers including C-reactive protein (CRP),[3] interleukin 6 (IL-6),[4] interleukin 18 (IL-18),[5] and fibrinogen.[6] However, the biological and genetic mechanisms behind the smoking-induced inflammatory response are not well understood.

DNA methylation has been found to be responsive to environmental changes and plays an important role in regulation of gene expression levels.[7] In the previous study (Aim 2), we investigated the role of DNA methylation as a mediator of the association between smoking and inflammation. We identified specific CpG sites that had differential methylation levels by smoking status, and demonstrated the direction of the association from smoking to DNA methylation for cg03636183 in coagulation factor II (thrombin) receptor-like 3 (*F2RL3*) gene, cg19859270 in G protein-coupled receptor 15 (*GPR15*) gene, cg13500388 in core-binding factor beta subunit (*CBFB*) gene, and cg04983977 in G protein-coupled receptor 25 (*GPR25*) gene using Mendelian randomization (MR), an approach for utilizing a genetic variant as an instrument to examine a causal association in observational studies. The genetic variants were

single nucleotide polymorphisms (SNPs). In addition, we examined the associations between DNA methylation levels of the identified CpG sites and levels of inflammatory markers, and demonstrated that the changes in DNA methylation levels are associated with the levels of inflammatory markers such as CRP and IL-18 using the two-step epigenetic MR approach.[8]

Although the two-step MR is useful to assess an indirect effect that causally mediates the exposure-outcome association, it is limited in the evaluation of the magnitude of the indirect effect relative to the total effect. If there are several pathways from smoking to inflammation, it is likely that the estimates from the previous approach could be biased because it assumes the only path through the mediator of interest. Therefore, it is necessary to evaluate the mediated effect by DNA methylation in consideration of the possible direct effect. In this study, we combine the two-step epigenetic MR approach with causal mediation analysis, which is also known as the network MR approach.[9] We also extended the network MR approach for more complex causal paths using structural equation modeling (SEM).

In this study, we first investigated the direct and indirect effects of smoking on inflammatory markers through DNA methylation. Following that, the causal paths of smoking, DNA methylation, and inflammatory markers were extended to include subclinical markers of cardiovascular disease.

Methods

Study population

The Genetic Epidemiology Network of Arteriopathy (GENOA) was established as a part of the Family Blood Pressure Program by the National Heart Lung and Blood Institute in 1996.[10] From its inception, GENOA's long-term objective has been to elucidate the genetics of hypertension

and its arteriosclerotic target-organ damage, including both atherosclerotic (macrovascular) and arteriolosclerotic (microvascular) complications involving the heart, brain, kidneys, and peripheral arteries. GENOA is a community based study of sibships that incudes individuals with a personal or family history of hypertension. The African-American cohort of GENOA, which is the focus of Aims 2 and 3, is from Jackson, Mississippi.[11] In Jackson, the sibships were recruited through hypertensive probands who had participated in the Atherosclerosis Risk in Communities (ARIC) study, which is a probability sample of 45 to 64-year-old African American residents of the community.[12] At the enrollment stage of GENOA (Phase I: 1995-2000), 1,854 subjects who were members of 683 sibships with at least two individuals diagnosed with essential hypertension before age sixty were recruited into the study. All siblings in the sibship were invited to participate, both normotensive and hypertensive. Thus, everyone had a personal or family history of hypertension. Exclusion criteria of the GENOA study were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. Eighty percent (1,482 subjects) of the initial study population returned for the second examination (Phase II: 2000-2005). Study visits were made in the morning after an overnight fast of at least eight hours. Data on demographic information, medical history, clinical characteristics, lifestyle factors, and blood samples were collected in each phase. Written informed consent was obtained from all subjects and approval was granted by participating institutional review boards. Because both DNA methylation levels and echocardiograms were measured at the second examination, only data from Phase II were used in this study.

Genotypes

For 1,263 participants, genotype data was obtained using the Affymetrix® Genome-Wide Human SNP Array 6.0 platform for about one million SNPs. For 269 samples, genotype data was

obtained using Illumina® Human 1M-Duo BeadChip for about one million SNPs. For 508 samples, genotype data was obtained using the Affymetrix® Genome-Wide Human SNP Array 6.0 platform for about one million SNPs, separately. Quality control was performed based on the following exclusion criteria in order: <95% SNP call rate, <1% MAF, and <95% sample call rate.

To get more dense markers, we imputed the genotype data using the 1,000 Genomes Project data.[13] The three sets of samples were pre-phased and imputed separately. The DNA strands were aligned to positive strand. A physical map was updated from NCBI build 36 (hg 18) to build 37 (hg 19) using LiftOver software (http://genome.sph.umich.edu/wiki/LiftOver). Pre-phasing was done using the Segmented HAPlotype Estimation and Imputation Tool (SHAPEIT) version 2.[14] Pre-phasing was done separately for each chromosome. Imputation was performed using IMPUTE version 2[15] and reference panels from the 1,000 Genomes project Phase I integrated variant set release (v3) in NCBI build 37 (hg19) coordinates (released on March 2012). For imputation, each chromosome was split into 5 mega base pair (Mbp) long segments, resulting in 565 segments across the genome. Outlier samples from principal component analyses (>6 standard deviations) were excluded. Monomorphic markers were also excluded. Finally, there were 30,022,375 imputed SNPs for 1599 samples.

DNA methylation

DNA methylation levels were measured from blood samples collected in Phase II. Peripheral blood leukocytes were isolated from blood samples and used to measure DNA methylation levels. The EZ DNA Methylation Gold Kit (Zymo Research, Orange CA) was used for bisulfite conversion. The methylation assay was performed at the Mayo Clinic Advanced Genomics Technology Center using Illumina® Infinium HumanMethylation27 BeadChips and Illumina BeadXpress reader. DNA methylation was measured from blood samples (obtained in the second

examination) on a single array, in two different color channels (red and green). A total of 1,008 individuals were measured for 27,578 CpG sites.

As a quality control, seven samples were excluded from analysis due to poor bisulfite conversion efficiency (intensity < 4,000). An additional 28 samples were removed because of poor background signals. Finally, there were 973 samples left. The *lumi* package[16] in R software was used for background adjustment, color balance adjustment, and quantile normalization. After quality control and normalization, the distribution of methylation intensity became similar across the samples. The CpG site intensities were used to calculate beta-values and M-values, a ratio of methylated allele intensity and total intensity+100, and a log2 ratio of methylated and unmethylated probe intensities, respectively.[17] Cell proportions for CD8 T lymphocytes, CD4 T lymphocytes, natural killer cells, B cells, monocytes, granulocytes were estimated using Houseman's method.[18]

Smoking variable

Current smoking status was obtained during Phase II from questionnaire asking a question "Do you now smoke cigarettes?".

Inflammatory markers

Inflammatory markers were measured from blood samples obtained during Phase II. Circulating C-reactive protein (CRP) levels were measured by a highly sensitive immunoturbidimetric assay.[19] Fibrinogen level was measured from citrated plasma using the clotting time based Clauss method.[20] Interleukin-6 (IL-6) and interleukin-18 (IL-18) were measured using a contracted service with SearchLight™ Technologies (Boston, MA).[21] CRP, IL-6, and IL-18 were log-transformed to reduce skewness.

Echocardiogram

Echocardiograms were performed at the second examination in Jackson. Two-dimensional, M-mode, and Doppler echocardiography was performed at the Echocardiography Laboratory in the Mayo Clinic using an Acuson 128XP echo machine (Acuson, Malvern, PA).[22, 23] The parasternal long axis view was used to measure left ventricular ejection fraction. Penn and American Society of Echocardiography (ASE) conventions and the cube function formula was used to measure left ventricular mass.[24] Left ventricular mass index was obtained by normalizing the left ventricular mass to height$^{2.7}$.[25]

Data Analysis

The allele dosage from imputation was used for SNPs assuming an additive inheritance model. For DNA methylation levels, M-values were used as a primary measure because M-values have better statistical validity than beta-values.[17] To take into account the correlation among siblings, mixed effects models with a random intercept were implemented. Assumption checking was done using the nlme R package and SAS® 9.4. Application of network MR was done using Stata 13.

Causal mediation analysis with instrumental variables

We assessed the indirect effect mediated by DNA methylation in comparison to the direct effect of smoking on inflammatory response by implementing a causal mediation analysis with instrumental variables (**Figure 4.1**).

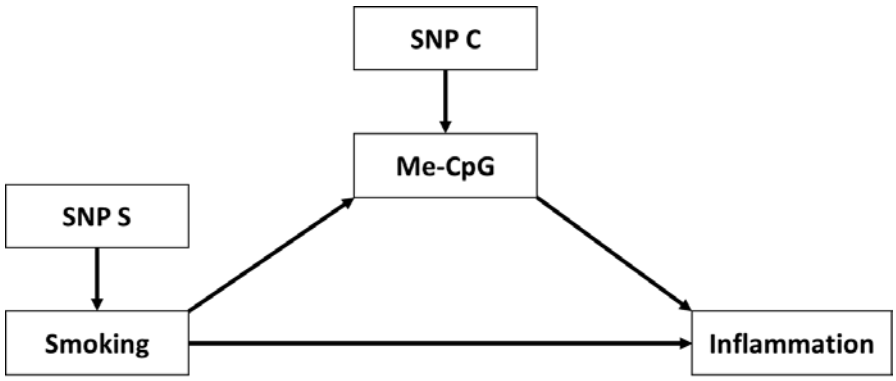

Figure 4.1. Application of Network Mendelian randomization on smoking, DNA methylation (Me-CpG), and inflammation. Rectangular boxes represent variables of interest and arrows represent causal relationships. SNP S is an instrument for smoking status and SNP C is an instrument for DNA methylation of a CpG site.

Assumption checking

There are three sets of assumptions for this approach: assumptions for instrumental variable analysis[26], assumptions for causal mediation analysis,[27] and assumptions to combine instrumental variable analysis with causal mediation analysis.[28] The assumptions for instrumental variable analysis were checked in Chapter 3, so here we will discuss the other sets of assumptions.

B. Assumptions for causal mediation analysis[27]

   a. No bidirectional associations: the bidirectional associations, inflammatory responses to DNA methylation, and from inflammatory responses to smoking were assessed using the reciprocal Mendelian randomization approach.[29] The reciprocal MR approach is similar to the original MR approach except that it is used to investigate whether there is an inverse association. To examine associations from inflammatory markers to smoking, and from inflammatory markers to DNA methylation, we first identified appropriate

instrument SNPs for log(CRP), log(IL-6), log(IL-18), and fibrinogen. Genome-wide association studies (GWAS) have identified SNPs that are associated with serum CRP[30-32], IL-6[33, 34], IL-18[35, 36], and fibrinogen levels[37-40]. The SNPs consistently found in GWAS were located in *CRP*[30-32]; *IL6*, and *IL6R*[32, 33]; *IL18*, and *BCO2*[35, 36]; *FGA*, *FGB*, and *FGG*[37-40] genes, respectively. We investigated the genes ±10kb to identify the most significantly associated SNPs with each inflammatory marker level, and used them as instrument SNPs for the reciprocal MR approach.

b. No interaction effect of exposure and mediator on outcome: to check this assumption, the interaction effect of smoking and DNA methylation on inflammatory markers was assessed using the linear mixed effects model (Equation 4.1). The significance of the interaction term was assessed using a Wald test.

$$Inflammatory\ marker = \beta_0 + \beta_1 * Smoking_{ij} + \beta_2 * DNA\ methylation\ levels_{ij} + \beta_3 * Smoking_{ij} * DNA\ methylation\ levels_{ij} + \beta_4 * age_{ij} + \beta_5 * gender_{ij} + \beta_6 * CD8T_{ij} + \beta_7 * CD4T_{ij} + \beta_8 * NK_{ij} + \beta_9 * Bcell_{ij} + \beta_{10} * Mono_{ij} + \beta_{11} * Gran_{ij} + u_j + \varepsilon_{ij} \quad (Equation\ 4.1)$$

where inflammatory markers are log(CRP), log(IL-6), log(IL-18), and fibrinogen.

C. Assumptions for combining instrumental variable analysis with causal mediation analysis[41]

a. Sequential ignorability[28] means that the exposure should be statistically independent of mediators and outcomes, and the mediator should be statistically independent of the outcome given exposure and confounders. Using instrumental variables for the exposure and the mediator satisfies the first condition but not the second condition necessarily. If there is another mediator (X) in the pathway which either has a causal effect on or is

influenced by the mediator of interest, the second condition is not met (**Figure 4.2**). It is possible that other DNA methylation sites mediating smoking and inflammation are associated with the DNA methylation site of interest. The variable X can also be a confounder between the mediator and the outcome that is affected by the exposure. Cigarette smoking is found to be associated with an elevated level of peripheral blood leukocytes.[42] Hence, cell proportions, which are affected by cigarette smoking and are confounders in the association between DNA methylation and inflammation, are also an example of X.

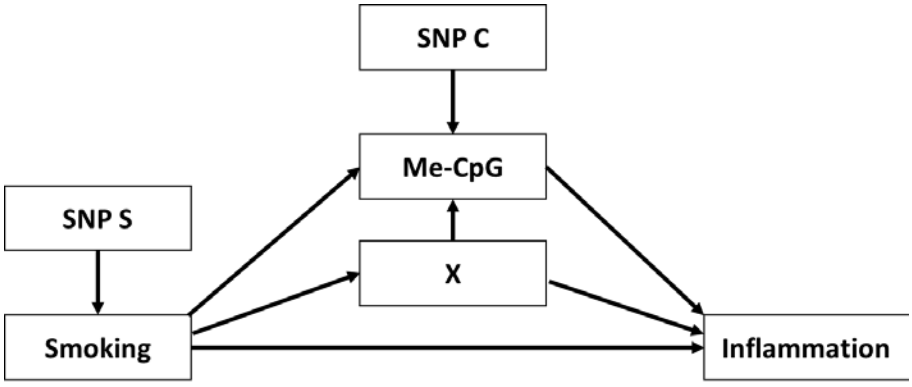

Figure 4.2. The example of violation of sequential ignorability. Rectangular boxes represent variables of interest and arrows represent causal relationships. SNP S is an instrument for smoking status and SNP C is an instrument for DNA methylation of a CpG site (Me-CpG).

When there is another mediator (X) in the pathway which either has a causal effect on or is influenced by the mediator of interest, the error terms of the mediator of interest and the outcome will not be independent. The violation of this assumption can be assessed using a sensitivity analysis that tests whether the correlation of the two error terms is significantly different from zero.[28]

Standardization

Because of poor convergence of the structural equation modeling (SEM) method, the effects of age and gender were eliminated by obtaining residuals from either logistic mixed effect modeling or linear mixed effect modeling. All the residuals were standardized to follow the normal distribution with mean of zero and standard deviation of one, *N(0,1),* for more intuitive interpretation.

Structural equation modeling (SEM)

The path diagram can be transformed to a path model using software for SEM. The network MR method was implemented using Stata 13.[43]

Results

Study population descriptive

The total of 848 African-Americans in GENOA have both measured genotypes and DNA methylation levels and were included in the analysis. The mean age was 67 years (standard deviation (SD) = 7 years), and 72% of the population were women. In this study population, 11% were current smokers, 30% were previous smokers, and 59% were never smokers. The majority of the study participants have been diagnosed for hypertension (77%), as expected (**Table 4.1**).

Assumption checking

B. Assumptions for causal mediation analysis[27]
    a. No bidirectional association
        i. No bidirectional association between smoking and DNA methylation: we investigated the bidirectional effect using reciprocal MR. After applying a Bonferroni correction

(threshold = 0.05/15 = 0.0033), cg13633560 on *LRRC32* gene, cg03636183 on *F2RL3* gene, cg19859270 on *LRRC32* gene, and cg17791651 on *POU3F1* gene were significantly associated with current smoking status (**Table 4.2**).

ii. No bidirectional association between DNA methylation and inflammatory markers including log(CRP), log(IL-6), log(IL-18), and fibrinogen: After applying Bonferroni correction (threshold = 0.05/4 = 0.0125), log(CRP) levels were significantly associated with the DNA methylation levels of cg19859270 on *GPR15* gene (**Table 4.3**). Even though the p-values did not satisfy the threshold, log(IL-6) levels were marginally associated with the DNA methylation levels of cg19859270 on *GPR15* gene (p=0.06), and log(IL-18) levels were associated with cg03636183 on *F2RL3* gene (p=0.04) and cg14223444 on *NCBP1* gene (p=0.05) (**Table 4.4-4.6**).

iii. No bidirectional association between smoking and inflammatory markers. After applying Bonferroni correction (threshold = 0.05/4 = 0.0125), none of the inflammatory markers were associated with current smoking status (**Table 4.7**). However, all four inflammatory markers were marginally associated with current smoking status.

b. No interaction effect of exposure and mediator on outcome: we tested the interaction effects of smoking and DNA methylation on inflammatory markers including log(CRP), log(IL-6), log(IL-18), and fibrinogen. After applying Bonferroni correction (thereshold = 0.05/15 = 0.0033), none of the interaction terms were significant (**Table 4.8-4.11**). However, the interaction between smoking status and DNA methylation levels of cg14223444 on *NCBP1* gene was close to being significant for log(IL-18) (P=0.006, **Table 4.10**)

C. Assumption for combining instrumental variable analysis with causal mediation analysis
    a. Sequential ignorability: the correlation between the error terms of DNA methylation and inflammatory marker levels were checked. None of them reject the null hypothesis of zero correlation with an alpha level of 0.05.

Direct and indirect effects

We evaluated both the direct and indirect effects of smoking on inflammatory markers considering DNA methylation as a mediator. In our previous study (Aim 2), we identified the associations between smoking status and DNA methylation levels of cg03636183 on *F2RL3* gene, and cg03636183 and log(IL-18) levels. In this study, we reevaluated the effect in consideration of both direct and indirect effects through DNA methylation levels of cg03636183 (**Table 4.12**). The direct effects of smoking were significant for log(CRP), log(IL-6), and fibrinogen but not for log(IL-18) levels. The indirect effects of smoking through DNA methylation were significant for all the four inflammatory markers. However, when the indirect effect was decomposed to the effects from smoking to DNA methylation and from DNA methylation to inflammatory markers, the second effect was only significant for log(IL-18), which is consistent with the results from the Aim 2. The estimate of the indirect effect on log(IL-18) was twice the estimate for the direct effect.

We also reevaluated the effect of smoking on inflammatory markers in consideration of both direct and indirect effects through DNA methylation levels of cg13500388 on *CBFB* gene (**Table 4.13**). In our previous study (Aim2), we identified the associations from smoking status to DNA methylation levels of cg13500388 on *CBFB* gene, and from the CpG site to log(CRP) levels. When both direct and indirect effects were considered, there was a significant direct effect of smoking for log(CRP). This suggests that smoking confounds the association between the

DNA methylation levels of cg13500388 and log(CRP) levels. Hence, when the direct effect of smoking was considered, the effect from the DNA methylation levels of cg13500388 to log(CRP) levels was no longer statistically significant.

Extension to subclinical markers of cardiovascular diseases

The identified path of smoking, cg03636183 on *F2RL3*, and log(IL-18) was extended to include selected subclinical markers of cardiovascular disease such as left ventricular ejection fraction (LVEF) and left ventricular mass index (LVMI). The indirect path did not have a significant effect on LVEF (**Figure 4.3**). On the other hand, the indirect path had a significant effect on the LVMI (**Figure 4.4**).

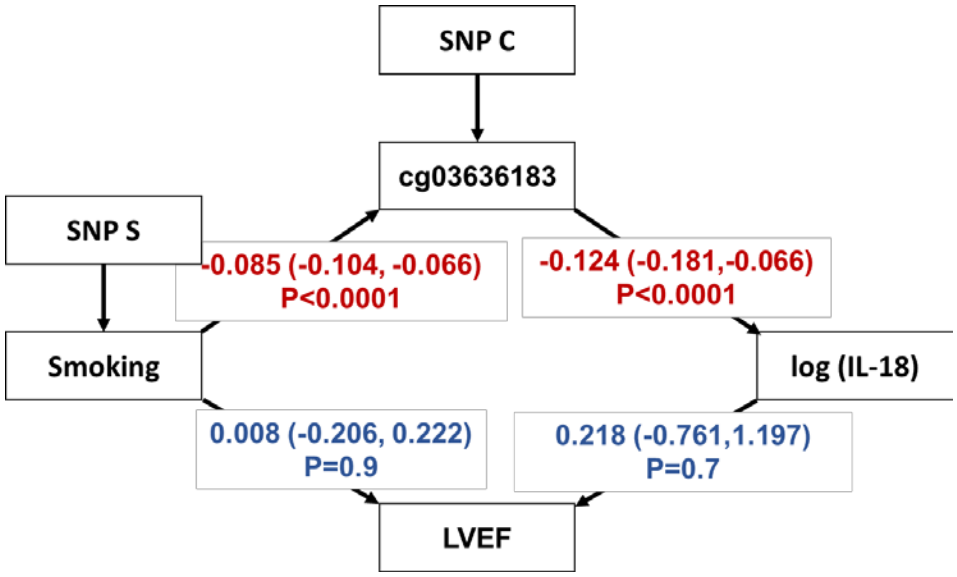

Figure 4.3 Estimates of the path analysis of smoking, cg03636183 on *F2RL3*, log(IL-18), and left ventricular ejection fraction (LVEF). Boxes represent each of the variables considered and arrows represent directed effects. Smoking, cg03636183, log(IL-18) and LVEF were standardized to follow N(0,1). The numbers in the arrow represent estimate, 95% confidence interval, and p-value. Red color indicates significant estimates and blue color indicates non-significant estimates.

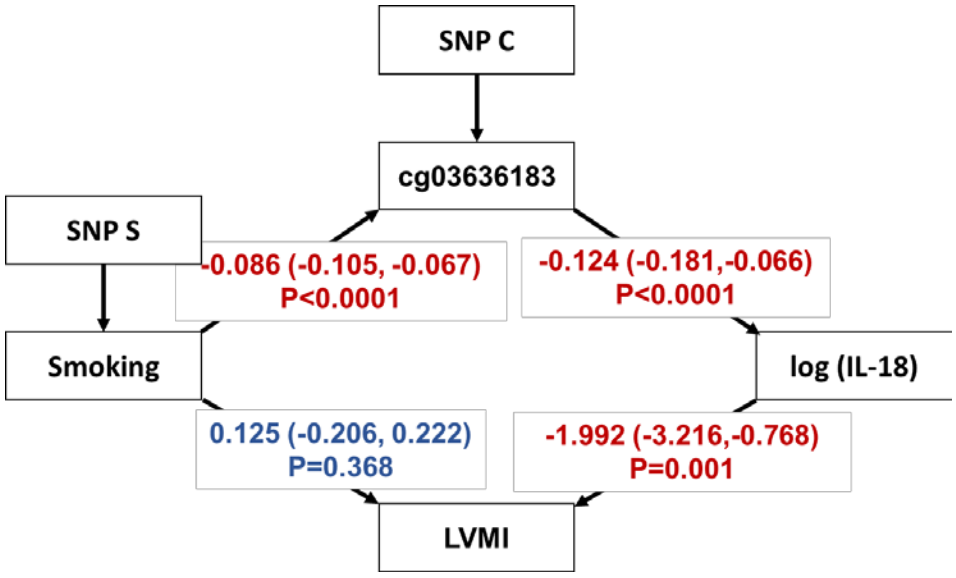

Figure 4.4 Estimates of the path analysis of smoking, cg03636183 on *F2RL3*, log(IL-18), and left ventricular mass index (LVMI). Boxes represent each of the variables considered and arrows represent directed effects. Smoking, cg03636183, log(IL-18) and LVMI were standardized to follow N(0,1). The numbers in the arrow represent estimate, 95% confidence interval, and p-value. Red color indicates significant estimates and blue color indicates non-significant estimates.

When we investigated the total effect of smoking on LVEF and LVMI, the total effect was not significant (**Table 4.14**). We would be able to identify the indirect path using this approach that would not be identified using conventional regression based approaches.

Discussion

We investigated both direct and indirect effects from smoking to inflammatory markers mediated through DNA methylation. The significant mediated effect of cg03636183 on *F2RL3* was consistent with the results from the two-step epigenetic MR approach[8], and the mediated effect was also associated with LVMI. However, we did not find other significant mediated effects. It may be because of violation of assumptions required for causal mediation analysis, because of low statistical power, or because there are no other mediated effects.

When we checked the assumptions, we identified some bidirectional effects using reciprocal MR. The inflammatory markers had bidirectional effects on the smoking status. It may be because individuals who experience symptoms or who are diagnosed for a disease are often recommended to or voluntarily stop smoking. Both DNA methylation and inflammatory markers change due to external exposures. Therefore, it is also possible that inflammatory markers also have bidirectional effects on DNA methylation. It is less likely, but possible, that DNA methylation levels have bidirectional effects on smoking status. In fact, behavior such as substance use, learning, parenting, and stress can be affected by epigenetic modifications.[44] Recent findings in animal studies showed that epigenetic modification has an effect on behavior as a response to environmental exposure.[45]

We also checked the interaction effects between smoking status and DNA methylation levels. The DNA methylation level of cg14223444 on *NCBP1* gene had a significant interaction with smoking status on log(IL-18) levels. The slopes of the association between cg14223444 levels and log(IL-18) levels were significantly different between smokers and non-smokers (estimate = -0.45, p=0.006). This is a good example of gene by environment interaction and it suggests that DNA methylation is not just a mediator but may interact with environmental exposure. It suggests that the role of DNA methylation is not limited to a mediator, and other roles and interaction effects should also be carefully considered.

The linearity assumption in the relationship among variables is required for SEM. However, the relationships between variables in genetic or epidemiology studies are often not linear. For example, SNPs that were used as instruments may have dominant or recessive effects rather than the assumed additive effects, and the age effect is often quadratic in epidemiology studies. To avoid bias in estimates, the basic assumptions should be carefully checked.

When the assumptions are satisfied, this approach provides extensibility. By using SEM, this approach can incorporate diverse path models. Taking advantage of SEM, we extended our study to subclinical markers of cardiovascular disease and we identified the directed path of smoking->cg03636183->log(IL-18)->LVMI. Implementing this approach to epidemiologic and genetic data may further explain genetic and/or biological mechanisms behind complex diseases such as cardiovascular disease.

GENOA is a cohort study of individuals with a personal history or family history of hypertension. In the current study, we considered only the African American participants in GENOA and used data only from the second examination, thus, the data are cross-sectional. To provide confirmation of the causal path, this should be further evaluated in a longitudinal study setting. In addition, the results may not be generalizable to other ethnic groups.

In conclusion, the application of network MR and the extension to include more complex paths provides the opportunity to better understand the relations among various factors that influence subclinical measures of CVD. These approaches can be applied to other risk factors in addition to smoking and inflammation for CVD to better understand the role of DNA methylation.

Tables

Table 4.1 Descriptive characteristics of GENOA Jackson study population (N=848) at Phase 2 Examination.

Continuous variables	**N**	**Mean**	**SD**
Age at examination (years)	848	67	7
Height (cm)	844	168	9
Weight (kg)	844	88	18
Body Mass Index (kg/m^2)	844	31	6
Pack years	348	25	22
Age started smoking	348	20	6
Age quitted smoking	252	48	12
Alcohol (number of drinks/week)	122	4	4
Education (years)	848	12	4
Systolic blood pressure (sitting, mmHg)	848	154	22
Diastolic blood pressure (sitting, mmHg)	848	92	12
Pulse pressure (sitting, mmHg)	848	62	18
C reactive protein (mg/L)	821	6	7
Fibrinogen (mg/dL)	823	369	81
Interleukin-6 (pg/ml)	741	10	7
Interleukin-18 (pg/ml)	735	72	45
High density lipoprotein (mg/dL)	848	59	18
Low density lipoprotein (mg/dL)	848	121	38
Lipid lp(a) (mg/dL)	823	63	51
Left ventricular ejection fraction	821	61	8
Left ventricular mass	822	163	48
Left ventricular mass index	819	40	11
Number of siblings diagnosed for hypertension before age of 60	682	3	2
Categorical variables		**N**	**%**
Female		610	72%
Current smoker		96	11%
Previous smoker		252	30%

Never smoker	500	59%
Currently taking blood pressure control medication	634	75%
Ever diagnosed for hypertension	661	78%
Ever diagnosed for heart attack or MI	43	5%
Ever diagnosed for diabetes	249	29%
Ever diagnosed for cancer	71	8%

Table 4.2 Results of reciprocal Mendelian randomization from DNA methylation to smoking status.

CpG site	Gene	Instrument SNP	F-value	Estimate	SE	P
cg13633560	*LRRC32*	rs4945097	40	-1.88	0.48	0.0001
cg03636183*	*F2RL3*	rs2227341	505	-0.58	0.17	0.0008
cg19859270	*GPR15*	rs6790026	3	-6.98	2.10	0.0010
cg17791651	*POU3F1*	rs34860389	31	-1.28	0.39	0.0011
cg03330058	*ABTB1*	rs7627836	155	-0.98	0.34	0.0046
cg26259865*	*ZG16B*	rs11076844	31	-2.04	0.75	0.0072
cg01500140	*LIM2*	rs10409027	14	0.78	0.38	0.0410
cg04983977	*GPR25*	rs2292099	4	1.87	0.98	0.0563
cg00353953*	*ZNF384*	rs35787939	22	-1.54	0.92	0.0961
cg13500388	*CBFB*	rs8048014	630	-0.52	0.33	0.1148
cg03340878*	*OR2B6*	rs276364	9	-0.96	0.65	0.1442
cg13668129*	*HNRNPUL1*	rs12980267	7	-1.42	1.21	0.2407
cg11314684*	*AKT3*	rs10157763	31	-0.78	0.85	0.3563
cg14223444*	*NCBP1*	rs60028882	237	-0.51	0.64	0.4230
cg13745870*	*SPATA12*	rs9857168	8	-0.44	0.76	0.5626

*Because the SAS GLIMMIX procedure did not converge, the sibship effect was modeled as repeated measurements instead of a random effect. *ABTB1*: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 4.3 Results of reciprocal Mendelian randomization from log(CRP) to DNA methylation.

CpG site	Gene	Estimate	SE	Pr > |t|
cg19859270	*GPR15*	-0.083	0.03	0.01
cg03340878	*OR2B6*	0.039	0.03	0.15
cg14223444	*NCBP1*	-0.034	0.03	0.25
cg13745870	*SPATA12*	-0.023	0.02	0.30
cg13668129	*HNRNPUL1*	0.023	0.02	0.31
cg04983977	*GPR25*	0.033	0.03	0.32
cg13500388	*CBFB*	-0.038	0.05	0.42
cg26259865	*LOC124220*	0.026	0.03	0.44
cg03330058	*ABTB1*	0.041	0.05	0.45
cg01500140	*LIM2*	-0.022	0.03	0.48
cg11314684	*AKT3*	-0.014	0.03	0.67
cg13633560	*LRRC32*	0.014	0.04	0.75
cg03636183	*F2RL3*	0.031	0.10	0.75
cg00353953	*ZNF384*	0.009	0.04	0.80
cg17791651	*POU3F1*	0.007	0.05	0.90

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 4.4 Results of reciprocal Mendelian randomization from log(IL-6) to DNA methylation.

CpG site	Gene	Estimate	SE	Pr > |t|
cg19859270	*GPR15*	-0.184	0.10	0.06
cg01500140	*LIM2*	0.128	0.09	0.15
cg13745870	*SPATA12*	-0.092	0.06	0.15
cg11314684	*AKT3*	0.116	0.10	0.24
cg13500388	*CBFB*	-0.158	0.14	0.26
cg14223444	*NCBP1*	0.083	0.09	0.35
cg13668129	*HNRNPUL1*	-0.059	0.07	0.38
cg03330058	*ABTB1*	-0.117	0.16	0.48
cg04983977	*GPR25*	0.054	0.10	0.58
cg17791651	*POU3F1*	-0.069	0.16	0.67
cg13633560	*LRRC32*	0.034	0.13	0.78
cg03636183	*F2RL3*	-0.039	0.30	0.90
cg26259865	*LOC124220*	0.012	0.10	0.91
cg03340878	*OR2B6*	0.004	0.08	0.96
cg00353953	*ZNF384*	-0.001	0.11	0.99

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 4.5 Results of reciprocal Mendelian randomization from log(IL-18) to DNA methylation.

CpG site	Gene	Estimate	Std	Pr > |t|
cg03636183	*F2RL3*	-0.613	0.29	0.04
cg14223444	*NCBP1*	0.168	0.09	0.05
cg03340878	*OR2B6*	0.093	0.08	0.24
cg26259865	*LOC124220*	-0.100	0.10	0.31
cg17791651	*POU3F1*	-0.135	0.16	0.39
cg13633560	*LRRC32*	0.104	0.12	0.40
cg13668129	*HNRNPUL1*	0.052	0.07	0.43
cg11314684	*AKT3*	0.073	0.10	0.45
cg00353953	*ZNF384*	0.055	0.10	0.59
cg03330058	*ABTB1*	0.076	0.16	0.64
cg13745870	*SPATA12*	0.023	0.06	0.72
cg19859270	*GPR15*	-0.023	0.09	0.81
cg04983977	*GPR25*	0.020	0.10	0.84
cg01500140	*LIM2*	0.012	0.09	0.89
cg13500388	*CBFB*	0.001	0.14	0.99

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 4.6 Results of reciprocal Mendelian randomization from fibrinogen to DNA methylation.

CpG site	Gene	Estimate	Std	Pr > |t|
cg14223444	*NCBP1*	-0.0009	0.0005	0.08
cg03340878	*OR2B6*	0.0008	0.0004	0.09
cg19859270	*GPR15*	-0.0009	0.0005	0.11
cg03636183	*F2RL3*	-0.0018	0.0017	0.28
cg13668129	*HNRNPUL1*	0.0004	0.0004	0.35
cg11314684	*AKT3*	0.0005	0.0005	0.39
cg17791651	*POU3F1*	-0.0007	0.0009	0.41
cg26259865	*LOC124220*	0.0004	0.0006	0.51
cg13745870	*SPATA12*	-0.0002	0.0004	0.54
cg00353953	*ZNF384*	0.0003	0.0006	0.59
cg13500388	*CBFB*	0.0004	0.0008	0.62
cg01500140	*LIM2*	0.0002	0.0005	0.68
cg03330058	*ABTB1*	-0.0001	0.0009	0.89
cg04983977	*GPR25*	0.0001	0.0006	0.89
cg13633560	*LRRC32*	-0.0001	0.0007	0.94

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 4.7 Results of reciprocal Mendelian randomization from inflammatory markers to smoking status.

Inflammatory marker	Instrumental SNP	F-value of SNP	Estimate	Std	Pr > |t|
log(CRP)	rs9628671	38	0.704	0.324	0.03
log(IL-6)	rs2069837	6	1.268	0.753	0.09
log(IL-18)	rs11214098	4	2.042	0.902	0.02
Fibrinogen	rs28577061	8	0.010	0.005	0.04

CRP: C-reactive protein, IL-6: interleukin-6, IL-18: interleukin-18

Table 4.8 Estimates of the interaction effect of smoking status and DNA methylation on log(CRP) levels.

CpG site	Gene	Estimate	Std	P
cg04983977	*GPR25*	0.54	0.32	0.10
cg00353953	*ZNF384*	0.38	0.25	0.13
cg14223444	*NCBP1*	0.37	0.27	0.17
cg13745870	*SPATA12*	0.46	0.38	0.22
cg26259865	*ZG16B*	0.25	0.23	0.28
cg01500140	*LIM2*	0.20	0.21	0.33
cg03330058	*ABTB1*	-0.19	0.20	0.34
cg03636183	*F2RL3*	-0.07	0.11	0.52
cg13500388	*CBFB*	-0.13	0.21	0.53
cg19859270	*GPR15*	-0.18	0.29	0.54
cg13668129	*HNRNPUL1*	-0.27	0.44	0.54
cg13633560	*LRRC32*	0.15	0.24	0.54
cg11314684	*AKT3*	-0.17	0.30	0.55
cg03340878	*OR2B6*	0.10	0.37	0.79
cg17791651	*POU3F1*	-0.02	0.18	0.92

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 4.9 Estimates of the interaction effect of smoking status and DNA methylation on log(IL-6) levels.

CpG site	Gene	Estimate	Std	P
cg13668129	*HNRNPUL1*	-0.44	0.23	0.06
cg03636183	*F2RL3*	0.10	0.06	0.09
cg01500140	*LIM2*	-0.18	0.11	0.12
cg17791651	*POU3F1*	0.10	0.10	0.30
cg26259865	*ZG16B*	0.12	0.13	0.34
cg19859270	*GPR15*	0.15	0.16	0.35
cg13500388	*CBFB*	0.09	0.11	0.41
cg11314684	*AKT3*	-0.13	0.16	0.41
cg03330058	*ABTB1*	-0.08	0.11	0.47
cg13745870	*SPATA12*	0.09	0.21	0.66
cg04983977	*GPR25*	-0.06	0.18	0.75
cg14223444	*NCBP1*	0.04	0.15	0.81
cg00353953	*ZNF384*	0.03	0.14	0.84
cg13633560	*LRRC32*	-0.02	0.13	0.89
cg03340878	*OR2B6*	0.01	0.20	0.96

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 4.10 Estimates of the interaction effect of smoking status and DNA methylation on log(IL-18) levels.

CpG site	Gene	Estimate	Std	P
cg14223444	*NCBP1*	-0.45	0.16	0.006
cg13633560	*LRRC32*	-0.33	0.15	0.02
cg17791651	*POU3F1*	-0.24	0.11	0.03
cg03330058	*ABTB1*	-0.23	0.12	0.06
cg13500388	*CBFB*	-0.23	0.12	0.07
cg13668129	*HNRNPUL1*	-0.43	0.27	0.11
cg26259865	*ZG16B*	-0.21	0.14	0.15
cg13745870	*SPATA12*	0.19	0.23	0.42
cg11314684	*AKT3*	-0.12	0.18	0.51
cg19859270	*GPR15*	0.11	0.18	0.52
cg00353953	*ZNF384*	0.07	0.15	0.65
cg01500140	*LIM2*	0.06	0.13	0.67
cg03340878	*OR2B6*	-0.09	0.23	0.69
cg03636183	*F2RL3*	0.02	0.07	0.75
cg04983977	*GPR25*	0.04	0.20	0.86

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 4.11 Estimates of the interaction effect of smoking status and DNA methylation on fibrinogen levels.

CpG site	Gene	Estimate	Std	P
cg13668129	*HNRNPUL1*	-0.43	0.24	0.07
cg03636183	*F2RL3*	0.10	0.06	0.09
cg01500140	*LIM2*	-0.18	0.11	0.12
cg17791651	*POU3F1*	0.10	0.10	0.30
cg26259865	*ZG16B*	0.12	0.13	0.34
cg19859270	*GPR15*	0.15	0.16	0.35
cg13500388	*CBFB*	0.09	0.11	0.41
cg11314684	*AKT3*	-0.13	0.16	0.41
cg03330058	*ABTB1*	-0.08	0.11	0.47
cg13745870	*SPATA12*	0.09	0.21	0.66
cg04983977	*GPR25*	-0.06	0.18	0.75
cg14223444	*NCBP1*	0.04	0.15	0.81
cg00353953	*ZNF384*	0.03	0.14	0.84
cg13633560	*LRRC32*	-0.02	0.13	0.89
cg03340878	*OR2B6*	0.01	0.20	0.96

ABTB1: ankyrin repeat and BTB (POZ) domain containing 1, *AKT3*:v-akt murine thymoma viral oncogene homolog 3, *CBFB*: core-binding factor beta subunit, *F2RL3*: coagulation factor II (thrombin) receptor-like 3, *GPR15*: G protein-coupled receptor 15, *GPR25*: G protein-coupled receptor 25, *HNRPUL1*: heterogeneous nuclear ribonucleoprotein U-like 1, *LIM2*: lens intrinsic membrane protein 2 (19kDa), *ZG16B*: zymogen granule protein 16B, *LRRC32*: leucine rich repeat containing 32, *NCBP1*: nuclear cap binding protein subunit 1 (80kDa), *OR2B6*: olfactory receptor, family 2, subfamily B, member 6, *POU3F1*: POU class 3 homeobox 1, *SPATA12*: spermatogenesis associated 12, *ZNF384*: zinc finger protein 384

Table 4.12 Total, direct and indirect effects of smoking on inflammatory markers mediated by cg03636183 on *F2RL3* gene.

Inflammatory marker	Effect	Estimate	Std.	P	95%	CI
log(CRP)	Total effect	0.0329	0.0142	0.0200	0.0052	0.0606
	Direct effect	0.0296	0.0141	0.0360	0.0019	0.0574
	Indirect effect	0.0033	0.0004	<0.0001	0.0026	0.0040
	Smoking -> Methylation	-0.0861	0.0095	<0.0001	-0.1047	-0.0676
	Methylation -> log(CRP)	-0.0381	0.0490	0.4370	-0.1341	0.0580
log(IL6)	Total effect	0.0211	0.0082	0.0100	0.0050	0.0371
	Direct effect	0.0176	0.0082	0.0320	0.0015	0.0336
	Indirect effect	0.0035	0.0004	<0.0001	0.0027	0.0043
	Smoking -> Methylation	-0.0863	0.0095	<0.0001	-0.1050	-0.0676
	Methylation -> log(IL6)	-0.0407	0.0294	0.1670	-0.0983	0.0170
log(IL18)	Total effect	0.0147	0.0086	0.0880	-0.0022	0.0316
	Direct effect	0.0049	0.0085	0.5660	-0.0118	0.0216
	Indirect effect	0.0098	0.0011	<0.0001	0.0076	0.0119
	Smoking -> Methylation	-0.0850	0.0096	<0.0001	-0.1038	-0.0662
	Methylation -> log(IL18)	-0.1151	0.0307	0.0002	-0.1752	-0.0549
Fibrinogen	Total effect	3.4017	1.0751	0.0020	1.2945	5.5090
	Direct effect	3.4081	1.0751	0.0020	1.3009	5.5153
	Indirect effect	-0.0064	0.0007	<0.0001	-0.0077	-0.0050
	Smoking -> Methylation	-0.0862	0.0095	<0.0001	-0.1047	-0.0676
	Methylation -> Fibrinogen	0.0739	3.7185	0.9840	-7.2143	7.3620

CRP: C-reactive protein, IL6: interleukin-6, and IL18: interleukin-18

Table 4.13 Total, direct and indirect effects of smoking on inflammatory markers mediated by cg13500388 on *CBFB* gene.

Inflammatory marker	Effect	Estimate	Std.	P	95%	CI
log(CRP)	Total effect	0.0329	0.0135	0.0150	0.0064	0.0595
	Direct effect	0.0333	0.0135	0.0140	0.0067	0.0598
	Indirect effect	-0.0004	0.0001	0.0001	-0.0006	-0.0002
	Smoking -> Methylation	-0.0156	0.0040	0.0001	-0.0235	-0.0077
	Methylation -> log(CRP)	0.0236	0.0949	0.8030	-0.1623	0.2095
log(IL6)	Total effect	0.0213	0.0078	0.0060	0.0060	0.0367
	Direct effect	0.0203	0.0078	0.0090	0.0050	0.0356
	Indirect effect	0.0011	0.0003	0.0002	0.0005	0.0017
	Smoking -> Methylation	-0.0152	0.0041	0.0002	-0.0232	-0.0072
	Methylation -> log(IL6)	-0.0719	0.0575	0.2110	-0.1846	0.0408
log(IL18)	Total effect	0.0150	0.0082	0.0690	-0.0012	0.0311
	Direct effect	0.0138	0.0082	0.0940	-0.0024	0.0299
	Indirect effect	0.0012	0.0003	0.0001	0.0006	0.0018
	Smoking -> Methylation	-0.0159	0.0041	0.0001	-0.0241	-0.0078
	Methylation -> log(IL18)	-0.0756	0.0595	0.2040	-0.1923	0.0411
Fibrinogen	Total effect	3.3943	1.0290	0.0010	1.3775	5.4112
	Direct effect	3.4460	1.0289	0.0010	1.4294	5.4627
	Indirect effect	-0.0517	0.0135	0.0001	-0.0781	-0.0253
	Smoking -> Methylation	-0.0155	0.0040	0.0001	-0.0234	-0.0076
	Methylation -> Fibrinogen	3.3431	7.2040	0.6430	-10.7765	17.4627

CRP: C-reactive protein, IL6: interleukin-6, and IL18: interleukin-18

Table 4.14 Total, direct and indirect effects of smoking on LVEF and LVMI mediated by cg03636183 on *F2RL3* gene and log(IL-18) levels.

Subclinical outcome	Effect	Estimate	Std.	P	95% CI	
LVEF	Total effect	0.0103	0.1094	0.925	-0.2041	0.2246
	Direct effect	0.0080	0.1094	0.942	-0.2064	0.2224
	Indirect effect	0.0023	0.0003	<0.0001	0.0018	0.0028
LVMI	Total effect	0.1039	0.1391	0.455	-0.1686	0.3765
	Direct effect	0.1252	0.1390	0.368	-0.1473	0.3977
	Indirect effect	-0.0212	0.0024	<0.0001	-0.0260	-0.0165

LVEF: left ventricular ejection fraction, LVMI: left ventricular mass index.

References

1. Benowitz NL. Cigarette smoking and cardiovascular disease: Pathophysiology and implications for treatment. *Progress in cardiovascular diseases*. 2003;46:91-111
2. Ambrose JA, Barua RS. The pathophysiology of cigarette smoking and cardiovascular disease: An update. *Journal of the American College of Cardiology*. 2004;43:1731-1737
3. Tracy RP, Psaty BM, Macy E, Bovill EG, Cushman M, Cornell ES, Kuller LH. Lifetime smoking exposure affects the association of c-reactive protein with cardiovascular disease risk factors and subclinical disease in healthy elderly subjects. *Arteriosclerosis, thrombosis, and vascular biology*. 1997;17:2167-2176
4. Helmersson J, Larsson A, Vessby B, Basu S. Active smoking and a history of smoking are associated with enhanced prostaglandin f(2alpha), interleukin-6 and f2-isoprostane formation in elderly men. *Atherosclerosis*. 2005;181:201-207
5. McKay A, Komai-Koma M, MacLeod KJ, Campbell CC, Kitson SM, Chaudhuri R, Thomson L, McSharry C, Liew FY, Thomson NC. Interleukin-18 levels in induced sputum are reduced in asthmatic and normal smokers. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*. 2004;34:904-910
6. Wannamethee SG, Lowe GD, Shaper AG, Rumley A, Lennon L, Whincup PH. Associations between cigarette smoking, pipe/cigar smoking, and smoking cessation, and haemostatic and inflammatory markers for cardiovascular disease. *European heart journal*. 2005;26:1765-1773
7. Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature genetics*. 2003;33 Suppl:245-254
8. Relton CL, Davey Smith G. Two-step epigenetic mendelian randomization: A strategy for establishing the causal role of epigenetic processes in pathways to disease. *International journal of epidemiology*. 2012;41:161-176
9. Burgess S, Daniel RM, Butterworth AS, Thompson SG, the E-IC. Network mendelian randomization: Using genetic variants as instrumental variables to investigate mediation in causal pathways. *International journal of epidemiology*. 2014
10. Investigators F. Multi-center genetic study of hypertension: The family blood pressure program (fbpp). *Hypertension*. 2002;39:3-9
11. Daniels PR, Kardia SL, Hanis CL, Brown CA, Hutchinson R, Boerwinkle E, Turner ST, Genetic Epidemiology Network of Arteriopathy s. Familial aggregation of hypertension treatment and control in the genetic epidemiology network of arteriopathy (genoa) study. *The American journal of medicine*. 2004;116:676-681
12. investigators TA. The atherosclerosis risk in communities (aric) study: Design and objectives. *American journal of epidemiology*. 1989;129:687-702
13. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56-65
14. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nature methods*. 2013;10:5-6
15. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics*. 2012;44:955-959

16. Du P, Kibbe WA, Lin SM. Lumi: A pipeline for processing illumina microarray. *Bioinformatics*. 2008;24:1547-1548
17. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of beta-value and m-value methods for quantifying methylation levels by microarray analysis. *BMC bioinformatics*. 2010;11:587
18. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics*. 2012;13:86
19. Kullo IJ, Seward JB, Bailey KR, Bielak LF, Grossardt BR, Sheedy PF, 2nd, Peyser PA, Turner ST. C-reactive protein is related to arterial wave reflection and stiffness in asymptomatic subjects from the community. *American journal of hypertension*. 2005;18:1123-1129
20. Khaleghi M, Singletary LA, Kondragunta V, Bailey KR, Turner ST, Mosley TH, Jr., Kullo IJ. Haemostatic markers are associated with measures of vascular disease in adults with hypertension. *Journal of human hypertension*. 2009;23:530-537
21. Kim CX, Bailey KR, Klee GG, Ellington AA, Liu G, Mosley TH, Jr., Rehman H, Kullo IJ. Sex and ethnic differences in 47 candidate proteomic markers of cardiovascular disease: The mayo clinic proteomic markers of arteriosclerosis study. *PloS one*. 2010;5:e9065
22. Fox ER, Klos KL, Penman AD, Blair GJ, Blossom BD, Arnett D, Devereux RB, Samdarshi T, Boerwinkle E, Mosley TH, Jr. Heritability and genetic linkage of left ventricular mass, systolic and diastolic function in hypertensive african americans (from the genoa study). *American journal of hypertension*. 2010;23:870-875
23. Coutinho T, Al-Omari M, Mosley TH, Jr., Kullo IJ. Biomarkers of left ventricular hypertrophy and remodeling in blacks. *Hypertension*. 2011;58:920-925
24. Devereux RB, Alonso DR, Lutas EM, Gottlieb GJ, Campo E, Sachs I, Reichek N. Echocardiographic assessment of left ventricular hypertrophy: Comparison to necropsy findings. *The American journal of cardiology*. 1986;57:450-458
25. de Simone G, Daniels SR, Devereux RB, Meyer RA, Roman MJ, de Divitiis O, Alderman MH. Left ventricular mass and body size in normotensive children and adults: Assessment of allometric relations and impact of overweight. *Journal of the American College of Cardiology*. 1992;20:1251-1260
26. Hernan MA, Robins JM. Instruments for causal inference: An epidemiologist's dream? *Epidemiology*. 2006;17:360-372
27. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992;3:143-155
28. Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Stat Sci*. 2010;25:51-71
29. Timpson NJ, Nordestgaard BG, Harbord RM, Zacho J, Frayling TM, Tybjaerg-Hansen A, Smith GD. C-reactive protein levels and body mass index: Elucidating direction of causation through reciprocal mendelian randomization. *International journal of obesity*. 2011;35:300-308
30. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, Pellikka N, Wallaschofski H, Kettunen J, Henneman P, Baumert J, Strachan DP, Fuchsberger C, Vitart V, Wilson JF, Pare G, Naitza S, Rudock ME, Surakka I, de Geus EJ, Alizadeh BZ, Guralnik J, Shuldiner A, Tanaka T, Zee RY, Schnabel RB, Nambi V, Kavousi M, Ripatti

S, Nauck M, Smith NL, Smith AV, Sundvall J, Scheet P, Liu Y, Ruokonen A, Rose LM, Larson MG, Hoogeveen RC, Freimer NB, Teumer A, Tracy RP, Launer LJ, Buring JE, Yamamoto JF, Folsom AR, Sijbrands EJ, Pankow J, Elliott P, Keaney JF, Sun W, Sarin AP, Fontes JD, Badola S, Astor BC, Hofman A, Pouta A, Werdan K, Greiser KH, Kuss O, Meyer zu Schwabedissen HE, Thiery J, Jamshidi Y, Nolte IM, Soranzo N, Spector TD, Volzke H, Parker AN, Aspelund T, Bates D, Young L, Tsui K, Siscovick DS, Guo X, Rotter JI, Uda M, Schlessinger D, Rudan I, Hicks AA, Penninx BW, Thorand B, Gieger C, Coresh J, Willemsen G, Harris TB, Uitterlinden AG, Jarvelin MR, Rice K, Radke D, Salomaa V, Willems van Dijk K, Boerwinkle E, Vasan RS, Ferrucci L, Gibson QD, Bandinelli S, Snieder H, Boomsma DI, Xiao X, Campbell H, Hayward C, Pramstaller PP, van Duijn CM, Peltonen L, Psaty BM, Gudnason V, Ridker PM, Homuth G, Koenig W, Ballantyne CM, Witteman JC, Benjamin EJ, Perola M, Chasman DI. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for c-reactive protein levels. *Circulation*. 2011;123:731-738

31. Doumatey AP, Chen G, Tekola Ayele F, Zhou J, Erdos M, Shriner D, Huang H, Adeleye J, Balogun W, Fasanmade O, Johnson T, Oli J, Okafor G, Amoah A, Eghan BA, Agyenim-Boateng K, Acheampong J, Adebamowo C, Gerry NP, Christman MF, Adeyemo A, Rotimi CN. C-reactive protein (crp) promoter polymorphisms influence circulating crp levels in a genome-wide association study of african americans. *Human molecular genetics*. 2012;21:3063-3072
32. Reiner AP, Beleza S, Franceschini N, Auer PL, Robinson JG, Kooperberg C, Peters U, Tang H. Genome-wide association and population genetic analysis of c-reactive protein in african american and hispanic american women. *American journal of human genetics*. 2012;91:502-512
33. Shah T, Zabaneh D, Gaunt T, Swerdlow DI, Shah S, Talmud PJ, Day IN, Whittaker J, Holmes MV, Sofat R, Humphries SE, Kivimaki M, Kumari M, Hingorani AD, Casas JP. Gene-centric analysis identifies variants associated with interleukin-6 levels and shared pathways with other inflammation markers. *Circulation. Cardiovascular genetics*. 2013;6:163-170
34. Tekola Ayele F, Doumatey A, Huang H, Zhou J, Charles B, Erdos M, Adeleye J, Balogun W, Fasanmade O, Johnson T, Oli J, Okafor G, Amoah A, Eghan BA, Jr., Agyenim-Boateng K, Acheampong J, Adebamowo CA, Herbert A, Gerry N, Christman M, Chen G, Shriner D, Adeyemo A, Rotimi CN. Genome-wide associated loci influencing interleukin (il)-10, il-1ra, and il-6 levels in african americans. *Immunogenetics*. 2012;64:351-359
35. He M, Cornelis MC, Kraft P, van Dam RM, Sun Q, Laurie CC, Mirel DB, Chasman DI, Ridker PM, Hunter DJ, Hu FB, Qi L. Genome-wide association study identifies variants at the il18-bco2 locus associated with interleukin-18 levels. *Arteriosclerosis, thrombosis, and vascular biology*. 2010;30:885-890
36. Matteini AM, Li J, Lange EM, Tanaka T, Lange LA, Tracy RP, Wang Y, Biggs ML, Arking DE, Fallin MD, Chakravarti A, Psaty BM, Bandinelli S, Ferrucci L, Reiner AP, Walston JD. Novel gene variants predict serum levels of the cytokines il-18 and il-1ra in older adults. *Cytokine*. 2014;65:10-16
37. Dehghan A, Yang Q, Peters A, Basu S, Bis JC, Rudnicka AR, Kavousi M, Chen MH, Baumert J, Lowe GD, McKnight B, Tang W, de Maat M, Larson MG, Eyhermendy S, McArdle WL, Lumley T, Pankow JS, Hofman A, Massaro JM, Rivadeneira F, Kolz M,

Taylor KD, van Duijn CM, Kathiresan S, Illig T, Aulchenko YS, Volcik KA, Johnson AD, Uitterlinden AG, Tofler GH, Gieger C, Wellcome Trust Case Control C, Psaty BM, Couper DJ, Boerwinkle E, Koenig W, O'Donnell CJ, Witteman JC, Strachan DP, Smith NL, Folsom AR. Association of novel genetic loci with circulating fibrinogen levels: A genome-wide association study in 6 population-based cohorts. *Circulation. Cardiovascular genetics*. 2009;2:125-133

38. Danik JS, Pare G, Chasman DI, Zee RY, Kwiatkowski DJ, Parker A, Miletich JP, Ridker PM. Novel loci, including those related to crohn disease, psoriasis, and inflammation, identified in a genome-wide association study of fibrinogen in 17 686 women: The women's genome health study. *Circulation. Cardiovascular genetics*. 2009;2:134-141
39. Lovely RS, Yang Q, Massaro JM, Wang J, D'Agostino RB, Sr., O'Donnell CJ, Shannon J, Farrell DH. Assessment of genetic determinants of the association of gamma' fibrinogen in relation to cardiovascular disease. *Arteriosclerosis, thrombosis, and vascular biology*. 2011;31:2345-2352
40. Sabater-Lleal M, Huang J, Chasman D, Naitza S, Dehghan A, Johnson AD, Teumer A, Reiner AP, Folkersen L, Basu S, Rudnicka AR, Trompet S, Malarstig A, Baumert J, Bis JC, Guo X, Hottenga JJ, Shin SY, Lopez LM, Lahti J, Tanaka T, Yanek LR, Oudot-Mellakh T, Wilson JF, Navarro P, Huffman JE, Zemunik T, Redline S, Mehra R, Pulanic D, Rudan I, Wright AF, Kolcic I, Polasek O, Wild SH, Campbell H, Curb JD, Wallace R, Liu S, Eaton CB, Becker DM, Becker LC, Bandinelli S, Raikkonen K, Widen E, Palotie A, Fornage M, Green D, Gross M, Davies G, Harris SE, Liewald DC, Starr JM, Williams FM, Grant PJ, Spector TD, Strawbridge RJ, Silveira A, Sennblad B, Rivadeneira F, Uitterlinden AG, Franco OH, Hofman A, van Dongen J, Willemsen G, Boomsma DI, Yao J, Swords Jenny N, Haritunians T, McKnight B, Lumley T, Taylor KD, Rotter JI, Psaty BM, Peters A, Gieger C, Illig T, Grotevendt A, Homuth G, Volzke H, Kocher T, Goel A, Franzosi MG, Seedorf U, Clarke R, Steri M, Tarasov KV, Sanna S, Schlessinger D, Stott DJ, Sattar N, Buckley BM, Rumley A, Lowe GD, McArdle WL, Chen MH, Tofler GH, Song J, Boerwinkle E, Folsom AR, Rose LM, Franco-Cereceda A, Teichert M, Ikram MA, Mosley TH, Bevan S, Dichgans M, Rothwell PM, Sudlow CL, Hopewell JC, Chambers JC, Saleheen D, Kooner JS, Danesh J, Nelson CP, Erdmann J, Reilly MP, Kathiresan S, Schunkert H, Morange PE, Ferrucci L, Eriksson JG, Jacobs D, Deary IJ, Soranzo N, Witteman JC, de Geus EJ, Tracy RP, Hayward C, Koenig W, Cucca F, Jukema JW, Eriksson P, Seshadri S, Markus HS, Watkins H, Samani NJ, Consortium VTE, Consortium S, Wellcome Trust Case Control C, Consortium CD, Consortium CA, Wallaschofski H, Smith NL, Tregouet D, Ridker PM, Tang W, Strachan DP, Hamsten A, O'Donnell CJ. Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. *Circulation*. 2013;128:1310-1324
41. Imai K, Keele L, Tingley D, Yamamoto T. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *Am Polit Sci Rev*. 2011;105:765-789
42. Lavi S, Prasad A, Yang EH, Mathew V, Simari RD, Rihal CS, Lerman LO, Lerman A. Smoking is associated with epicardial coronary endothelial dysfunction and elevated white blood cell count in patients with chest pain and early coronary artery disease. *Circulation*. 2007;115:2621-2627

43. StataCorp. Stata statistical software: Release 13. 2013
44. Lester BM, Tronick E, Nestler E, Abel T, Kosofsky B, Kuzawa CW, Marsit CJ, Maze I, Meaney MJ, Monteggia LM, Reul JM, Skuse DH, Sweatt JD, Wood MA. Behavioral epigenetics. *Annals of the New York Academy of Sciences*. 2011;1226:14-33
45. Day JJ, Childs D, Guzman-Karlsson MC, Kibe M, Moulden J, Song E, Tahir A, Sweatt JD. DNA methylation regulates associative reward learning. *Nature neuroscience*. 2013;16:1445-1452

Chapter 5

Conclusions and future directions

Overall conclusions

Complex diseases, such as CVD, are influenced by a combination and interaction of various factors. Risk of CVD is affected by many factors including genetic variants, epigenetic modification, and environmental exposures. Most studies that tend to focus on genetic variants ignore other factors or include them as confounders. In contrast, many studies that focus on environmental exposures or behaviors tend to ignore genetic susceptibility among individuals. While a major assumption of many studies is that there is no gene by environment interaction, evidence exists for such interactions. Evidence for gene by environment interaction implies that the outcome for a specific genotype depends on environmental exposure or the outcome for environmental exposure differs by genotypes. The environment may be measured directly, for example by cumulative lead exposure or self-reported smoking, or more indirectly by epigenetic modification in response to environmental exposures. Epigenetic modification occurs as a response to changes in environmental and lifestyle factors.[1] Thus, epigenetic modification is known to explain, in part, the interaction between genes and environmental factors in development of various diseases including CVD.[2-8] Epigenetic modification may be seen as an intermediate biomarker which is a result of the interaction between genes and environmental factors.

To further understand biological mechanisms of CVD using gene by environment interaction, this dissertation utilized effect modification by candidate genetic polymorphisms and mediation analysis using epigenetic modification on measures of inflammation including CRP, IL-6, IL-18, and fibrinogen as well as CVD risk markers including pulse pressure, LVEF, and LVMI. In Chapter 2, we investigated whether there is evidence for an interaction between *VDR* genetic polymorphisms and cumulative bone lead levels on pulse pressure in unrelated non-Hispanic White men. In Chapter 3, we examined whether there was evidence for a mediated effect of DNA methylation on the relation between smoking and inflammatory markers in African American siblings. In Chapter 4, we investigated the causal paths of smoking, DNA methylation, and inflammatory markers and extended the approach to include subclinical markers of CVD in African American siblings.

Chapter 2 focused on lead exposure as the environmental measure. Lead (Pb) is known to increase the risk of CVD. Lead exposure may also induce endothelial injury and atherosclerosis.[9-11] The study by Perlstein et al[12] suggests that lead accumulation may contribute to the increase in pulse pressure, a measure of arterial stiffness, and with clinical cardiovascular events. Vitamin D and VDR are involved in arterial stiffness and arterial aging.[13, 14] The complex of vitamin D and VDR acts as a transcription factor regulating the gene expression of calcium-binding receptors. Because lead is a divalent cation, lead competes with calcium to bind to calcium-binding receptors.[15] The *VDR* genetic variants have been identified as potential genetic factors that can influence the absorption, retention and accumulation of lead in the human body.[16] In Chapter 2, we investigated effect modification by the *VDR* gene on the association between cumulative lead exposure measured by bone lead levels and pulse pressure, a marker of arterial stiffness. The study was conducted with longitudinal measures (3,100 observations

obtained during up to 20 years of follow up) on 727 participants of the NAS. With an interquartile range (IQR) increase in tibia lead (15µg/g), pulse pressure was 2.5 mmHg (95% CI: 0.4-4.7) greater for the participants with at least one minor allele on *Bsm1* compared with the participants without the minor frequency allele. With an IQR increase in patella lead (20µg/g), pulse pressure was 1.9 mmHg (95% CI: 0.1-3.8) greater for the participants with at least one copy of the minor frequency allele in *Bsm1* compared with the participants without the minor frequency allele. Similar results were found for *Taq1*. In spite of the changing association between bone lead levels and pulse pressure over time, the effect modifications by *VDR Bsm1* and *Taq1* genotypes on the association between bone lead levels and pulse pressure persists over time. In the present study, we found that individuals with at least one *VDR Bsm1* variant had a 2.5 mmHg greater pulse pressure in relation to every 15 µg/g increase in cumulative (tibia) lead exposure. We interpret this as suggesting that individuals with at least one copy of the *VDR Bsm1* variant allele may have 2.5% to 6% greater risks for CVD mortality for every 15 µg/g increase in cumulative lead exposure. Our findings suggest the importance of restricting early exposure to lead to avoid its persistent adverse health effects in the subjects' later life. Since the NAS is an older cohort of predominantly white male participants, the findings may not be generalizable to women, younger individuals, and other ethnicities. Our observation suggests that the *VDR* gene may modify the toxicokinetics of lead and shed light on biological mechanisms behind gene by lead interaction on CVD risk. These findings support the public health initiatives to limit lead exposure to protect even people who are susceptible to greater lead-induced elevation in pulse pressure (those who carry *VDR Bsm1* or *Ta1q1* variants) since lead stays in the body for a long period of time and is related to higher levels of arterial stiffness. Importantly, in our study if we had ignored lead exposure in this cohort, there would have been no association

between *VDR* polymorphisms and pulse pressure. Thus, incorporating *VDR* polymorphisms by lead exposure interactions provided information about the role of these polymorphisms.

Chapter 3 focused on cigarette smoking and DNA methylation. Cigarette smoking is a well-known risk factor for CVD.[17] Activation of inflammation has been suggested as a potential mechanism behind the smoking-induced cardiovascular risk.[18] Efforts to identify genetic components in the biological mechanisms behind cigarette smoking-induced inflammation have been extended to EWAS.[19-23] Several EWAS identified differentially methylated CpG sites related to cigarette smoking status.[19-24] However, the role of DNA methylation levels as a mediator in the association between cigarette smoking and inflammatory responses has not been investigated yet. In Chapter 3, we hypothesized that cigarette smoking is associated with DNA methylation levels, and DNA methylation levels are associated with inflammatory responses. We implemented the two-step epigenetic MR approach[25] to study African American siblings from GENOA. We identified CpG sites that respond to smoking exposure and mediate the effect of smoking on inflammation. We demonstrated that the current smoking status is associated with changes in DNA methylation levels of CpG sites on *F2RL3*, *GPR15*, *CBFB*, and *GPR25* genes. The methylation level changes of CpG sites on *F2RL3* and *CBFB* genes were associated with the changes in the levels of inflammatory markers log(IL-18) and log(CRP), respectively. The *F2RL3* gene has been found to be associated with platelet activation and perioperative myocardial injury,[26] post infectious irritable bowel syndrome,[27] and gastric cancer.[28] The *CBFB* gene makes a beta subunit of core-binding transcription factor and it binds to Runt-related transcription factors (*RUNX*1, 2, and 3) that are involved in hematopoiesis and osteogenesis.[29] The *CBFB* gene has been found to be associated with breast cancer, prostate and ovarian cancer, and acute myeloid leukemia.[30-32] The two-step epigenetic MR approach enabled us to identify

CpG sites that are involved in the association between an environmental exposure and an outcome. This approach can expand our understanding on the role of DNA methylation in disease development and provides information on biological mechanisms behind it. The DNA methylation levels of the *F2RL3* and *CBFB* genes may provide additional information to assess genetically-influenced subject-specific smoking exposure levels.

Chapter 4 extended the approach of the chapter 3 to include direct effects and subclinical markers of CVD. Although the two-step MR is useful to assess an indirect effect that causally mediates the exposure-outcome association, it is limited to evaluate the magnitude of the indirect effect relative to the total effect. If there are several pathways from smoking to inflammation, it is likely that the estimates from the previous approach could be biased because that approach assumes the only path is through the mediator of interest. Therefore, it is necessary to evaluate the mediated effect by DNA methylation in consideration of the possible direct effect. In Chapter 4, we combined the two-step epigenetic MR approach with causal mediation analysis, which is also known as the network MR approach.[33] We also extended the network MR approach for more complex causal paths using structural equation modeling (SEM). The direct effects of smoking were significant for log(CRP), log(IL-6), and fibrinogen but not for log(IL-18) levels. The indirect effects of smoking through DNA methylation were significant for all the four inflammatory markers. However, when the indirect effect was decomposed into the effects from smoking to DNA methylation and from DNA methylation to inflammatory markers, the second effect was only significant for log(IL-18), which is consistent with the results from Chapter 3. The estimate of the indirect effect on log(IL-18) was twice the estimate for the direct effect. The identified path of smoking, cg03636183 on *F2RL3*, and log(IL-18) was extended to include selected subclinical markers of cardiovascular disease such as LVEF and LVMI. The indirect

path had a significant effect on LVMI. The additional information on the DNA methylation levels of *F2RL3* gene may provide more accurate subject-specific smoking exposure level information and help to identify a higher risk group among smokers for CVD.

Studies on gene by environment interaction may shed light on identifying biological mechanisms underlying CVD development. The studies of gene by environment interaction in this dissertation, using both directly measured environmental exposures as well as indirectly measured environmental exposures, together helped us improve our understanding of the epigenetic factors that affect CVD risk markers.

Limitations

Replication studies are required for many genetic epidemiologic studies. Continued advances in genotyping and sequencing during the last decade enables us to examine tens of millions of dense genetic markers simultaneously. This has resulted in extensive numbers of tests and multiple testing issues. However, even with a conservative multiple testing correction such as the Bonferroni correction, there are still false positive findings due to other factors such as confounding. Replication studies are a good way to confirm the findings here. However, there are not many longitudinal studies with measurements on bone lead levels that can be used to replicate findings from Chapter 2. There are not many studies with measurements on genetics, epigenetics, inflammation and measures from echocardiograms that can be used to replicate findings from Chapter 3 or 4. We will continue to seek any available data for replication recognizing that it is more likely we will find studies for replication for Aims 2 and 3 than for Aim 1.

To assess the causal relationship and to distinguish a cause and an effect, observations over time are critical. Even though GENOA is a longitudinal cohort study, our study on DNA

methylation was done cross-sectionally using Phase II data. DNA methylation information is also available in Phase I. However, Phase I data was obtained from the Illumina HumanMethylation 450K BeadChip ® using both Type I and Type II probes,[34] while Phase II data was obtained from Illumina HumanMethylation 27K BeadChip ® using only Type I probes.[35] The Type I and Type II probes have different biochemical properties,[36] therefore, we cannot compare DNA methylation levels from Phase I to the those from Phase II. It may be possible to compare them by extracting only Type I probes (used for 135,501 CpG sites in the 450K chip) and normalize them separately from Type II probes. However, it is still questionable whether DNA methylation levels obtained from different beadchips are comparable. Another limitation of this approach is that we obtained DNA methylation information on only about 400 subjects in Phase I, compared to 1008 subjects in Phase II. However, it may be still useful for a simple analysis that does not require dense methylation markers or a large sample size.

Future directions

In Chapters 3 and 4, we restricted the role of DNA methylation to a mediator. However, we found evidence suggesting that DNA methylation interacts with genetic markers and there is an exchange feedback with inflammatory markers. There is also accumulating evidence from other work suggesting an interaction between DNA methylation, histone modification, and its role in regulation of gene expression.[37] In the future, our study could be extended to investigate the diverse role of DNA methylation in disease development by considering other regulators.

Our approach in Aims 2 and 3 can be extended to study gene by environment interaction. When we observe significantly different relationships between an environmental exposure and an outcome by certain genotypes, it indicates evidence of gene by environment interaction. Similarly, we can study gene by epigenetically-mediated-environment interaction by

investigating whether there exists significant difference in effect sizes of the mediating path by genotypes of interest. This approach may extend our understanding of the relationship between genotype and DNA methylation as a response to environmental exposures.

DNA methylation changes overtime and it is a reversible change.[38] We still have a limited understanding on when, where, and how DNA methylation occurs, and when, where and how DNA methylation reverses in response to environmental exposures. Even though we could not obtain a good resolution of more detailed DNA methylation status in our study, sequencing techniques to detect the states during reverse DNA methylation exist.[39] Information on reverse DNA methylation may improve our understanding on the role of DNA methylation in disease development and its role as a therapeutic target.

Information on gene expression levels can provide useful evidence on whether the changes in DNA methylation levels have an impact on the mRNA levels. Even though mRNA level is not a final phenotype, and there are many post-transcriptional and post-translational modifications, mRNA levels are useful to confirm the effect of DNA methylation on gene expression. Therefore, incorporating mRNA level data might provide better support for the findings.

Genetic studies are moving toward capturing a full picture of the genetic complexity. In the future, to extend Aim1, genetic polymorphisms other than *VDR*, such as genes in the iron metabolism pathway and the inflammation pathway can be studied.[40] In Aims 2 and 3, we utilized both genetic and epigenetic information. However, phenotypic variation is determined by various factors which are not limited to SNPs and DNA methylation. Histone methylation and acetylation also plays an important role in regulation of gene expression.[41] There are also many other regulators in post-transcriptional and post-translational modifications including

microRNAs,[42, 43] and lncRNAs.[44, 45] In the future, we will move toward integrating information on various key regulators, which will improve our understanding of complex mechanisms in gene expression and their role in disease development.

In particular, CVD is a common and complex disease with multiple risk factors including genetic and environmental factors. CVD is complex not only in terms of number of risk factors but also in terms of the relationship among the risk factors. The relationship is not limited to direct or indirect effects but to feedback effects and more complex forms of relationships. There have been efforts to integrate genetic variation, gene expression, and clinical phenotypes using systems-based approach.[46-48] In the future, systems-based approach may play more important role in unraveling the molecular networks underlying complex disease development.

The approaches used in all three aims can be applied to other exposures, other genetic markers, other measures of DNA methylation, other measures of inflammation as well as other measures associated with disease risk, and finally other subclinical measures of disease.

References

1. Tammen SA, Friso S, Choi SW. Epigenetics: The link between nature and nurture. *Molecular aspects of medicine*. 2013;34:753-764
2. van Vliet J, Oates NA, Whitelaw E. Epigenetic mechanisms in the context of complex diseases. *Cellular and molecular life sciences : CMLS*. 2007;64:1531-1538
3. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature*. 2007;447:433-440
4. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nature reviews. Genetics*. 2007;8:253-262
5. Abdolmaleky HM, Smith CL, Faraone SV, Shafa R, Stone W, Glatt SJ, Tsuang MT. Methylomics in psychiatry: Modulation of gene-environment interactions may be through DNA methylation. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 2004;127B:51-59
6. Lu Q, Qiu X, Hu N, Wen H, Su Y, Richardson BC. Epigenetics, disease, and therapeutic interventions. *Ageing research reviews*. 2006;5:449-467
7. Foley DL, Craig JM, Morley R, Olsson CA, Dwyer T, Smith K, Saffery R. Prospects for epigenetic epidemiology. *Am J Epidemiol*. 2009;169:389-400
8. Udali S, Guarini P, Moruzzi S, Choi SW, Friso S. Cardiovascular epigenetics: From DNA methylation to micrornas. *Molecular aspects of medicine*. 2013;34:883-901
9. Revis NW, Zinsmeister AR, Bull R. Atherosclerosis and hypertension induction by lead and cadmium ions: An effect prevented by calcium ion. *Proceedings of the National Academy of Sciences of the United States of America*. 1981;78:6494-6498
10. Kaji T, Suzuki M, Yamamoto C, Mishima A, Sakamoto M, Kozuka H. Severe damage of cultured vascular endothelial cell monolayer after simultaneous exposure to cadmium and lead. *Archives of environmental contamination and toxicology*. 1995;28:168-172
11. Vaziri ND. Mechanisms of lead-induced hypertension and cardiovascular disease. *American journal of physiology. Heart and circulatory physiology*. 2008;295:H454-465
12. Perlstein T, Weuve J, Schwartz J, Sparrow D, Wright R, Litonjua A, Nie H, Hu H. Cumulative community-level lead exposure and pulse pressure: The normative aging study. *Environmental health perspectives*. 2007;115:1696-1700
13. Cozzolino M, Stucchi A, Rizzo MA, Soldati L, Cusi D, Ciceri P, Brenna I, Elli F, Gallieni M. Reprint of: Vitamin d receptor activation and prevention of arterial ageing. *Proceedings and abstracts from the 8th International Symposium on Amino Acid and Protein Metabolism in Health and Disease*. 2013;23, Supplement 1:S31-S36
14. Giallauria F, Milaneschi Y, Tanaka T, Maggio M, Canepa M, Elango P, Vigorito C, Lakatta EG, Ferrucci L, Strait J. Arterial stiffness and vitamin d levels: The baltimore longitudinal study of aging. *The Journal of Clinical Endocrinology & Metabolism*. 2012;97:3717-3723
15. Rosen JF, Pounds JG. Quantitative interactions between pb2+ and ca2+ homeostasis in cultured osteoclastic bone cells. *Toxicology and applied pharmacology*. 1989;98:530-543
16. Onalaja AO, Claudio L. Genetic susceptibility to lead poisoning. *Environmental health perspectives*. 2000;108 Suppl 1:23-28

17. Ockene IS, Miller NH. Cigarette smoking, cardiovascular disease, and stroke: A statement for healthcare professionals from the american heart association. American heart association task force on risk reduction. *Circulation*. 1997;96:3243-3247
18. Wannamethee SG, Lowe GD, Shaper AG, Rumley A, Lennon L, Whincup PH. Associations between cigarette smoking, pipe/cigar smoking, and smoking cessation, and haemostatic and inflammatory markers for cardiovascular disease. *European heart journal*. 2005;26:1765-1773
19. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27k discovery and replication. *American journal of human genetics*. 2011;88:450-457
20. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, Strauch K, Waldenberger M, Illig T. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PloS one*. 2013;8:e63812
21. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, Belvisi MG, Brown R, Vineis P, Flanagan JM. Epigenome-wide association study in the european prospective investigation into cancer and nutrition (epic-turin) identifies novel genetic loci associated with smoking. *Human molecular genetics*. 2013;22:843-851
22. Dogan MV, Shields B, Cutrona C, Gao L, Gibbons FX, Simons R, Monick M, Brody GH, Tan K, Beach SR, Philibert RA. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from african american women. *BMC genomics*. 2014;15:151
23. Harlid S, Xu Z, Panduri V, Sandler DP, Taylor JA. Cpg sites associated with cigarette smoking: Analysis of epigenome-wide data from the sister study. *Environmental health perspectives*. 2014;122:673-678
24. Sun YV, Smith AK, Conneely KN, Chang Q, Li W, Lazarus A, Smith JA, Almli LM, Binder EB, Klengel T, Cross D, Turner ST, Ressler KJ, Kardia SL. Epigenomic association analysis identifies smoking-related DNA methylation sites in african americans. *Human genetics*. 2013;132:1027-1037
25. Relton CL, Davey Smith G. Two-step epigenetic mendelian randomization: A strategy for establishing the causal role of epigenetic processes in pathways to disease. *International journal of epidemiology*. 2012;41:161-176
26. Muehlschlegel JD, Perry TE, Liu KY, Fox AA, Smith SA, Lichtner P, Collard CD, Shernan SK, Hartwig JH, Body SC, Hoffmeister KM. Polymorphism in the protease-activated receptor-4 gene region associates with platelet activation and perioperative myocardial injury. *American journal of hematology*. 2012;87:161-166
27. Han W, Wang Z, Lu X, Guo C. Protease activated receptor 4 status of mast cells in post infectious irritable bowel syndrome. *Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society*. 2012;24:113-119, e182
28. Zhang Y, Yu G, Jiang P, Xiang Y, Li W, Lee W, Zhang Y. Decreased expression of protease-activated receptor 4 in human gastric cancer. *The international journal of biochemistry & cell biology*. 2011;43:1277-1283
29. Chuang LS, Ito K, Ito Y. Runx family: Regulation and diversification of roles through interacting proteins. *International journal of cancer. Journal international du cancer*. 2013;132:1260-1271

30. Mendoza-Villanueva D, Deng W, Lopez-Camacho C, Shore P. The runx transcriptional co-activator, cbfbeta, is essential for invasion of breast cancer cells. *Molecular cancer*. 2010;9:171
31. Davis JN, Rogers D, Adams L, Yong T, Jung JS, Cheng B, Fennell K, Borazanci E, Moustafa YW, Sun A, Shi R, Glass J, Mathis JM, Williams BJ, Meyers S. Association of core-binding factor beta with the malignant phenotype of prostate and ovarian cancer cells. *Journal of cellular physiology*. 2010;225:875-887
32. O'Reilly J, Chipper L, Springall F, Herrmann R. A unique structural abnormality of chromosome 16 resulting in a cbf beta-myh11 fusion transcript in a patient with acute myeloid leukemia, fab m4. *Cancer genetics and cytogenetics*. 2000;121:52-55
33. Burgess S, Daniel RM, Butterworth AS, Thompson SG, the E-IC. Network mendelian randomization: Using genetic variants as instrumental variables to investigate mediation in causal pathways. *International journal of epidemiology*. 2014
34. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R. High density DNA methylation array with single cpg site resolution. *Genomics*. 2011;98:288-295
35. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL. Genome-wide DNA methylation profiling using infinium(r) assay. *Epigenomics*. 2009;1:177-200
36. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing illumina 450k methylation array data. *BMC genomics*. 2013;14:293
37. Vaissiere T, Sawan C, Herceg Z. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. *Mutat Res*. 2008;659:40-48
38. Ramchandani S, Bhattacharya SK, Cervoni N, Szyf M. DNA methylation is a reversible biological signal. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;96:6107-6112
39. Wu H, Zhang Y. Reversing DNA methylation: Mechanisms, genomics, and biological functions. *Cell*. 2014;156:45-68
40. Park SK, Hu H, Wright RO, Schwartz J, Cheng Y, Sparrow D, Vokonas PS, Weisskopf MG. Iron metabolism genes, low-level lead exposure, and qt interval. *Environ Health Perspect*. 2009;117:80-85
41. Zhang Y, Reinberg D. Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails. *Genes & development*. 2001;15:2343-2360
42. Pasquinelli AE. Microrna s and their targets: Recognition, regulation and an emerging reciprocal relationship. *Nature reviews. Genetics*. 2012;13:271-282
43. Gurtan AM, Sharp PA. The role of mirnas in regulating gene expression networks. *Journal of molecular biology*. 2013;425:3582-3600
44. Kim ED, Sung S. Long noncoding rna: Unveiling hidden layer of gene regulatory networks. *Trends in plant science*. 2012;17:16-21
45. Wierzbicki AT. The role of long non-coding rna in transcriptional gene silencing. *Current opinion in plant biology*. 2012;15:517-522
46. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardson A, Castellini LW, Wang S, Champy MF, Zhang B, Emilsson V, Doss S, Ghazalpour A, Horvath S, Drake TA, Lusis AJ, Schadt EE. Variations in DNA elucidate molecular networks that cause disease. *Nature*. 2008;452:429-435

47. Lusis AJ, Weiss JN. Cardiovascular networks: Systems-based approaches to cardiovascular disease. *Circulation*. 2010;121:157-170
48. MacLellan WR, Wang Y, Lusis AJ. Systems-based approaches to cardiovascular disease. *Nature reviews. Cardiology*. 2012;9:172-184