Gene by Diet Interactions on Epigenetic Changes Modulating Cardiovascular Disease Risk Factors

A dissertation

Presented to the Faculty of

Friedman School of Nutrition Science and Policy

By

Yiyi Ma

In Partial Fulfilment of the requirements for

the Degree of Doctor of Philosophy

TUFTS UNIVERSITY

August 2014

Advisor: Jose M. Ordovas, PhD

Other committee members: Caren E. Smith, DVM; Paul F. Jacques, DSci.; Jean-Marc Zingg, PhD; Sang-Woon Choi, PhD

ACKNOWLEDGMENTS

I am sincerely grateful to my thesis advisor Dr. Jose M. Ordovas for his generous support and tremendous encouragement for me to conduct this novel but risky research. This thesis would never have been completed without his magnificent support to introduce my proposals to the Genetics of Lipid Lowering Drugs and Diet Network study (GOLDN) and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, from which the relevant data were obtained to accomplish the thesis research.

I am also deeply grateful to my committee member Dr. Caren E. Smith for her persistent patience and huge dedications to my research project by critically reviewing each proposal and manuscript, and by her significant support and guidance for my communications with all the participating cohorts in the CHARGE consortium. I would also like to thank my committee members Dr. Paul F. Jacques, Dr. Sang Woon Choi, and Dr. Jean-Marc Zingg for their dedications of their precious time, knowledgeable guidance, and critical review of each step of the development of my thesis.

I would also like to thank all the past and current members, especially Dr. Chao-Qiang Lai, Dr. Laurence D. Parnell, Yu-Chi Lee, and Lucia Pham from the Nutrition and Genomics Laboratory at Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University for their inspiring discussions and critical reviews. Thanks to my academic advisor Dr. Nirupa Matthan for her overall advice and guidance. Thanks to Dr. Stefania Lamon-Fava for her patience, support, and inspiring discussions regarding career in research. I would like to express my gratitude to the program advisors Dr. Lynne M. Ausman, Dr. Sarah L. Booth, and Dr. Edward Saltzman from the Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy at

Tufts University for giving me such a wonderful opportunity to start my exciting and inspiring PhD journey which helps me to become mature and prepared for my dream career in research. Thanks to all the classmates and friends, who have given me a memorable experience at Tufts University.

I would like to thank all the investigators and participants for the International HapMap Consortium, the Encyclopedia of DNA Elements (ENCODE) project, and the 1000 Genome Project, and the whole genetic community to facilitate scientific researches and findings by making their data publically available. I would like to thank all the collaborating investigators and participants in the GOLDN study, and all the other participating cohorts in the CHARGE consortium, including the Cardiovascular Health Study (CHS), the Invecchiare in Chianti (InCHIANTI), the Atherosclerosis Risk in Communities (ARIC) Study, the Multi-Ethnic Study of Atherosclerosis (MESA) Study, the Three-City Study (3C), and the Women's Genome Health Study (WGHS). Particularly, I would like to thank Dr. Rozenn Lemaitre, the coordinator of the CHARGE fatty acids working group, for her patient coordination and instructive guidance for the project. I also would like to thank Dr. Toshiko Tanaka, Dr. David S Siscovick, and Dr. Dariush Mozaffarian for their guidance and support for the CHARGE project.

Finally, I would like to give my most special gratitude to my husband, Fei Wang, for his always accompany, considerate understanding and care for my life, and full support for my research especially during our difficult time. Of course, I would also like to thank my parents, who have provided me with invaluable and unconditional love, and unbelievable endowment of full freedom for me to pursue my dream research career.

ABSTRACT

Background: Evidence has been steadily accumulating to document gene-by-environment (G×E) interactions for cardiovascular disease (CVD) related traits. However, the underlying mechanisms are still unclear. DNA methylation, one of the epigenetic mechanisms directly affecting genetic nucleotides and their interaction with regulatory proteins, may represent one of the potential mechanisms for the observed G×E interactions, based on its role as the interface between the 'nature' and 'nurture'.

Objectives: To explore the interplay among DNA methylation, genetic variants, and environmental factors, we examined 1) the association between single nucleotide polymorphism (SNP) and DNA methylation; and 2) the role that DNA methylation plays in G×E interactions.

Methodology: We applied a genome-wide approach with an integrated bioinformatics analysis to publicly available datasets of both genotypes (the HapMap project) and methylation patterns in B lymphocyte cell line (the Encyclopedia of DNA Elements (ENCODE) project) to explore the relationship between SNPs and DNA methylation patterns. A candidate gene approach was utilized to explore the potential mechanistic role that DNA methylation plays in significant G×E interactions at CVD-related loci, including *APOE*, *IL6*, *ABCA1*, *APOA5*, *PCSK9*, *HMGCR* and *HNF1A*. DNA methylation was measured by the Infinium Human Methylation 450K BeadChip in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study. Meta-analysis with 7 cohorts in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium was conducted to explore the G×E interactions for blood lipids. Correlation between DNA methylation and gene expressions across 17 available cell lines in the ENCODE consortium were analyzed.

Results: On a genome-wide scale, DNA methylation patterns are associated with haplotypes of multiple CpG-related SNPs (CGSs) within the same linkage disequilibrium (LD) block ($P < 0.0001$). At the *APOE* locus, the promoter SNP rs405509 interacts with age in the GOLDN population to modulate the methylation of the promoter CpG site cg01032398 ($P = 0.03$). At the *IL6* locus, erythrocyte N3 polyunsaturated fatty acids (PUFAs) interact with the promoter SNP rs2961298 to modulate the methylation of a promoter CpG site cg01770232 ($P = 0.02$), which was suggested as a potentially functional methylation site based on its consistent correlation with *IL6* gene expression in ENCODE ($P = 0.0005$) and plasma concentration of IL6 in GOLDN ($P = 0.03$). Meta-analysis with 7 cohorts in the CHARGE consortium found nominal interactions between circulating eicosapentaenoic acid (EPA) and the *ABCA1* promoter SNP rs2246293 for blood high-density lipoprotein (HDL) cholesterol level ($P = 0.006$), and between circulating alpha-linolenic acid (ALA) and *APOE* promoter SNP rs405509 for plasma triglyceride (TG) ($P = 0.01$). Analysis with methylation in GOLDN and gene expression in ENCODE suggested that the genotype-dependent methylation of CpG site cg14019050 ($P = 3.51\times10^{-18}$ and 0.007 for association and interaction analysis, respectively) and cg04406254 ($P = 0.008$ and 0.009 for association and interaction analysis, respectively) may be mechanistically linked to the observed interactions of loci of *ABCA1* and *APOE*, respectively.

Conclusion: DNA methylation patterns are associated with haplotypes of multiple CGSs within the same LD block. Genotype-dependent methylation may account, in part, for the mechanisms underlying observed G×E interactions in *APOE*, *IL6*, and *ABCA1*. Our studies call for further demonstration with interventional studies and molecular mechanistic experiments, with the ultimate goal of providing fundamental evidence to support genetically-based strategies for the development of personalized medical care.

Table of Contents

LIST OF ABBREVIATIONS

3C: the Three-City Study;
A: adenosine;
A^{vy}: agouti viable yellow;
AA: Arachidonic acid;
AA: African American;
ABCA1: ATP-binding cassette transporter 1;
APOA1, ApoA-I: apolipoprotein A1;
APOA5: apolipoprotein 5;
APOE, ApoE: apolipoprotein E
ARIC: the Atherosclerosis Risk in Communities Study;
ASE: alternatively spliced exon;
ASM: allele specific methylation;
ASW: African ancestry in Southwest USA;
BMI: body mass index;
BPRHS: Boston Puerto Rican Health Study;
C, cytosine;
C/EBPδ: CCAAT/enhancer-binding protein;
Cbs: cystathionine beta-synthase;
CEU: Utah residents with Northern and Western European Ancestry;
CGS: CpG related SNP;
CGS-C: CpG related SNP with the allele to create CpG dinucleotides;
CGS-Ci: index CGS-C;
CGS-Cp: proxy CGS-C;
CGS-D: CpG related SNP with the allele to disrupt CpG dinucleotides;
CGS-Di: index CGS-D;
CGS-Dp: proxy CGS-D;
CHARGE: the Cohorts for Heart and Aging Research in Genomic Epidemiology;
ChIP: chromatin immune-precipitation;
CHS: the Cardiovascular Health Study;
COX: cyclooxygenase;
CRP: C-reactive protein;
CTCF: CCCTC-binding factor;
CVD: cardiovascular disease;
DAC: 5-aza-2'-deoxycytidine;
DAP: death-associated protein;
DHA: docosahexaenoic acid;
DRD4: dopamine receptor 4 gene;
DS-DMR: developmental stage differential methylation region;
AD: Alzheimer's disease;
AdoHcy: S-adenosylhomocysteine;
AdoMet: S-adenosylmethionine;
AHA: American Heart Association;
ANCOVA: Analysis of covariance;
ANOVA, Analysis of variance;
EA: European American;
EDARADD: Edar associated death domain;
EMSA: electrophoretic mobility shift assay;
ENCODE: Encyclopedia of DNA Elements;
EPA: eicosapentaenoic acid;
eQTL: quantitative trait loci for gene expression;
ER: oestrogen receptor;
ESR1: estrogen receptor alpha;
ESE: exonic splicing enhancer;
Exp: expected probability;
F2RL3: coagulation factor II (thrombin) receptor-like 3;
F3: tissue factor;
Fads2: fatty acid desaturase 2;
FFQ: food frequency questionnaire;
FOXP3: forkhead box P3;
FTO: fat mass and obesity associated gene
G: guanosine;
G×E: gene-by-environment;
GOLDN: Genetics of Lipid Lowering Drugs and Diet Network;
GR: glucocorticoid receptor;
GSTM1: Glutathione S-transferase isoform mu1;
GSTM5: Glutathione S-transferase isoform mu5;
GWAS: genome-wide association studies;
HDL: high density lipoprotein;
HDL-C: high density lipoprotein cholesterol;
HEP: Human Epigenome Project;
HGP: Human Genome Project;
HHcy: hyperhomocysteinemia;
HMGCR: 3-hydroxy-3-methylglutaryl-coenzyme A reductase;
HNF1A: hepatocyte nuclear factor 1 homeobox A;
HNF4, *Hnf4a*: hepatocyte nuclear factor 4 a;

hsCRP: High-sensitivity C-reactive protein;
HWE: Hardy-Weinberg Equilibrium;
ICAM1: intercellular adhesion molecule 1;
IFNγ: interferon gamma;
IGF2: insulin-like growth factor 2;
InCHIANTI: the Invecchiare in Chianti;
IL1β : interleukin 1β;
IL2sR-α: interleukin-2 soluble receptor α;
IL-6, *IL6*: interleukin-6;
IL8: interleukin-8;
KCC3, SLC12A6: potassium-chloride co-transporter 3;
LD: linkage disequilibrium;
LDL: low density lipoprotein;
LDL-C: low density lipoprotein cholesterol;
LIPC: hepatic lipase;
LOX: lipoxygenase;
LPL: lipoprotein lipase;
LPS: lipopolysaccharide;
LXR: liver X receptors;
MAF: minor allele frequency;
MAOA: monoamine oxidase A;
MCP-1: monocyte chemoattractant protein-1;
MeCP2, *MECP2*: methyl-CpG-binding protein 2;
MESA: the Multi-Ethnic Study of Atherosclerosis;
MMP1: matrix metalloproteinase 1;
mQTL: quantitative trait loci for DNA methylation;
MSNP: methylation-sensitive SNP analysis;
MUFA: monounsaturated fatty acid;
MXL: Mexican ancestry in Los Angeles;
NCEP: National Cholesterol Education Program
NF-κB: nuclear factor kappa-light-chain enhancer of activated B cells;
NO2: nitrogen dioxide;
NPTX2: neuronal pentraxin II;
OA: oleic acid;
P16: cyclin-dependent kinase inhibitor 2A;
P(x>0): percentage of CpG sites with methylation level greater than 0;
PA: Palmitic acid;
PBMC: peripheral blood mononuclear cell;
PCSK9: proprotein convertase subtilisin/kexin type 9;
PGC-1α: peroxisome proliferator-activated receptor γ coactivator-1 α;
PM: particulate matter;
PPAR: proliferator-activated receptor;
PPARα: peroxisomal proliferator-activated receptor alpha;
PPARβ: peroxisomal proliferator-activated receptor beta;
PPARγ: peroxisomal proliferator-activated receptor gamma;
PPRE: peroxisome proliferator-activated receptors response element;
PUFA: polyunsaturated fatty acid;
RefSeq: Reference Sequence;
RRBS: reduced representative bisulfite sequencing;
RXR: retinoid X receptor;
SERT/SLC6A4: serotonin transporter gene;
SFA: saturated fatty acid;
SNP: single nucleotide polymorphism;
SREBP: sterol regulatory element-binding protein;
T: thymines;
TC: total cholesterol;
T-DMR: tissue differential methylation region;
TG: triglyceride;
TLR2: toll-like receptor 2;
TNF-α*,TNFα*: tumor necrosis factor alpha;
TOM1L1: target of myb1 (chicken)-like 1;
TSS: transcription start site;
UTR: untranslated region;
VCAM1: vascular cell adhesion molecule-1;
WC: waist circumference;
WGHS: the Women's Genome Health Study.
WHO: World Health Organization;
YRI: Yoruba in Ibadan, Nigeria;

LIST OF TABLES AND FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

APPENDIX

CHAPTER 1

Introduction

I. **Overview**

Cardiovascular disease (CVD) is the leading cause of total mortality not only in USA but also globally. According to the American Heart Association (AHA) report based on the 2010 data [1], the overall rate of death attributed to CVD was 235.5 per 100,000 USA population. Also, CVD accounted for approximately 1 out of 3 deaths in USA, and more than 2150 Americans die of CVD each day with an average of 1 death every 40 seconds. According to the World Health Organization (WHO) Fact Sheet, it was estimated that 17.3 million people die from CVDs in 2008, representing 30% of all global deaths. Over 80% of CVD deaths take place in low- and middle-income countries. The most important is that CVD are projected to remain the single leading cause of death worldwide.

With the goal to cure and prevent CVD, numerous risk factors have been identified, including dyslipidemia, inflammation, obesity, hypertension, smoking, and diabetes [2]. Dyslipidemia and chronic inflammation are two critical ones. Dyslipidemia refers to abnormalities of concentrations of a set of blood lipids. High concentrations of triglyceride (TG), total cholesterol (TC), and low density lipoprotein cholesterol (LDL-C), and low concentration of high density lipoprotein cholesterol (HDL-C) constitute the atherogenic dyslipidemia. According to the American Heart Association (AHA) [3], the prevalence of adults (age ≥ 20 y) having high triglyceride (TG **>** 150 mg/dl), high total cholesterol (TC≥ 200 mg/dl), high LDL-C (LDL-C≥130 mg/dl), and low HDL-cholesterol (HDL-C≤ 40 mg/dl) is 33%, 44.4%, 31.9%, and 18.9% respectively. Also, inflammation is part of the complex biological response to harmful stimuli,

which is common to a number of chronic diseases. However both dyslipidemia and inflammation are preventable or reversible by having a healthy lifestyle.

Of the factors that define a healthy lifestyle, diet is one of the most important components and the potential benefits of maintaining a healthy diet are well-established. Specifically, dietary fatty acids are associated with risk factors for CVD and different types of fatty acids have different effects. For example, although it is still under debate, in general, unsaturated fatty acids tend to increase HDL-C[4], reduce TG [5], and decrease IL-6 [6] compared to saturated fatty acids (SFA). Although monounsaturated fatty acids (MUFAs) and polyunsaturated fatty acids (PUFAs) differ in the magnitude of these beneficial effects [7], anti-atherosclerosis effect has been demonstrated for N3 not N6 PUFAs [8-18]. In addition, individuals exhibit different physiological responses to dietary fatty acids, reflecting, in part, the contributions of genetic variability [19].

The role of genetic factors in contributing to these inter-individual differences in lipid responses to dietary fatty acids has been widely studied. Our group has found that the association between dietary intake of total fat and plasma HDL-C was modified by the genetic variants located within hepatic lipase gene (*LIPC*) [20], and that the association between dietary PUFA intake and plasma fasting TG is modified by the genetic variants located within *APOA5* gene [21]. The effect of PUFA on HDL may differ according to different genotypes of several genes such as apolipoprotein A5 (*APOA5*), apolipoprotein A1 (*APOA1*), interleukin 6 (*IL6*), nuclear factor kappa-light-chain enhancer of activated B cells (*NF-κB*), tumor necrosis factor alpha (*TNF-α*)

[22-25]. In addition, there is also a genetically-based difference in TG response to N3 PUFAs [26, 27]. As a result of wide availability of new genetic technologies such as Genome-wide Association Studies (GWAS) and next generation sequencing, an enriched catalogue of common or rare single nucleotide polymorphisms (SNPs) has been formulated. However, the variation explained by all these genetic variants only account for less than 20%, indicating the existence of other sources of variability, such as epigenetic mechanisms.

Epigenetics has recently emerged as a research area of intense interest and growth. The definition of epigenetics underwent a series of changes as biological knowledge expanded. In 1940, 'epigenetics' was first defined as "… the interactions of genes with their environment which bring the phenotype into being…" by developmental biologists [28]. In the 1990s, epigenetics was described as the study of changes in gene expression which were not a result of changes in the DNA sequence [29]. Recently, inspired by genome-wide technologies, a new term 'epigenomics' has been coined, targeting the study of all factors contributing to changes in genome-wide chromatin structure including DNA methylation, histone modification, and chromatin remodeling [30].

Compared to the other two epigenetic mechanisms, DNA methylation is the most stable one and is the only one with direct relationships with DNA residues. DNA methylation is the addition of a methyl group directly onto DNA residues such as cytosine and adenine [31] and the C^5-methylcytosine modification is the major form in eukaryotes. DNA methylation can occur in different regions of the genome such as repetitive sequences, gene body, promoter related CpG

island and CpG island shore, which are located up to 2kb upstream of the CpG island [30]. DNA methylation patterns in different regions present different functions. For example, gene silencing is correlated with hypermethylation in promoter regions rather than in the gene body [32]. Also, cancer and aging are correlated with hypomethylation of repetitive elements while this is not the case for methylation of specific genes. Considering the different functionalities of DNA methylation in different regions, studies of DNA methylation occurring in specific sites of specific genes could provide more interpretable and meaningful explanations.

Similar to all the other epigenetic mechanisms, DNA methylation may act as a biomarker of the effect of environmental factors on the genetic structures. A wide array of environmental factors have been identified to affect DNA methylation patterns, including aging [33, 34], dietary fatty acids [35-37], malnutrition [38-40], dietary protein [41, 42], methyl-donors [43-45], chemical pollutants [46-48], sun exposure [49], and smoking [50, 51]. The connection of aging with DNA methylation was first observed in the candidate tumor suppressor genes, of which the methylation is increased with age, leading to the gene silencing [52]. Later, it was reported that the aging effects on DNA methylation is a prevalent phenomenon across the whole genome based on studies with monozygotic and dizygotic twins, which showed that the variation in DNA methylation increase significantly with age [33, 34]. Also, dietary fatty acids were suggested to regulate DNA methylation patterns. The intervention of a high fat diet was found to increase the DNA methylation of a metabolically related gene, peroxisome proliferator-activated receptor gamma, coactivator 1alpha (*PPARGC1A*); however, after the intervention was withdrawn, DNA methylation of *PPARGC1A* returned back to its baseline level [53]. The methylation of the same gene, *PPARGC1A,* was further reported to be affected by palmitic acid (PA) and oleic acid (OA)

[36]. Arachidonic acid (AA) and docosahexaenoic acid (DHA) were shown to affect DNA methylation of fatty acid desaturase 2 (*Fads2*) in mice liver [35]. In addition, eicosapentaenoic acid (EPA) was found to have demethylation effect on the tumor suppressor gene [37].

Besides numerous evidence of the effects of environmental factors, DNA methylation is shown to be related with different phenotypes. For instance, DNA methylation has been proposed as one mechanism of atherosclerosis [54]. In apolipoprotein E (*ApoE)* knock out mice, DNA methylation changes were shown to precede any histological sign of atherosclerosis [55]. In addition, the same study also found associations between global DNA hypermethylation and dyslipidemia, characterized by the atherogenic lipoproteins. An in vitro oligonucleotide binding assay found that a CG-rich 17-nucleotide sequence could bind to *ApoA1* [56], suggesting the relationship between lipoproteins and DNA methylation target sites, CpG dinucleotides. Besides affecting lipid concentrations, DNA methylation is also involved in inflammation. IL-6 is an acute phase protein induced during inflammation that functions as an inducer of differentiation of inflammatory helper T cells [57, 58]. DNA methylation has been identified as one mechanism of transcription regulation of *IL6*. For example, methylation of the promoter region in *IL6* is negatively correlated with gene expression [59] in peripheral blood mononuclear cells (PBMCs) and the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine (DAC) induces *IL6* transcription [60] in cancer cells. This silencing of *IL6* expression may be due to the binding of methyl-CpG-binding protein 2 (MeCP2) to the hypothetical binding sites in *IL6* gene, which is close to its transcription start site.

Based on the potential interplay between genetic variations and epigenetic mechanisms to modulate CVD risk factors, we aimed to test the interplay among genetic variants, DNA methylation, and environmental factors to modulate risk factors of CVD.

The significance of this research is related to its eventual translation into the arena of public health. The traditional concept of 'one size fits all' is limited, and the study of epigenetics will facilitate knowledge to further the development of personalized medical care. In this case, it is necessary to generate a more complete understanding of both genetic and epigenetic mechanisms contributing to the substantial inter-individual variations of response to environmental challenges. Through this research, we will also expand our knowledge of the molecular mechanism of gene-environment interaction and provide more solid evidence to promote new dietary guidelines.

Reference

1. Go, A.S., et al., *Executive summary: heart disease and stroke statistics--2014 update: a report from the American Heart Association.* Circulation, 2014. **129**(3): p. 399-410.
2. Grundy, S.M., et al., *Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations: a statement for healthcare professionals from the American Heart Association and the American College of Cardiology.* Circulation, 1999. **100**(13): p. 1481-92.
3. Roger, V.L., et al., *Heart disease and stroke statistics--2011 update: a report from the American Heart Association.* Circulation, 2010. **123**(4): p. e18-e209.
4. Hunter, J.E., J. Zhang, and P.M. Kris-Etherton, *Cardiovascular disease risk of dietary stearic acid compared with trans, other saturated, and unsaturated fatty acids: a systematic review.* Am J Clin Nutr, 2010. **91**(1): p. 46-63.
5. Lopez-Alvarenga, J.C., et al., *Polyunsaturated fatty acids effect on serum triglycerides concentration in the presence of metabolic syndrome components. The Alaska-Siberia Project.* Metabolism, 2010. **59**(1): p. 86-92.
6. Masson, C.J. and R.P. Mensink, *Exchanging saturated fatty acids for (n-6) polyunsaturated fatty acids in a mixed meal may decrease postprandial lipemia and markers of inflammation and endothelial activity in overweight men.* J Nutr, 2011. **141**(5): p. 816-21.

7. Bos, M.B., et al., *Effect of a high monounsaturated fatty acids diet and a Mediterranean diet on serum lipids and insulin sensitivity in adults with mild abdominal obesity.* Nutr Metab Cardiovasc Dis, 2010. **20**(8): p. 591-8.
8. Hansen, J.B., et al., *Effects of highly purified eicosapentaenoic acid and docosahexaenoic acid on fatty acid absorption, incorporation into serum phospholipids and postprandial triglyceridemia.* Lipids, 1998. **33**(2): p. 131-8.
9. Sudheendran, S., C.C. Chang, and R.J. Deckelbaum, *N-3 vs. saturated fatty acids: effects on the arterial wall.* Prostaglandins Leukot Essent Fatty Acids, 2010. **82**(4-6): p. 205-9.
10. Li, Z., et al., *Fish consumption shifts lipoprotein subfractions to a less atherogenic pattern in humans.* J Nutr, 2004. **134**(7): p. 1724-8.
11. Fakhrzadeh, H., et al., *The effects of low dose n-3 fatty acids on serum lipid profiles and insulin resistance of the elderly: a randomized controlled clinical trial.* Int J Vitam Nutr Res, 2010. **80**(2): p. 107-16.
12. Mori, T.A., et al., *Dietary fish as a major component of a weight-loss diet: effect on serum lipids, glucose, and insulin metabolism in overweight hypertensive subjects.* Am J Clin Nutr, 1999. **70**(5): p. 817-25.
13. Illingworth, D.R., W.S. Harris, and W.E. Connor, *Inhibition of low density lipoprotein synthesis by dietary omega-3 fatty acids in humans.* Arteriosclerosis, 1984. **4**(3): p. 270-5.
14. Kalogeropoulos, N., et al., *Unsaturated fatty acids are inversely associated and n-6/n-3 ratios are positively related to inflammation and coagulation markers in plasma of apparently healthy adults.* Clin Chim Acta, 2010. **411**(7-8): p. 584-91.
15. Farzaneh-Far, R., et al., *Inverse association of erythrocyte n-3 fatty acid levels with inflammatory biomarkers in patients with stable coronary artery disease: The Heart and Soul Study.* Atherosclerosis, 2009. **205**(2): p. 538-43.
16. Perunicic-Pekovic, G.B., et al., *Effect of n-3 fatty acids on nutritional status and inflammatory markers in haemodialysis patients.* Nephrology (Carlton), 2007. **12**(4): p. 331-6.
17. Sundrarjun, T., et al., *Effects of n-3 fatty acids on serum interleukin-6, tumour necrosis factor-alpha and soluble tumour necrosis factor receptor p55 in active rheumatoid arthritis.* J Int Med Res, 2004. **32**(5): p. 443-54.
18. Lopez-Garcia, E., et al., *Consumption of (n-3) fatty acids is related to plasma biomarkers of inflammation and endothelial activation in women.* J Nutr, 2004. **134**(7): p. 1806-11.
19. Schaefer, E.J., et al., *Individual variability in lipoprotein cholesterol response to National Cholesterol Education Program Step 2 diets.* Am J Clin Nutr, 1997. **65**(3): p. 823-30.
20. Ordovas, J.M., et al., *Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on high-density lipoprotein metabolism: evidence of a strong dose effect in this gene-nutrient interaction in the Framingham Study.* Circulation, 2002. **106**(18): p. 2315-21.
21. Lai, C.Q., et al., *Dietary intake of n-6 fatty acids modulates effect of apolipoprotein A5 gene on plasma fasting triglycerides, remnant lipoprotein concentrations, and lipoprotein particle size: the Framingham Heart Study.* Circulation, 2006. **113**(17): p. 2062-70.
22. Ordovas, J.M., et al., *Polyunsaturated fatty acids modulate the effects of the APOA1 G-A polymorphism on HDL-cholesterol concentrations in a sex-specific manner: the Framingham Study.* Am J Clin Nutr, 2002. **75**(1): p. 38-46.
23. Zhou, Q., et al., *Association of interleukin-6 gene -572 C > G polymorphism with dietary intake of n-3 fatty acids on plasma HDL-c level in Chinese male adults.* Asia Pac J Clin Nutr, 2010. **19**(4): p. 506-12.

24. Fontaine-Bisson, B., et al., *NF-kappaB -94Ins/Del ATTG polymorphism modifies the association between dietary polyunsaturated fatty acids and HDL-cholesterol in two distinct populations.* Atherosclerosis, 2009. **204**(2): p. 465-70.
25. Fontaine-Bisson, B., et al., *Genetic polymorphisms of tumor necrosis factor-alpha modify the association between dietary polyunsaturated fatty acids and fasting HDL-cholesterol and apo A-I concentrations.* Am J Clin Nutr, 2007. **86**(3): p. 768-74.
26. Lindi, V., et al., *Impact of the Pro12Ala polymorphism of the PPAR-gamma2 gene on serum triacylglycerol response to n-3 fatty acid supplementation.* Mol Genet Metab, 2003. **79**(1): p. 52-60.
27. Tai, E.S., et al., *Polyunsaturated fatty acids interact with the PPARA-L162V polymorphism to affect plasma triglyceride and apolipoprotein C-III concentrations in the Framingham Heart Study.* J Nutr, 2005. **135**(3): p. 397-403.
28. Waddington, C., *Organisers and Genes*. 1940: Cambridge University Press, Cambridge.
29. Wolffe, A.P. and M.A. Matzke, *Epigenetics: regulation through repression.* Science, 1999. **286**(5439): p. 481-6.
30. Portela, A. and M. Esteller, *Epigenetic modifications and human disease.* Nat Biotechnol, 2010. **28**(10): p. 1057-68.
31. Cheng, X., *Structure and function of DNA methyltransferases.* Annu Rev Biophys Biomol Struct, 1995. **24**: p. 293-318.
32. Hellman, A. and A. Chess, *Gene body-specific methylation on the active X chromosome.* Science, 2007. **315**(5815): p. 1141-3.
33. Fraga, M.F., et al., *Epigenetic differences arise during the lifetime of monozygotic twins.* Proc Natl Acad Sci U S A, 2005. **102**(30): p. 10604-9.
34. Wong, C.C., et al., *A longitudinal study of epigenetic variation in twins.* Epigenetics, 2010. **5**(6): p. 516-26.
35. Devlin, A.M., et al., *Hypermethylation of Fads2 and altered hepatic fatty acid and phospholipid metabolism in mice with hyperhomocysteinemia.* J Biol Chem, 2007. **282**(51): p. 37082-90.
36. Barres, R., et al., *Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density.* Cell Metab, 2009. **10**(3): p. 189-98.
37. Ceccarelli, V., et al., *Eicosapentaenoic acid demethylates a single CpG that mediates expression of tumor suppressor CCAAT/enhancer-binding protein delta in U937 leukemia cells.* J Biol Chem, 2011. **286**(31): p. 27092-102.
38. Waterland, R.A., et al., *Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles.* PLoS Genet, 2010. **6**(12): p. e1001252.
39. Heijmans, B.T., et al., *Persistent epigenetic differences associated with prenatal exposure to famine in humans.* Proc Natl Acad Sci U S A, 2008. **105**(44): p. 17046-9.
40. Tobi, E.W., et al., *DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific.* Hum Mol Genet, 2009. **18**(21): p. 4046-53.
41. Sandovici, I., et al., *Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the Hnf4a gene in rat pancreatic islets.* Proc Natl Acad Sci U S A, 2011. **108**(13): p. 5449-54.
42. Lillycrop, K.A., et al., *Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring.* J Nutr, 2005. **135**(6): p. 1382-6.
43. Hoyo, C., et al., *Methylation variation at IGF2 differentially methylated regions and maternal folic acid use before and during pregnancy.* Epigenetics, 2011. **6**(7): p. 928-36.

44. Steegers-Theunissen, R.P., et al., *Periconceptional maternal folic acid use of 400 microg per day is related to increased methylation of the IGF2 gene in the very young child.* PLoS One, 2009. **4**(11): p. e7845.
45. Waterland, R.A. and R.L. Jirtle, *Transposable elements: targets for early nutritional effects on epigenetic gene regulation.* Mol Cell Biol, 2003. **23**(15): p. 5293-300.
46. Baccarelli, A., et al., *Rapid DNA methylation changes after exposure to traffic particles.* Am J Respir Crit Care Med, 2009. **179**(7): p. 572-8.
47. Bollati, V., et al., *Changes in DNA methylation patterns in subjects exposed to low-dose benzene.* Cancer Res, 2007. **67**(3): p. 876-80.
48. Langevin, S.M., et al., *The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood.* Epigenetics, 2011. **6**(7): p. 908-19.
49. Gronniger, E., et al., *Aging and chronic sun exposure cause distinct epigenetic changes in human skin.* PLoS Genet, 2010. **6**(5): p. e1000971.
50. Breitling, L.P., et al., *Tobacco-smoking-related differential DNA methylation: 27K discovery and replication.* Am J Hum Genet, 2011. **88**(4): p. 450-7.
51. Belinsky, S.A., et al., *Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers.* Cancer Res, 2002. **62**(8): p. 2370-7.
52. Ahuja, N., et al., *Aging and DNA methylation in colorectal mucosa and cancer.* Cancer Res, 1998. **58**(23): p. 5489-94.
53. Brons, C., et al., *Deoxyribonucleic acid methylation and gene expression of PPARGC1A in human muscle is influenced by high-fat overfeeding in a birth-weight-dependent manner.* J Clin Endocrinol Metab, 2010. **95**(6): p. 3048-56.
54. Dong, C., W. Yoon, and P.J. Goldschmidt-Clermont, *DNA methylation and atherosclerosis.* J Nutr, 2002. **132**(8 Suppl): p. 2406S-2409S.
55. Lund, G., et al., *DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E.* J Biol Chem, 2004. **279**(28): p. 29147-54.
56. Panin, L.E., F.V. Tuzikov, and O.I. Gimautdinova, *Tetrahydrocortisol-apolipoprotein A-I complex specifically interacts with eukaryotic DNA and GCC elements of genes.* J Steroid Biochem Mol Biol, 2003. **87**(4-5): p. 309-18.
57. Veldhoen, M., et al., *TGFbeta in the context of an inflammatory cytokine milieu supports de novo differentiation of IL-17-producing T cells.* Immunity, 2006. **24**(2): p. 179-89.
58. Bettelli, E., et al., *Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells.* Nature, 2006. **441**(7090): p. 235-8.
59. Nile, C.J., et al., *Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis.* Arthritis Rheum, 2008. **58**(9): p. 2686-93.
60. Dandrea, M., et al., *MeCP2/H3meK9 are involved in IL-6 gene silencing in pancreatic adenocarcinoma cell lines.* Nucleic Acids Res, 2009. **37**(20): p. 6681-90.

II. Literature review:

1. **Gene-by-environment (G×E) interaction**

The connection between environment and phenotype is not as straightforward as what we expected, and there is a substantial variability in the individuals' response to the same environmental factor [1-4]. A well-controlled randomized clinical trial with low fat diets according to National Cholesterol Education Program (NCEP) Step 2 found that there was a large variability in lipid response to the diet. For example, the changes in low density lipoprotein (LDL) cholesterol ranged from +3% to -55% in men and from +13% to -39% in women [1]. A systematic analysis showed a normal distribution of the inter-individual variability in the lipid response to dietary interventions, including hypo-responders, hyper-responders, and normal-responders [2]. An intervention study with supervised cycle ergometer exercise found a marked variability in the high density lipoprotein (HDL) cholesterol response, within the range of approximately 30% [3]. Also, a cross-sectional study with 1,143 adults suggested that the magnitude of day-to-day variability of the cortisol awaking response increased with age among men [4]. The existence of such inter-individual variability suggested a modification role of genetic factors.

A wide array of evidence has been accumulated to indicate that genetic variants may contribute to the observed substantial inter-individual variability in the responses to different types of environmental factors. In the Framingham study, we have observed that the association between dietary total fat intake and plasma HDL cholesterol could be modified by the genetic variant located within promoter region of hepatic lipase gene (*LIPC*) [5]. Later, in the same well-

phenotyped population, we identified that the association between dietary polyunsaturated fatty acids (PUFAs) and the fasting plasma level of triglyceride (TG) may be modified by the promoter variant of apolipoprotein A5 (*APOA5*) [6]. Also, the aging effect on the Alzheimer's disease was found to be modified by the genetic variants characterizing different isoforms of apolipoprotein E (ApoE) [7]. Physical activity was found to have interaction with the genetic variants located within the *FTO* (fat mass and obesity associated gene) locus to modulate obesity in both African Americans and European Americans [8]. Smoking has been reported to interact with genetic variants within the inflammation related genes to modulate serum concentrations of inflammation markers [9-11]. Having alcohol intake greater than 20 g per day was found to interact with steatosis to affect liver fibrosis progression in 142 untreated patients showing positive for anti-hepatic C virus tests [12]. Despite extensive examples of such G×E interactions [13-15], our knowledge of the underlying mechanism to explain the observed G×E interaction is still limited. Epigenetic mechanism may partially contribute to the observed G×E interaction based on its established relationships with both genetics and environment.

2. Epigenetics and DNA methylation

2.1 Overview of epigenetics

Epigenetics acts as the cross-talk between the genome and environment because it studies the genetic response to the environmental factors. There are three major epigenetic mechanisms, including DNA methylation, histone modification, and chromatin remodeling. DNA methylation mechanism adds a methyl group onto DNA nucleotide such as cytosine and adenine [16]. With respect to histone modifications, a wide array of modifications were added to the histone tails, such as methylation, acetylation, phosphorylation, ubiquitylation, sumoylation, ADP ribosylation,

deamination, proline isomerization, crotonylation, propionylation, butyrylation, formylation, hydroxylation, and O-GlcNAcylation [17]. In terms of the chromatin remodeling mechanism, ATP-dependent enzymes remodel and control chromatin structure and assembly to make it become active or inactive to the extrinsic stimulus [18]. All of these epigenetic mechanisms have been suggested to play a critical role in developmental biology and complex diseases.

2.2 DNA methylation and DNA sequence

2.2.1 DNA methylation, CpG dinucleotides, CpG island

DNA methylation has the most direct contact with DNA nucleotide compared to the other two epigenetic mechanisms because the methyl-group was added directly on top of DNA nucleotide. In mammalian cells, most DNA methylations occur on the CpG dinucleotides, the major components for the CpG islands [19]. It is widely agreed that CpG islands are CpG dinucleotide enriched regions. However, the definitions of CpG islands have undergone an evolution from sliding window based methodology to statistics based test and currently to the experimental data based machine learning prediction. In 1987, Gardiner-Garden and Frommer [20] made the first definition of CpG islands based on vertebrate genomes sequence characteristics. They defined CpG island as " a stretch of DNA sequence where moving average of % G+C was greater than 50, and the moving average of ratio of the observed to expected CpG was greater than 0.6." These calculations are based on 100 bp window with sliding across the sequence at 1 bp intervals. Due to the lack of considerations of repetitive elements, the first definition has great false positive. In this case, Takai and Jones set up more stringent criteria for CpG islands, including criteria of having %G+C greater than 55, ratio of observed to expected CpG greater than 0.65

and sequence length being ≥ 500 bp. With Takai and Jones' criteria, a web page service algorithm "CpGIS" was developed [21]. Furthermore, Ponger [22] extended their criteria to estimate the transcription start sites associated CpG islands with algorithm "CpGProD". However, both "CpGIS" and "CpGProD" are subjective and computationally inefficient for the analysis of the genome-wide DNA sequences, so a new definition, named "CpGcluster" [23], was proposed. This algorithm is based on statistical test of the physical distance of neighboring CpG sites. It used only integer arithmetic algorithm which makes it fast and computationally efficient than the sliding window methods. However, it has the disadvantage of low sensitivity. Recently, a new algorithm "CpG_MI" [24] was developed to take into account more variability of the test such as different locations of CpG dinucleotide among different CpG islands. With the growing availability of the experimental results of the DNA methylation, the prediction of DNA methylation based on machine learning approach is possible. More specifically speaking, "EpiGRAPH" [25] algorithm for prediction of DNA methylation was trained by the wet-lab experiments data to predict the methylation probability of another stretch of DNA sequence.

2.2.2 DNA methylation and genetic variants

Recently, DNA methylation was suggested to be determined by the local DNA sequence. In one genetic manipulation study, 10 promoters with different endogenous DNA methylation patterns were inserted into one uniform DNA sequence shown to be epigenetically inert [26]. However, after insertion, most of these 10 promoters recapitulate their endogenous DNA methylation patterns, indicating a deterministic role of DNA sequence for the control of DNA methylation.

Meantime, a number of evidence has been accumulated in terms of the regulation of DNA methylation by the single nucleotide polymorphisms (SNPs) [27-33]. For example, the C allele of a SNP located within the promoter region of matrix metalloproteinase 1 (*MMP1*) was shown to have significantly higher DNA methylation status than the corresponding T allele [27]. Also, the G allele of one SNP located within the potassium-chloride co-transporter 3 (*KCC3*, *SLC12A6*) was found to be methylated at the adjacent C nucleotide [28].

Later, systematic analyses of the whole human genome have identified a wide array of such genetic variants having regulatory effects on DNA methylation patterns, indicating that the genetic regulations on DNA methylation are prevalent across the whole genome. For instance, a genomic survey using methylation-sensitive SNP analysis (MSNP) showed that 16 SNP-tagged loci were confirmed to have allele-specific DNA methylation (ASM) events [33]. Also, in the brain samples, approximately 10% of the CpG sites included in the analysis were found to be affected by the genotypes of the SNPs located in cis-position, while 0.1% of the analyzed CpG sites were regulated by the genotypes of the SNPs in the trans-position [30]. Furthermore, it was suggested by studies with 16 human pluripotent and adult cell lines that approximately one out of three (23% to 37%) heterozygous SNPs in the human genome may regulate DNA methylation patterns [29], and a big proportion of the observed loci with ASM events (38% to 88%) is dependent on the allele status of CpG related SNPs (CGSs), a type of SNPs with one allele to disrupt and the other allele to create CpG dinucleotides [29]. Finally, the effect of genetic variants outweighed the influence of imprinting on DNA methylation because it was shown that the number of methylation loci affected by genetic variants were much more than those loci influenced by the gender of parent of origin [31].

2.2.3 DNA methylation and genetic functions

DNA methylations have different genetic functions mostly depending on genetic locations. For example, DNA methylations within the promoter regions are more likely to regulate gene transcription [34, 35]; while, DNA methylations within the gene body tend to modify the alternative promoters and splicing events [36-40].

2.2.3.1 DNA methylation within promoter regions and gene transcription

The negative correlation between DNA methylation and gene transcription is striking and common to most genetic regions across the whole genome with rare exceptions [35, 41]. The first experiment indicating the transcription-regulatory effects of DNA methylation was conducted with restriction enzymes with different sensitivity to methylation at genetic locus of *β*-globin genes across different tissues of chicken [42]. Restriction enzyme *Hpa*II (CCGG) cannot cut the sequences with internal 5-methylated-cytosine, while enzyme *Msp*I (CCGG) can cleave the same DNA sequence regardless of their methylation status, so the treatment with both enzymes provide a sensitive tool to annotate the methylation status of DNA sequence of CCGG. By utilizing the differential cleavage of both enzymes *Hpa*II and *Msp*I, the CCGG sequence near the ends of the *β*-globin gene seem to be completely unmethylated in the cells expressing or have expressed the gene, including adult reticulocytes and erythrocytes; while the CCGG sequence were at least partially methylated in those cells not expressing the *β*-globin gene, such as oviduct, brain and embryonic red blood cells. Since the first indication, a large body of evidence has been accumulated to verify the inverse correlation between DNA methylation and gene transcription

for most genes, including but not limited to house-keeping genes [43], genes located on the inactive X chromosome [44-46], imprinted genes [33, 47], tumor suppressor genes or oncogenes [48-50], cellular differentiation and development related genes [51-54], metabolic genes [55-59], and inflammation related genes [60-64]. However, under rare circumstances, DNA methylation was shown to have a positive correlation with gene expressions [65-69]. Most of such transcriptional regulation effects were related to DNA methylations within the promoter regions [34] by direct blocking the binding of transcriptional activators or indirect recruitment of methyl-binding proteins and co-repressor complexes to facilitate the formation of heterochromatin in a cooperative way [70].

2.2.3.2 DNA methylation within gene bodies and alternative promoter and splicing events

Besides promoter regions, DNA methylations are also found with CpG sites located within gene bodies [38, 71-73], indicating their potential genetic functions other than gene transcriptions. By comparing differential DNA methylation patterns on a genome-wide scale across different tissues (brain, heart, liver, and testis) and different developmental stages (15 day embryo, new born, 12 week adult) of mice, approximately 16% of the identified tissue differential methylation regions (T-DMR) or developmental stage differential methylation regions (DS-DMR) were located within intragenic regions [73]. Also, it was found that the majority of methylated CpG sites were located within gene bodies, indicating their possible more important genetic functions regarding DNA methylations [38, 71, 72, 74]. For example, according to analysis with human normal tissues (whole blood, monocyte, granulocyte, skeletal muscle, spleen, and brain), 15.4% CpG islands located within the gene bodies were found to be methylated, which is higher than

the proportion of methylated CpG islands within 5' promoter region (7.8%) and the whole genetic region (10.6%) [72]. With samples of human brain tissues, Maunakea [38] generated high-resolution methylome maps with dense coverage of 24.7 million of the 28 million CpG sites across the whole genome. They found that 34% of all intragenic CpG islands were methylated, whereas only 2% of the CpG islands located within the 5' promoter regions were methylated, so they concluded that "DNA methylation may serve a broader role in intragenic compared to 5' promoter CpG islands in the human brain". Again, the altered DNA methylation in the immune system were shown to occur predominantly at CpG islands within gene bodies based on the analysis with both mouse cells within hematopoietic lineage [71] and human B cells [74].

The methylations within gene bodies may be related to alternative promoters [38] and alternative splicing events [36, 37, 39, 40]. According to the high-resolution and dense coverage methylome maps of human brain tissues, differentially methylated intragenic CpG islands have features of promoters, and novel transcripts have been found to be initiated from the identified differentially methylated and evolutionarily conserved intragenic promoters, indicating that intragenic methylation functions to regulate cell context-specific alternative promoters in gene bodies [38]. With a computational analysis of human chromosome 6, 20, and 22 based on datasets from the Human Epigenome Project (HEP) and the Human Genome Project (HGP), hypermethylated CpG sites were found to be prevalent in alternatively spliced sites, and the frequency of methylation increases in loci harboring multiple putative exonic splicing enhancers (ESEs) [40]. According to the analysis of data from RNA-seq experiments and methylome data with single nucleotide resolution of human cell lines, DNA methylation was found to be enriched in included alternatively spliced exons (ASEs), and inhibition of DNA methylation leaded to aberrant

splicing of ASEs. Further, they found that the alternative splicing may be because of the alternative definitions of exons via recruitment of methylated CpG site binding protein 2 (MeCP2) to the methylated CpG sites [36]. Another potential mechanism for the regulation of DNA methylation on alternative splicing events may be the fact that DNA methylation patterns affect chromatin structure [37]. Finally, a DNA methylation related protein, CCCTC-binding factor (CTCF), was shown to promote alternative splicing events on a genome-wide scale, providing potential links between DNA methylation and alternative splicing events [39].

2.3 DNA methylation and environmental factors

2.3.1 DNA methylation and aging

DNA methylation was indicated to be affected by aging process because of its intimate relationship with development. DNA methylation patterns change during each stage of development [75]. Before implantation, almost all DNA methylations were erased except for those imprinting regions. During implantation, the entire genome gets methylated except for the CpG islands. After implantation, pluripotency genes are de novo methylated and tissue-specific genes are demethylated in the cell types for their expression.

The correlations between aging and DNA methylation were also suggested by the in vitro studies. For example, compared to the immortal cell lines, normal diploid fibroblasts were found to have a dramatic decrease in their 5-methylcytosine contents during their growth in culture [76]. Furthermore, the observation that the decrease rate in mouse primary diploid fibroblasts were

faster than hamsters and humans and the fact that mouse has the shortest lifespan compared to hamsters and humans suggested that the rate of loss of 5-methylcytosine is positively correlated with growth potential. Also, the treatment of human diploid fibroblasts with DNA methylation inhibitors, azacytidine (5-aza-CR) and azadeoxycytidine (5-aza-CdR), were shown to inhibits the initial cellular growth [77].

Recently, a series of epidemiological analysis have indicated the potential relationships between aging and DNA methylation patterns. A cross-sectional study with monozygotic twins [78] found that younger twins have significantly lower levels of 5-methylated cytosines than older twins, and that the variance of DNA methylation of the older twins was significantly greater than that of the younger twins. The observed differences in DNA methylation was consistent with the findings with gene expression by showing that the 50-year-old twins had dramatically different expression profiles while the 3-year-old twins had almost identical ones. The observed discordance of DNA methylation with age was consistent across different tissues within the analysis, including lymphocytes, epithelial mouth cells, intra-abdominal fat, and skeletal muscle biopsies. Later, another study with 34 male monozygotic twins with age ranging from 21 to 55 year-old identified 88 sites located within or near 80 genes of which DNA methylation patterns were significantly correlated with age [79]. Three genes from that list of 80 genes were further validated and replicated with the analysis of their correlations with age in a population-based sample of 31 males and 29 females with age ranging from 18 to 70 years old, which are Edar associated death domain (*EDARADD*), target of myb1 (chicken)-like 1 (*TOM1L1*), and neuronal pentraxin II (*NPTX2*). Interestingly, all of these three genes have been reported to be associated with a wide array of age-related phenotypes, such as wound healing [80], Parkinson disease [81],

cancers [82, 83], and loss of teeth, hair, and sweat glands [84]. Also, a longitudinal study found that DNA methylation differs by age because methylation patterns of candidate genetic loci, such as the dopamine receptor 4 gene (DRD4), the serotonin transporter gene (SERT/SLC6A4), and the X-linked monoamine oxidase A gene (MAOA), were shown to change during the period when these children grew from 5 years to 10 years old.

Finally, changes in DNA methylation patterns have been reported to be associated with a series of age-related diseases. For instance, an extensive body of evidence has suggested that a global hypomethylation and gene specific promoter hypermethylation were associated with different types of cancer. It was found that the number of a subpopulation of cells in human colonic mucosa increase with age, and the promoter of oestrogen receptor (*ER*) gene in this subpopulation of cells becomes hypermethylated. And this age-related hypermethylation of *ER* was found in all cells in colorectal tumors examined [85]. Also, age-dependent methylation of estrogen receptor alpha (ESR1) was shown to be associated with prostate cancer [86]. The hypermethylation of several tumor suppressor genes have been suggested as the biomarkers of lung cancer [87]. Alzheimer's disease (AD), a demonstrated age-related disease, was found to be correlated with DNA methylation of CpG sites located near or within the genetic loci reported to harbor genetic susceptible risk variants for AD [88]. Compared to the normal retinas, retinas of patients with age-related macular degeneration were found to have hypermethylation and gene repression of Glutathione S-transferase isoform mu1 (*GSTM1*) and mu5 (*GSTM5*) [89].

2.3.2 DNA methylation and fatty acids

The connections between dietary fatty acids and DNA methylation were strongly indicated by the striking effects of fatty acids on gene expressions, one of the major genetic functions of DNA methylations as mentioned above. Fatty acids affect expressions of a wide array of genes by acting as the important ligands for transcription factors, such as peroxisome proliferator-activated receptors (PPARs), the liver X receptors (LXRs), retinoid X receptor (RXR), hepatocyte nuclear factor 4 (HNF4), sterol regulatory element-binding proteins (SREBPs), nuclear factor kappa-light-chain enhancer of activated B cells (NF-κB), cyclooxygenase (COX), and lipoxygenase (LOX). PPARs and LXR are members of the nuclear hormone receptor superfamily of transcription factors, which bind to specific motifs within the promoters of genes as heterodimers with the RXR. There are three isoforms of PPARs, including PPARα, PPARβ, and PPARγ. In general, PPARs bind with both saturated and unsaturated fatty acids with a relatively more potent binding with N6 and N3 polyunsaturated fatty acids (PUFAs) and their derivatives to regulate expressions of genes that control lipid and glucose homeostasis and inflammation. Regarding LXRs, there are two family members, LXRα and LXRβ. As a sensor of cholesterol in the nucleus, LXRs can be activated by increased intracellular cholesterol concentrations. Also, the biding of long chain fatty acids to LXRs [90] was shown to regulate expressions of genes involved in sterol and fatty acid metabolism [91], lipogenesis [92-95], carbohydrate metabolism [96, 97]. HNF4α is an orphan member of the steroid hormone receptor superfamily and functions by binding with the activated (CoA) form of fatty acids to regulate expressions of genes participating in the lipid and lipoprotein metabolisms [98, 99] and glucose metabolism [100, 101]. SREBPs have three isoforms, which are SREBP-1a, SREBP-1c, and SREBP-2, and all of them are the transcription factors playing a critical role in controlling

synthesis of fatty acids, triglycerides, and cholesterol [102]. PUFAs were found to lower the mature form of the protein levels of SREBPs by raising cellular cholesterol levels, or by reducing SREBP mRNA stability and SREBP transcription, or by promoting degradation of SREBP protein [103-106]. COX and LOX function to convert N6 and N3 PUFAs into pro- and anti-inflammatory signaling molecules to regulate activity of transcription factors of inflammation such as NF-κB [107].

The effect of fatty acids on DNA methylation was also indirectly suggested by a study with mice heterozygous for disruption of cystathionine beta-synthase (*Cbs*+/-) [58], which could be induced to have hyperhomocysteinemia (HHcy), providing an indirect evidence because of the potential modifications on DNA methylation by homocysteine through its participating in the one-carbon metabolism. In that study, a dosage of HHcy (normal, mild, and moderate) was developed by treating the mice (Cbs+/+) with control diet (normal), treating the mice (Cbs+/+) with diet to induce HHcy (mild), and treating the mice (Cbs+/-) with diet to induce HHcy (moderate). The potential relationship between homocysteine and DNA methylation was supported by the significantly inverse correlation between total homocysteine levels and liver methylation capacity, measured by the ratio of S-adenosylmethionine (AdoMet) to S-adenosylhomocysteine (AdoHcy). Correspondingly, mice with moderate HHcy had higher methylation of candidate CpG sites within the promoter region of fatty acid desaturase 2 (*Fads2*) in liver, leading to lower gene expression of *Fads2* and lower protein activity of δ(6)-desaturase (encoded by *Fads2*) in liver, compared to mice with mild and normal HHcy. Also, mice with moderate HHcy have lowest level of arachidonic acid (AA) and docosahexaenoic acid (DHA) in total liver than those mice with mild and normal HHcy.

Direct evidence for the link between fatty acids and DNA methylation were conducted with *in vitr*o and *in vivo* studies. It was found that incubation of human skeletal muscle cells with 48 hours treatment with free fatty acids, such as palmitate and oleate, can increase DNA methylation levels of the promoter region of peroxisome proliferator-activated receptor γ (PPARγ) coactivator-1 α (*PGC-1α*), leading to suppression of its gene expression [108]. Also, *in vitro* treatment of U937 leukemia cells with eicosapentaenoic acid (EPA) was found to decrease methylation of the promoter regions of a myeloid lineage-specific transcription factor CCAAT/enhancer-binding protein (*C/EBPδ*), a tumor suppressor gene, resulting in an increased gene expression [109]. One *in vivo* study with rats found that feeding a diet high in N3 PUFAs, mainly with EPA and DHA, could significantly decrease global DNA methylation levels [110].

A randomized control trial with high-fat overfeeding in young adults with low or normal birth weight confirm the potential relationship between fatty acids and DNA methylation. It was reported that having high-fat overfeeding (+50% calories) for five days increased DNA methylation in the promoter region of a metabolic gene, *PGC-1α*, measured in the skeletal muscle cells extracted from healthy young men with low birth weight [59]. The observed induction of DNA methylation in *PGC-1α* was found to be reversible because DNA methylation returned to its baseline level after the high-fat diets were withdrawn. Although DNA methylation of *PGC-1α* was not found to have significant correlation with its gene expression, high-fat challenge in the subjects with low birth weight were shown to induce peripheral insulin resistance and decrease gene expression of *PGC-1α*.

2.3.3 DNA methylation and other environmental factors

Besides aging and dietary fatty acids, DNA methylation patterns are modifiable by several other environmental factors, including nutrition status, air pollution, weather, and smoking. Supplementations of methyl donors during gestational stage of mice was shown to have a dose-response relationship with the methylation of viable yellow agouti (A^{vy}) locus and the browness of coat color in the offspring [111]. Calorie restriction *in utero* was found to decrease the overall methylations and changes in the methylation patterns of imprinted loci in mice [112]. Similarly in humans, those subjects having experienced famine prenatally because of their *in utero* exposure to the Dutch Hunger Winter were shown to have less DNA methylation of the imprinted gene, insulin-like growth factor 2 (*IGF2*) [113, 114]. It was found that the increased concentrations of ozone and components of fine article mass were associated with hypomethylation of tissue factor (*F3*), intercellular adhesion molecule 1 (*ICAM1*), and toll-like receptor 2 (*TLR2*), and hypermethylation of interferon gamma (*IFNγ*) and interleukin 6 (*IL6*) [115]. Increased exposures to air pollutants, such as nitrogen dioxide (NO2), particulate matter 10 (PM10), PM2.5, and ozone, were found to decrease global DNA methylation in whole blood [116]. With a genome-wide scale analysis followed by an independent replication study, those individuals with smoking were suggested to have decreased level of DNA methylation of a single CpG site located within the coagulation factor II (thrombin) receptor-like 3 (*F2RL3*) [117]. Also, methylation of tumor suppressor genes, cyclin-dependent kinase inhibitor 2A (p16) and death-associated protein (DAP) kinase might lead to lung cancer [118]. Sun exposure was indicated to be associated with the phenotypic changes related with skin aging by their modifications of DNA methylation across the genome [119]. Finally, individuals with *in utero*

exposure to rainy season in rural Gambia were shown to increase methylation of genetic regions contributing to the dramatic and systemic inter-individual variations in epigenetic regulations [120].

2.4 DNA methylation and CVD risk factors

2.4.1 DNA methylation and inflammation

Accumulating evidence have suggested that DNA methylation patterns were associated with inflammatory markers, such as interleukin 6 (*IL6*) [60, 61, 121-123], interleukin 1β (*IL1β*) [60], and interleukin 8 (*IL8*) [60], high sensitivity C-reactive protein (*hsCRP*) [124], and vascular cell adhesion molecule-1 (*VCAM1*) [125]. A case control study found that patients with Rheumatoid Arthritis have lower DNA methylation levels of a CpG site, which was located at -1099 bp to the transcription start site of *IL6*, measured in peripheral blood monuclear cells (PBMCs). In the macrophages from healthy control subjects, lower methylation of the previously identified CpG site was in line with the higher *IL6* expression stimulated by lipopolysaccharide (LPS). Experiments with electrophoretic mobility shift assay (EMSA) provided potential mechanistic explanation for the observed associations by identifying the methylation-dependent affinity of protein-DNA interactions [61]. The *in vitro* treatment of 5-aza-2'-deoxycytidine was capable to activate *IL6* expression in all six analyzed human pancreatic adenocarcinoma cell lines, indicating an important role of DNA methylation at *IL6* genetic locus [126]. Also, the chromatin immune-precipitation (ChIP) assays with the same cell lines identified a potential response element to the binding of MeCP2, located from -666 to -426 bp to the transcription start sites, providing potentially mechanistic explanations for the DNA methylation of *IL6* [126]. A cross-

sectional study with blood leukocyte found that workers living in the industrial estate had the lowest methylation levels of the second intron of *IL6*, with rural and urban residents had the highest and intermediate levels [121]. Another cross-sectional study with white blood cells found that a prudent diet, characterized by a high intake of vegetables and fruits, was associated with DNA methylation levels of the promoter region of *IL6* [122]. According to the analysis of DNA methylation patterns of *IL6* in periodontal tissues, patients with periodontitis were found to have lower methylation and higher gene expression [123]. An *in vitro* study with cultured human lung cells showed that the DNA methylation levels of promoter regions of a panel of inflammation related genes (*IL6*, *IL1β*, and *IL8*) were higher in cancer cells than normal ones, and the higher methylations went along with the lower gene expressions [60]. A study with patients with pediatric obstructive sleep apnea found that DNA methylation of forkhead box P3 (*FOXP3*) had significantly positive correlations with serum levels of hsCRP [124]. A cross-sectional study with blood samples from 742 community-dwelling elderly individuals found that hypomethylation of repetitive element LINE-1 was associated with increased levels of serum VCAM1 [125]. Finally, a study with samples of peripheral blood leukocytes from 966 African American identified that DNA methylations of 257 CpG sites within 240 genes contribute to serum levels of CRP [127].

2.4.2 DNA methylation and dyslipidemia

DNA methylation patterns have been shown to be related with dyslipidemia [128-131]. After stimulation with lipoprotein mixture (68.8 ug/ml VLDL, 32.1 ug/ml LDL, 91.1 ug/ml HDL), the global levels of 5-methylated cytosines within the differentiated human monocyte-macrophage

cell line THP-1 was significantly increased [128]. According to a genome-wide DNA methylation analysis with samples of CD4+ cells from 991 individuals of the Genetics of Lipid-lowering Drugs and Diet Network (GOLDN) Study, four CpG sites located within the intron 1 of carnitine palmitoyltransferase 1 A (*CPT1A*) were found to be associated with fasting levels of VLDL cholesterol and triglyceride (TG). DNA methylation of the CpG site with top findings was further found to be associated with *CPT1A* expression. The observed association between DNA methylation, gene expression, and fasting TG was replicated in an independent population, Framingham Heart Study [130]. Also, a higher methylation pattern of the promoter region of ATP-binding cassette A1 (*ABCA1*) in samples of whole blood was found to be associated with a lower circulating HDL cholesterol and HDL2-phospholipid levels in 97 patients with familial hypercholesterolemia [131]. Similarly in patients with familial hypercholesterolemia, leukocyte DNA methylations of lipoprotein lipase (*LPL*) had positive correlations with HDL cholesterol and HDL particle size, whereas DNA methylations of cholesteryl ester transfer protein (*CETP*) had a negative association with LDL cholesterol in all the participants and negative associations with HDL cholesterol, HDL-TG levels, and HDL partical size [132]. Further, the methylations of *LPL* in visceral adipose tissue extracted from 30 men with severe obesity were found to have negative correlations with HDL cholesterol and gene expression of *LPL* [132]. The potential mechanism for the effects of lipoproteins on DNA methylation is unknown. The modifications of chromatin structure may account as one potential mechanism, because it was found that apolipoprotein AI (ApoA-I) can physically bind to a CG-rich oligonucleotide *in vitro*, leading to the remodeling of chromatin structure [129].

3. Conclusion:

Cardiovascular disease (CVD) constitute as a major public health problem because of its leading role to the death rate in not only USA but also globally. Dyslipidemia and inflammation are two major risk factors for CVD. Meantime, a wide array of environmental factors have been identified to contribute to these two risk factors, including diet, physical activity, alcohol drinking, and smoking. However, the connections between environmental factors and phenotypes are not as straight-forwarded as what we have expected. Actually, there are many steps from environmental factors to the final phenotypes. For example, in 1990s, a substantial inter-individual variability in terms of lipid response to the same dietary intervention program has been identified although almost all the subjects started with the same levels of blood lipids. After approximately 10 years' research, scientists suggested that genetic variants may be the underlying factor to explain the observed substantial inter-individual variability. Despite numerous evidence have been accumulated regarding gene-by-environment (G×E) interactions, the underlying mechanism is still unclear. Based on the established roles with both nature and nurture, DNA methylation, one epigenetic mechanisms, may account, at least in part, for the observed G×E interactions.

4. References:

1. Schaefer, E.J., et al., *Individual variability in lipoprotein cholesterol response to National Cholesterol Education Program Step 2 diets.* Am J Clin Nutr, 1997. **65**(3): p. 823-30.
2. Jacobs, D.R., Jr., et al., *Variability in individual serum cholesterol response to change in diet.* Arteriosclerosis, 1983. **3**(4): p. 349-56.
3. Leon, A.S., et al., *Variability in the response of HDL cholesterol to exercise training in the HERITAGE Family Study.* Int J Sports Med, 2002. **23**(1): p. 1-9.
4. Almeida, D.M., J.R. Piazza, and R.S. Stawski, *Interindividual differences and intraindividual variability in the cortisol awakening response: an examination of age and gender.* Psychol Aging, 2009. **24**(4): p. 819-27.

5. Ordovas, J.M., et al., *Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on high-density lipoprotein metabolism: evidence of a strong dose effect in this gene-nutrient interaction in the Framingham Study.* Circulation, 2002. **106**(18): p. 2315-21.
6. Lai, C.Q., et al., *Dietary intake of n-6 fatty acids modulates effect of apolipoprotein A5 gene on plasma fasting triglycerides, remnant lipoprotein concentrations, and lipoprotein particle size: the Framingham Heart Study.* Circulation, 2006. **113**(17): p. 2062-70.
7. Chen, J., et al., *The Interaction of APOE Genotype by Age in Amnestic Mild Cognitive Impairment: A Voxel-Based Morphometric Study.* J Alzheimers Dis, 2014.
8. Demerath, E.W., et al., *Interaction of FTO and physical activity level on adiposity in African-American and European-American adults: the ARIC study.* Obesity (Silver Spring), 2011. **19**(9): p. 1866-72.
9. Rosner, S.A., et al., *Interaction between inflammation-related gene polymorphisms and cigarette smoking on the risk of myocardial infarction in the Physician's Health Study.* Hum Genet, 2005. **118**(2): p. 287-94.
10. Saijo, Y., et al., *Effects of the Interaction between Interleukin-6 -634C/G Polymorphism and Smoking on Serum C-Reactive Protein Concentrations.* Hypertens Res, 2007. **30**(7): p. 593-9.
11. Sunyer, J., et al., *Interaction between smoking and the interleukin-6 gene affects systemic levels of inflammatory biomarkers.* Nicotine Tob Res, 2009. **11**(11): p. 1347-53.
12. Serfaty, L., et al., *Effect of the interaction between steatosis and alcohol intake on liver fibrosis progression in chronic hepatitis C.* Am J Gastroenterol, 2002. **97**(7): p. 1807-12.
13. Corella, D. and J.M. Ordovas, *Nutrigenomics in cardiovascular medicine.* Circ Cardiovasc Genet, 2009. **2**(6): p. 637-51.
14. Ordovas, J.M., et al., *Gene-diet interaction in determining plasma lipid response to dietary intervention.* Atherosclerosis, 1995. **118 Suppl**: p. S11-27.
15. Hegele, R.A., *Gene-environment interactions in atherosclerosis.* Mol Cell Biochem, 1992. **113**(2): p. 177-86.
16. Cheng, X., *Structure and function of DNA methyltransferases.* Annu Rev Biophys Biomol Struct, 1995. **24**: p. 293-318.
17. Dawson, M.A. and T. Kouzarides, *Cancer epigenetics: from mechanism to therapy.* Cell, 2012. **150**(1): p. 12-27.
18. Ho, L. and G.R. Crabtree, *Chromatin remodelling during development.* Nature, 2010. **463**(7280): p. 474-84.
19. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-22.
20. Gardiner-Garden, M. and M. Frommer, *CpG islands in vertebrate genomes.* J Mol Biol, 1987. **196**(2): p. 261-82.
21. Takai, D. and P.A. Jones, *The CpG island searcher: a new WWW resource.* In Silico Biol, 2003. **3**(3): p. 235-40.
22. Ponger, L. and D. Mouchiroud, *CpGProD: identifying CpG islands associated with transcription start sites in large genomic mammalian sequences.* Bioinformatics, 2002. **18**(4): p. 631-3.
23. Hackenberg, M., et al., *CpGcluster: a distance-based algorithm for CpG-island detection.* BMC Bioinformatics, 2006. **7**: p. 446.
24. Su, J., et al., *CpG_MI: a novel approach for identifying functional CpG islands in mammalian genomes.* Nucleic Acids Res, 2010. **38**(1): p. e6.
25. Bock, C., et al., *EpiGRAPH: user-friendly software for statistical analysis and prediction of (epi)genomic data.* Genome Biol, 2009. **10**(2): p. R14.

26. Lienert, F., et al., *Identification of genetic elements that autonomously determine DNA methylation states.* Nat Genet, 2011. **43**(11): p. 1091-7.
27. Wang, H., et al., *Genetic and epigenetic mechanisms combine to control MMP1 expression and its association with preterm premature rupture of membranes.* Hum Mol Genet, 2008. **17**(8): p. 1087-96.
28. Moser, D., et al., *Functional analysis of a potassium-chloride co-transporter 3 (SLC12A6) promoter polymorphism leading to an additional DNA methylation site.* Neuropsychopharmacology, 2009. **34**(2): p. 458-67.
29. Shoemaker, R., et al., *Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome.* Genome Res, 2010. **20**(7): p. 883-9.
30. Zhang, D., et al., *Genetic control of individual differences in gene-specific methylation in human brain.* Am J Hum Genet, 2010. **86**(3): p. 411-9.
31. Gertz, J., et al., *Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation.* PLoS Genet, 2011. **7**(8): p. e1002228.
32. Qu, W., et al., *Genome-wide genetic variations are highly correlated with proximal DNA methylation patterns.* Genome Res, 2012. **22**(8): p. 1419-25.
33. Kerkel, K., et al., *Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation.* Nat Genet, 2008. **40**(7): p. 904-8.
34. Portela, A. and M. Esteller, *Epigenetic modifications and human disease.* Nat Biotechnol, 2010. **28**(10): p. 1057-68.
35. Eden, S. and H. Cedar, *Role of DNA methylation in the regulation of transcription.* Curr Opin Genet Dev, 1994. **4**(2): p. 255-9.
36. Maunakea, A.K., et al., *Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition.* Cell Res, 2013. **23**(11): p. 1256-69.
37. Gelfman, S. and G. Ast, *When epigenetics meets alternative splicing: the roles of DNA methylation and GC architecture.* Epigenomics, 2013. **5**(4): p. 351-3.
38. Maunakea, A.K., et al., *Conserved role of intragenic DNA methylation in regulating alternative promoters.* Nature, 2010. **466**(7303): p. 253-7.
39. Shukla, S., et al., *CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.* Nature, 2011. **479**(7371): p. 74-9.
40. Anastasiadou, C., et al., *Human epigenome data reveal increased CpG methylation in alternatively spliced sites and putative exonic splicing enhancers.* DNA Cell Biol, 2011. **30**(5): p. 267-75.
41. Razin, A. and H. Cedar, *DNA methylation and gene expression.* Microbiol Rev, 1991. **55**(3): p. 451-8.
42. McGhee, J.D. and G.D. Ginder, *Specific DNA methylation sites in the vicinity of the chicken beta-globin genes.* Nature, 1979. **280**(5721): p. 419-20.
43. Stein, R., et al., *Pattern of methylation of two genes coding for housekeeping functions.* Proc Natl Acad Sci U S A, 1983. **80**(9): p. 2422-6.
44. Toniolo, D., et al., *Specific methylation pattern at the 3' end of the human housekeeping gene for glucose 6-phosphate dehydrogenase.* EMBO J, 1984. **3**(9): p. 1987-95.
45. Yen, P.H., et al., *Differential methylation of hypoxanthine phosphoribosyltransferase genes on active and inactive human X chromosomes.* Proc Natl Acad Sci U S A, 1984. **81**(6): p. 1759-63.
46. Wolf, S.F., et al., *Complete concordance between glucose-6-phosphate dehydrogenase activity and hypomethylation of 3' CpG clusters: implications for X chromosome dosage compensation.* Nucleic Acids Res, 1984. **12**(24): p. 9333-48.
47. Orozco, L.D., et al., *Intergenerational genomic DNA methylation patterns in mouse hybrid strains.* Genome Biol, 2014. **15**(5): p. R68.

48. Yamada, N., et al., *MUC1 expression is regulated by DNA methylation and histone H3 lysine 9 modification in cancer cells.* Cancer Res, 2008. **68**(8): p. 2708-16.
49. Wang, J., et al., *Delta DNMT3B variants regulate DNA methylation in a promoter-specific manner.* Cancer Res, 2007. **67**(22): p. 10647-52.
50. Harada, K., et al., *Aberrant promoter methylation and silencing of the RASSF1A gene in pediatric tumors and cell lines.* Oncogene, 2002. **21**(27): p. 4345-9.
51. Lu, R., et al., *Inhibition of the extracellular signal-regulated kinase/mitogen-activated protein kinase pathway decreases DNA methylation in colon cancer cells.* J Biol Chem, 2007. **282**(16): p. 12249-59.
52. Wei, D. and M.R. Loeken, *Increased DNA Methyltransferase 3b (Dnmt3b) -mediated CpG Island Methylation Stimulated by Oxidative Stress Inhibits Expression of a Gene Required for Neural Tube and Neural Crest Development in Diabetic Pregnancy.* Diabetes, 2014.
53. Lee, S.T., et al., *A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network.* Nucleic Acids Res, 2012. **40**(22): p. 11339-51.
54. Nishino, K., et al., *DNA methylation-mediated control of Sry gene expression in mouse gonadal development.* J Biol Chem, 2004. **279**(21): p. 22306-13.
55. Laker, R.C., et al., *Exercise prevents maternal high-fat diet-induced hypermethylation of the Pgc-1alpha gene and age-dependent metabolic dysfunction in the offspring.* Diabetes, 2014. **63**(5): p. 1605-11.
56. Barres, R., et al., *Acute exercise remodels promoter methylation in human skeletal muscle.* Cell Metab, 2012. **15**(3): p. 405-11.
57. Ling, C., et al., *Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion.* Diabetologia, 2008. **51**(4): p. 615-22.
58. Devlin, A.M., et al., *Hypermethylation of Fads2 and altered hepatic fatty acid and phospholipid metabolism in mice with hyperhomocysteinemia.* J Biol Chem, 2007. **282**(51): p. 37082-90.
59. Brons, C., et al., *Deoxyribonucleic acid methylation and gene expression of PPARGC1A in human muscle is influenced by high-fat overfeeding in a birth-weight-dependent manner.* J Clin Endocrinol Metab, 2010. **95**(6): p. 3048-56.
60. Tekpli, X., et al., *DNA methylation at promoter regions of interleukin 1B, interleukin 6, and interleukin 8 in non-small cell lung cancer.* Cancer Immunol Immunother, 2013. **62**(2): p. 337-45.
61. Nile, C.J., et al., *Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis.* Arthritis Rheum, 2008. **58**(9): p. 2686-93.
62. Fernandez-Alvarez, A., et al., *Evaluation of epigenetic modulation of cyclooxygenase-2 as a prognostic marker for hepatocellular carcinoma.* Oncogenesis, 2012. **1**: p. e23.
63. Turcot, V., et al., *DPP4 gene DNA methylation in the omentum is associated with its gene expression and plasma lipid profile in severe obesity.* Obesity (Silver Spring), 2011. **19**(2): p. 388-95.
64. El Gazzar, M., et al., *Epigenetic silencing of tumor necrosis factor alpha during endotoxin tolerance.* J Biol Chem, 2007. **282**(37): p. 26857-64.
65. Bell, A.C. and G. Felsenfeld, *Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene.* Nature, 2000. **405**(6785): p. 482-5.
66. Morey, S.R., et al., *DNA methylation pathway alterations in an autochthonous murine model of prostate cancer.* Cancer Res, 2006. **66**(24): p. 11659-67.
67. Wu, H., et al., *Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes.* Science, 2010. **329**(5990): p. 444-8.

68. Dias, R.P., et al., *Multiple segmental uniparental disomy associated with abnormal DNA methylation of imprinted Loci in silver-russell syndrome.* J Clin Endocrinol Metab, 2012. **97**(11): p. E2188-93.
69. Bahar Halpern, K., T. Vana, and M.D. Walker, *Paradoxical Role of DNA Methylation in Activation of FoxA2 Gene Expression during Endoderm Development.* J Biol Chem, 2014. **289**(34): p. 23882-92.
70. Klose, R.J. and A.P. Bird, *Genomic DNA methylation: the mark and its mediators.* Trends Biochem Sci, 2006. **31**(2): p. 89-97.
71. Deaton, A.M., et al., *Cell type-specific DNA methylation at intragenic CpG islands in the immune system.* Genome Res, 2011. **21**(7): p. 1074-86.
72. Illingworth, R., et al., *A novel CpG island set identifies tissue-specific methylation at developmental gene loci.* PLoS Biol, 2008. **6**(1): p. e22.
73. Liang, P., et al., *Genome-wide survey reveals dynamic widespread tissue-specific changes in DNA methylation during development.* BMC Genomics, 2011. **12**(1): p. 231.
74. Rauch, T.A., et al., *A human B cell methylome at 100-base pair resolution.* Proc Natl Acad Sci U S A, 2009. **106**(3): p. 671-8.
75. Cedar, H. and Y. Bergman, *Programming of DNA methylation patterns.* Annu Rev Biochem, 2012. **81**: p. 97-117.
76. Wilson, V.L. and P.A. Jones, *DNA methylation decreases in aging but not in immortal cells.* Science, 1983. **220**(4601): p. 1055-7.
77. Holliday, R., *Strong effects of 5-azacytidine on the in vitro lifespan of human diploid fibroblasts.* Exp Cell Res, 1986. **166**(2): p. 543-52.
78. Fraga, M.F., et al., *Epigenetic differences arise during the lifetime of monozygotic twins.* Proc Natl Acad Sci U S A, 2005. **102**(30): p. 10604-9.
79. Bocklandt, S., et al., *Epigenetic predictor of age.* PLoS One, 2011. **6**(6): p. e14821.
80. Langton, A.K., S.E. Herrick, and D.J. Headon, *An extended epidermal response heals cutaneous wounds in the absence of a hair follicle stem cell contribution.* J Invest Dermatol, 2008. **128**(5): p. 1311-8.
81. Moran, L.B., et al., *Neuronal pentraxin II is highly upregulated in Parkinson's disease and a novel component of Lewy bodies.* Acta Neuropathol, 2008. **115**(4): p. 471-8.
82. Park, J.K., et al., *Quantitative analysis of NPTX2 hypermethylation is a promising molecular diagnostic marker for pancreatic cancer.* Pancreas, 2007. **35**(3): p. e9-15.
83. Qi, Y., et al., *Decreased Srcasm expression in esophageal squamous cell carcinoma in a Chinese population.* Anticancer Res, 2010. **30**(9): p. 3535-9.
84. Yan, M., et al., *Identification of a novel death domain-containing adaptor molecule for ectodysplasin-A receptor that is mutated in crinkled mice.* Curr Biol, 2002. **12**(5): p. 409-13.
85. Issa, J.P., et al., *Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon.* Nat Genet, 1994. **7**(4): p. 536-40.
86. Li, L.C., et al., *Age-dependent methylation of ESR1 gene in prostate cancer.* Biochem Biophys Res Commun, 2004. **321**(2): p. 455-61.
87. Belinsky, S.A., *Gene-promoter hypermethylation as a biomarker in lung cancer.* Nat Rev Cancer, 2004. **4**(9): p. 707-17.
88. De Jager, P.L., et al., *Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci.* Nat Neurosci, 2014.
89. Hunter, A., et al., *DNA methylation is associated with altered gene expression in AMD.* Invest Ophthalmol Vis Sci, 2012. **53**(4): p. 2089-105.
90. Janowski, B.A., et al., *Structural requirements of ligands for the oxysterol liver X receptors LXRalpha and LXRbeta.* Proc Natl Acad Sci U S A, 1999. **96**(1): p. 266-71.

91. Geyeregger, R., M. Zeyda, and T.M. Stulnig, *Liver X receptors in cardiovascular and metabolic disease.* Cell Mol Life Sci, 2006. **63**(5): p. 524-39.
92. Peet, D.J., et al., *Cholesterol and bile acid metabolism are impaired in mice lacking the nuclear oxysterol receptor LXR alpha.* Cell, 1998. **93**(5): p. 693-704.
93. Hannah, V.C., et al., *Unsaturated fatty acids down-regulate srebp isoforms 1a and 1c by two mechanisms in HEK-293 cells.* J Biol Chem, 2001. **276**(6): p. 4365-72.
94. Yoshikawa, T., et al., *Identification of liver X receptor-retinoid X receptor as an activator of the sterol regulatory element-binding protein 1c gene promoter.* Mol Cell Biol, 2001. **21**(9): p. 2991-3000.
95. Yoshikawa, T., et al., *Polyunsaturated fatty acids suppress sterol regulatory element-binding protein 1c promoter activity by inhibition of liver X receptor (LXR) binding to LXR response elements.* J Biol Chem, 2002. **277**(3): p. 1705-11.
96. Laffitte, B.A., et al., *Activation of liver X receptor improves glucose tolerance through coordinate regulation of glucose metabolism in liver and adipose tissue.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5419-24.
97. Cao, G., et al., *Antidiabetic action of a liver x receptor agonist mediated by inhibition of hepatic gluconeogenesis.* J Biol Chem, 2003. **278**(2): p. 1131-6.
98. Naiki, T., et al., *Analysis of gene expression profile induced by hepatocyte nuclear factor 4alpha in hepatoma cells using an oligonucleotide microarray.* J Biol Chem, 2002. **277**(16): p. 14011-9.
99. Krapivner, S., et al., *DGAT1 participates in the effect of HNF4A on hepatic secretion of triglyceride-rich lipoproteins.* Arterioscler Thromb Vasc Biol, 2010. **30**(5): p. 962-7.
100. Boj, S.F., et al., *A transcription factor regulatory circuit in differentiated pancreatic cells.* Proc Natl Acad Sci U S A, 2001. **98**(25): p. 14481-6.
101. Odom, D.T., et al., *Control of pancreas and liver gene expression by HNF transcription factors.* Science, 2004. **303**(5662): p. 1378-81.
102. Goldstein, J.L., R.A. DeBose-Boyd, and M.S. Brown, *Protein sensors for membrane sterols.* Cell, 2006. **124**(1): p. 35-46.
103. Thewke, D.P., S.R. Panini, and M. Sinensky, *Oleate potentiates oxysterol inhibition of transcription from sterol regulatory element-1-regulated promoters and maturation of sterol regulatory element-binding proteins.* J Biol Chem, 1998. **273**(33): p. 21402-7.
104. Worgall, T.S., et al., *Polyunsaturated fatty acids decrease expression of promoters with sterol regulatory elements by decreasing levels of mature sterol regulatory element-binding protein.* J Biol Chem, 1998. **273**(40): p. 25537-40.
105. Tabor, D.E., et al., *Identification of conserved cis-elements and transcription factors required for sterol-regulated transcription of stearoyl-CoA desaturase 1 and 2.* J Biol Chem, 1999. **274**(29): p. 20603-10.
106. Yahagi, N., et al., *A crucial role of sterol regulatory element-binding protein-1 in the regulation of lipogenic gene expression by polyunsaturated fatty acids.* J Biol Chem, 1999. **274**(50): p. 35840-4.
107. Patterson, W.L., 3rd and P.T. Georgel, *Breaking the cycle: the role of omega-3 polyunsaturated fatty acids in inflammation-driven cancers.* Biochem Cell Biol, 2014: p. 1-8.
108. Barres, R., et al., *Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density.* Cell Metab, 2009. **10**(3): p. 189-98.
109. Ceccarelli, V., et al., *Eicosapentaenoic acid demethylates a single CpG that mediates expression of tumor suppressor CCAAT/enhancer-binding protein delta in U937 leukemia cells.* J Biol Chem, 2011. **286**(31): p. 27092-102.
110. Kulkarni, A., et al., *Effects of altered maternal folic acid, vitamin B12 and docosahexaenoic acid on placental global DNA methylation patterns in Wistar rats.* PLoS One, 2011. **6**(3): p. e17706.

111. Waterland, R.A. and R.L. Jirtle, *Transposable elements: targets for early nutritional effects on epigenetic gene regulation.* Mol Cell Biol, 2003. **23**(15): p. 5293-300.
112. Chen, P.Y., et al., *Intrauterine calorie restriction affects placental DNA methylation and gene expression.* Physiol Genomics, 2013. **45**(14): p. 565-76.
113. Heijmans, B.T., et al., *Persistent epigenetic differences associated with prenatal exposure to famine in humans.* Proc Natl Acad Sci U S A, 2008. **105**(44): p. 17046-9.
114. Tobi, E.W., et al., *DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific.* Hum Mol Genet, 2009. **18**(21): p. 4046-53.
115. Bind, M.A., et al., *Air pollution and gene-specific methylation in the Normative Aging Study: association, effect modification, and mediation analysis.* Epigenetics, 2014. **9**(3): p. 448-58.
116. De Prins, S., et al., *Influence of ambient air pollution on global DNA methylation in healthy adults: a seasonal follow-up.* Environ Int, 2013. **59**: p. 418-24.
117. Breitling, L.P., et al., *Tobacco-smoking-related differential DNA methylation: 27K discovery and replication.* Am J Hum Genet, 2011. **88**(4): p. 450-7.
118. Belinsky, S.A., et al., *Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers.* Cancer Res, 2002. **62**(8): p. 2370-7.
119. Gronniger, E., et al., *Aging and chronic sun exposure cause distinct epigenetic changes in human skin.* PLoS Genet, 2010. **6**(5): p. e1000971.
120. Waterland, R.A., et al., *Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles.* PLoS Genet, 2010. **6**(12): p. e1001252.
121. Peluso, M., et al., *DNA methylation differences in exposed workers and nearby residents of the Ma Ta Phut industrial estate, Rayong, Thailand.* Int J Epidemiol, 2012. **41**(6): p. 1753-60; discussion 1761-3.
122. Zhang, F.F., et al., *White blood cell global methylation and IL-6 promoter methylation in association with diet and lifestyle risk factors in a cancer-free population.* Epigenetics, 2012. **7**(6): p. 606-14.
123. Stefani, F.A., et al., *Expression, polymorphism and methylation pattern of interleukin-6 in periodontal tissues.* Immunobiology, 2013. **218**(7): p. 1012-7.
124. Kim, J., et al., *DNA methylation in inflammatory genes among children with obstructive sleep apnea.* Am J Respir Crit Care Med, 2012. **185**(3): p. 330-8.
125. Baccarelli, A., et al., *Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA normative aging study.* Epigenetics, 2010. **5**(3): p. 222-8.
126. Dandrea, M., et al., *MeCP2/H3meK9 are involved in IL-6 gene silencing in pancreatic adenocarcinoma cell lines.* Nucleic Acids Res, 2009. **37**(20): p. 6681-90.
127. Sun, Y.V., et al., *Gene-specific DNA methylation association with serum levels of C-reactive protein in African Americans.* PLoS One, 2013. **8**(8): p. e73480.
128. Lund, G., et al., *DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E.* J Biol Chem, 2004. **279**(28): p. 29147-54.
129. Panin, L.E., F.V. Tuzikov, and O.I. Gimautdinova, *Tetrahydrocortisol-apolipoprotein A-I complex specifically interacts with eukaryotic DNA and GCC elements of genes.* J Steroid Biochem Mol Biol, 2003. **87**(4-5): p. 309-18.
130. Irvin, M.R., et al., *Epigenome-Wide Association Study of Fasting Blood Lipids in the Genetics of Lipid Lowering Drugs and Diet Network Study.* Circulation, 2014.
131. Guay, S.P., et al., *ABCA1 gene promoter DNA methylation is associated with HDL particle profile and coronary artery disease in familial hypercholesterolemia.* Epigenetics, 2012. **7**(5): p. 464-72.
132. Guay, S.P., et al., *DNA methylation variations at CETP and LPL gene promoter loci: new molecular biomarkers associated with blood lipid profile variability.* Atherosclerosis, 2013. **228**(2): p. 413-20.

III. Central hypothesis

The environmental factors (dietary fatty acids and age) have a genotype-dependent dose-response relationship with DNA methylation patterns of the CpG sites within promoter regions of the candidate genes for CVD risk factors (dyslipidemia and inflammation) in humans (**Figure 1.1**).

IV. Specific aims:

Specific aim #1: To explore the association between single nucleotide polymorphisms (SNPs) and DNA methylation patterns by genome-wide approach with integrated bioinformatics analysis. The working hypothesis is that SNPs could affect DNA methylation by either cis-regulation or disturbing formation of CpG dinucleotides. The approach used to test this hypothesis is an integrated bioinformatics analysis of publically available datasets on the genome-wide scale. We expect to find a significant associations between SNPs and DNA methylation patterns.

Specific aim #2: To explore the role of DNA methylation plays in the observed gene-by-environment (G×E) interactions by candidate gene approach with focus on *APOE* (sub-aim #2-a), *IL6* (sub-aim #2-b), and blood lipids-related genetic loci (sub-aim #2-c). The working hypothesis is that DNA methylation may be the biomarkers of the effects of environmental factors on genetic loci. The approach used to test this hypothesis is the genetic association and interaction analysis with population-based cohorts, combined with the analysis of the relationship between DNA methylation and gene expression of candidate genetic loci across different cell lines based

on publically available datasets. We expect to see that DNA methylation may at least partially explain the observed G×E interactions by the consistent findings to suggest the reactions from dietary fatty acids / age to DNA methylation and then to gene expression and then to phenotypes.

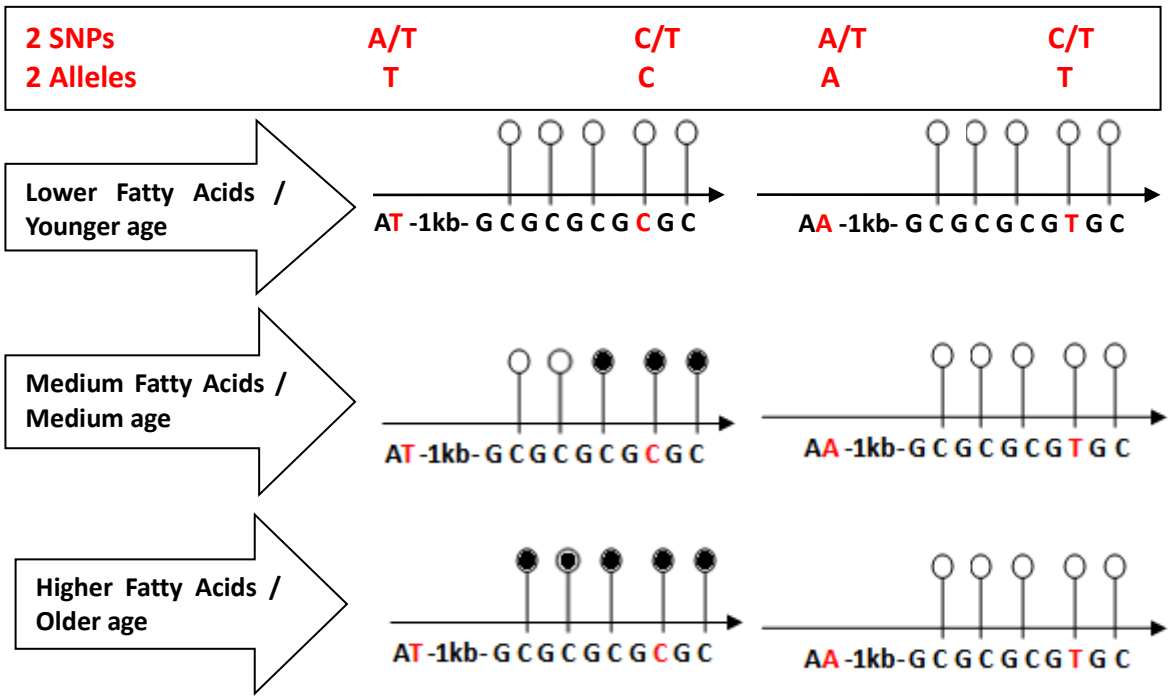


Figure 1.1 Central hypothesis. Arrow represents locus of interest. Open lollipops presents hypomethylated CpG sites and filled ones represent methylated sites.

CHAPTER 2

Genome-wide analysis of associations between CpG related SNPs and DNA methylation patterns

2.1 Abstract:

Background:
DNA methylation occurs on CpG dinucleotides. Single nucleotide polymorphisms (SNPs) may affect DNA methylation by changing the formation of CpG dinucleotides. In this study, we defined those SNPs which could change the formation of CpG dinucleotides as "CpG related SNPs" (CGS). Each CGS has two types of alleles, the allele to create CpG dinucleotides (CGS-C) and the allele to disrupt CpG dinucleotides (CGS-D).

Methods:
We applied a genome-wide scale and integrated bioinformatics analysis to publicly available datasets of both genotypes (HapMap project) and methylation patterns in B lymphocyte cell line (ENCODE project) to explore the relationship between these CGSs and DNA methylation from three perspectives: (1) whether the genotype of a single CGS affects the DNA methylation of that locus; (2) whether two CGSs in high linkage disequilibrium (LD) tend to be with the same type of allele (create or disrupt the CpG); (3) whether the haplotype consisting of multiple CGSs in high LD is associated with DNA methylation pattern of that region.

Results:
Approximately 80% of CGS-Cs can be methylated. In addition, when two CGSs are in high LD, they tend to act as the same type of allele, which means that if the allele of one CGS is to create the CpG then the allele of another CGS in high LD tend to create the CpG also. This finding is highly consistent in both unrelated individuals and family members of HapMap CEU population. Finally, the haplotype of one LD block, consisting of multiple CGSs close to both CpG islands and promoter regions, is correlated with DNA methylation pattern according to both categorical and continuous analysis.

Conclusion:
CGSs which are close to CpG islands and promoters may affect DNA methylation pattern in the form of haplotype.

Key words:
Methylation, haplotype, SNP analysis/discovery, bioinformatics, genome scan.

2.2 Introduction:

Many technological advances have facilitated greatly the cataloging of millions of single nucleotide polymorphisms (SNPs). In spite of these, an understanding of the functional implications of single base pair changes remains limited. For example, genome projects including the Human Genome Project, HapMap project [1], and 1000 Genomes project have sequenced approximately 6 giga-base pairs of DNA nucleotides to identify 15 million SNPs [2, 3]. Accompanying the discovery of novel SNPs is growing recognition of associations between SNPs and human diseases, many through genome-wide association studies (GWAS). However, despite this vast growth in genome databases, mechanistic understanding of how SNPs alter phenotypes is largely unknown.

Recent findings have suggested that links between epigenetic status and genetic variants may underlie the functionality of SNPs. Of the major types of epigenetic processes [4], DNA methylation has been most frequently linked to human diseases including cardiovascular diseases [5], diabetes [6], obesity [7], dyslipidemia [8], and cancer [9]. Recently, extensive studies have identified the genetic contribution to DNA methylation. Genetic manipulation in mouse stem cells demonstrated that local sequences regulate DNA methylation in both a necessary and sufficient manner [10]. SNPs are commonly occurring and well-documented example of DNA sequence variation, and SNP regulation of DNA methylation is widespread in humans [11-16]. For example, 80% of the variation in DNA methylation can be predicted by overall genotype of 40 sequence-dependent autosomal regions [15]. Additionally, sequence-based regulation of DNA methylation is flexible, occurring either in *cis* or in *trans* [11]. Further, methylation related SNPs have been shown to affect gene expression [11], alternative splicing [14], and binding by certain transcription factors [14, 17], all of which are potential mechanisms by which SNPs may alter phenotypes.

CpG related SNPs (CGSs) constitute a group of SNPs with a particular relationship to DNA methylation. By definition, CGSs referred to those SNPs which can change the formation of CpG dinucleotides, which has been established as the primary target site of DNA methylation. For example, studies with human differentiated cells found that 99.98% of the methylation occurs on CpG dinucleotides [18]. CGSs have been found to contribute a significant fraction (38% ~ 88%) of allele specific methylation regions in the human genome [13, 15]. Also, over 80% CGSs were shown to play a regulation role in DNA methylation [19].

However, almost all current research has focused on the effects on DNA methylation by a single SNP, rather than the combination of multiple SNPs within one region. With the growth of knowledge in genetics, the concept of region with multiple variants outweighs the concept of single variant in terms of functional significance. In addition, SNPs in close proximity tend to be in linkage disequilibrium (LD) because of their less chance for recombination compared to those SNPs far away from each other. As nearby SNPs often are linked from the perspective of variant alleles, the methylation status of adjacent CpG sites also tends to be similar [13]. In this case, we hypothesize that CGSs affect DNA methylation patterns in the form of haplotype.

To test our hypothesis, we conducted a pilot study on genome-wide scale to explore the potential relationships between DNA methylation and CGSs in three perspectives: (1) whether the genotype of a CGS is correlated with the DNA methylation level of that CpG dinucleotides; (2) whether two CGSs in high LD tend to carry the same type of allele, either creating or disrupting the CpG site; (3) whether the haplotype consisting of multiple CGSs in high LD is associated with the DNA methylation pattern of that LD region. (**Table 2.1**).

2.3 Materials and Methods

2.3.1 Definition of CGS, CGS-C and CGS-D:

In this study, CGSs are defined as those SNPs which could change the formation of CpG dinucleotides. Each CGS has two types of alleles, which can either create (CGS-C) or disrupt (CGS-D) CpG dinucleotides. For example, for a SNP with adenosine/cytosine (A/C) allele substitution, if the nucleotide on the 3' of this SNP is guanosine (G), then this SNP can be defined as CGS because the allele cytosine (C) of this SNP and the 3' guanosine (G) can form a CpG dinucleotides site. In this case, allele cytosine (C) is the CGS-C allele that will create while allele adenosine (A) is the CGS-D allele that will disrupt CpG dinucleotides.

Considering the possibility of methylation when the CpG site was present, the expected probabilities of CGS-C allele and CGS-D allele were differently defined in this study, depending on the level of data available, on the level of allele or genotype. If the data were on the level of allele, the expected probabilities of each allele would be accurately estimated as 50% (or 1/2), assuming an equal mutation rate across all SNPs. However, if the data available were on the level of genotype, we defined the heterozygotes as CGS-C rather than CGS-D, because one allele of the CpG site has the potential to be methylated, which resulted in the over-estimation of the probability of CGS-C allele and under-estimation of the probability of CGS-D allele. By this definition, thus, there are 67% (or 2/3) CGSs that are with CGS-C allele and 33% (or 1/3) with CGS-D allele, assuming an equal mutation rate across all SNPs. Taken the above SNP with A/C allele substation as one example, individuals with genotype of CC or AC were grouped altogether into CGS-C allele, while those with genotype of AA were grouped into CGS-D allele.

2.3.2 Retrieval of CGSs:

As a first step, all types of variants were downloaded from the SNP135 database from UCSC human genome browser [20]. Since our interest was limited to SNPs, we extracted just those variants with class defined as "single" and a difference between the starting and ending coordinates of 1. Furthermore, we excluded ambiguous SNPs which had multiple different coordinates in one database. Then, the bioinformatics tool Galaxy [21] was used to fetch human hg19 sequence data provided by the UCSC genome browser in order to retrieve one adjacent nucleotide on the 5' and 3' ends of each SNP. From this data set, all CGSs on the genome-wide scale were extracted. In addition, a subgroup of extracted CGSs were further determined, which are located within 3kb up- and down-stream of both the transcription start site (TSS) of annotated gene, identified by the NCBI Reference Sequence (RefSeq) project from "refGene" database (as of 02-02-2012) [22], and CpG islands from "cpgIslandExt" database [23] , both of which were downloaded from UCSC genome browser. An R script was created to implement all these procedures.

2.3.3 Data sources for methylation, phased allele, and genotypes:

Three sources of methylation data were used in this study. One provides the methylation values by a single nucleotide based on the method of reduced representation bisulfite sequencing (RRBS) [24], called Methyl-RRBS. The other two datasets, Methyl-seq and Methyl-450, provide the methylation levels for 50 nucleotides on average. Methyl-seq methodology combines DNA digestion by HpaII and MspI with the Illumina DNA sequencing platform [25]. Methyl-450 measures DNA methylation by bisulfite conversion and then is assayed with methylated and unmethylated probes on Illumina Infinium Human Methylation 450 Bead Array platform. A

detailed description of the methodology can be found on the UCSC website. Within the above described data, one site in the Methyl-RRBS database equaled one nucleotide, while one site within the Methl-seq and Methyl-450 databases represented on average approximately 50 nucleotides, which have been taken into consideration for the calculation of hypermethylated nucleotides for each respective region as described below. All three databases were downloaded from the Encyclopedia of DNA Elements (ENCODE) project [26] through the UCSC genome browser. According to UCSC genome browser criteria, hypermethylated sites are those with >=50% of reads showing methylation at the given position in the genome for Methyl-RRBS, a score of 1000 for Methyl-seq, or a score >=600 for Methyl-450.

We downloaded genotype information for three individuals of European ancestry: GM12878, GM12891, and GM12892, whose genomes have been made publicly available via the 1000 Genomes Project high-coverage pilot study [3]. These data were accessed from the UCSC genome browser. Genotype information for a fourth individual, NA06990, was downloaded from NCBI. In order to get the exact haplotype information, we downloaded phase III phased allele data (released in Feb. 2009) of Utah residents with Northern and Western European Ancestry (CEU) in HapMap project. Their corresponding genotype data were also downloaded, which were released in Aug. 2010 with release number 28. There were 113 individuals in the phased allele dataset, while 174 individuals were available in the unphased genotype dataset [1, 27].

2.3.4 LD region development:

Each CGS within 3 kb distance from TSS and CpG islands was defined as an index SNP. Proxy SNPs for each index SNP were obtained using SNAP software [28] with the following searching

parameters: (1) the SNP data set was the 1000 Genomes Pilot 1; (2) the population panel was CEU; (3) the r^2 threshold was 1.0; and (4) the distance limit was 500 kb. Those CGS pairs with the same index CGS were grouped and the LD region (block) was defined as the region starting from the CGS at the most 5' side and proceeding to the one on the most 3' side. And duplicated LD regions were removed.

2.3.5 Statistical analysis:

Both data of phased allele and unphased genotype were used to test the haplotype defined by two highly linked CGSs: (1) both CGSs carry the alleles to create CpG dinucleotides, i.e. index CGS-C with proxy CGS-C (CGS-Ci & CGS-Cp); (2) both CGSs carry the alleles to disrupt CpG dinucleotides, i.e. index CGS-D with proxy CGS-D (CGS-Di & CGS-Dp); (3) the index CGS carries the allele to create while the proxy CGS carries the allele to disrupt CpG dinucleotides (CGS-Ci & CGS-Dp); and (4) the index CGS carries the allele to disrupt while the proxy CGS carries the allele to create CpG dinucleotides (CGS-Di & CGS-Cp). One-way Chi-square analysis was applied to test whether the observed probabilities of these four haplotypes were different from their corresponding expected probabilities, all of which should be equaled to 25% (or 1/4) because the expected probabilities of CGS-C and CGS-D are both 50% (or 1/2), based on the assumption of independence. Considering the fact that gametic phase information was not available for DNA methylation data used by this study, and that there were limited amount of individuals with both phased allele data and DNA methylation to analyze the association between haplotype and DNA methylation, an indirect estimation of haplotype based on unphased genotype information was further studied. Heterozygotes were determined to be included in the analysis because of statistical power, and they were further categorized into the CGS-C group

not CGS-D group. In this case, the probability of CGS-C allele was over-estimated and their expected values were 67% (or 2/3), higher than the accurate estimation of 50% (or 1/2), while the probability of CGS-D allele was under-estimated with expected values of 33% (or 1/3), lower than the accurate estimation of 50% (or 1/2). And the corresponding expected probabilities were 44.4% (or 4/9), 11.1% (or 1/9), 22.2% (or 2/9), and 22.2% (or 2/9) for haplotype CGS-Ci & CGS-Cp, CGS-Di & CGS-Dp, CGS-Ci & CGS-Dp, and CGS-Di & CGS-Cp, respectively. In order to further know to which haplotype two CGSs tend to belong, the observed probabilities of the haplotype when both CGSs are with the same type of allele was compared to the random probability assuming that two CGSs show no interdependence, which was 50% based on right estimation and 56% based on both over-estimation of CGS-C and under-estimation of CGS-D.

To conduct the association with DNA methylation in a given region, the haplotype of multiple CGSs, derived with genotype data, within one LD region was analyzed as both categorical and continuous variables. For the categorical analysis, LD regions were dichotomized into two groups based on the fact whether or not there were more than 50% CGSs with CGS-C allele. Analysis of variance (ANOVA) was used to test whether those LD regions with >50% CGS-Cs tend to have more hypermethylated nucleotides compared to those LD regions with <=50% CGS-Cs. In the continuous analysis, the exposure variable is the number of CGSs with CGS-C allele within one LD region. The dependent variable is the sum of the nucleotides with hypermethylation from Methyl-seq, Methyl-450, and Methyl-RRBS datasets within that LD region. Pearson correlation and linear regression were applied to analyze whether the haplotype of multiple CGSs, defined as the number of CGS-Cs, within one LD region was related to the number of hypermethylated nucleotides in that region. In order to reduce potential bias, an adjusted number of hypermethylated nucleotides and an adjusted number of CGSs in one LD

region were calculated by dividing the length of the LD region from these measurements, and then multiplied by 1Mb. Because the total number of CGSs and CpG sites with methylation measurements in one LD region was determined by the technological resolution of genotyping platform and methylation chip, these two parameters were further adjusted in the regression model. All data were analyzed using SAS (version 9.2 for Windows; SAS Institute, Inc. Cary, NC, USA). A two-tailed *P*-value of <0.05 was considered statistically significant.

2.4 Results and discussion:

2.4.1 Distribution of CGSs:

There were 12,023,433 CGSs from a total 54,212,080 variants within the SNP135 dataset after excluding those polymorphisms which were not SNPs (6,091,064, typically insertions and deletions), those ambiguous variants (4,178,554, variants with different coordinates in the database), and those SNPs do not change CpG dinucleotides (31,919,029). Among these CGSs, 602,700 were close to the TSS of annotated genes as identified by the NCBI RefSeq project and 359,151 were also close to CpG islands.

2.4.2 Genotype of single CGS and its methylation

There were approximately 1000 sites across the human genome with both genotype data and methylation measurements at the level of the single nucleotide as assessed by the method of Methyl-RRBS (1184 in GM12878, 1076 in GM12891, 956 in GM12892, and 701 in NA06990) (**Figure 2.1**). The means of DNA methylation level for CGS-C sites were above 40% (43.94% in GM12878, 47.52% in GM12891, 46.52% in GM12892, and 51.68% in NA06990), while those

means were less than 3% for CGS-D sites (0.96% in GM12878, 2.33% in GM12891, 2.31% in GM12892, and 1.02% in NA06990). The variation of DNA methylation level for CGS-C sites was greater than those for CGS-D sites (35.29% vs. 8.47% in GM12878, 38.47% vs. 12.6% in GM12891, 37.44% vs. 10.87% in GM12892, and 37.05% vs. 6.48% in NA06990). Less than 10% of the CGS-D sites were methylated (7% in GM12878, 6% in GM12891, 10% in GM12892, and 5% in NA06990). But the majority of CGS-C sites (approximately 80%) were with methylation level greater than 0 (82% in GM12878, 76% in GM12891, 79% in GM12892, and 85% in NA06990).

This observation is highly consistent with the findings that most DNA methylation occurs on CpG dinucleotides in humans [18]. Interestingly, our findings further indicate that methylation is prevalent when CpG dinucleotides were present. However, approximately 60% of CpG sites in Figure 2.1 were neither completely methylated nor completely unmethylated. This is in line with the widespread existence of allelic skewing of DNA methylation in the human genome, a phenomenon in which the allele-specific methylation is relatively subtle rather than clear-cut [11]. Also, 7.6% of CpG islands in polyclonal or monoclonal cell lines were found to be predominant with CpG dinucleotides with intermediate levels of methylation (25% ~ 75%) [29]. It has been hypothesized that the usual small effect sizes and failure of replications in genetic association studies were because of the existence of these CpG sites with intermediate methylation levels [11].

2.4.3 Haplotype of two highly linked CGSs

The phased allele data provided direct test for the haplotype of two highly linked CGSs. In most CEU individuals, the observed probabilities of four haplotypes were significantly different from

their expected ones with assumption of independency ($P < 0.05$) (**Figure 2.2A**). Furthermore, the probability of the condition when both CGSs with CGS-C allele or CGS-D allele is significantly higher than the random probability of 50% ($P < 0.001$) (**Figure 2.2B**). Figure 2.2 A and B showed the results of the analysis with transmitted allele and the analysis with untransmitted allele yielded similar results (data not shown).

Consistent to the results with phased allele data, tests with unphased genotype information also showed not only the significant differences between the observed and expected probabilities of four haplotypes ($P < 0.05$) (**Figure 2.2C**), but also a higher than expected probability of haplotypes of CGS-Ci & CGS-Cp and CGS-Di & CGS-Dp ($P < 0.0001$) (**Figure 2.2D**), although the corresponding expected probability was over-estimated from the original 50% up towards 56%. Although the probability of CGS-C was over-estimated by genotype data, the similarity of the significant findings from unphased genotype data to those from phased allele data ameliorated the effect of the over-estimation and validated the usage of genotype information to derive haplotype structure.

Both tests indicate that two CGSs tend to contain the same type of allele (e.g., either both create or both disrupt the formation of CpG dinucleotides). The potential mechanism for the observed haplotype may be related to the similar selection pressure and mutation rate introduced by correlated DNA methylation levels among closely located CpG dinucleotides. Methylated cytosines at CpG dinucleotides can be mutated to thymines (T) through deamination [30-33]. If the neighboring CpG dinucleotides were methylated, then the incidence of SNP significantly increased by ~50% within the region within 10bp of a CpG site [16]. Additionally, CpG sites closely located to each other tend to have similar methylation levels. The correlation coefficients (R^2) for the methylation levels across 30% ~ 48% of the CpG sites were greater than 0.3 [13].

Also, the methylation state of CpG sites located in proximal promoter regions were found to be highly correlated [34]. In this case, it is plausible for two highly linked CGSs to contain the same type of allele, either both will create or both will disrupt CpG dinucleotides.

2.4.4 Haplotype of multiple CGSs and methylation

According to categorical analysis (**Figure 2.3**), those LD regions with more than 50% CGS-Cs have more hypermethylated nucleotides than those regions with less than 50% CGS-Cs ($P < 0.05$). This pattern is highly consistent across not only four individuals (GM12878, GM12891, GM12892, and NA06990) but also across three different methods to measure DNA methylation pattern (Methyl-seq, Methyl-450, and Methyl-RRBS).

Based on continuous analysis **(Figure 2.4),** there was a positive relationship between the adjusted number of CGS-Cs and the adjusted number of nucleotides with hypermethylation per 1Mb region across the entire genome. The correlation coefficients range from 0.2 to 0.5 and all of these correlations are statistically significant ($P < 0.0001$) across four individuals (GM12878, GM12891, GM12892, and NA06990).

Two example LD regions for the association between haplotype of multiple highly linked CGSs and DNA methylation pattern within one LD region are depicted in **Figure 2.5**. The LD region with methylation-susceptible haplotype was shown on the left side of the figure, in which all four CGSs are CGS-Cs. In this region, about 1000 nucleotides have measurements for methylation and most of these nucleotides showed hypermethylation. In contrast, the right side of the figure shows the LD region with methylation-resistant haplotype, consisting of two CGS-Ds and most nucleotides in this LD region exhibit a hypomethylation pattern.

Our finding is consistent with previous studies showing that CpG density within the methylation-determining regions is one important factor for regulation of DNA methylation [10]. However, the direction of the effect of CpG density on methylation based on our observations is not the same as that previous study. Lienert, et al. found a significant negative correlation ($r = -0.49$) between the hypermethylation level and the number of CpG sites within methylation-determining regions of length less than 1 kb. In contrast, our study found that the more CpG sites introduced by CGS-Cs in one LD region, the greater the number of hypermethylated nucleotides. A lack of measurement of exact CpG density in our study may account for this discrepancy. All the observations in our study were based on those nucleotides determined by the chip for methylation and genotyping, so not all the CpG sites and CGSs within the region of interest were measured. In this case, although the adjustment by length of each LD region was taken into account, the exact value of CpG density is still unknown. In addition, differences in the length of region of interest may explain, in part, this conflict. The regions of interest in the previous study are fixed with size less than 1 kb, while the length of regions in our study was flexible depending on the distance that encompassed those highly linked CGSs, which spans from 100 bp to 1 Mb. Discrepancies may also be related to species differences, since the previous study was based on mice data and our findings were derived from humans. Nonetheless, the positive relationship between haplotype and DNA methylation is consistent with our previous finding that 80% of the CGSs can be methylated when these SNPs contain the allele to create CpG dinucleotides, indicating the susceptibility to methylation of CGSs.

2.4.5 Strength and limitation

An important strength of this study is the replication not only across multiple individuals but also across different experimental methodologies. The consistency of the replication validates the findings of our study. Additionally, this study was performed on a genome-wide scale, providing a more systematic and comprehensive view of the whole genome. Finally, this study is based on publically available data sets, whose quality has been vetted by numerous groups conducting a myriad of other analyses. Although some experimental bias may remain, replication across different methodologies negates these biases to a significant degree.

This study does have some limitations. All findings are from B lymphocytes, so the identified relationship between CGSs and DNA methylation might not be generalized to other tissue types. We are limited by a relatively small sample size of individuals for which both genotype and DNA methylation data were available. In addition, all DNA methylation datasets were derived from cell lines rather than primary cells, which are more directly applicable to any inferred mechanistic consequences and which indeed may have different methylation patterns than those observed here. Also, the methylation datasets available did not allow us to analyze the differences in gametic phases.

2.5 Conclusion remarks:

To our knowledge, this is the first study to focus specifically on CpG-related SNPs (CGSs) and their relationship with DNA methylation through genome-wide scale and integrated bioinformatics analysis of publicly available datasets. Our study is one of the few to examine the hypothesis that CGSs are capable of altering the formation of CpG dinucleotides, the target site for DNA methylation. We found that approximately 80% of CGSs were methylated when they carry the allele to create CpG dinucleotides. In addition, when two CGSs are in high LD, they tend to act in a coordinate fashion, meaning that if the allele of one CGS creates the CpG site, then the allele of another CGS in high LD also tends to create another CpG site. This finding is highly consistent in all 113 individuals with phased allele data and 174 individuals with unphased genotype data of the HapMap CEU population. Finally, the haplotype of one LD block, consisting of multiple CGSs close to both CpG islands and promoter regions, is correlated with the DNA methylation patterns according to both categorical and continuous analyses, showing that the more CGS-C one LD block have, the more nucleotides will undergo hypermethylation. The impact of this study resides not only in providing a candidate functional mechanism to link SNPs and DNA methylation, but also in its potential contributions to personalized medicine that relies on knowledge of functional genomic regions.

2.6 References:

1. Consortium, I.H., *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
2. Clarke, L., et al., *The 1000 Genomes Project: data management and community access.* Nat Methods, 2012. **9**(5): p. 459-62.
3. Altshuler D, D.R., Abecasis GR, et al, *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
4. Portela, A. and M. Esteller, *Epigenetic modifications and human disease.* Nat Biotechnol, 2010. **28**(10): p. 1057-68.
5. Ordovas, J.M. and C.E. Smith, *Epigenetics and cardiovascular disease.* Nat Rev Cardiol, 2010. **7**(9): p. 510-9.
6. Hidalgo, B., et al., *Epigenome-wide Association Study of Fasting Measures of Glucose, Insulin, and HOMA-IR in GOLDN.* Diabetes, 2013.
7. Milagro, F.I., et al., *CLOCK, PER2 and BMAL1 DNA methylation: association with obesity and metabolic syndrome characteristics and monounsaturated fat intake.* Chronobiol Int, 2012. **29**(9): p. 1180-94.
8. Guay, S.P., et al., *DNA methylation variations at CETP and LPL gene promoter loci: new molecular biomarkers associated with blood lipid profile variability.* Atherosclerosis, 2013. **228**(2): p. 413-20.
9. Weisenberger, D.J., *Characterizing DNA methylation alterations from The Cancer Genome Atlas.* J Clin Invest, 2014. **124**(1): p. 17-23.
10. Lienert, F., et al., *Identification of genetic elements that autonomously determine DNA methylation states.* Nat Genet, 2011. **43**(11): p. 1091-7.
11. Schalkwyk, L.C., et al., *Allelic skewing of DNA methylation is widespread across the genome.* Am J Hum Genet, 2010. **86**(2): p. 196-212.
12. Zhang, D., et al., *Genetic control of individual differences in gene-specific methylation in human brain.* Am J Hum Genet, 2010. **86**(3): p. 411-9.
13. Shoemaker, R., et al., *Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome.* Genome Res, 2010. **20**(7): p. 883-9.
14. Bell, C.G., et al., *Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus.* PLoS One, 2010. **5**(11): p. e14040.
15. Gertz, J., et al., *Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation.* PLoS Genet, 2011. **7**(8): p. e1002228.
16. Qu, W., et al., *Genome-wide genetic variations are highly correlated with proximal DNA methylation patterns.* Genome Res, 2012. **22**(8): p. 1419-25.
17. Paliwal, A., et al., *Comparative anatomy of chromosomal domains with imprinted and non-imprinted allele-specific DNA methylation.* PLoS Genet, 2013. **9**(8): p. e1003622.
18. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-22.
19. Zhi, D., et al., *SNPs located at CpG sites modulate genome-epigenome interaction.* Epigenetics, 2013. **8**(8): p. 802-6.
20. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.
21. Goecks, J., et al., *Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.* Genome Biol, 2010. **11**(8): p. R86.
22. Pruitt, K.D., T. Tatusova, and D.R. Maglott, *NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.* Nucleic Acids Res, 2007. **35**(Database issue): p. D61-5.

23. Gardiner-Garden, M. and M. Frommer, *CpG islands in vertebrate genomes.* J Mol Biol, 1987. **196**(2): p. 261-82.
24. Meissner, A., et al., *Genome-scale DNA methylation maps of pluripotent and differentiated cells.* Nature, 2008. **454**(7205): p. 766-70.
25. Brunner, A.L., et al., *Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver.* Genome Res, 2009. **19**(6): p. 1044-56.
26. Myers, R.M., et al., *A user's guide to the encyclopedia of DNA elements (ENCODE).* PLoS Biol, 2011. **9**(4): p. e1001046.
27. Altshuler, D.M., et al., *Integrating common and rare genetic variation in diverse human populations.* Nature, 2010. **467**(7311): p. 52-8.
28. Johnson, A.D., et al., *SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap.* Bioinformatics, 2008. **24**(24): p. 2938-9.
29. Deng, J., et al., *Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming.* Nat Biotechnol, 2009. **27**(4): p. 353-60.
30. Lindahl, T. and B. Nyberg, *Rate of depurination of native deoxyribonucleic acid.* Biochemistry, 1972. **11**(19): p. 3610-8.
31. Coulondre, C., et al., *Molecular basis of base substitution hotspots in Escherichia coli.* Nature, 1978. **274**(5673): p. 775-80.
32. Cooper, D.N. and M. Krawczak, *Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes.* Hum Genet, 1989. **83**(2): p. 181-8.
33. Sved, J. and A. Bird, *The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4692-6.
34. Guay, S.P., et al., *ABCA1 gene promoter DNA methylation is associated with HDL particle profile and coronary artery disease in familial hypercholesterolemia.* Epigenetics, 2012. **7**(5): p. 464-72.

Table 2.1 Hypotheses of study in Chapter 2

N	SNP	DNA methylation of	Hypothesis
1	Single CGS[1]	Single nucleotide	Genotype of single CGS affects DNA methylation status at that site.
2	Two CGSs in high LD[2]	/	CGSs in high LD tend to carry the same type of allele, i.e. CGS-Ci[3] & CGS-Cp[4] or CGS-Di[5] & CGS-Dp[6].
3	Multiple CGSs in high LD	thousands nucleotides	Haplotype of multiple CGSs in high LD is associated with DNA methylation pattern of that LD region

Abbreviation: [1]CGS, CpG related SNP; [2]LD, linkage disequilibrium; [3]CGS-Ci, index CGS with the allele to create CpG dinucleotides; [4]CGS-Cp, proxy CGS with the allele to create CpG dinucleotides; [5]CGS-Di, index CGS with the allele to disrupt CpG dinucleotides; [6]CGS-Dp, proxy CGS with the allele to disrupt CpG dinucleotides.

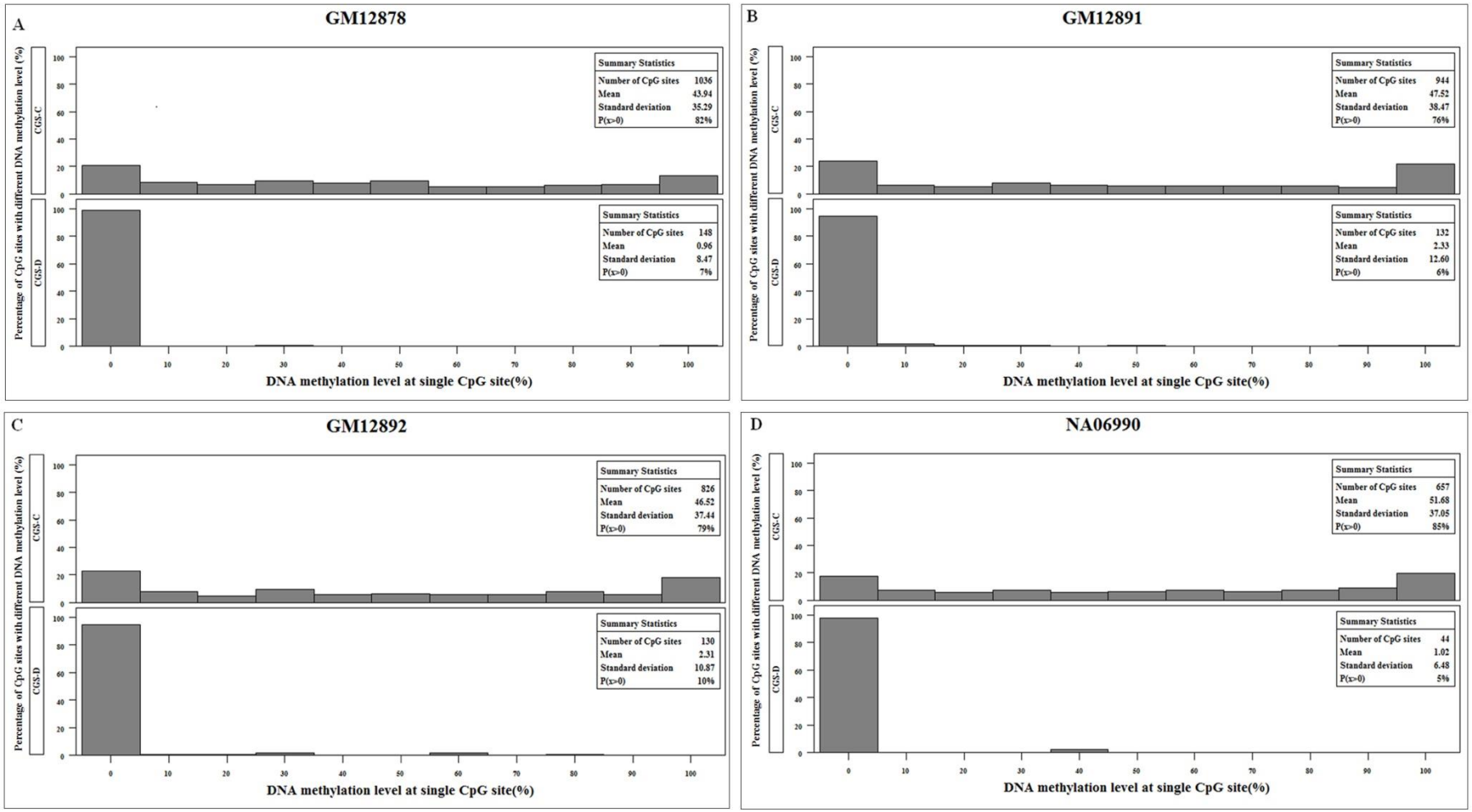


Figure 2.1 Methylation status of CGSs with different types of alleles. Histograms of methylation status for CGS-C sites (upper panel) and CGS-D sites (lower panel) in human blood lymphocyte cell lines. Data from GM12878 (A), GM12891 (B), GM12892 (C), NA06990 (D) were shown. The summary statistics of methylation status, including number of CpG sites, mean, standard deviation, and percentage of CpG sites with methylation level greater than 0 (P(x>0)), were presented on the right upper corner of each panel. Abbreviation: CGS, CpG related SNPs; CGS-C, CpG related SNP with the allele to create CpG dinucleotides; CGS-D, CpG related SNP with the allele to disrupt CpG dinucleotides; P(x>0), percentage of CpG sites with methylation level greater than 0.

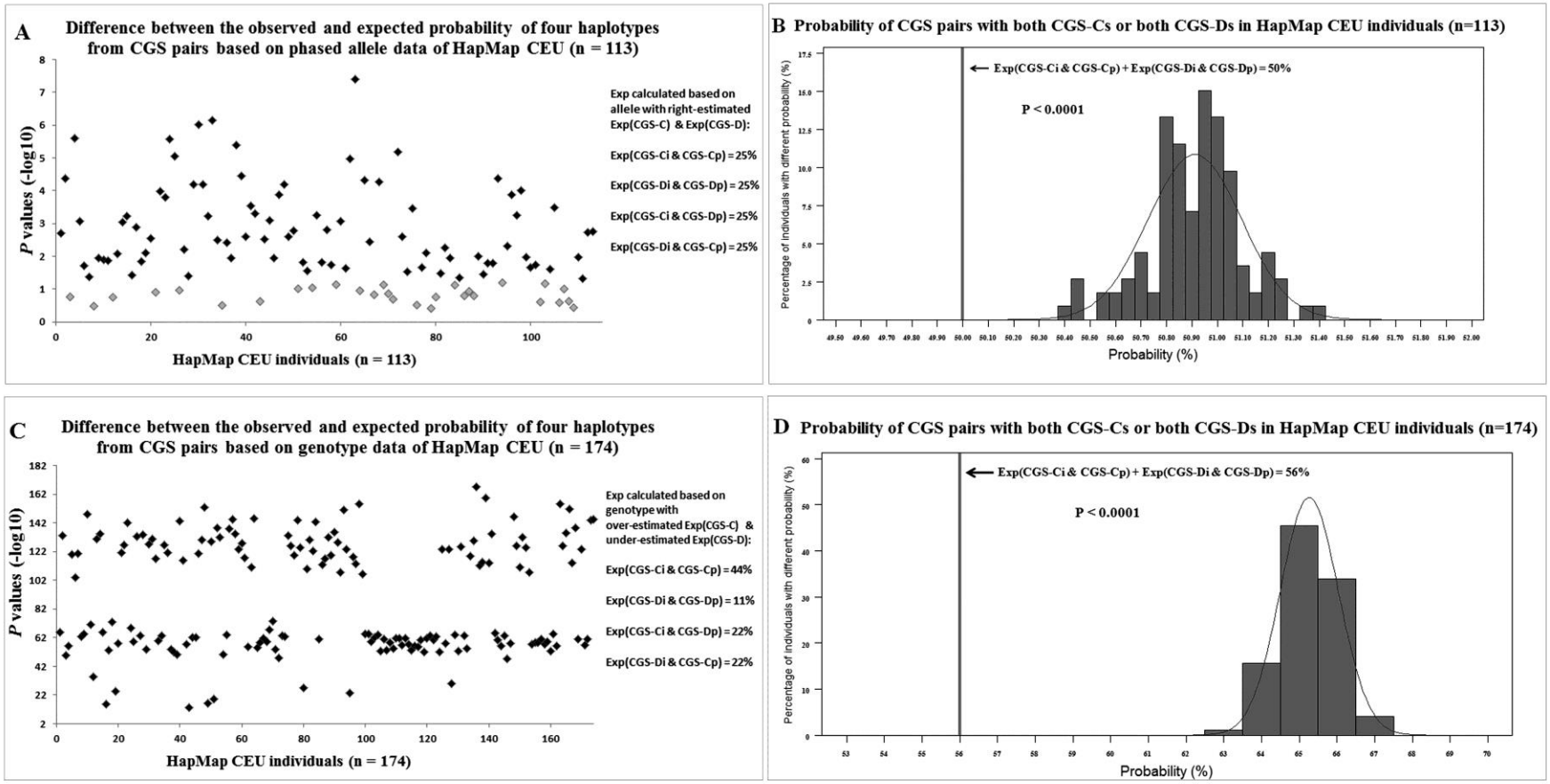


Figure 2.2 Haplotype of two CGSs with high LD in HapMap CEU. Phased allele data (n=113) (panel A and B) and unphased genotype data (n=174) (panel C and D) were used to measure the haplotype of two linked CGSs. (A,C) Scatter plot of all *P* values (-log10) from Chi-square tests of the difference between the observed and expected probability of four haplotypes derived from CGS pairs in each HapMap CEU individual, shown as diamond (black for $P < 0.05$ and gray for $P >= 0.05$). (B,D) Histogram of the distribution of the probability of the haplotypes when both CGSs are with the same type of alleles, both with CGS-Cs or both with CGS-Ds. Vertical lines represent the expected probability, which was 50% for the estimation based on phased allele and 56% for the estimation based on unphased genotype. Abbreviation: CGS, CpG related SNP; LD, linkage disequilibrium; CEU, Utah residents with ancestry from northern and western Europe; CGS-C, CpG related SNP with the allele to create CpG dinucleotides; CGS-D, CpG related SNP with the allele to disrupt CpG dinucleotides; Exp, expected probability; CGS-Ci, index CGS-C; CGS-Cp, proxy CGS-C; CGS-Di, index CGS-D; CGS-Dp, proxy CGS-D.

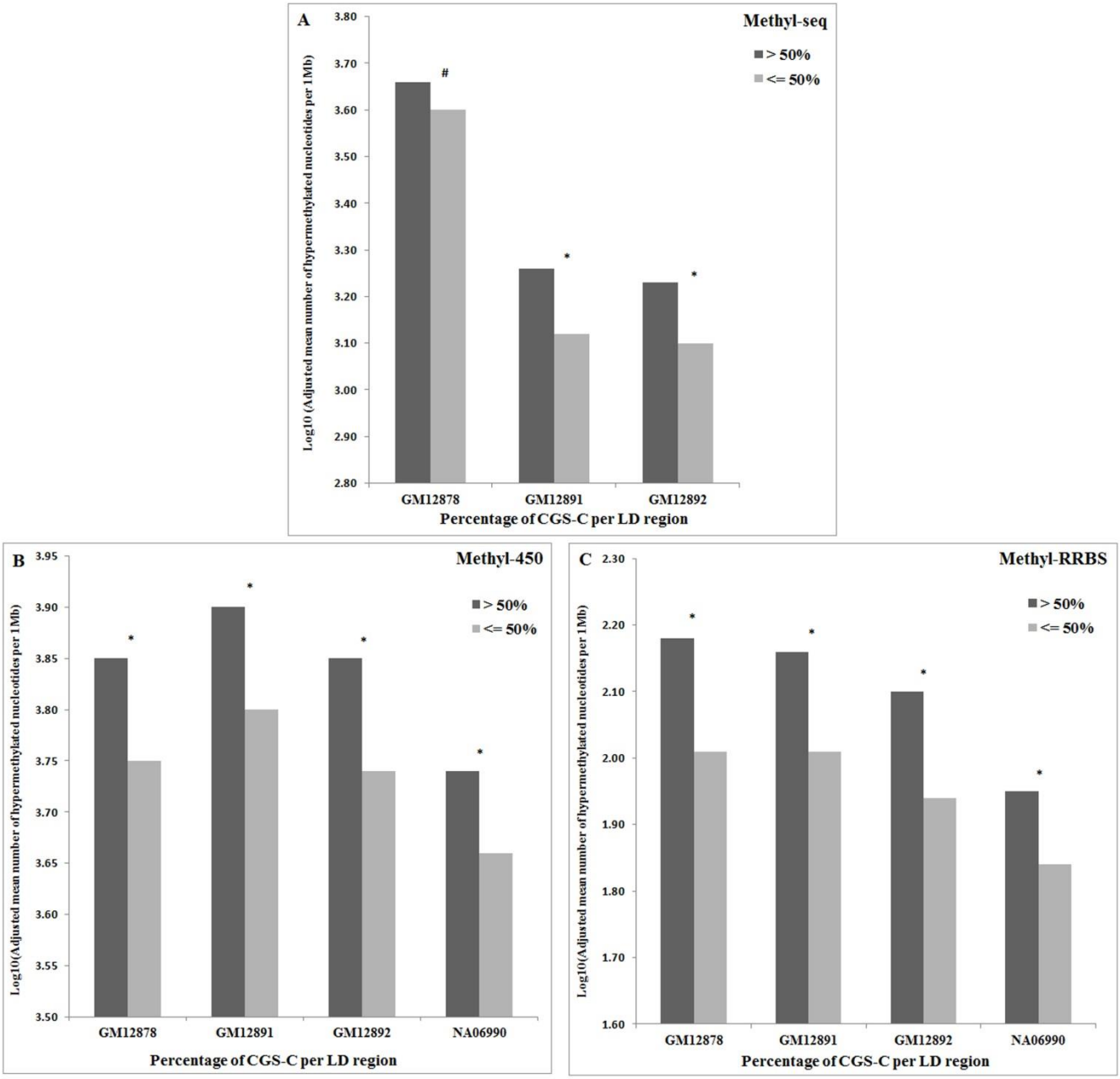


Figure 2.3 Categorical analysis of association between CGS haplotype and DNA methylation. Adjusted mean number of hypermethylated nucleotides per 1Mb region were compared between those LD regions with >50% CGS-C (dark gray) and those with <= 50% CGS-C (light gray), with different techniques to measure DNA methylation: Methyl-seq (A), Metyl-450 (B), and Methyl-RRBS (C). The model was adjusted for total number of proxy CGSs and total number of nucleotides with methylation measurement ($^{\#}P < 0.05$ and * $P < 0.01$). Abbreviation: LD, linkage disequilibrium; CGS-C, CpG related SNP with the allele to create CpG dinucleotides.

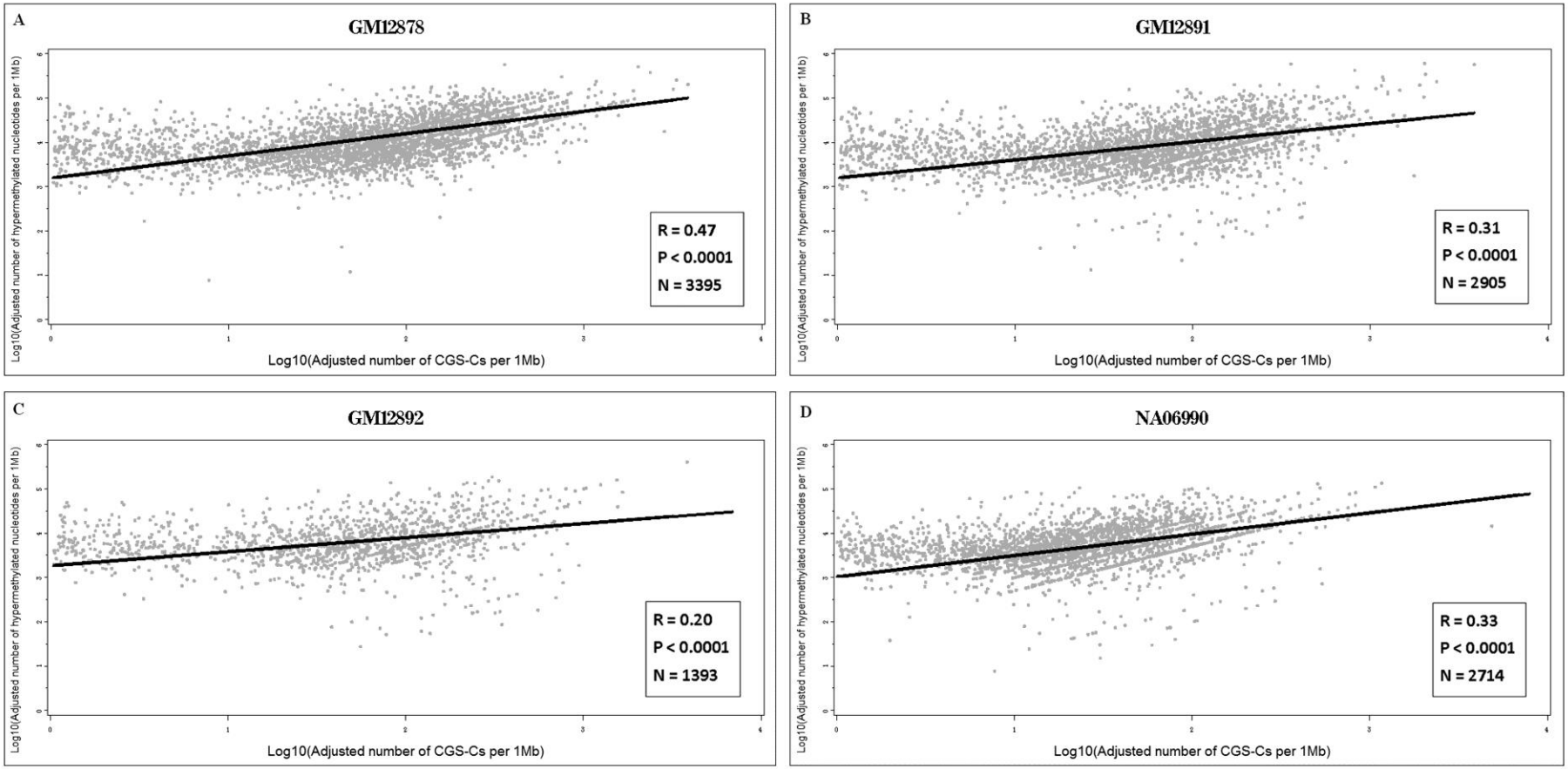


Figure 2.4 Continuous Analysis of association between CGS haplotype and DNA methylation. Correlation between adjusted number of CGS-Cs and adjusted number of hypermethylated nucleotides per 1Mb in GM12878 (A), GM12891 (B), GM12892 (C), and NA06990 (D) are presented. A gray dot represents each LD region. The black line is the predicted correlation line adjusting for total number of proxy CGSs and total number of nucleotides with methylation measurement. R and P represent the Pearson correlation coefficient and the corresponding *P* value for its significance test, respectively. N represents the total number of LD regions in each individual. Abbreviation: CGS-C, CpG related SNP with the allele to create CpG dinucleotides; LD, linkage disequilibrium.

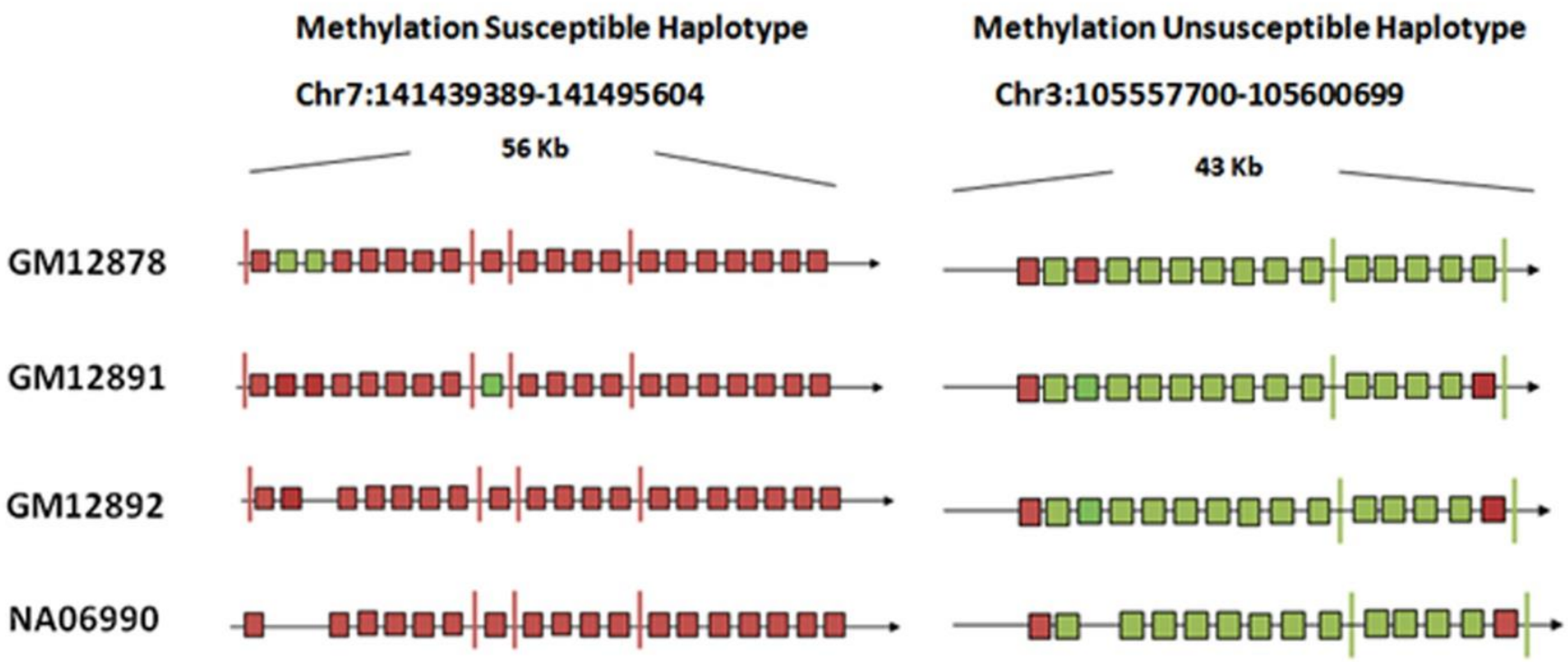


Figure 2.5 Examples of association between CGS haplotype and DNA methylation. Two examples of regions with methylation susceptible haplotype and methylation unsusceptible haplotype in GM12878, GM12891, GM12892, and NA06990 were displayed. On the left, the LD region occupies 56kb in length and contains four CGS-Cs (red dash), hypermethylation regions (red square) and hypomethylation regions (green square). On the right, the LD region occupies 43 kb in length and contains two CGS-Ds (green dash), hypermethylation regions (red square) and hypomethylation regions (green squre). Each square represents about 50 bp and each dash represents 1bp. Abbreviations: CGS, CpG related SNP; CGS-C, CpG related SNP with the allele to create CpG dinucleotides; CGS-D, CpG related SNP with the allele to disrupt CpG dinucleotides.

CHAPTER 3

Age, SNPs, and DNA methylation at *APOE* locus

3.1 Abstract

Although apolipoprotein E (*APOE*) variants are associated with age-related diseases, the underlying mechanism is unknown and DNA methylation may be a potential one. With methylation data, measured by Methyl450, from both 993 participants (age ranging from 18 to 87 y) in the Genetics of Lipid Lowering Drugs and Diet Network study (GOLDN) study and from Encyclopedia of DNA Elements (ENCODE) consortium, we described the methylation pattern of 13 CpG sites within *APOE* locus, their correlations with gene expression across different cell types, and their relationships with 1) age, 2) plasma lipids, and 3) sequence variants. Based on methylation levels and regions of the gene, we categorized the 13 *APOE* CpG sites into three groups: Group 1 showed hypermethylation (>50%) and were located in the promoter region, Group 2 exhibited hypomethylation (<50%) and were located in the first two exons and introns, and Group 3 showed hypermethylation (>50%) and were located in the exon 4. Methylation at most CpG sites was negatively correlated with gene expression (minimum r = -0.66 with P = 0.004). *APOE* methylation was significantly associated with age (minimum P = 2.06E-08) and plasma total cholesterol (minimum P = 0.005). Finally, *APOE* methylation patterns differed across *APOE* ε variants (minimum P = 7.44E-05) and the promoter variant rs405509 (minimum P = 0.03), which further showed a significant interaction with age (P = 0.03). These findings suggest that methylation may be a potential mechanistic explanation for *APOE* functions related to aging and call for further molecular mechanistic studies.

Key words: apolipoprotein E, aging, DNA methylation, variants, epidemiology

3.2 Introduction:

Apolipoprotein E (ApoE, encoded by the *APOE* gene), a protein involved in both exogenous and endogenous lipid metabolism [1], plays a significant role in the process of age-related diseases, including cardiovascular diseases, Alzheimer's disease, and age-related macular disease [2]. While the vast majority of studies have investigated relationships between *APOE* sequence variants and age-related diseases [3-6], current studies on the relationship between aging and methylation pattern *APOE* are limited. One small-scale study suggested that the differences in *APOE* methylation between brains with late-onset Alzheimer's disease and normal brains increase with age [7]. The effect of aging on *APOE* methylation is highly plausible based on the general link between DNA methylation and aging. For example, nearly every step of cellular development and differentiation involves DNA methylation changes [8]. DNA methylation has also been shown to be associated with age-related diseases [9]. Furthermore, methylation of *APOE* was shown to be functional since it is modified by environmental factors such as folate [10, 11], and correlates with clinical phenotypes in some [12] but not all studies [13].

In light of the potential shared relationship with aging that appear to link *APOE* sequence variants and *APOE* methylation, along with the observation that sequence variants may actually alter methylation status [14], we propose that studying the two phenomenon in combination may be especially informative. Specifically, three functional single nucleotide polymorphisms (SNPs) may have the potential to modify DNA methylation at the *APOE* locus. The first SNP, rs405509, is located in the promoter region. The variant of this SNP was postulated to increase DNA methylation based on its demonstrated decreasing effect on gene transcription [15], the main functional effect of DNA methylation [16]. The other two SNPs are rs429358 and rs7412, which define the ε2/ε3/ε4 isoforms of ApoE and both are located within exon 4. These two SNPs are hypothesized to change DNA methylation as well, not only because they are located within the CpG island contained in the exon 4, but also because both are CpG related SNPs (eg. the cytosine allele forms a CpG dinucleotide while the thymine allele disrupts it).

Using data from 933 participants of the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study and data from Encyclopedia of DNA Elements (ENCODE) consortium, we described the methylation pattern of the 13 CpG sites within the *APOE* locus, which were available in Methyl450 array. With the data from ENCODE consortium, we analyzed the relationship between *APOE* methylation and gene expression across different cell types. Finally, we utilized GOLDN population to explore 1) whether age is associated with *APOE* methylation, 2) whether the previously observed age-associated changes in *APOE* methylation can be further linked to changes in plasma lipids, the main functional phenotype of *APOE*, and 3) whether the effect of age on *APOE* methylation can be modulated by methylation-related genetic variants in *APOE* locus.

3.3 Methodology

Study population

The GOLDN study was designed to evaluate genetic factors that modulate plasma lipid responses to a diet intervention (consumption of a high-fat meal) and fenofibrate treatment. GOLDN participants are of European ancestry and were enrolled from the National Heart, Lung, and Blood Institute Family Heart Study [17]. This GOLDN analysis included 475 men and 518 women (age ranged from 18 to 87y) who have baseline data for all required variables. The detailed design and methodology of GOLDN has been described [18]. Of relevance for this analysis, GOLDN required all subjects with a history of antilipemic drug use to be off all antilipemic medications at least 4 weeks prior to their study visit. The protocol for this study was approved by the Human Studies Committee of Institutional Review Board at the University of Minnesota, University of Utah, and Tufts University/New England Medical Center. Written informed consent was obtained from all participants. Fasting blood samples were collected to measure the lipid profile and detailed methodology was described previously [19]. Dietary intake was estimated using the diet history questionnaire [20, 21]. Physical activity was assessed by a questionnaire containing questions on the number of hours per day dedicated to different levels (heavy, slight, and sedentary) of activity [22].

Genotyping and methylation measurements

Genomic DNA for sequence genotyping was extracted from blood samples using Gentra Puregene Blood Kits (Gentra Systems, Inc., Minneapolis, MN). Genotypes of rs429358 and rs7412 were obtained using TaqMan assays on a ABI 7900HT system (Applied Biosystems, Foster City, CA). *APOE* genotypes were then called on the basis of the guidelines of Hixson and Vernier [23]. The promoter SNP rs405509 was genotyped using the Affymetrix Genome-Wide Human SNP array 6.0 (Affymetrix, Santa Clara, CA) [24].

Detailed methodology to measure DNA methylation was described previously [25]. CD4+ T-cells for methylation measurement were extracted from baseline frozen buffy coat samples isolated from peripheral blood using positive selection (Invitrogen, Grand Island, NY) followed by sorting of subsets by flow cytometry (FACSAriaII, BD Biosciences, San Jose, CA). Cells were then lysed and DNA was extracted using QIAGEN DNAeasy kits (QIAGEN, Germantown, MD). DNA sample (500 ng) were treated with sodium bisulfite using Zymo EZ DNA methylation kit (Zymo Research Corporation, Irvine, CA). DNA methylation was measured by the Infinium Human Methylation 450K BeadChip (Illumina, San Diego, CA) through amplification, hybridization, and imaging steps. Intensity files were generated and analyzed with Illumina's GenomeStudio, through which beta scores and "detection *P*-values" were generated. These beta scores represent the proportion of total signal from the methylation-specific probe or color channel. The "detection *P*-values" were defined as the probability that the total intensity

for a given probe falls within the background signal intensity. Those CpG probes with "detection *P*-values" greater than 0.01 and with more than 10% of samples that failed to yield adequate intensity were eliminated from further analysis. Those samples with more than 1.5% missing data points across ~470,000 autosomal CpGs were removed. After quality control, 13 CpG sites related to *APOE* remained. Start from the 5', the first four CpG sites (cg14123992, cg04406254, cg01032398, cg26190885) were located within 1.5 kb before the transcription start site, the fifth CpG site (cg12049787) was within the first exon, the next three CpG sites (cg08955609, cg18768621, cg19514613) were within the first intron, the ninth CpG site (cg06750524) was within the second intron, and the last four CpG sites (cg16471933, cg05501958, cg18799241, cg21879725) were within the fourth exon.

Gene expression and DNA methylation in ENCODE

Methylation levels of all 13 CpG sites for 62 cell lines in ENCODE were downloaded (09-20-2013) from UCSC genome browser HAIB Methyl450 track (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeHaibMethyl450/) and represented by a heat map using R (www.r-project.org). The score of the methylation value associated with each CpG site was defined as the beta value multiplied by 1000, with the beta value in turn defined as the proportion of the intensity value from the methylated bead type from the sum of the intensity values from both methylated and unmethylated bead type plus 100.

Gene expression data for *APOE* in 17 cell lines, which were also with methylation data, in ENCODE were downloaded (11-08-2013) from UCSC genome browser Duke Affymetrix Exon Array track (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeDukeAffyExon/). This track displayed exon array data which had been aggregated to the gene level for those probes that had been linked to genes. The expression score for each cell line represented a linearly scaled value for that particular cell type multiplied by 100, which ranged from 0 to 1000.

Statistical methods

According to ENCODE data, Pearson correlation analysis was conducted to test the correlations between methylation of each CpG site and gene expression of *APOE*. In GOLDN, Mantel-Haenszel χ^2 tests and ANOVA tests were used to examine the trend of significance in characteristics of the study population, as categorical and continuous variables respectively, by age, categorized in quintiles. Generalized estimating equations (GEE) methods were used to test the association of methylation of each CpG site with age (continuous variable), blood lipids, and *APOE* genotypes, and the interaction between age (continuous variable) and selected variants. Each CpG site was included into the model separately. *APOE* ε variants were coded into three

categories, ε3/ε3, ε2 carriers (including ε2/ε2 and ε2/ε3), and ε4 carriers (including ε4/ε4 and ε3/ε4). Individuals with the ε2/ ε4 genotype (n=46) were excluded from the analysis to distinguish the distinct role of each variant. The primary analysis was adjusted for pedigree, sex, study center, and the first principal component of both cellular purity and population structure. To test the effect of other potential confounders, the secondary analysis was adjusted for smoking (never smoker past smoker, and current smoker), drinking (ever drink alcohol or not), total energy intake (kcal/d), physical activity (hours of total physical activity / d), vitamin B12 intake (mcg/d), folate intake (mcg/d), hormone replacement therapy in women, and a history of use of antilipemic medication. Likelihood ratio tests were conducted to analyze whether the effects of age on plasma TC is partially through methylation of *APOE*. Each continuous variable was tested for normality, and log transformation were performed for those variables not following a normal distribution. Correlation analysis with data from ENCODE consortium and all data analysis with GOLDN population were analyzed using SAS (version 9.3 for Windows; SAS Institute, Inc. Cary, NC, USA). A two-tail *P*-value of <0.05 was considered statistically significant.

3.4 Results

(1) Population characteristics

Population characteristics were compared across age quintiles (**Table 3.1**). Gender distribution did not differ across age quintiles. Compared to the younger age quintiles, the older quintiles tended to contain fewer smokers ($P < 0.05$) but more with a past history of taking antilipemic medication ($P < 0.0001$). Also, those individuals in the older age quintiles tended to consume less total energy, vitamin B12 and folate ($P < 0.005$), and to have higher plasma triglycerides (TG), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-c), and low-density lipoprotein cholesterol (LDL-c) ($P < 0.0001$) than those in the younger quintiles.

(2) *APOE* methylation status in GOLDN and ENCODE data

We examined the methylation status of 13 CpG sites in *APOE* using both GOLDN and ENCODE data. The genetic structure of *APOE* locus was described in **Figure 3.1**. Based on both the methylation levels in GOLDN and genetic locations, three groups of CpG sites can be distinguished. Specifically, the first three sites comprised a group (Group 1) that were both hypermethylated (all sites > 50% methylation) and located within the promoter region. The second group of CpG sites (Group 2, sites 4 to 9) were both hypomethylated (< 50%) and located in the 5' part of the gene. The third group (Group 3, sites 10 to 13) were both hypermethylated (>50%) and located at the 3' end of the gene. Examination of *APOE* methylation patterns using 62 cell lines in ENCODE data confirms that these 13 CpG sites can be categorized into three groups based on their methylation status and genetic locations.

(3) Correlation between methylation of each CpG site and *APOE* gene expression using ENCODE data:

We extended our analyses in ENCODE by evaluating the relationship between CpG methylation and gene expression across 17 cell types with data available for both gene expression and methylation. Methylation of four CpG sites, distributed across three methylation groups identified above, was negatively correlated with gene expression (**Table 3.2**). Two CpG sites within Group 1, CpG 2 (cg04406254) and CpG 3 (cg01032398), showed the strongest correlation with gene expression, with Pearson correlation coefficients of -0.66 ($P = 0.004$) and -0.62 ($P = 0.008$), respectively. In Group 2, CpG 7 (cg18768621) had a borderline significance ($P = 0.05$) with a coefficient of -0.48. In Group 3, CpG 12 (cg18799241) was negatively correlated with a coefficient of -0.51 ($P = 0.04$).

(4) Age is associated with *APOE* methylation in GOLDN

We next evaluated relationships between age and *APOE* methylation by examining methylation of the 13 CpG sites occurring within the three distinct methylation groups (Group 1, Group 2, Group 3), identified above. Age was significantly associated with methylation of at least one CpG site in each group (**Table 3.3**). For example, age was significantly negatively associated with DNA methylation values for all three CpG sites in Group 1, a hypermethylated group within promoter region ($P < 0.0001$), including CpG 1 (cg14123992), CpG 2 (cg04406254), and CpG 3 (cg01032398). However, age was significantly positively associated with methylation of CpG 8 (cg19514613) ($P = 0.004$), one CpG site in Group 2, a hypomethylated group within the 5' part of the gene. For Group 3, which was hypermethylated and at the 3' end of the gene, age was significantly negatively associated with methylation of CpG 10 (cg16471933) ($P = 0.04$). Significant relationships were not altered in secondary analyses adjusted for smoking, drinking, total energy intake, physical activity, vitamin B12 intake, folate intake, hormone replacement therapy in women, and history of antilipemic medication (data not shown).

(5) Age-associated *APOE* methylation is associated with blood lipids in GOLDN

We next examined whether the age-related *APOE* methylation previously observed is associated with plasma lipids (TC, LDL-C, TG and high density lipoprotein cholesterol (HDL-C)), the main phenotype of *APOE* (**Table 3.4**). Methylation of five CpG sites was significantly associated with plasma TC. Specifically, methylation of CpG 7 (cg18768621) and CpG 9 (cg06750524) were positively associated with TC ($P = 0.03$). Methylation of CpG 10 (cg16471933), CpG 12 (cg18799241), and CpG 13 (cg21879725) was negatively associated with TC ($P < 0.05$). For LDL-C, methylation of CpG 9 (cg06750524) had a significant positive association ($P = 0.01$).

To examine whether the association of age on TC could be mediated in part through DNA methylation mechanisms, we conducted both indirect (**Figure 3.2**) and direct analysis (**Table 3.5**). With respect to the indirect evidence, we plotted the beta values for age regressed on methylation of each CpG site (**Figure 3.2A**) and those for methylation of each CpG site regressed on TC (**Figure 3.2B**). The pattern of the association between age and methylation is moderately parallel to that between methylation and TC, which showed negative associations for CpG sites in both Group 1 and Group 3, both of which demonstrated hypermethylation, and a positive association for CpG sites in Group 2 which showed hypomethylation. In terms of the direct evidence, we conducted a sensitivity analysis to compare the regression coefficients and statistical significance of TC on age with model without and with inclusion of methylation values of each CpG site (**Table 3.5**). Compared to the model without inclusion of each CpG site, the regression coefficients and statistical significance were reduced in the model that adjusted for the methylation levels of each CpG sites. According to the likelihood ratio tests, these two models

significantly differ with the inclusion of each following CpG site: CpG 4 (cg01032398) (P < 0.0001), CpG 10 (cg16471933) (P = 0.04), and CpG 13 (cg21879725) (P = 0.03).

(6) *APOE* variants are associated with methylation in GOLDN

We next explored relationships between *APOE* methylation and *APOE* sequence variants. Five CpG sites were significantly associated with *APOE* ε variants (**Table 3.6**). For CpG 1 (cg14123992) and CpG 3 (cg01032398), individuals with ε3/ε3 genotype have lowest methylation levels compared to the ε2 carriers and ε4 carriers (P = 0.03 and 0.003 respectively). However, the order for the betas of the associations with methylation level is ε2 carriers < ε3/ε3 < ε4 carriers at CpG 8 (cg19514613) (P = 0.03), CpG 9 (cg06750524) (P = 7.44E-5), CpG 10 (cg16471933) (P = 0.05), and CpG 13 (cg21879725) (P = 0.02).

Promoter SNP rs405509 had significant association with methylation of CpG 2 (cg0406254) and CpG 10 (cg16471933) (P = 0.03) (**Table 3.6**). Homozygotes of the minor allele (AA) had the highest methylation levels while the homozygotes of major allele (CC) had the lowest, with the values for the heterozygotes in the middle.

(7) *APOE* variant interacts with age to modulate methylation in GOLDN

Finally, we examined genetic variants as a potential modulator of the effect of age on methylation of each CpG site at the *APOE* locus. Promoter SNP rs405509 significantly interacted with age to modulate methylation of CpG 3 (cg01032398) (P for interaction = 0.03) (**Figure 3.3**). For major allele carriers (CC and AC), methylation of CpG 3 significantly decreased with age (P for trend = 0.01 and 0.004 respectively). However, methylation was unaffected by age in homozygotes for the minor allele (AA) (P for trend = 0.97).

3.5 Discussion

In the current study, we described the methylation patterns of the *APOE* locus and their correlations with gene expression, and observed associations between age, *APOE* methylation, and blood lipids, and an interaction between age and a methylation-associated promoter variant. This is the first study to explore *APOE* methylation at single nucleotide resolution with a population of a thousand individuals.

Our finding that age is associated with *APOE* methylation is consistent with a previous study by Wang et.al. [7] showing that the differences in *APOE* methylation between individuals with and without Alzheimer's disease increases with age. Other studies demonstrated that age affects global DNA methylation [26]. Compared to younger monozygous twins, older twins exhibited greater differences in DNA methylation. This may be a consequence of the fact that DNA methylation is modifiable by environmental factors [8], which accumulate gradually or change continuously with age. In our study, we expected *APOE* methylation to change with age because *APOE* is considered an age-related gene. This is based on the genetic associations between *APOE* variants and many age-related diseases, such as coronary heart disease [3], atherosclerosis [4], age-related macular degeneration (AMD) [5], and Alzheimer's disease [6]. We also observed that the direction of the age-associated differences in methylation levels appeared to be associated with the existing degree of methylation. Specifically, we observed that greater age was associated with less methylation in hypermethylated regions and with greater methylation in hypomethylated regions. Finally, we observed moderate parallel patterns between the association of age with methylation and associations of methylation with blood lipids, which may imply biological connections between aging, methylation and lipid concentrations.

Two hypothetical mechanisms for the observed differences in *APOE* methylation by *APOE* variants are differential allelic gene expression and changes in the DNA sequence that affect CpG site formation. With respect to the first mechanism, the minor allele of promoter SNP rs405509 (A) has been reported to exhibit lower gene transcription compared to the major allele (C) [15]. For the *APOE* ε variants, mRNA expression for ε3 allele was shown to be greater than that of ε4 [27]. Based on the established relationship between DNA methylation and reduced gene expression [16], our findings of higher methylation levels for carriers of the rs405509 A allele and ε4 allele in GOLDN are consistent with the lower expression reported in previous studies. The second potential mechanism by which genotype may alter methylation is based on the creation and disruption of CpG sites by nucleotide changes that determine the *APOE* ε variants. The two SNPs that constitute *APOE* ε variants are both CpG-related SNPs. That means that ε3 carriers have two CpG dinucleotides because both SNPs contain the C allele that is needed to create CpG dinucleotides, while ε2/ ε4 have one CpG dinucleotide because one SNP contains a C allele but the other SNP contains a T allele. Based on the recognition that most DNA methylation in the mammalian genome occurs on CpG dinucleotides [28], the density of CpG sites is likely to affect the local DNA methylation patterns.

In light of our findings that both age and *APOE* genotype were related to methylation, a logical next step was to examine whether these two factors might act in combination to alter methylation. Our observation that promoter SNP rs405509 significantly modified the association between age and promoter methylation is novel and plausible. Greater age was associated with less promoter methylation in the carriers of the major allele (CC and AC) but not in the homozygotes of the minor allele (AA), such that methylation of those in the AA group remained high regardless of age. This age-related allelic difference in methylation level may provide mechanistic support for previous findings linking the A allele of rs405509 to greater risk of myocardial infarction [29], premature coronary heart disease [30], and Alzheimer disease [31], but lower plasma concentration of ApoE [29].

This study had a number of limitations. Based on the cross-sectional study design, we cannot establish causality. Further mechanistic studies are necessary. Also, the measurement of *APOE* methylation in circulating CD4+ T-cells may provide a limited perspective on a protein with multiple functions. However, the consistent methylation pattern across different types of cell lines observed with ENCODE data increases the generalizability of our findings.

In summary, we characterized thirteen CpG sites at the *APOE* locus into three groups based on their genetic locations and methylation status in both GOLDN and ENCODE, and observed that most of these sites were negatively correlated with *APOE* gene expression based on ENCODE data. With a large population in GOLDN, we found that age was indeed associated with *APOE* methylation and linked those associations to changes in blood lipid profile. Furthermore, we observed that methylation-associated genetic variants of *APOE* modified the aging effect on methylation. Our findings are novel and consistent with the previous evidence from genetic studies, and may provide potential mechanistic explanations for aging related functions of *APOE*.

3.6 References:

1. Mahley, R.W., *Apolipoprotein E: cholesterol transport protein with expanding role in cell biology.* Science, 1988. **240**(4852): p. 622-30.
2. Ang, L.S., et al., *Apolipoprotein E, an important player in longevity and age-related diseases.* Exp Gerontol, 2008. **43**(7): p. 615-22.
3. Ward, H., et al., *APOE genotype, lipids, and coronary heart disease risk: a prospective population study.* Arch Intern Med, 2009. **169**(15): p. 1424-9.
4. Davignon, J., R.E. Gregg, and C.F. Sing, *Apolipoprotein E polymorphism and atherosclerosis.* Arteriosclerosis, 1988. **8**(1): p. 1-21.
5. Klaver, C.C., et al., *Genetic association of apolipoprotein E with age-related macular degeneration.* Am J Hum Genet, 1998. **63**(1): p. 200-6.
6. Saunders, A.M., et al., *Specificity, sensitivity, and predictive value of apolipoprotein-E genotyping for sporadic Alzheimer's disease.* Lancet, 1996. **348**(9020): p. 90-3.
7. Wang, S.C., B. Oelze, and A. Schumacher, *Age-specific epigenetic drift in late-onset Alzheimer's disease.* PLoS One, 2008. **3**(7): p. e2698.
8. Cedar, H. and Y. Bergman, *Programming of DNA methylation patterns.* Annu Rev Biochem, 2012. **81**: p. 97-117.
9. Johnson, A.A., et al., *The role of DNA methylation in aging, rejuvenation, and age-related disease.* Rejuvenation Res, 2012. **15**(5): p. 483-94.
10. Yi-Deng, J., et al., *Folate and ApoE DNA methylation induced by homocysteine in human monocytes.* DNA Cell Biol, 2007. **26**(10): p. 737-44.
11. Glier, M.B., T.J. Green, and A.M. Devlin, *Methyl nutrients, DNA methylation, and cardiovascular disease.* Mol Nutr Food Res, 2014. **58**(1): p. 172-82.
12. Turan, N., et al., *DNA methylation differences at growth related genes correlate with birth weight: a molecular signature linked to developmental origins of adult disease?* BMC Med Genomics, 2012. **5**: p. 10.
13. Sharma, P., et al., *Detection of altered global DNA methylation in coronary artery disease patients.* DNA Cell Biol, 2008. **27**(7): p. 357-65.
14. Zhi, D., et al., *SNPs located at CpG sites modulate genome-epigenome interaction.* Epigenetics, 2013. **8**(8): p. 802-6.
15. Artiga, M.J., et al., *Allelic polymorphisms in the transcriptional regulatory region of apolipoprotein E gene.* FEBS Lett, 1998. **421**(2): p. 105-8.
16. Lindahl, T., *DNA methylation and control of gene expression.* Nature, 1981. **290**(5805): p. 363-4.
17. Higgins, M., et al., *NHLBI Family Heart Study: objectives and design.* Am J Epidemiol, 1996. **143**(12): p. 1219-28.
18. Corella, D., et al., *The -256T>C polymorphism in the apolipoprotein A-II gene promoter is associated with body mass index and food intake in the genetics of lipid lowering drugs and diet network study.* Clin Chem, 2007. **53**(6): p. 1144-52.
19. Tsai, M.Y., et al., *Effect of influenza vaccine on markers of inflammation and lipid profile.* J Lab Clin Med, 2005. **145**(6): p. 323-7.
20. Thompson, F.E., et al., *Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study.* J Am Diet Assoc, 2002. **102**(2): p. 212-25.
21. Subar, A.F., et al., *Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires : the Eating at America's Table Study.* Am J Epidemiol, 2001. **154**(12): p. 1089-99.
22. Corella, D., et al., *A high intake of saturated fatty acids strengthens the association between the fat mass and obesity-associated gene and BMI.* J Nutr, 2011. **141**(12): p. 2219-25.

23. Hixson, J.E. and D.T. Vernier, *Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI.* J Lipid Res, 1990. **31**(3): p. 545-8.
24. Aslibekyan, S., et al., *A genome-wide association study of inflammatory biomarker changes in response to fenofibrate treatment in the Genetics of Lipid Lowering Drug and Diet Network.* Pharmacogenet Genomics, 2012. **22**(3): p. 191-7.
25. Absher, D.M., et al., *Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations.* PLoS Genet, 2013. **9**(8): p. e1003678.
26. Fraga, M.F., et al., *Epigenetic differences arise during the lifetime of monozygotic twins.* Proc Natl Acad Sci U S A, 2005. **102**(30): p. 10604-9.
27. Lambert, J.C., et al., *Distortion of allelic expression of apolipoprotein E in Alzheimer's disease.* Hum Mol Genet, 1997. **6**(12): p. 2151-4.
28. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-22.
29. Lambert, J.C., et al., *Independent association of an APOE gene promoter polymorphism with increased risk of myocardial infarction and decreased APOE plasma concentrations-the ECTIM study.* Hum Mol Genet, 2000. **9**(1): p. 57-61.
30. Viitanen, L., et al., *Apolipoprotein E gene promoter (-219G/T) polymorphism is associated with premature coronary heart disease.* J Mol Med (Berl), 2001. **79**(12): p. 732-7.
31. Lambert, J.C., et al., *A new polymorphism in the APOE promoter associated with risk of developing Alzheimer's disease.* Hum Mol Genet, 1998. **7**(3): p. 533-40.

Table 3.1. Subject characteristics of GOLDN by age quintile[a].

Variable	Age quintiles					
	Q1	Q2	Q3	Q4	Q5	*P* for trend[c]
Age (y)[b]	24 (18-34)	40 (35-43)	47 (44-51)	57 (52-64)	71 (65-87)	
n	200	185	204	209	195	
men	88 (44)	98 (52.97)	96 (47.06)	99 (47.37)	94 (48.21)	0.64
current smoker	17 (8.5)	18 (9.73)	21 (10.34)	13 (6.22)	4 (2.05)	0.004
drinker	79 (39.5)	101 (54.59)	104 (50.98)	102 (48.8)	95 (48.72)	0.16
Total energy intake (kcal/day)	2448 (1758)	2228 (1202)	2169 (981)	2037 (925)	1746 (809)	<0.0001
Vitamin B12 intake (mcg/day)	6.08 (4.41)	5.39 (3.33)	5.31 (3.03)	5.31 (2.93)	4.68 (3.29)	0.0001
Folate intake (mcg/day)	466.95 (315.66)	411.93 (237.7)	398.22 (171.36)	398.46 (187.07)	362.91 (166.53)	<0.0001
Have history to take antilipemic medication	0 (0)	2 (1.09)	1 (0.49)	14 (6.7)	20 (10.31)	<0.0001
TG (mg/dL)	100.11 (69.93)	128.06 (95.93)	133 (87.82)	148.96 (86.16)	150.69 (97.61)	<0.0001
TC (mg/dL)	162.04 (31.96)	186.39 (31.66)	192.21 (30.45)	198.71 (35.68)	189.66 (39.49)	<0.0001
HDL-c (mg/dL)	46.65 (11.27)	48.48 (13.09)	49.79 (13.48)	49.54 (14.88)	50.08 (14.67)	0.009
LDL-c (mg/dL)	102.65 (28.02)	119.86 (28.6)	121.93 (27.49)	127.13 (30.11)	116.8 (31.92)	<0.0001

[a]Data are means (standard deviation) or n (%);
[b]Data are median age (minimum age – maximum age) within each quintile;
[c]Mantel-Haenszel $\chi 2$ tests and ANOVA tests were applied to obtain *P* values for trend of categorical and continuous variables, irrespectively, according to the median of age in each quintile.

Table 3.2. Correlations between methylation of each CpG site with *APOE* gene expression in ENCODE

CpG group*	CpG #	CpG Name	Pearson Correlation Coefficients	*P*
Group 1	CpG 1	cg14123992	-0.45	0.07
	CpG 2	cg04406254	-0.66	0.004
	CpG 3	cg01032398	-0.62	0.008
Group 2	CpG 4	cg26190885	-0.17	0.52
	CpG 5	cg12049787	-0.33	0.2
	CpG 6	cg08955609	-0.16	0.53
	CpG 7	cg18768621	-0.48	0.05
	CpG 8	cg19514613	-0.4	0.11
	CpG 9	cg06750524	-0.28	0.28
Group 3	CpG 10	cg16471933	0.39	0.12
	CpG 11	cg05501958	0.21	0.43
	CpG 12	cg18799241	-0.51	0.04
	CpG 13	cg21879725	0.3	0.25

*CpG groups were defined according to both the methylation level and region of the gene.

Table 3.3. Methylation of *APOE* CpG sites by age quintile in GOLDN[a]

CpG group	CpG #	CpG Name	Age quintiles[b]					*P*[c]
			Q1 24 (18-34) y n = 200	Q2 40 (35-43) y n = 185	Q3 47 (44-51) y n = 204	Q4 57 (52-64) y n = 209	Q5 71 (65-87) y n = 195	
Group 1	CpG 1	cg14123992	88.13 (0.45)	85.32 (0.83)	86.07 (0.61)	85.72 (0.42)	84.46 (0.62)	8.00E-05
	CpG 2	cg04406254	72.36 (0.24)	70.62 (0.35)	70.75 (0.33)	70.29 (0.34)	69.90 (0.27)	2.06E-08
	CpG 3	cg01032398	82.80 (0.18)	81.58 (0.45)	82.22 (0.22)	81.96 (0.23)	81.39 (0.23)	1.16E-04
Group 2	CpG 4	cg26190885	9.17 (0.10)	9.30 (0.09)	9.32 (0.08)	9.44 (0.08)	9.33 (0.08)	0.11
	CpG 5	cg12049787	6.63 (0.24)	7.13 (0.42)	6.94 (0.35)	6.80 (0.19)	7.02 (0.23)	0.44
	CpG 6	cg08955609	3.79 (0.11)	5.38 (0.77)	4.83 (0.59)	4.69 (0.44)	4.69 (0.42)	0.19
	CpG 7	cg18768621	11.39 (0.86)	14.77 (1.23)	11.98 (0.81)	12.76 (1.12)	13.70 (1.10)	0.30
	CpG 8	cg19514613	13.13 (0.15)	13.21 (0.22)	13.40 (0.15)	13.92 (0.20)	13.61 (0.18)	0.004
	CpG 9	cg06750524	24.78 (0.25)	25.75 (0.51)	25.51 (0.42)	25.35 (0.35)	24.66 (0.27)	0.60
Group 3	CpG 10	cg16471933	80.97 (0.35)	79.51 (0.49)	80.01 (0.50)	80.28 (0.37)	79.54 (0.39)	0.04
	CpG 11	cg05501958	96.29 (0.11)	95.17 (0.60)	95.41 (0.50)	95.65 (0.37)	95.98 (0.09)	0.62
	CpG 12	cg18799241	87.27 (0.87)	83.57 (1.27)	86.23 (0.86)	86.49 (1.08)	85.62 (1.08)	0.68
	CpG 13	cg21879725	84.22 (1.15)	80.46 (1.37)	83.64 (1.02)	86.74 (1.20)	82.12 (1.24)	0.62

[a]Data are means (standard error of the means) of DNA methylation percentages (%) adjusted for the covariates below;
[b]Median age (minimum age – maximum age) within each quintile were presented;
[c]*P*: *P* value for the association between age (quintile) and DNA methylation adjusting for pedigree, sex, center, and the first principle component of cellular purity and population structure.

Table 3.4. Associations between methylation of *APOE* CpG sites with lipids in GOLDN.

CpG group	CpG #	CpG Name	TC			HDL-c			LDL-c			TG (log)		
			beta[b]	sem[b]	*P*[c]	beta[b]	sem[b]	*P*[c]	beta[b]	sem[b]	*P*[c]	beta[a]	sem[a]	*P*[c]
Group 1	CpG 1	cg14123992	-0.20	0.10	0.08	-0.09	0.05	0.14	-0.13	0.11	0.25	4.4E-04	0.002	0.84
	CpG 2	cg04406254	-0.26	0.20	0.24	-0.04	0.09	0.64	-0.18	0.19	0.38	-0.003	0.005	0.46
	CpG 3	cg01032398	0.09	0.25	0.70	-0.12	0.09	0.34	0.17	0.22	0.40	0.004	0.005	0.52
Group 2	CpG 4	cg26190885	0.01	0.85	0.99	0.25	0.31	0.42	0.06	0.74	0.93	-9.5E-06	0.014	1.00
	CpG 5	cg12049787	0.33	0.22	0.14	0.15	0.12	0.25	0.23	0.22	0.32	-0.002	0.004	0.57
	CpG 6	cg08955609	0.12	0.14	0.43	0.03	0.04	0.48	0.15	0.14	0.33	-0.001	0.003	0.60
	CpG 7	cg18768621	0.17	0.07	0.03	-0.02	0.03	0.57	0.12	0.07	0.09	0.003	0.001	0.08
	CpG 8	cg19514613	-0.65	0.41	0.14	-0.17	0.17	0.31	-0.12	0.37	0.76	-0.007	0.008	0.33
	CpG 9	cg06750524	0.38	0.16	0.03	-0.02	0.06	0.73	0.47	0.15	0.01	0.001	0.003	0.64
Group 3	CpG 10	cg16471933	-0.41	0.15	0.03	0.003	0.07	0.97	-0.19	0.13	0.18	-0.005	0.004	0.13
	CpG 11	cg05501958	-0.29	0.11	0.09	-0.04	0.05	0.40	-0.29	0.11	0.09	-1.1E-04	0.003	0.97
	CpG 12	cg18799241	-0.18	0.07	0.02	-0.01	0.03	0.79	-0.10	0.06	0.14	-0.002	0.001	0.20
	CpG 13	cg21879725	-0.18	0.06	0.005	0.005	0.02	0.85	-0.11	0.05	0.06	-0.002	0.001	0.06

[a]Beta and sem represent changes in log transformed TG (mg/dL) corresponding to 1% increase in DNA methylation;
[b]Beta and sem represent changes in lipids (mg/dL) corresponding to 1% increase in DNA methylation;
[c]*P*: *P* value for the association between DNA methylation (%) and lipids adjusting for pedigree, sex, center, and the first principal component of cellular purity and population structure.

Table3.5. Sensitivity analysis to determine whether the methylation level of each CpG site could be mediating factor between age and blood total cholesterol[a]

CpG group	CpG #	CpG Name	Not include CpG site		Include CpG site		Likelihood Ratio Tests for two models
			Beta(SEM)	*P*	Beta(SEM)	*P*	
Group 1	CpG 1	cg14123992	0.574(0.08)	4.62454E-08	0.562(0.08)	6.4164E-08	0.56
	CpG 2	cg04406254	0.574(0.08)	4.62454E-08	0.563(0.08)	6.521E-08	0.59
	CpG 3	cg01032398	0.574(0.08)	4.62454E-08	0.576(0.08)	4.5111E-08	0.41
Group 2	CpG 4	cg26190885	0.574(0.08)	4.62454E-08	0.568(0.08)	1.4086E-07	<0.0001
	CpG 5	cg12049787	0.574(0.08)	4.62454E-08	0.573(0.08)	5.0233E-08	0.3
	CpG 6	cg08955609	0.574(0.08)	4.62454E-08	0.573(0.08)	4.7008E-08	0.27
	CpG 7	cg18768621	0.574(0.08)	4.62454E-08	0.569(0.08)	4.9486E-08	0.15
	CpG 8	cg19514613	0.574(0.08)	4.62454E-08	0.583(0.08)	4.4032E-08	0.11
	CpG 9	cg06750524	0.574(0.08)	4.62454E-08	0.576(0.08)	3.7557E-08	0.13
Group 3	CpG 10	cg16471933	0.574(0.08)	4.62454E-08	0.566(0.08)	5.8425E-08	0.04
	CpG 11	cg05501958	0.574(0.08)	4.62454E-08	0.574(0.08)	4.4807E-08	0.3
	CpG 12	cg18799241	0.574(0.08)	4.62454E-08	0.572(0.08)	4.882E-08	0.09
	CpG 13	cg21879725	0.574(0.08)	4.62454E-08	0.572(0.08)	4.877E-08	0.03

[a]Data are the regression coefficients (standard errors) of the association between age and plasma total cholesterol adjusted for pedigree, sex, center, and the first principal components of cellular purity and population structure.

Table 3.6. Methylation of *APOE* CpG sites by carriers of three different ε alleles in GOLDN[a]

CpG group	CpG #	CpG Name	*APOE* rs405509				*APOE* ε variants			
			CC	AC	AA	P^b	ε2 carriers	ε3/ ε3	ε4 carriers	P^b
			n = 256	n = 500	n = 236		n = 99	n = 588	n = 257	
Group 1	CpG 1	cg14123992	85.87 (0.56)	86.19 (0.35)	85.52 (0.62)	0.66	86.84 (0.32)	85.70 (0.41)	86.14 (0.49)	0.03
	CpG 2	cg04406254	69.96 (0.37)	71.07 (0.18)	71.10 (0.26)	0.03	70.36 (0.34)	70.84 (0.20)	70.85 (0.30)	0.67
	CpG 3	cg01032398	81.80 (0.19)	81.98 (0.22)	82.23 (0.21)	0.51	82.07 (0.30)	81.78 (0.19)	82.43 (0.21)	0.003
Group 2	CpG 4	cg26190885	9.22 (0.08)	9.34 (0.05)	9.35 (0.08)	0.48	9.28 (0.11)	9.38 (0.05)	9.23 (0.07)	0.31
	CpG 5	cg12049787	6.97 (0.30)	6.87 (0.18)	6.89 (0.28)	0.92	6.57 (0.12)	7.01 (0.18)	6.83 (0.28)	0.31
	CpG 6	cg08955609	5.11 (0.57)	4.61 (0.31)	4.31 (0.38)	0.59	4.00 (0.08)	4.80 (0.32)	4.69 (0.47)	0.10
	CpG 7	cg18768621	13.45 (0.99)	13.05 (0.60)	11.93 (0.74)	0.54	13.56 (1.87)	13.37 (0.60)	11.51 (0.59)	0.08
	CpG 8	cg19514613	13.20 (0.17)	13.52 (0.11)	13.61 (0.14)	0.19	13.17 (0.29)	13.37 (0.11)	13.72 (0.13)	0.03
	CpG 9	cg06750524	24.41 (0.38)	25.52 (0.21)	25.44 (0.26)	0.08	23.85 (0.29)	25.05 (0.19)	26.25 (0.33)	7.44E-05
Group 3	CpG 10	cg16471933	79.53 (0.37)	80.02 (0.28)	80.78 (0.30)	0.03	79.07 (0.57)	80.07 (0.26)	80.46 (0.29)	0.05
	CpG 11	cg05501958	95.28 (0.49)	95.71 (0.24)	96.15 (0.10)	0.12	96.15 (0.11)	95.67 (0.24)	95.58 (0.38)	0.19
	CpG 12	cg18799241	85.34 (0.99)	85.69 (0.62)	86.84 (0.74)	0.51	85.15 (1.80)	85.45 (0.61)	87.06 (0.58)	0.08
	CpG 13	cg21879725	82.35 (1.12)	82.41 (0.72)	84.45 (0.82)	0.22	81.57 (2.16)	82.21 (0.72)	84.68 (0.66)	0.02

[a]Data are means (standard error of the means) of DNA methylation percentages (%) adjusted for the covariates below;
[b]*P*: *P* value for the association between carriers of different *APOE* ε isoforms and DNA methylation (%) adjusting for pedigree, sex, center, and the first principal components of cellular purity and population structure.

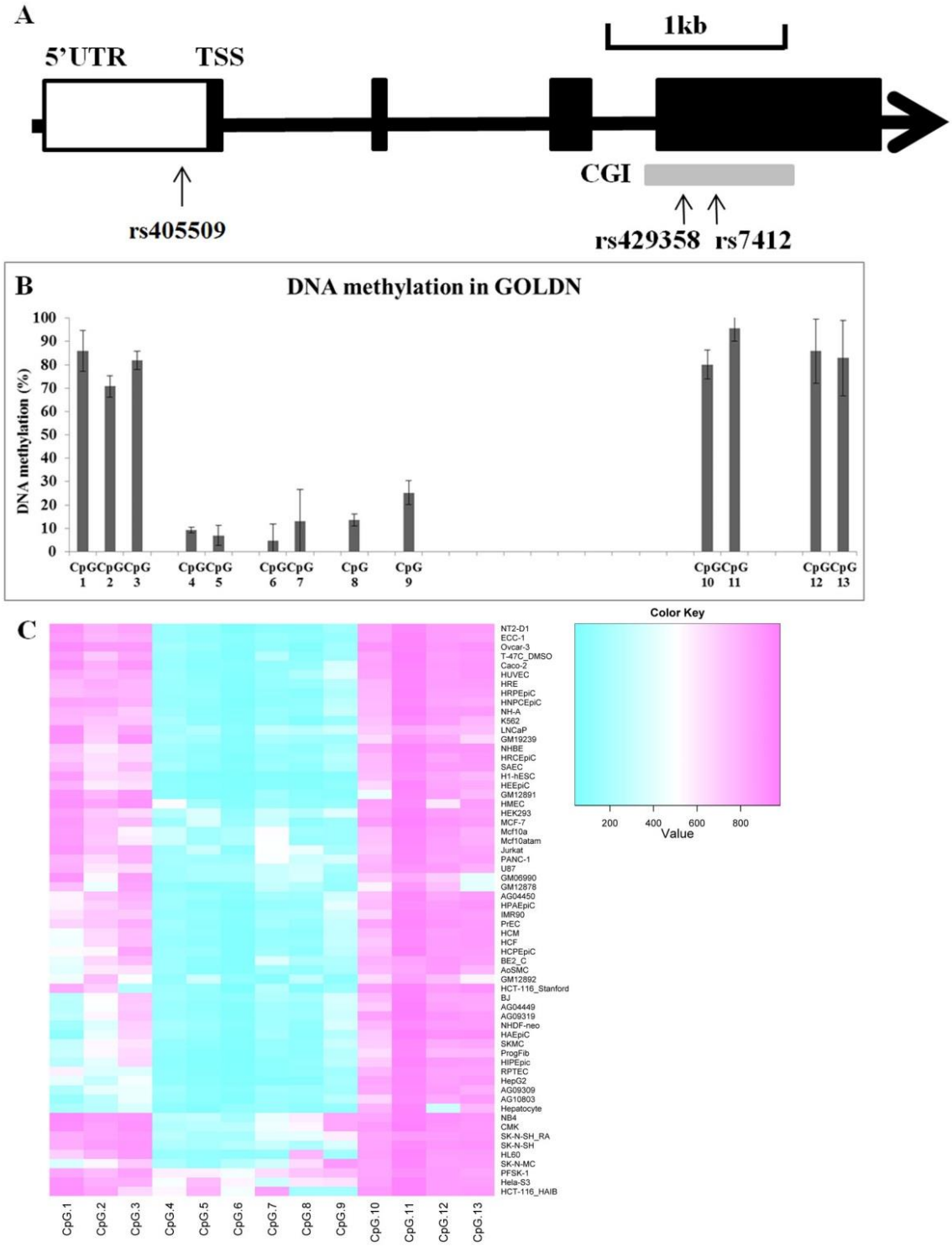


Figure 3.1 Gene structure and methylation pattern of *APOE* in GOLDN and ENCODE. The top panel shows the structure of *APOE* gene. The arrow represents the direction of the gene and filled rectangles represent exons. The panel in the middle shows the mean of DNA methylation (%) for each of the 13 CpG sites in GOLDN. The locations of these sites are corresponding to the gene structure in the top panel. The bottom panel is the heatmap of the methylation values of these 13 CpG sites in different cell lines according to ENCODE. The pink represent hypermethylation (>50%) and the green represent hypomethylation (<50%). Abbreviations: 5'UTR, 5' untranslated region; TSS, transcription start site; CGI, CpG island.

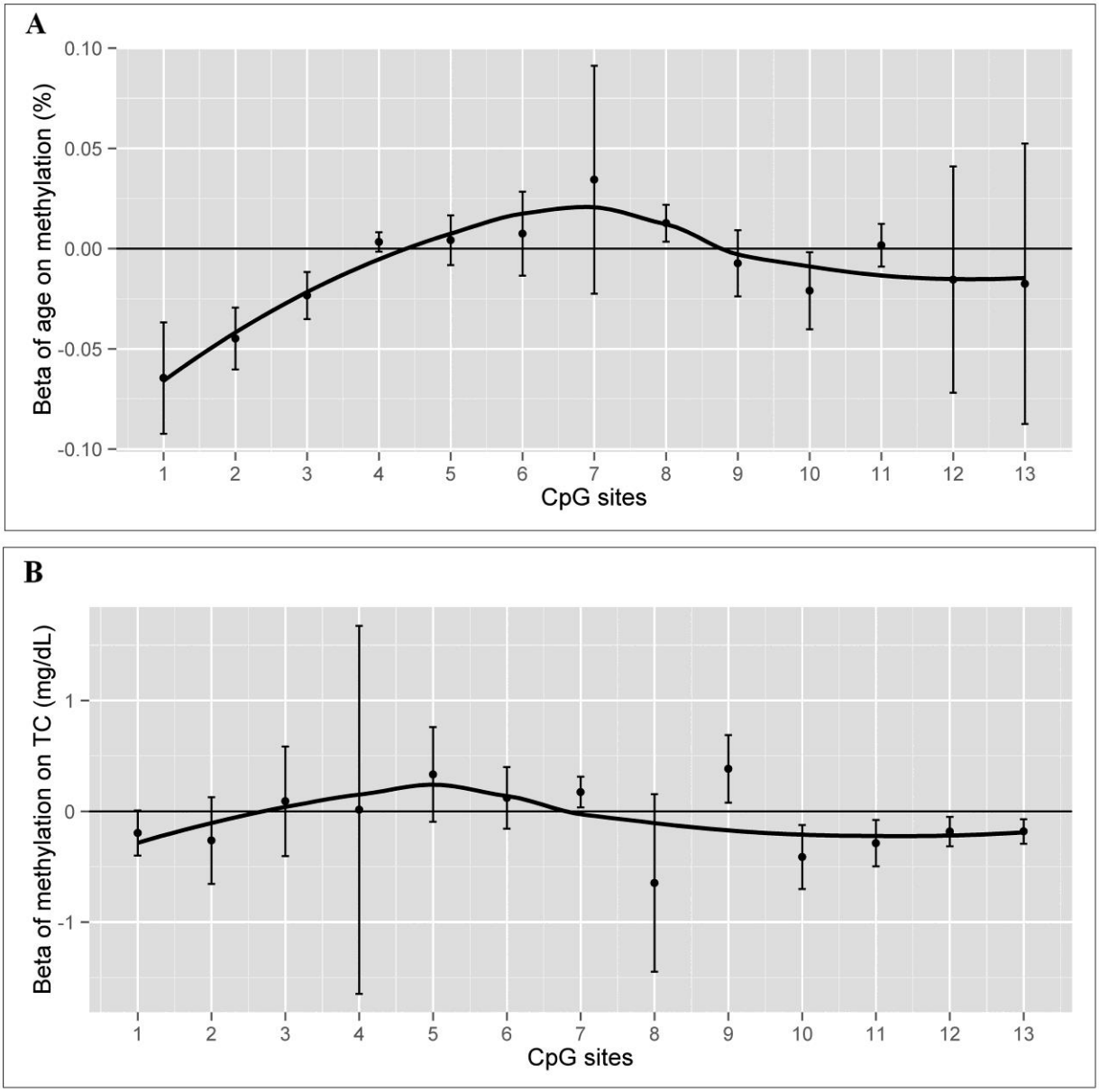


Figure 3.2 Parallel patterns of the ApoE methylation changes with age and total cholesterol in GOLDN. Thirteen CpG sites are listed as x-axis and the y-axis are the betas of age on methylation (%) (A) and the betas of methylation on TC (mg/dL) (B). Black dot represent point estimate of beta for each CpG site from generalized linear models adjusting pedigree, gender, study center, and the first principal component of cellular purity and population structure. Lower and upper bars represent the lower and higher values of 95% confidence intervals for each beta. Red line is the fitted line for the pattern of all 13 CpG sites.

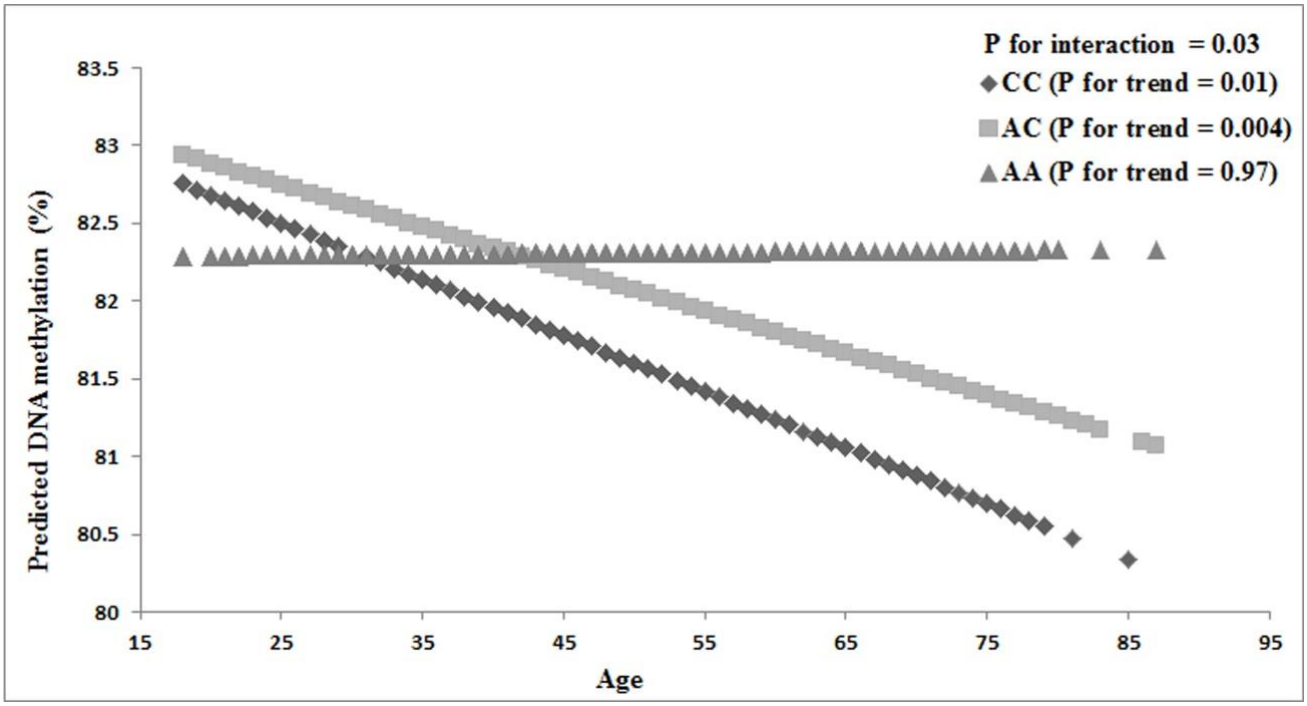


Figure 3.3 Interaction between rs405509 and age for methylation of CpG 3 (cg01032398) in GOLDN.
Predicted methylation level of cg01032398 by genotype of rs405509 were plotted against age, adjusted for pedigree, gender, center, and the first principal component of cellular purity and population structure. *P* values indicate the statistical significance of the adjusted interaction term and adjusted regression coefficients in the regression line corresponding to three genotype groups of rs405509 (diamond for CC, square for AC, and triangle for AA).

CHAPTER 4

Fatty acids, SNPs, and DNA methylation at *IL6* locus

4.1 Abstract

N3 polyunsaturated fatty acids (N3 PUFAs) ameliorate inflammation with regulations on interleukin-6 (*IL6*). However, the molecular mechanism for this regulation is unclear and DNA methylation may represent a potential one. Using both population data from the Genetics of Lipid Lowering Drugs and Diet Network study (GOLDN) and cell lines data from the Encyclopedia of DNA Elements (ENCODE) consortium, we explored the potential interplay among DNA methylation , single nucleotide polymorphisms (SNPs), and N3 PUFAs within the *IL6* locus. Our findings suggest a CpG site (cg01770232) within *IL6* may be a potentially functional methylation site based on its significant correlation with IL-6 plasma concentration in GOLDN ($P = 0.03$) and gene expression in ENCODE ($R = 0.8$, $P = 0.0003$). Erythrocyte level of total N3 PUFAs was associated with cg01770232 methylation ($P = 0.007$) and plasma IL-6 concentration ($P = 0.02$). SNP rs2961298 was shown to have significant association with cg01770232 methylation ($P = 2.55\times10^{-7}$), as well as to modify the association between N3 PUFAs and cg01770232 methylation (P for interaction = 0.02). Higher total N3 PUFAs was associated with lower cg01770232 methylation in the heterozygotes (P for interaction = 0.04) but not the homozygotes of SNP rs2961298 ($P > 0.05$). To conclude, N3 PUFAs may affect *IL6* through methylation of its potential functional site cg01770232 in the promoter region, and this effect may be further modified by *IL6* SNP rs2961298. These findings may increase mechanistic understanding of the link between N3 PUFAs and *IL6*.

Key words:

n3 polyunsaturated fatty acids, interleukin-6, DNA methylation, SNPs, epidemiology

4.2 Introduction:

The interleukin-6 (*IL6*) gene has been demonstrated to be regulated by DNA methylation, and increased understanding of *IL6* methylation may be relevant to amelioration of diseases that are mediated by inflammation.[1-5] For example, methylation of the *IL6* promoter region was found to be associated with inflammation-related diseases, such as rheumatoid arthritis,[1] chronic periodontitis,[3] and cancer.[4] In addition to its etiologic relevance in inflammatory processes, methylation of the promoter region was also found to affect gene expression of *IL6*.[1-4] The observed effect of *IL6* methylation on its gene expression may be related to the presence of potential binding sites for the methyl-CpG-binding protein 2 (MeCP2), which are located from positions -666 to -426 relative to the transcription start site.[2] The functional relevance of *IL6* methylation sites is not limited to the promoter region; for example, methylation of the second intron of *IL6* was shown to be correlated with a biomarker of DNA damage. [5]

As evidence of the connections between IL-6 concentration, *IL6* methylation, and human disease accumulates, research to identify environmental factors that alter *IL6* methylation will be particularly informative. N3 polyunsaturated fatty acids (PUFAs) may represent an environmental factor that modifies *IL6* methylation, based on their demonstrated amelioration of systematic inflammation through reduction of IL-6 concentration, of which the higher value exhibit deleterious effects.[6, 7] One cross-sectional study found that erythrocyte N3 PUFAs were negatively associated with plasma IL-6 concentration.[8] Randomized clinical trials showed reduction of plasma IL-6 concentration by N3 PUFA supplementation. For example, after a 12-week fish oil supplementation, IL-6 production by peripheral blood mononuclear cells was significantly decreased.[9] Supplementation with docosahexaenoic acid (DHA) for 13 weeks decreased the concentration of IL-6 by ~20%.[10] Also, supplementation with both

eicosapentaenoic acid (EPA) and DHA for 8 weeks was found to reduce not only the plasma concentration of IL-6 but also its gene expression in adipose tissue.[11]

Adding to the complexity of environmental modulation of *IL6* methylation is the role of genetic sequence variation. For example, single nucleotide polymorphisms (SNPs) were found to affect DNA methylation across the whole genome.[12] Moreover, *IL6* SNPs were shown to interact with different environmental factors, including diet,[13] smoking,[14, 15] and social position to determine plasma IL-6 concentration.[16] This evidence demonstrates the modifying role of *IL6* sequence variants on the effect of environmental factors on plasma IL-6 concentration. Therefore, the objective of this study is to explore the relationship between N3 PUFAs, *IL6* SNPs, and *IL6* methylation.

To achieve this objective, we first explored potential functional methylation sites within the *IL6* locus based on DNA methylation-expression relationship in datasets from the Encyclopedia of DNA Elements (ENCODE) consortium, and the DNA methylation-protein relationship in data from participants of the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study. We further investigated the association between N3 PUFAs and methylation levels of the identified functional CpG site, and evaluated whether the observed associations could be further modified by *IL6* SNPs in GOLDN participants.

4.3 Methodology:

Study population:

The GOLDN study, designed to evaluate genetic factors that modulate lipid responses to diet and fenofibrate treatment, recruited participants from the National Heart, Lung, and Blood Institute Family Heart Study.[17] The study design and methodology were described previously.[18] The protocol for this study was approved by the Human Studies Committee of Institutional Review Board at the University of Minnesota, University of Utah, and Tufts University/New England Medical Center. Written informed consent was obtained from all participants. Validated diet history questionnaire was applied to collect dietary intakes.[19] Total hours of physical activity were obtained by relevant questionnaire.[20]

Biochemical measurements:

Blood samples from each participant were collected, stored frozen at -70 degrees C, and analyzed at the same time to eliminate inter-assay variability. IL-6, interleukin-2 soluble receptor (IL2sR)-α, tumor necrosis factor (TNF)-α, and monocyte chemoattractant protein-1 (MCP-1) were measured using quantitative sandwich enzyme immunoassay techniques (ELISA kit assays, R&D System Inc., Minneapolis, MN) as described previously.[19] High-sensitivity C-reactive protein (hsCRP) was measured using a latex particle enhanced immunoturbidimetric assay (Kamiya Biomedical Company, Seattle, WA) as described previously.[21] Plasma adiponectin was measured using competitive RIA (Linco Research, St Charles, MO, USA) as described previously.[22] Fatty acids in the erythrocyte membranes were measured by a capillary Varian CP7420 100-m column with a Hewlett Packard 5890 gas chromatograph equipped with a HP6890A autosampler.[23] The measurements of fatty acids were reliable and have been validated against a diet history questionnaire.[24, 25]

Genotyping and DNA methylation

IL6 is a gene with five exons (~5 kb) and located on chromosome 7 (**Figure 4.1 A**). Approximately 60 kb upstream of transcription start site of *IL6* gene, there was a CpG island of 598 bp in length. Since no known genes are located between the CpG island and *IL-6* gene, the region of interest for this study, referred as *IL6* locus, was defined as the region from 1 kb upstream of the CpG island to 1 kb downstream of the 3' unstranslated region (UTR) of *IL6* gene. Within the *IL6* locus, there are 39 SNPs with measurements of genotypes. Two CpG sites, referred to as cg01770232 and cg26061582, were included into the analysis because they were located in the potential binding site for the MeCP2.[2]

Genotypes for *IL6* SNPs were obtained using the Affymetrix Genome-Wide Human SNP array 6.0 (Affymetrix, Santa Clara, CA) with the genomic DNA extracted from blood samples using Gentra Puregene Blood Kits (Gentra Systems, Inc., Minneapolis, MN).

CD4+ T-cells were obtained from frozen buffy coat samples isolated from peripheral blood using positive selection (Invitrogen, Grand Island, NY) followed by sorting of subsets by flow cytometry (FACSAriaII, BD Biosciences, San Jose, CA). Methylation of CpG sites within *IL6* locus was measured using Infinium Human Methylation 450K BeadChip (Illumina, San Diego, CA) [26] with 500 ng sodium bisulfite treated DNA (Zymo Research Corporation, Irvine, CA) extracted from CD4+ T-cells (QIAGEN, Germantown, MD). For each CpG site, a beta score and a detection p-value were generated through the analysis of the intensity files with Illumina's GenomeStudio. The beta score represented the proportion of total signal from the methylation-specific probe or color channel. The detection p-value was defined as the probability that the total intensity for a given probe falls within the background signal intensity. During quality

control, those CpG probes with detection p-values greater than 0.01 and with more than 10% of samples that failed to yield adequate intensity were eliminated, and those samples with more than 1.5% missing data points across ~470,000 autosomal CpGs were removed. The quality control procedures yielded 21 CpG sites within the *IL6* locus, with two of them (cg01770232 and cg26061582) included into the analysis due to their location within the potential binding site for the MeCP2.[2]

Gene expression and DNA methylation in ENCODE

Methylation levels of CpG sites within the *IL6* locus from all 17 cell lines, with available datasets of both methylation and gene expression, were downloaded (01-09-2014) from UCSC genome browser HAIB Methyl450 track (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeHaibMethyl450/). The methylation status of a CpG site was assayed by single base-pair extension with a Cy3 or Cy5 labeled nucleotide on oligo-beads specific for the methylated or unmethylated state. The methylation level of each CpG site was represented by a score, which was 1000 times the proportion of the intensity value from the methylated bead type of the sum of the intensity values from both methylated and unmethylated bead type plus 100. The range of the score was from 0 to 1000.

Gene expression data for *IL6* in all 17 cell lines with available datasets of both methylation and gene expression in ENCODE were downloaded (01-09-2014) from UCSC genome browser Duke Affymetrix Exon Array track (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeDukeAffyExon/). All probes linked to *IL6* gene were aggregated and the expression of *IL6* was represented by a score,

which was 100 times the signal value linearly scaled for that particular cell type and with ranges from 0 to 1000.

Statistical methods:

Pearson correlation analysis was conducted between methylation of the candidate CpG sites and *IL6* gene expression across different cell lines in ENCODE. In GOLDN, the population for analysis consisted of 877 individuals (473 men and 404 women) after excluding those who reported taking hormone replacement therapies (n=113) because of their mixed effects on IL-6.[27, 28] χ^2 tests, ANOVA and ANCOVA analyses were conducted to examine the differences in population characteristics and potential confounding factors by erythrocyte total N3 PUFAs, categorized in quartiles. Generalized linear models were applied to test the main associations and interactions among the methylation levels of both CpG sites, plasma IL-6 concentration, genotypes of *IL6* SNPs, and erythrocyte N3 PUFAs, including total, EPA, and DHA. The methylation levels of CpG sites were represented as quartiles for the exposure variables and as continuous variables for the outcome variable. The potential confounding factors adjusted by in this study included pedigree, principal components of cellular purity [29] and population structure, sex, study center, smoking, alcohol intake, total energy intake, physical activity, vitamin B12 intake, folate intake, acute inflammatory conditions (infection or fever), chronic diseases known to affect IL-6 such as abdominal obesity,[30] cardiovascular disease (CVD),[6, 7] diabetes,[31] and hypertension,[32] and other inflammatory markers known to affect IL-6, including hsCRP,[24] TNF-α,[33] MCP-1,[34] IL2sR-α,[24] and adiponectin.[35] Log transformation was performed for those variables not following a normal distribution. All data were analyzed using SAS (version 9.3 for Windows; SAS Institute, Inc. Cary, NC, USA). A two-tail *P*-value of <0.05 was considered statistically significant.

4.4 Results:

(1) Population characteristics

Population characteristics were compared across quartiles of erythrocyte total N3 PUFAs (**Table 4.1**). Compared to the individuals in the quartiles with lower total N3 PUFAs, those in the quartiles with higher level tended to be older, consume less total energy intake, and have lower concentration of MCP-1 ($P < 0.05$). Also, the quartiles with higher levels of N3 PUFAs tended to contain less smokers, and more individuals with abdominal obesity than those with lower levels ($P < 0.05$). Finally, individuals in the quartiles with higher N3 PUFAs tended to have lower plasma IL-6 concentration ($P = 0.02$).

(2) Cg01770232 may be the functional DNA methylation site within *IL6* locus

Methylation level of cg01770232 was positively correlated with *IL6* gene expression across different cell lines in ENCODE (R = 0.75 and $P = 0.0005$) (**Figure 4.1 B**). In GOLDN, methylation level of cg01770232 was also positively associated with plasma IL-6 concentration ($P = 0.03$) (**Figure 4.1 C**). We did not find any significance for cg26061582 (data not shown).

(3) N3 PUFAs are associated with both plasma IL-6 and methylation level of cg01770232

In order to indirectly test whether the known effect of N3 PUFAs on IL-6 was mediated through changes in DNA methylation, we compared associations of N3 PUFAs with the methylation level of cg01770232, and also with plasma IL-6 concentration in GOLDN (**Figure 4.2**). Total N3 PUFAs and DHA were significantly negatively associated with plasma IL-6 ($P < 0.05$), while the negative correlation with EPA did not reach statistical significance. In parallel with the association pattern observed between total N3 PUFAs, EPA and DHA for the outcome of plasma IL-6 concentrations, the same fatty acids (total N3 PUFAs, EPA, and DHA) were also negatively associated with methylation level of cg01770232 ($P < 0.05$).

(4) Rs2961298 within *IL6* locus is associated with methylation level of cg01770232

Evidence for associations between all 39 SNPs within *IL6* locus and the methylation level of cg01770232 is presented in **Table 4.2** and **Figure 4.3**. Rs2961298, a SNP close to the CpG island upstream of *IL6* gene, exhibited the most robust association with the methylation level of cg01770232 ($P = 2.55\times10^{-7}$).

(5) Rs2961298 interacts with N3 PUFAs to modulate methylation level of cg01770232

Because of the identified associations of both SNP rs2961298 and N3 PUFAs with the methylation level of cg01770232, the potential interactions between fatty acids and SNPs were further explored (**Figure 4.4**). Rs2961298 was shown to significantly modify the association of total N3 PUFAs (*P* for interaction = 0.02), EPA (*P* for interaction = 0.01), and DHA (*P* for interaction = 0.05) with the methylation level of cg01770232. Our findings suggest that N3 PUFAs were associated with lower methylation level of cg01770232 in the rs2961298 heterozygotes (AC) (*P* for trend = 0.04, 0.06, and 0.05 for total N3 PUFAs, EPA, and DHA, respectively), but not in the homozygotes (AA and CC).

4.5 Discussion

In the current study, we identified a potentially functional CpG site within the *IL6* locus and its association with N3 PUFAs in erythrocyte membrane, which was further modified by *IL6* SNPs. This is the first study to explore the effect of N3 PUFAs on *IL6* methylation and relevant modification by genetic variants within a large population. Based on the fact that most N3 PUFAs were taken from diet, our findings imply that *IL6* genotype alters the relationship between diet and methylation for the outcome of plasma IL-6, and this mechanism may underlie previously reported *IL6* genotype × diet interactions for health-related outcomes.

Our finding that methylation of CpG site cg01770232 is associated with *IL6* gene expression is consistent with previous studies, however, the direction of this effect in our study differs from that of earlier work.[1-4] In previous studies, methylation of *IL6* gene promoter region has been shown to have a negative correlation with its gene expression,[1-4] and MeCP2 was found to have gene silencing effect on *IL6* [2] whereas we observed consistently positive associations between methylation and the outcomes of both gene expression and plasma IL-6 concentration. Plausible explanations to account for the lack of agreement between the current findings and those of previous studies include differences in methylation measurement methodologies and the cell types used for methylation. Specifically, previous studies measured the methylation of hundreds of nucleotides within the promoter region, rather than the single nucleotide in our study. It is possible that the methylation of cg01770232 is negatively correlated with the methylation of the hundreds of promoter regions nucleotides investigated in earlier studies. Moreover, our decision to investigate cg01770232 is based on not only its location within the promoter region of *IL6*, but also on its specific position at -611 bp to the transcription start site, which is just within the region predicted to act as binding sites for the MeCP2. With respect to differences in cells types used for the current study vs. previous studies, the silencing

effect of MeCP2 in earlier studies was observed in cell types that did not actually express *IL6*, in which MeCP2 was bound to *IL6*. In contrast to these earlier studies, we used data generated from immune cells, in which *IL6* is expressed and produced. We hypothesize that in the IL-6 producing cell types, the methylated cg01770232 site does not bind with MeCP2 to inhibit *IL6* expression, but instead binds with other factors to induce expression.

Potential mechanisms to account for the observed association between N3 PUFAs and *IL6* methylation may be related to absence of MeCP2 binding to the methylated CpG site in the *IL6* expressing cells, as mentioned above. This mechanistic possibility is supported by previous studies reporting protective effects of N3 PUFAs on Rett Syndrome, a neurodevelopmental disorder mainly caused by mutations in *MECP2* gene.[36] In Rett Syndrome, sporadic mutations in *MECP2* lead to dysfunctional binding of MeCP2 protein with methylated CpG sites.[37] N3 PUFAs have been shown to partially rescue clinical symptoms of patients affected with Rett Syndrome,[38] by normalizing 16 disrupted acute phase proteins,[36] further indicating the anti-inflammatory effects of N3 PUFAs.

An important strength of the current study is that it provides evidence for potential mechanisms by which environmental factors (in this case diet) are modulating genotype-based variability in disease biomarkers. While previous studies reported the interactions between sequence variants and environmental factor that modulate plasma IL-6 concentration,[13-16] our work suggests that the modification effect of *IL6* SNPs on the relationship between N3 PUFAs may occur through *IL6* methylation. Specifically, SNP rs2961298 interacted significantly with total N3 PUFAs, EPA, and DHA to modulate methylation status of the identified functional CpG site, cg01770232. When evaluated in the same human population, N3 PUFAs were found to be

associated with methylation level of cg01770232 in a genotype-specific manner. The potential mechanism for the observed interactions may be related to different genetic structures. This is supported by our finding of the significant differences in methylation of cg01770232 by genotype of SNP rs2961298, indicating that different genotypes may lead to different genetic structures, which will be further differentially modified by the environmental factors, either methyl-donors or N3 PUFAs as suggested by our study.

Our study was limited by its cross-sectional design, in which only association rather than causality can be established. Further longitudinal and intervention studies are needed in order to confirm and solidify these findings. Additionally, lack of replication and inability to directly assess relationships with gene expression in GOLDN should also be taken into consideration.

In summary, we identified a potentially functional methylation site (cg01770232) within the *IL6* locus, which is not only associated with *IL6* gene expression according to ENCODE, but also associated with plasma IL-6 concentration in the GOLDN population. We further found that erythrocyte membrane levels of N3 PUFAs were associated with the methylation level of cg01770232 and the plasma concentration of IL-6. Moreover, the observed association between N3 PUFAs and methylation of cg01770232 was further modified by SNP rs2961298. Our findings may lead to novel mechanistic explanations of the link between N3 PUFAs and IL-6, and may also provide a basis for investigating genotype-altered methylation as a mediator of gene-diet interactions.

4.6 References

1. Nile, C.J., et al., *Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis.* Arthritis Rheum, 2008. **58**(9): p. 2686-93.
2. Dandrea, M., et al., *MeCP2/H3meK9 are involved in IL-6 gene silencing in pancreatic adenocarcinoma cell lines.* Nucleic Acids Res, 2009. **37**(20): p. 6681-90.
3. Stefani, F.A., et al., *Expression, polymorphism and methylation pattern of interleukin-6 in periodontal tissues.* Immunobiology, 2013. **218**(7): p. 1012-7.
4. Tekpli, X., et al., *DNA methylation at promoter regions of interleukin 1B, interleukin 6, and interleukin 8 in non-small cell lung cancer.* Cancer Immunol Immunother, 2013. **62**(2): p. 337-45.
5. Peluso, M., et al., *DNA methylation differences in exposed workers and nearby residents of the Ma Ta Phut industrial estate, Rayong, Thailand.* Int J Epidemiol, 2012. **41**(6): p. 1753-60; discussion 1761-3.
6. Cesari, M., et al., *Inflammatory markers and onset of cardiovascular events: results from the Health ABC study.* Circulation, 2003. **108**(19): p. 2317-22.
7. Su, D., et al., *Association between serum interleukin-6 concentration and mortality in patients with coronary artery disease.* Mediators Inflamm, 2013. **2013**: p. 726178.
8. Farzaneh-Far, R., et al., *Inverse association of erythrocyte n-3 fatty acid levels with inflammatory biomarkers in patients with stable coronary artery disease: The Heart and Soul Study.* Atherosclerosis, 2009. **205**(2): p. 538-43.
9. Trebble, T., et al., *Inhibition of tumour necrosis factor-alpha and interleukin 6 production by mononuclear cells following dietary fish-oil supplementation in healthy men and response to antioxidant co-supplementation.* Br J Nutr, 2003. **90**(2): p. 405-12.
10. Kelley, D.S., et al., *DHA supplementation decreases serum C-reactive protein and other markers of inflammation in hypertriglyceridemic men.* J Nutr, 2009. **139**(3): p. 495-501.
11. Itariu, B.K., et al., *Long-chain n-3 PUFAs reduce adipose tissue and systemic inflammation in severely obese nondiabetic patients: a randomized controlled trial.* Am J Clin Nutr, 2012. **96**(5): p. 1137-49.
12. Zhi, D., et al., *SNPs located at CpG sites modulate genome-epigenome interaction.* Epigenetics, 2013. **8**(8): p. 802-6.
13. Kanoni, S., et al., *Assessment of gene-nutrient interactions on inflammatory status of the elderly with the use of a zinc diet score--ZINCAGE study.* J Nutr Biochem, 2010. **21**(6): p. 526-31.
14. Shin, K.K., et al., *Influence of the IL-6 -572C>G polymorphism on inflammatory markers according to cigarette smoking in Korean healthy men.* Cytokine, 2007. **39**(2): p. 116-22.
15. Sunyer, J., et al., *Interaction between smoking and the interleukin-6 gene affects systemic levels of inflammatory biomarkers.* Nicotine Tob Res, 2009. **11**(11): p. 1347-53.
16. Sanderson, S.C., et al., *Association between IL6 gene variants -174G>C and -572G>C and serum IL-6 levels: interactions with social position in the Whitehall II cohort.* Atherosclerosis, 2009. **204**(2): p. 459-64.
17. Higgins, M., et al., *NHLBI Family Heart Study: objectives and design.* Am J Epidemiol, 1996. **143**(12): p. 1219-28.
18. Corella, D., et al., *The -256T>C polymorphism in the apolipoprotein A-II gene promoter is associated with body mass index and food intake in the genetics of lipid lowering drugs and diet network study.* Clin Chem, 2007. **53**(6): p. 1144-52.
19. Tsai, M.Y., et al., *Effect of influenza vaccine on markers of inflammation and lipid profile.* J Lab Clin Med, 2005. **145**(6): p. 323-7.
20. Corella, D., et al., *A high intake of saturated fatty acids strengthens the association between the fat mass and obesity-associated gene and BMI.* J Nutr, 2011. **141**(12): p. 2219-25.

21. Aslibekyan, S., et al., *A genome-wide association study of inflammatory biomarker changes in response to fenofibrate treatment in the Genetics of Lipid Lowering Drug and Diet Network.* Pharmacogenet Genomics, 2012. **22**(3): p. 191-7.
22. Shen, J., et al., *Interleukin1beta genetic polymorphisms interact with polyunsaturated fatty acids to modulate risk of the metabolic syndrome.* J Nutr, 2007. **137**(8): p. 1846-51.
23. Cao, J., et al., *Incorporation and clearance of omega-3 fatty acids in erythrocyte membranes and plasma phospholipids.* Clin Chem, 2006. **52**(12): p. 2265-72.
24. Kabagambe, E.K., et al., *Smoking, inflammatory patterns and postprandial hypertriglyceridemia.* Atherosclerosis, 2009. **203**(2): p. 633-9.
25. Kabagambe, E.K., et al., *Erythrocyte fatty acid composition and the metabolic syndrome: a National Heart, Lung, and Blood Institute GOLDN study.* Clin Chem, 2008. **54**(1): p. 154-62.
26. Absher, D.M., et al., *Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations.* PLoS Genet, 2013. **9**(8): p. e1003678.
27. Georgiadou, P. and E. Sbarouni, *Effect of hormone replacement therapy on inflammatory biomarkers.* Adv Clin Chem, 2009. **47**: p. 59-93.
28. Ferrari, S.L., et al., *Interactions of interleukin-6 promoter polymorphisms with dietary and lifestyle factors and their association with bone mass in men and women from the Framingham Osteoporosis Study.* J Bone Miner Res, 2004. **19**(4): p. 552-9.
29. Frazier-Wood, A.C., et al., *Methylation at CPT1A locus is associated with lipoprotein subfraction profiles.* J Lipid Res, 2014.
30. Brydon, L., et al., *Stress-induced cytokine responses and central adiposity in young women.* Int J Obes (Lond), 2008. **32**(3): p. 443-50.
31. Duncan, B.B., et al., *Low-grade systemic inflammation and the development of type 2 diabetes: the atherosclerosis risk in communities study.* Diabetes, 2003. **52**(7): p. 1799-805.
32. Zhang, W., et al., *Interleukin 6 underlies angiotensin II-induced hypertension and chronic renal damage.* Hypertension, 2012. **59**(1): p. 136-44.
33. Schindler, R., et al., *Correlations and interactions in the production of interleukin-6 (IL-6), IL-1, and tumor necrosis factor (TNF) in human blood mononuclear cells: IL-6 suppresses IL-1 and TNF.* Blood, 1990. **75**(1): p. 40-7.
34. Kim, C.S., et al., *Circulating levels of MCP-1 and IL-8 are elevated in human obese subjects and associated with obesity-related parameters.* Int J Obes (Lond), 2006. **30**(9): p. 1347-55.
35. Abke, S., et al., *Adiponectin-induced secretion of interleukin-6 (IL-6), monocyte chemotactic protein-1 (MCP-1, CCL2) and interleukin-8 (IL-8, CXCL8) is impaired in monocytes from patients with type I diabetes.* Cardiovasc Diabetol, 2006. **5**: p. 17.
36. De Felice, C., et al., *Effects of omega-3 polyunsaturated fatty acids on plasma proteome in Rett syndrome.* Mediators Inflamm, 2013. **2013**: p. 723269.
37. Ausio, J., A.M. Paz, and M. Esteller, *MeCP2: the long trip from a chromatin protein to neurological disorders.* Trends Mol Med, 2014.
38. De Felice, C., et al., *The role of oxidative stress in Rett syndrome: an overview.* Ann N Y Acad Sci, 2012. **1259**: p. 121-35.

Table 4.1. Population characteristics of GOLDN

Variable		Erythrocyte N3 PUFAs				*P*
		Q1 (n = 212)	Q2 (n = 212)	Q3 (n = 212)	Q4 (n = 212)	
Erythrocyte N3 PUFAs	Total N3 PUFAs (% total fatty acids)[a]	4.65	5.25	5.96	7.08	
	EPA (% total fatty acids)[b]	0.39 (0.09)	0.45 (0.10)	0.53 (0.12)	0.80 (0.48)	<0.0001
	DHA (% total fatty acids)[b]	2.10 (0.30)	2.59 (0.30)	3.10 (0.32)	4.12 (0.73)	<0.0001
Demographics	Age (y)[b]	42 (14)	45 (16)	51 (15)	57 (15)	<0.0001
	Men (n)[c]	124 (58)	119 (56)	107 (50)	109 (51)	0.29
Lifestyle parameters	Current smoker (n)[c]	27 (13)	24 (11)	12 (6)	4 (2)	0.0003
	Drinker (n)[c]	114 (54)	112 (53)	93 (44)	93 (44)	0.06
	Total energy intake (kcal/day)[b]	2313.95 (1496.83)	2322.94 (1325.29)	2182.88 (1193.17)	1898.38 (836.63)	0.001
	Vitamin B12 Intake (mcg)[b]	5.62 (3.74)	5.84 (3.76)	5.56 (3.33)	5.18 (3.47)	0.29
	Folate Intake (mcg)[b]	423.61 (235.84)	416.24 (263.21)	421.98 (229.78)	404.28 (198.43)	0.83
Disease status	Infection or fever (n)[d]	2 (1)	5 (2)	4 (2)	2 (1)	0.45
	Abdominal obesity (n)[d]	86 (41)	121 (57)	114 (54)	108 (51)	0.01
	Hypertension (n)[d]	30 (14)	52 (25)	54 (26)	83 (39)	0.40
	Diabetes (n)[d]	8 (4)	19 (9)	12 (6)	25 (12)	0.98
	CVD (n)[d]	5 (2)	10 (5)	11 (5)	24 (11)	0.29
Other inflammatory markers	IL6 (pg/ml)[e]	1.75 (1.13)	1.75 (1.12)	1.65 (1.11)	1.57 (1.12)	0.02
	hsCRP (mg/dl)[f]	0.11 (1.09)	0.15 (1.11)	0.11 (1.09)	0.12 (1.09)	0.69
	TNF-α (pg/ml)[f]	2.92 (1.03)	3.08 (1.03)	2.91 (1.03)	3.06 (1.05)	0.57
	MCP-1 (pg/ml)[f]	212.66 (1.02)	210.69 (1.03)	200.22 (1.02)	198.26 (1.02)	0.006
	Adiponectin (ng/ml)[f]	7762.81 (1.05)	6460.1 (1.05)	6930.59 (1.04)	6915.36 (1.05)	0.22
	IL2sR-α (pg/ml)[f]	1000.05 (1.03)	1020.96 (1.03)	959.39 (1.03)	964.97 (1.03)	0.16

[a]Data are medians of each quartile of total N3 PUFAs;
[b]Data are means (standard deviations) within each quartile of total N3 PUFAs and *P* values were obtained from ANOVA test;
[c]Data are n (%) within each quartile of total N3 PUFAs and *P* values were obtained from Chi-square test;
[d]Data are means (standard deviations) within each quartile of total N3 PUFAs and *P* values were obtained from logistic regression adjusting for the first four principal components for cellular purity and population structure, age, sex, center, smoking, alcohol intake, total energy intake, physical activity, intake of vitamin B12 and folate.
[e]Data are least squared adjusted means (standard errors) within each quartile of total N3 PUFAs and *P* values were obtained from ANCOVA test adjusting for the first four principal components for cellular purity and population structure, age, sex, center, pedigree, smoking, alcohol intake, total energy intake, physical activity, intake of vitamin B12 and folate, acute inflammatory conditions (infection or fever), chronic disease status (including abdominal obesity, CVD, diabetes, and hypertension), and other inflammatory markers (including hsCRP, TNFα, MCP-1, IL2sR-α, and adiponectin).
[f]Data are least square adjusted means (standard errors) within each quartile of total N3 PUFAs, *P* values were obtained from ANCOVA test adjusting for the first four principal components for cellular purity and population structure, age, sex, center, pedigree, smoking, alcohol intake, total energy intake, physical activity, intake of vitamin B12 and Folate.

Table 4.2. Genetic associations for methylation of cg01770232 in GOLDN[a]

SNP	Distance to TSS[b] (bp)	Effect Allele	Effect Allele Frequency	Beta	SEM	*P*
rs2905321	-62179	G	0.38	0.005	0.01	0.72
rs2157958	-60059	C	0.31	0.02	0.01	0.11
rs2961298	-58675	A	0.28	0.09	0.01	2.55E-07
rs2961299	-58596	G	0.28	0.05	0.01	0.001
rs2961300	-57982	C	0.28	0.04	0.01	0.003
rs2905324	-57763	C	0.32	0.07	0.01	4.71E-06
rs1006001	-57394	A	0.04	0.21	0.03	2.00E-05
rs2961304	-53549	T	0.04	0.25	0.05	7.58E-05
rs2961309	-43862	A	0.04	0.21	0.03	2.83E-05
rs2961310	-42359	T	0.29	0.08	0.01	5.84E-07
rs1548418	-41352	C	0.04	0.25	0.05	7.58E-05
rs1476483	-35567	C	0.20	0.06	0.02	0.003
rs2961312	-31812	T	0.04	0.20	0.03	2.68E-05
rs4722166	-28004	G	0.36	-0.02	0.02	0.17
rs4321884	-25307	A	0.47	-0.04	0.02	0.02
rs7383869	-18576	T	0.49	0.005	0.02	0.82
rs6946864	-14817	C	0.16	0.05	0.02	0.02
rs6969502	-14340	T	0.16	0.05	0.02	0.02
rs6952003	-14061	A	0.27	-0.04	0.02	0.03
rs10156056	-12678	C	0.11	-0.05	0.02	0.01
rs4719711	-11078	G	0.46	0.01	0.02	0.59
rs1404008	-10766	A	0.46	0.01	0.02	0.59
rs6963444	-9611	G	0.03	0.05	0.06	0.43
rs6963591	-9496	T	0.45	0.01	0.02	0.69
rs7801617	-8684	T	0.10	-0.01	0.03	0.58
rs7805828	-8204	A	0.43	-0.003	0.02	0.90
rs1880241	-7297	G	0.50	0.04	0.02	0.06
rs1880242	-7159	C	0.46	0.01	0.02	0.59
rs10499563	-6278	C	0.24	-0.03	0.02	0.06
rs12700386	-3757	C	0.19	0.05	0.02	0.02
rs2069824	-1534	G	0.07	-0.05	0.02	0.03
rs2069827	-1310	T	0.08	-0.06	0.02	0.005
rs1800795	-121	C	0.42	-0.03	0.02	0.08
rs2069837	1261	G	0.09	-0.04	0.02	0.08
rs2066992	1483	T	0.04	0.24	0.02	1.15E-05
rs2069840	1806	G	0.35	0.005	0.02	0.84
rs1548216	3007	C	0.03	0.06	0.06	0.31
rs2069852	5494	T	0.03	0.24	0.03	1.83E-04
rs7808204	8505	C	0.02	0.07	0.07	0.33

[a]Model adjusted for the first four principal components for celluar purity and population structure, age, sex, center, pedigree, smoking, alcohol intake, total energy intake, physical activity, intake of vitamin B12 and folate, acute inflammatory conditions (infection or fever), chronic diseases (including abdominal obesity, CVD, diabetes, and hypertension), other inflammatory markers (MCP-1, adiponectin, CRP, TNFα, and IL2sR-α).
[b]TSS represents transcription start site.

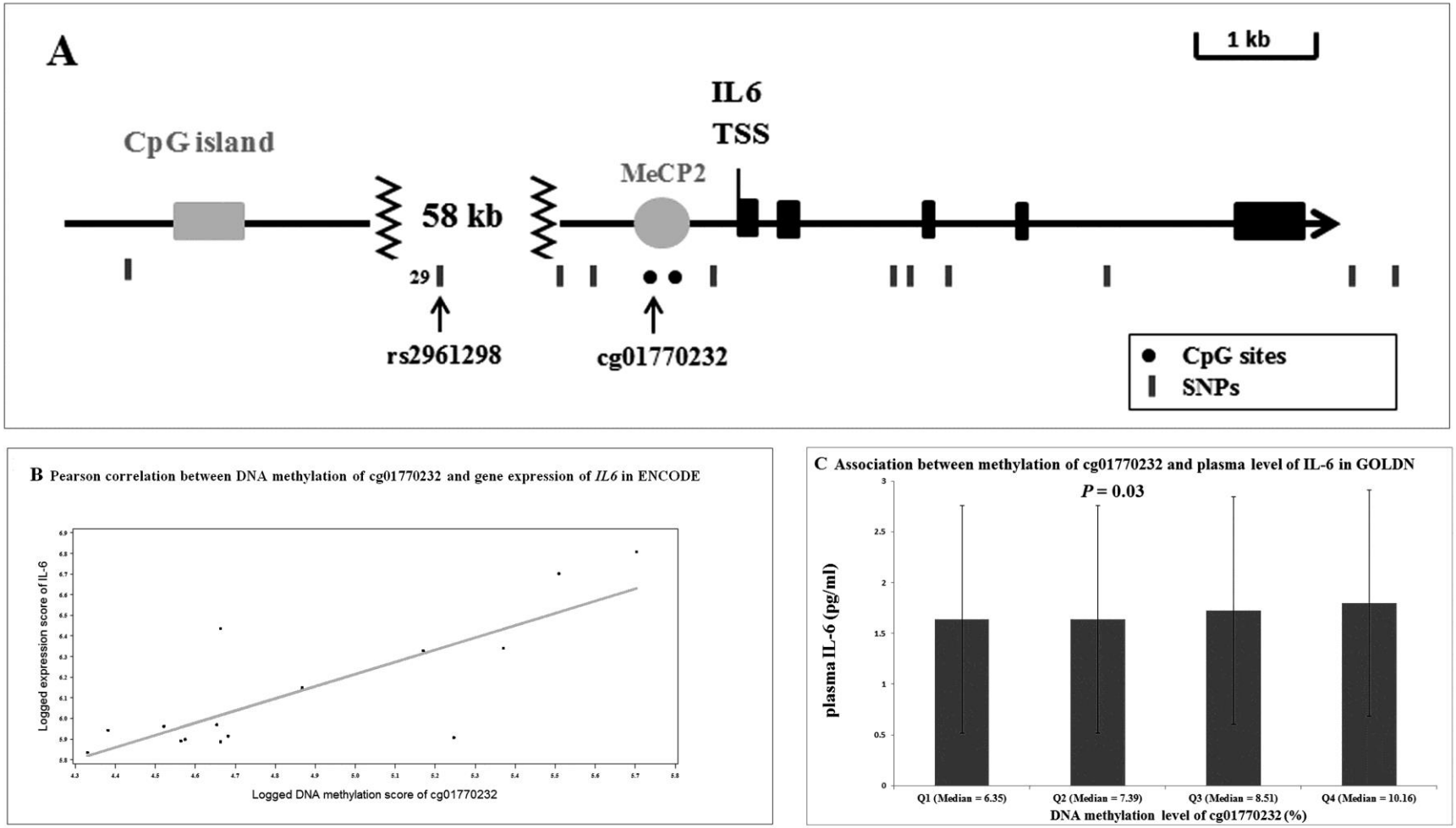


Figure 4.1. CpG site cg01770232 is the functional methylation site within the *IL6* locus. Genomic structure of *IL6* locus is shown in A, in which exons, CpG island, potential binding site for MeCP2, CpG sites, and SNPs are represented by black box, gray box, gray circle, black dots, and black bars, respectively. Pearson correlation between methylation level of cg01770232 and *IL6* gene expression across different cell lines in ENCODE is displayed in B, in which each black dot represents one cell line and the gray line represents the corresponding regression line, and the correlation coefficient (*R*) and its corresponding *P* value are shown at the bottom right corner. Association of quartiles of methylation level of cg01770232 with plasma IL-6 concentration (geometric means ± 95% CI) in GOLDN is shown in C. *P* value was obtained from the general linear model adjusting for first four principal components for cellular purity and population structure, age, sex, center, pedigree, smoking, alcohol intake, total energy intake, physical activity, intake of vitamin B12 and folate, acute inflammatory conditions (infection or fever), chronic diseases (abdominal obesity, CVD, diabetes, and hypertension), other inflammatory markers (MCP-1, adiponectin, hsCRP, TNF-α, and IL2sR-α).

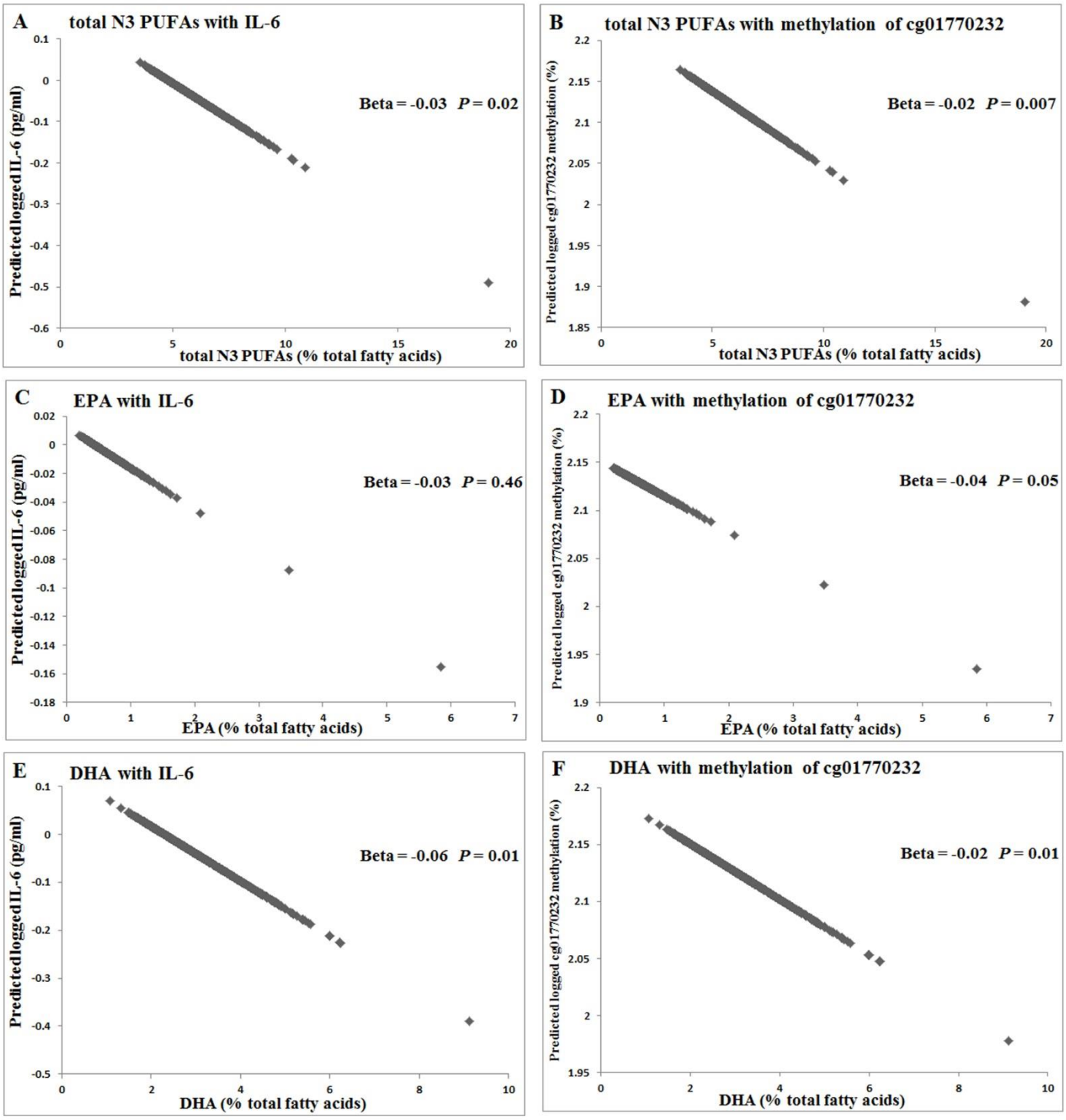


Figure 4.2 Associations of red blood cell N3 PUFAs with plasma IL-6 and DNA methylation level of cg01770232 in GOLDN. Predicted plasma concentration of IL-6 (log transformed) was plotted against total N3 PUFAs (A), EPA (C), and DHA (E), and predicted methylation level of cg01770232 was also plotted against total N3 PUFAs (B), EPA (D), and DHA (F). Betas and *P* values represent the regression coefficients and statistical significance, respectively, both of which were obtained from the general linear model adjusting for first four principal components for celluar purity and population structure, age, sex, center, pedigree, smoking, drinking, total energy intake, physical activity, intake of vitamin B12 and folate, acute inflammatory conditions (infection or fever), chronic disease status (including abdominal obesity, CVD, diabetes, and hypertension), other inflammatory markers (MCP-1, adiponectin, hsCRP, TNF-α, and IL2sR-α).

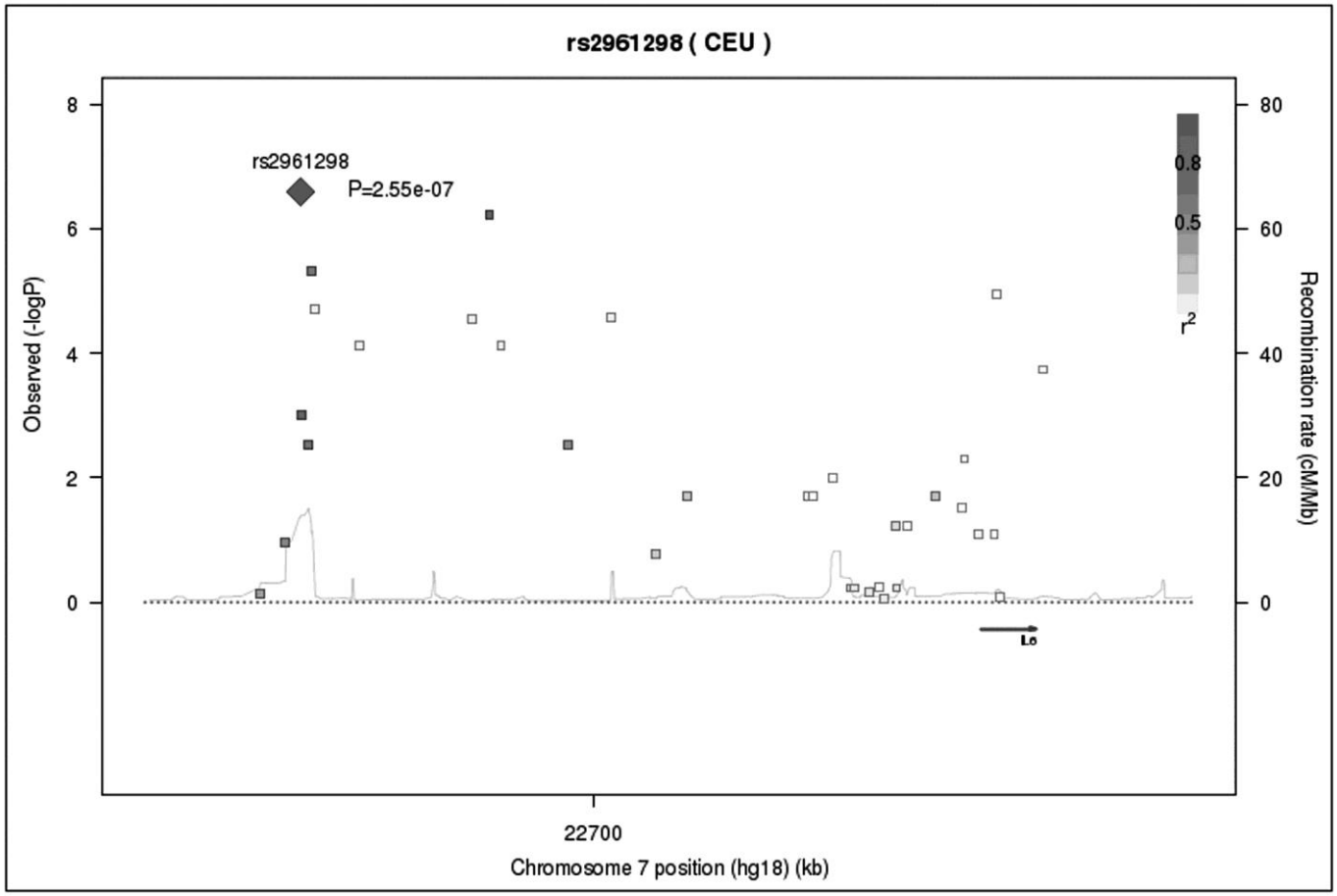


Figure 4.3 Genetic associations with the methylation level of cg01770232 in GOLDN within the *IL6* locus. SNPs were plotted by position within *IL6* locus against association with the methylation level of cg01770232 (-log10 *P*-value). Estimated recombination rates (from 1000 Genomes Pilot 1 CEU) are plotted in red gradients to reflect the local linkage disequilibrium structure. The SNPs surrounding the most significant SNP (rs2961298), represented by red diamonds, are plotted as color-coded squares to reflect their linkage disequilibrium with this SNP, which was estimated based on 1000 Genomes Pilot 1 CEU database.

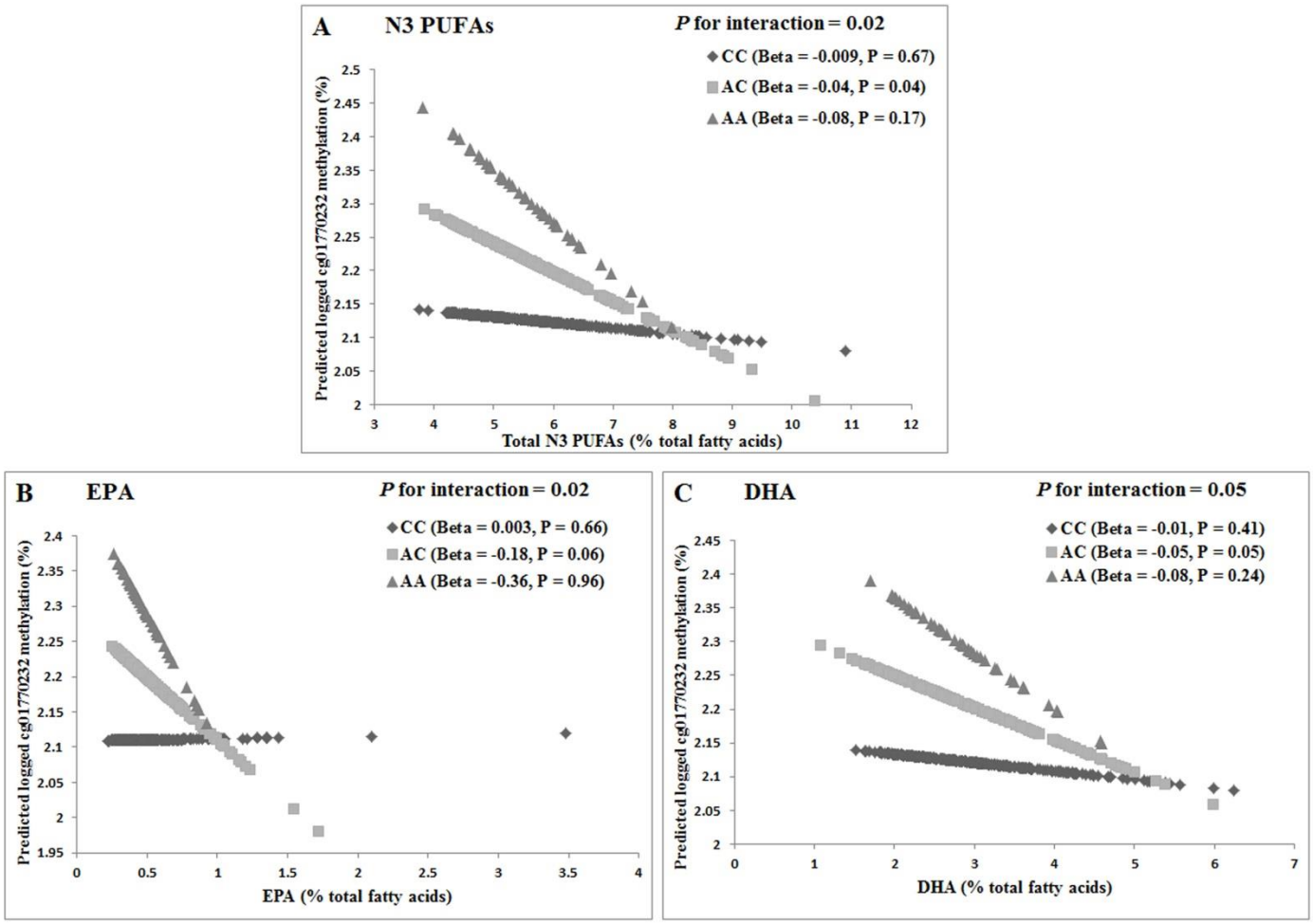


Figure 4.4 Interaction between N3 PUFAs and rs2961298 modulating the methylation level of cg01770232 in GOLDN. Predicted methylation levels of cg01770232 (log transformed) by genotype of rs2961298 were plotted against total N3 PUFAs (A), EPA (B), and DHA (C), adjusted for first four principal components for cellular purity and population structure, age, sex, center, pedigree, smoking, alcohol intake, total energy intake, physical activity, intake of vitamin B12 and folate, acute inflammatory conditions (infection or fever), chronic disease status (including abdominal obesity, CVD, diabetes, and hypertension), other inflammatory markers (MCP-1, adiponectin, hsCRP, TNF-α, and IL2sR-α). *P* values indicate the statistical significance of the adjusted interaction term and adjusted regression coefficients (represented as betas) in the regression line corresponding to three genotype groups of rs2961298 (diamond for CC, square for AC, and triangle for AA).

CHAPTER 5

Fatty acids, SNPs, and DNA methylation for blood lipids – meta-analysis of 7 studies in CHARGE consortium

5.1 Abstract

Background and aims

DNA methylation may represent a portion of the underlying molecular mechanisms for gene-by-environment interactions.

Methods and results

Based on multiple genome-wide association studies (GWAS) of blood lipids, 7 SNPs were selected for their predicted roles in the interplay among fatty acids, DNA methylation, and blood lipids, which were *APOE* rs405509, *ABCA1* rs2246293, *HMGCR* rs3761740, *APOA5* rs662799, *PCSK9* rs2479409, *HNF1A* rs1169288, and *HNF1A* rs1169287. Association and interactions of these 7 SNPs and circulating fatty acids for the outcome of blood triglycerides (TG) and high density lipoprotein (HDL) cholesterol were meta-analyzed using data from 7 cohorts participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. Significant findings were later supported by methylation analysis in the Genetics of Lipid Lowering Drugs and Diet Network study (GOLDN) and gene expression analysis with data from the Encyclopedia of DNA Elements (ENCODE) consortium.

Using a significance threshold corrected for multiple-testing, we observed significant associations between *APOE* rs405509 and TG ($P = 2.44\times10^{-4}$), and *APOA5* rs662799 for both TG ($P = 1.36\times10^{-18}$) and HDL cholesterol ($P = 2.49\times10^{-4}$). We did not observe significant circulating fatty acids by SNP interactions. However, two loci with interactions with two different circulating fatty acids approached significance ($P < 0.05$) were supported by methylation analyses. For *ABCA1* rs2246293 locus, the major G allele was associated with lower HDL ($P = 0.015$) and higher DNA methylation of cg14019050 ($P = 3.51\times10^{-18}$). The interaction term of SNP×circulating eicosapentaenoic acid (EPA) for HDL was positive ($\beta = 1.69$, $P = 0.006$) and it was negative for the methylation of cg14019050 ($\beta = -2.83$, $P = 0.007$). Methylation of cg14019050 was negatively correlated with HDL in GOLDN ($R = -0.12$, $P = 0.0002$) and *ABCA1* gene expression in ENCODE ($R = -0.61$, $P = 0.009$). At the *APOE* locus, the second locus showing nominally significant interaction between SNP and circulating fatty acids for a lipid outcome, the minor A allele of the promoter SNP rs405509 was associated with lower TG ($P = 2\times10^{-4}$) but higher DNA methylation of cg04406254 ($P = 0.008$). The interaction term of SNP×circulating α-linolenic acid (ALA) for TG was positive ($\beta = 0.16$, $P = 0.01$) but it was negative for methylation of cg04406254 ($\beta = -13.25$, $P = 0.009$). Methylation of cg04406254 was negatively correlated with TG in GOLDN ($R = -0.10$, $P = 0.002$) and *ABCA1* gene expression in ENCODE ($R = -0.66$, $P = 0.004$).

Conclusion

Methylation related SNPs may modify the effect of circulating fatty acids on blood lipids through their genotype-dependent associations with DNA methylation. These findings may increase mechanistic understanding of the interplay among environmental factors, genetic variants, epigenetic mechanisms, and phenotypes of interest.

Key words: fatty acids, DNA methylation, genetic variants, blood lipids, epidemiology

5.2 Introduction

There is accumulating evidence that the effects of dietary fatty acids on blood lipids depend on genotype. A modification effect of genetic variants on the relationship between dietary fatty acids and blood lipids was first suggested by the huge individual variability in lipid response to dietary fat intervention programs [1]. Later, epidemiological studies provided evidence of gene-by-dietary fat interactions. In the Framingham study, the association between dietary total fat and high density lipoprotein (HDL) cholesterol was shown to be modified by the genetic variants in the hepatic lipase gene promoter regions (*LIPC*) [2]. More recently, the associations between dietary n6-polyunsaturated fatty acids (PUFAs) and blood triglycerides (TG) were found to be modified by the genotypes of promoter variant in apolipoprotein A5 gene (*APOA5*) [3]. Additionally, the interaction between *APOE* variants and fatty acids were shown to modulate blood lipids [4]. Despite the growing evidence for gene-by-fatty acids interactions, the underlying mechanisms for the observed interactions are unknown.

DNA methylation may account as one potential mechanism to explain gene-fatty acids interactions. Compared to the other epigenetic mechanisms, DNA methylation has the most direct contact with nucleotides, and it has been suggested to act as the biomarker for the environmental factors. In cell systems, fatty acids have been shown to affect DNA methylation. Arachidonic acid (AA) and docosahexaenoic acid (DHA) were shown to affect methylation level of *Fads2* in mice liver [5]. Palmitic acid (PA) and oleic acid (OA) were found to induce methylation of peroxisome proliferator-activated receptor γ coactivator-1 α (*PGC-1α*) in human primary skeletal muscle cells [6]. Eicosapentaenoic acid (EPA) was shown to demethylate a CpG site of the tumor suppressor gene in human leukemia cell line [7]. In additional to a role of environmental factors, accumulating evidence suggests a relationship between DNA methylation and sequence variants. For instance, a genetic manipulation study demonstrated that methylation patterns are determined by the local sequence [8]. Genetic regulation of DNA methylation is widespread in humans [9-13]. Based on the evidence of the association of DNA methylation with both genetic variants and fatty acids, we propose our hypothesis that methylation related genetic variants modify the association of fatty acids with blood lipids through their genotype-dependent effects on DNA methylation.

To test our hypothesis, we selected candidate SNPs based on their relationship to circulating fatty acids and blood lipids and their potential to undergo DNA methylation. Meta-analysis of the association and interactions between these SNPs and circulating fatty acids for the outcome of blood lipids were conducted with 7 cohorts in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. Findings from the meta-analysis were further supported with methylation analysis in the Genetics of Lipid Lowering Drugs and Diet Network study (GOLDN) and gene expression analysis using data from Encyclopedia of DNA Elements (ENCODE) consortium.

5.3 Research design and methods

Study design

The flow chart in **Figure 5.1** depicts the three stages of the study. In stage I, candidate SNPs were selected based on their predicted relationship with DNA methylation, and their known associations with circulating fatty acids and blood lipids. Then, in stage II, the association and interaction of the selected SNPs with fatty acids on blood lipid levels were examined in cohorts in the CHARGE consortium and the results were meta-analyzed. In the final stage III, findings from stage II were validated by the analysis of methylation data in GOLDN and gene expression published by the ENCODE consortium.

Study populations

The study included 7 cohorts in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium: the Three-City Study (3C), the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN), the Invecchiare in Chianti (InCHIANTI), the Multi-Ethnic Study of Atherosclerosis (MESA), and the Women's Genome Health Study (WGHS). Details of these 7 cohorts are described in the **Supplementary Methods**. This study was approved by the institutional review board of each cohort and informed consent were received by all participants or their representatives.

SNP selection and genotyping in each cohort

Eight candidate SNPs were selected from the list of 323 SNPs located within 40 genes, which have been reported to be associated with blood lipids in genome-wide association studies (GWAS) [14-21]. We developed 8 inclusion criteria and 1 exclusion criteria to select the most relevant SNPs (**Figure 5.1**) with sufficient statistical power. SNPs meet the inclusion criteria (Criteria 1 to 8) will be at first included into the list, and then those SNPs meet the exclusion criteria (Criteria 9) will be excluded from the list. Criteria 1 required those SNPs to be located within the genes related with the phenotype of interest, blood lipids, by querying the gene name in PubMed-Gene database. Criteria 2 required those SNPs to be related to the exposure of interest, fatty acids, by searching the publications of the effects of fatty acids on the expression of genes containing the SNPs or the existence of the response element for fatty acids, peroxisome proliferator-activated receptors response elements (PPRE), within the genetic region of the genes covering the SNPs [22, 23]. Criteria 3 limited those SNPs with minor allele frequency greater than 1% to reach enough statistical power. SNPs were predicted to be related to DNA methylation if they meet one of the criteria for DNA methylation: Criteria 4, close to CpG island [23-27], or Criteria 5, within promoter region defined as within 3kb distance from the transcription start site downloaded from UCSC genome browser (02-02-2012), or Criteria 6, within regions reported to have tissue differential DNA methylation status [28], or Criteria 7, within region reported to have tissue differential chromatin status [29]. Criteria 8 required those SNPs with functional evidence, which was set up as optional considering the fact that these evidence are continuously growing. The exclusion criteria (Criteria 9) will exclude those SNPs in high linkage disequilibrium (R^2>0.8). As a result, 7 SNPs were included in the

analysis, which were *APOE* rs405509, *ABCA1* rs2246293, *HMGCR* rs3761740, *APOA5* rs662799, *PCSK9* rs2479409, *HNF1A* rs1169288, and *HNF1A* rs1169287. Details of genotyping methods in each cohort were described in **Supplementary Methods**.

Biochemical and circulating fatty acids measurements in each cohort.

Blood TG and high density lipoprotein (HDL) cholesterol were measured by enzymatic assays with detailed description in **Supplementary Methods**. We will focus on 7 circulating fatty acids, which are palmitic acid (PA), oleic acid (OA), linoleic acid (LA), arachidonic acid (AA), alpha-linolenic acid (ALA), eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA). Circulating fatty acids were measured in total plasma (3C and InCHIANTI), plasma phospholipids (ARIC, CHS, MESA), or erythrocytes (GOLDN and WGHS) as described in **Supplementary Methods**. Fatty acids were expressed as the percentage of total fatty acids. Information for the covariates in this study was collected through relevant questionnaires or measurements as described in **Supplementary Methods**.

DNA methylation in GOLDN

Detailed methodology for assessment of DNA methylation was previously described [30]. Briefly, frozen buffy coat samples from peripheral blood were used to extract CD4+ T-cells, from which the methylation of CpG sites within the genes covering the SNPs in the current study was measured using Infinium Human Methylation 450K BeadChip (Illumina, San Diego, CA, USA). CD4+ T-cells have been reported to be the suitable tissue for fasting blood lipids in the GOLDN population [31]. The measurement of methylation was expressed as the proportion of total signal from the methylation-specific probe or color channel, translated as the percentage of methylation.

DNA methylation and gene expression in ENCODE

Using databases of the ENCODE consortium, methylation and gene expression data across 17 cell lines were downloaded (06-20-2014) from the UCSC genome browser HAIB Methyl450 track (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeHaibMethyl450/) and Duke Affymetrix Array track (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeDukeAffyExon/), respectively. The methylation level of each CpG site was represented by a score, which was 1000 times the proportion of the intensity value from the methylated bead type from the sum of the intensity values from both methylated and unmethylated bead type plus 100. The range of the methylation score was from 0 to 1000. Gene expression value was represented by a score, which was 100 times the signal value linearly scaled for that particular cell type and with ranges from 0 to 1000.

Statistical analyses

In stage II, linear regression analysis was conducted by each cohort to generate regression coefficients (β) and robust standard errors of the association between the genotypes of the 7 SNPs and blood lipids (TG and HDL) and interaction between genotypes of 7 SNPs and circulating fatty acids (PA, OA, LA, AA, ALA, EPA, and DHA) for the outcome of blood lipids (TG and HDL).

Meta-analysis was preformed based on the regression coefficients and robust standard errors provided by each cohort, using an inverse variance-weighted, fixed effects approach. Two independent analysts performed the meta-analysis with R software and METAL (http://umich.edu/csg/abecasis/Metal/). Bonferroni correction for multiple testing was derived based on the 7 SNPs and the 7 fatty acids evaluated for interaction, and the corrected significance was 0.001 (α=0.05/49).

In stage III, for each SNP, we first conducted linear regression of the genotype and DNA methylation of the CpG sites located on the respective gene using data in the GOLDN study. Considering the dependent nature of methylation values of nearby CpG sites, a two-tailed *P*- value less than 0.05 was considered statistically significant. Then, using the methylation data on the CpG sites shown to have significant genetic associations, the interactions between SNPs and fatty acids with nominal significance with the outcome of blood lipids in stage II were further examined with the outcome of methylation. Finally, using the methylation data on the CpG sites with significance for both associations and interactions tests, Pearson correlation coefficients were obtained between methylation and blood lipids, and methylation and gene expression.

To limit variation of genetic structures in populations of different ethnicities, non-whites were excluded from the study. All the analyses were adjusted by age (continuous: years), sex (binary), study center (if applicable), population structure or pedigree (if applicable), and principle components for cellular purity and population structure (for methylation analysis) in Model 1. To account for potential confounding by other environmental factors, Model 2 included further adjustments of body mass index (BMI), smoking status (categorical: never vs. past vs. current smokers), physical activity (continuous, based on study-specific metric), alcohol intake (categorical: current vs. former/never), current estrogen therapy (categorical: yes/no), current lipid-lowering medication (categorical: yes/no), education level (categorical: cohort-specific metric) , total energy intake (continuous, kcal/day), dietary total fat intake (continuous, %total energy intake/day), glycemic load (if applicable) (continuous,g/day), dietary total folate intake (if applicable) (continuous, mcg/day), dietary VitB12 intake (if applicable) (continuous, mcg/day). Log transformation was performed for TG because it was not with normal distribution. SAS 9.2 (SAS Institute, Inc. Cary, NC) and STAT were used to conduct the analysis.

5.4 Results

(1) Population characteristics

Population characteristics, blood lipids, and circulating fatty acids are shown for each cohort in **Table 5.1**. Levels of blood lipids were similar across 7 cohorts. Fatty acids levels were similar among cohorts that measured fatty acids in the same compartment.

(2) Stage II: Meta-analysis of genetic associations and interactions with circulating fatty acids for blood lipids in CHARGE consortium

Meta-analysis of genetic associations with blood HDL and TG are shown in **Table 5.2**. The A allele of *APOE* rs405509 was associated with lower TG (P = 0.0002 and 0.00075 in Model 1 and Model 2, respectively). The G allele of *APOA5* rs662799 was associated with lower HDL (P = 0.0002 and 6.01×10^{-6} in Model 1 and Model 2, respectively) and higher TG ($P = 1.36\times10^{-18}$ and 7.07×10^{-19} in Model 1 and Model 2, respectively). The G allele of *ABCA1* rs2246293 was nominally associated with lower HDL (P = 0.07 and 0.015 in Model 1 and Model 2, respectively).

Meta-analysis of the genetic interactions with circulating fatty acids on blood HDL and TG are shown in **Table 5.3**. At the un-corrected threshold of significance ($P < 0.05$), three SNPs showed interactions. *HNF1A* rs1169287 interacted with circulating PA (P = 0.002 and 0.03 in Model 1 and Model 2, respectively) and ALA (P = 0.02 and 0.008 in Model 1 and Model 2 respectively) for the outcome of HDL. *APOE* rs405509 interacted with circulating ALA in the association with TG (P = 0.05 and 0.01 in Model 1 and Model 2, respectively). *ABCA1* rs2246293 interacted with circulating EPA to modulate HDL (P = 0.006 and 0.05 in Model 1 and Model 2, respectively). None of these three interactions met our a priori cut point for significance with correction for multiple hypotheses testing ($P < 0.001$).

(3) Stage III: Supportive evidence of methylation in GOLDN and gene expression in ENCODE

Two SNPs showed significant associations and interactions with the outcome of DNA methylation levels of the CpG sites within the respective genetic regions, which are *ABCA1* rs2246293 and *APOE* rs405509. The G allele of *ABCA1* rs2246293 was associated with higher methylation level of cg14019050 ($P = 3.51\times10^{-18}$ in Model 1) (**Figure 5.2B**), and this association was further modified by the circulating EPA (regression coefficient and P value for the interaction term in Model 1 was -2.83 and 0.007, irrespectively) (**Figure 5.2D**). Additionally, the A allele of *APOE* rs405509 was shown to be associated with higher methylation of cg04406254 (P = 0.008 in Model 1), which was previously reported with a different statistical model [CHAPTER 3], and this association was also modified by the circulating ALA (regression coefficient and P value for the interaction term in Model 1 was -13.25 and 0.009, irrespectively) (**Figure 5.3D**).

Methylation level of cg14019050 within the *ABCA1* was negatively correlated with plasma HDL cholesterol in the GOLDN study (R = -0.12, P = 0.0002) (**Figure 5.2E**) and *ABCA1* gene expression with data from ENCODE consortium (R = -0.61, P = 0.009) (**Figure 5.2F**). Additionally, methylation level of cg04406254 in the *APOE* was negatively associated with plasma TG (R = -0.10,

P = 0.002) (**Figure 5.3E, 5.3G**) and *APOE* gene expression with data from ENCODE consortium (*R* = -0.66, *P* = 0.004) (**Figure 5.2F**).

5.5 Discussion

Using the strict statistical significance threshold adjusted for multiple-testing, we did not observe interactions between 7 predicted methylation-related SNPs and 7 fatty acids to modulate blood lipids. However, two loci with nominal statistical significance were supported by the analysis of DNA methylation, suggesting their biological significance. Our consistent results for the outcome of blood lipids and DNA methylation measurements implied that *ABCA1* promoter SNP rs2246293 may interact with circulating EPA to modulate blood HDL through its genotype-dependent effect on DNA methylation of cg14019050, and the interaction between *APOE* promoter SNP rs405509 and circulating ALA for the outcome of blood TG may be contributed by the differential changes in DNA methylation of cg04406254. Our observations suggest a possible mechanistic role of epigenetic mechanism in observed gene-by-environment interactions.

According to our findings, the major G allele of *ABCA1* rs2246293 was associated with lower blood HDL cholesterol but higher methylation of cg14019050. Further, the regression coefficient for the interaction term between the SNP and circulating EPA for the outcome of HDL is positive, while it is negative for the outcome of methylation of cg14019050. This means that the difference, per copy of the G allele, of the difference in HDL associated with one unit higher circulating EPA is positive, while the difference, per copy of the G allele, of the difference in the methylation of cg14019050 associated with one unit higher circulating EPA is negative. These opposite findings for blood HDL and methylation of cg14019050 was consistent with their negative correlation, and also negative correlation between *ABCA1* gene expression and methylation of cg14019050. Our findings of the relationship of rs2246293 G allele – higher promoter methylation – lower *ABCA1* gene expression – lower blood HDL cholesterol was not only consistent with previous studies but also with the biological function of *ABCA1*. *ABCA1* encodes a membrane transporter involved in the reverse cholesterol transport pathway by promoting efflux of free cholesterol and phospholipids to form HDL particles [32]. Our group and others demonstrated that loss-of-functional mutations in *ABCA1* contribute to the etiology of Tangier disease, a disease characterized by extremely low level or even absence of HDL cholesterol [33-35]. Cg14019050 is a CpG site located within the promoter region of *ABCA1*. In previous studies, blood HDL was negatively correlated with methylation of *ABCA1* promoter in familial hypercholesterolemia [36].

The findings of the *APOE* locus were consistent with our own data and also with previous studies. The minor A allele of *APOE* promoter SNP rs405509 was found to be associated with lower TG but higher methylation of cg04406254. The regression coefficients for the interaction term SNP×ALA were also with the opposite sign for TG and methylation, of which was positive for TG but negative for methylation. This means that the difference, per copy of the A allele, of the difference in TG associated with one unit higher circulating ALA is positive, while the difference, per copy of the A allele, of the difference in the methylation of cg04406254 associated with one unit higher circulating ALA is negative. Being consistent to the opposite directions observed with the association and interaction studies,

methylation of cg04406254 was shown to have negative correlation with TG and also *APOE* expression. Compared to C allele of rs405509, A allele exhibited lower gene transcription [37], supporting our observations of the associations of A allele – higher methylation – lower gene expression.

Our study was limited by the observational study design, from which the association not the cause-effect relationship can be concluded. Our study calls for the intervention study and animal experiment to validate our hypothesis of the mechanistic role of DNA methylation for observed gene-by-environment interactions. From the perspective of statistical significance, the issue of under-power may account for the lack of significant interactions according to the multiple-testing corrected significance level. However, the consistent findings with blood lipids and with methylation provided biological supports for those interactions shown to be significant under nominal significance level. Not all studies measured fatty acids in the same compartment, which may also mask the potential significance for the interaction tests. The methylation analysis is only conducted in one study. Some of the fatty acids are biomarkers of intake while others are biomarkers of metabolism, so further work is needed with dietary fatty acids for confirmation.

The current study indicates that the molecular mechanisms for the observed gene-by-environment interaction studies may be related to epigenetic modifications. It suggests that methylation-related SNPs modify the effects of environmental factors on the phenotypes through genotype-dependent changes in DNA methylation. Our findings have clinical implications. It is impossible to change the genotypes an individual carries, but it is definitely possible to change DNA methylation status by identifying the corresponding modulators. In this case, the genotype-dependent deleterious effects will be ameliorated through the application of methylation modulators. Our study provides a novel and biologically important direction for future research of the mechanisms underlying gene-by-environment interactions.

5.6 Reference:

1. Schaefer, E.J., et al., *Individual variability in lipoprotein cholesterol response to National Cholesterol Education Program Step 2 diets.* Am J Clin Nutr, 1997. **65**(3): p. 823-30.
2. Ordovas, J.M., et al., *Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on high-density lipoprotein metabolism: evidence of a strong dose effect in this gene-nutrient interaction in the Framingham Study.* Circulation, 2002. **106**(18): p. 2315-21.
3. Lai, C.Q., et al., *Dietary intake of n-6 fatty acids modulates effect of apolipoprotein A5 gene on plasma fasting triglycerides, remnant lipoprotein concentrations, and lipoprotein particle size: the Framingham Heart Study.* Circulation, 2006. **113**(17): p. 2062-70.
4. Rudkowska, I., et al., *Gene-diet interactions on plasma lipid levels in the Inuit population.* Br J Nutr, 2013. **109**(5): p. 953-61.
5. Devlin, A.M., et al., *Hypermethylation of Fads2 and altered hepatic fatty acid and phospholipid metabolism in mice with hyperhomocysteinemia.* J Biol Chem, 2007. **282**(51): p. 37082-90.
6. Barres, R., et al., *Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density.* Cell Metab, 2009. **10**(3): p. 189-98.
7. Ceccarelli, V., et al., *Eicosapentaenoic acid demethylates a single CpG that mediates expression of tumor suppressor CCAAT/enhancer-binding protein delta in U937 leukemia cells.* J Biol Chem, 2011. **286**(31): p. 27092-102.
8. Lienert, F., et al., *Identification of genetic elements that autonomously determine DNA methylation states.* Nat Genet, 2011. **43**(11): p. 1091-7.
9. Bell, C.G., et al., *Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus.* PLoS One, 2010. **5**(11): p. e14040.
10. Gertz, J., et al., *Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation.* PLoS Genet, 2011. **7**(8): p. e1002228.
11. Qu, W., et al., *Genome-wide genetic variations are highly correlated with proximal DNA methylation patterns.* Genome Res, 2012. **22**(8): p. 1419-25.
12. Shoemaker, R., et al., *Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome.* Genome Res, 2010. **20**(7): p. 883-9.
13. Zhang, D., et al., *Genetic control of individual differences in gene-specific methylation in human brain.* Am J Hum Genet, 2010. **86**(3): p. 411-9.
14. Aulchenko, Y.S., et al., *Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts.* Nat Genet, 2009. **41**(1): p. 47-55.
15. Teslovich, T.M., et al., *Biological, clinical and population relevance of 95 loci for blood lipids.* Nature, 2010. **466**(7307): p. 707-13.
16. Wallace, C., et al., *Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia.* Am J Hum Genet, 2008. **82**(1): p. 139-49.
17. Willer, C.J., et al., *Newly identified loci that influence lipid concentrations and risk of coronary artery disease.* Nat Genet, 2008. **40**(2): p. 161-9.
18. Kathiresan, S., et al., *Common variants at 30 loci contribute to polygenic dyslipidemia.* Nat Genet, 2009. **41**(1): p. 56-65.
19. Diabetes Genetics Initiative of Broad Institute of, H., et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* Science, 2007. **316**(5829): p. 1331-6.
20. Ridker, P.M., et al., *Polymorphism in the CETP gene region, HDL cholesterol, and risk of future myocardial infarction: Genomewide analysis among 18 245 initially healthy women from the Women's Genome Health Study.* Circ Cardiovasc Genet, 2009. **2**(1): p. 26-33.
21. Sandhu, M.S., et al., *LDL-cholesterol concentrations: a genome-wide association study.* Lancet, 2008. **371**(9611): p. 483-91.
22. Podvinec, M., et al., *NUBIScan, an in silico approach for prediction of nuclear receptor response elements.* Mol Endocrinol, 2002. **16**(6): p. 1269-79.

23. Ponger, L. and D. Mouchiroud, *CpGProD: identifying CpG islands associated with transcription start sites in large genomic mammalian sequences.* Bioinformatics, 2002. **18**(4): p. 631-3.
24. Gardiner-Garden, M. and M. Frommer, *CpG islands in vertebrate genomes.* J Mol Biol, 1987. **196**(2): p. 261-82.
25. Takai, D. and P.A. Jones, *Comprehensive analysis of CpG islands in human chromosomes 21 and 22.* Proc Natl Acad Sci U S A, 2002. **99**(6): p. 3740-5.
26. Hackenberg, M., et al., *CpGcluster: a distance-based algorithm for CpG-island detection.* BMC Bioinformatics, 2006. **7**: p. 446.
27. Su, J., et al., *CpG_MI: a novel approach for identifying functional CpG islands in mammalian genomes.* Nucleic Acids Res, 2010. **38**(1): p. e6.
28. Meissner, A., et al., *Genome-scale DNA methylation maps of pluripotent and differentiated cells.* Nature, 2008. **454**(7205): p. 766-70.
29. Ernst, J., et al., *Mapping and analysis of chromatin state dynamics in nine human cell types.* Nature, 2011. **473**(7345): p. 43-9.
30. Absher, D.M., et al., *Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations.* PLoS Genet, 2013. **9**(8): p. e1003678.
31. Irvin, M.R., et al., *Epigenome-Wide Association Study of Fasting Blood Lipids in the Genetics of Lipid Lowering Drugs and Diet Network Study.* Circulation, 2014.
32. Tall, A.R., et al., *HDL, ABC transporters, and cholesterol efflux: implications for the treatment of atherosclerosis.* Cell Metab, 2008. **7**(5): p. 365-75.
33. Rust, S., et al., *Assignment of Tangier disease to chromosome 9q31 by a graphical linkage exclusion strategy.* Nat Genet, 1998. **20**(1): p. 96-8.
34. Brooks-Wilson, A., et al., *Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency.* Nat Genet, 1999. **22**(4): p. 336-45.
35. Brousseau, M.E., et al., *Novel mutations in the gene encoding ATP-binding cassette 1 in four tangier disease kindreds.* J Lipid Res, 2000. **41**(3): p. 433-41.
36. Guay, S.P., et al., *ABCA1 gene promoter DNA methylation is associated with HDL particle profile and coronary artery disease in familial hypercholesterolemia.* Epigenetics, 2012. **7**(5): p. 464-72.
37. Artiga, M.J., et al., *Allelic polymorphisms in the transcriptional regulatory region of apolipoprotein E gene.* FEBS Lett, 1998. **421**(2): p. 105-8.

Table 5.1. Population characteristics, blood lipids, and circulating fatty acids in each cohort*

	N	Age	% Women	HDL	TG	Circulating fatty acids, % total fatty acids							
						Compartment	PA	OA	LA	AA	ALA	EPA	DHA
Three-City (3C) Study	1240	74±5	60	60.8±15.4	113.2±56.8	plasma	28.3±5.8	20.4±3.9	24.7±5.5	6.7±1.9	0.4±0.2	1.0±0.6	2.4±0.8
Atherosclerosis Risk in Communities (ARIC) Study	3385	54±6	53	52.0±16.9	137.4±92.3	plasma phospholipids	25.4±1.7	8.6±1.1	22.0±2.7	11.5±2.0	0.1±0.1	0.6±0.3	2.8±0.9
Cardiovascular Health Study (CHS)	2399	72±5	62			plasma phospholipids	25.5±1.6	7.6±1.1	20.0±2.5	10.9±1.9	0.2±0.1	0.6±0.4	3.0±1.0
Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) Study	1120	48±16	52	47.1±13.1	138.8±115.8	erythrocytes	22.8±1.2	16.1±1.1	12.9±1.4	13.6±1.2	0.1±0.04	0.5±0.3	3.0±0.9
Multi-Ethnic Study of Atherosclerosis (MESA)	674	62±10	53	52.7±16.1	131.4±79.0	plasma phospholipids	25.8±1.9	8.3±1.2	20.9±3.0	12.1±2.2	0.2±0.1	1.0±0.7	3.7±1.4
Invescchiare in Chianti (InCHIANTI)	1002	68±15	55	56.3±15.0	123.4±65.6	plasma	22.5±2.4	25.8±3.7	24.9±3.9	8.0±1.9	0.5±0.2	0.6±0.2	2.3±0.8
Women's Genome Health Study (WGHS)	652	54±7	100	54.2±14.7	142.3±88.5	erythrocytes	23.4±2.3	15.0±1.5	12.3±1.4	12.7±2.3	0.2±0.1	0.5±0.2	3.3±1.0

*Data are mean ± SD.

Table 5.2. Meta-analysis of associations between SNPs and blood HDL and TG

SNP	Gene	Coded allele	HDL (mg/dl)						logged TG (mg/dl)					
			Model 1[1]			Model 2[2]			Model 1[1]			Model 2[2]		
			β[3] (95%CI)	*P*	N	β[3] (95%CI)	*P*	N	β[3] (95%CI)	*P*	N	β[3] (95%CI)	*P*	N
rs405509	*APOE*	A	0.10 (-0.37,0.58)	0.67	6378	0.18 (-0.28,0.65)	0.44	5924	-0.03 (-0.05,-0.01)	2.44E-04	6383	-0.03 (-0.05,-0.01)	7.51E-04	5928
rs2246293	*ABCA1*	G	-0.46 (-0.97,0.04)	0.07	6737	-0.60 (-1.08,-0.12)	0.015	6124	-0.01 (-0.02,0.01)	0.5222	6738	0.001 (-0.01,0.02)	0.8783	6124
rs3761740	*HMGCR*	A	0.14 (-0.51,0.78)	0.68	10115	0.16 (-0.46,0.78)	0.62	9464	0.01 (-0.02,0.03)	0.5628	10120	0.005 (-0.02,0.03)	0.6585	9468
rs662799	*APOA5*	G	-1.68 (-2.59,-0.78)	0.0002	8035	-2.00 (-2.86,-1.13)	6.01E-06	7420	0.15 (0.12,0.19)	1.36E-18	8036	0.15 (0.12,0.18)	7.07E-19	7420
rs2479409	*PCSK9*	C	0.03 (-0.56,0.62)	0.92	4878	-0.25 (-0.82,0.31)	0.38	4675	-0.01 (-0.03,0.01)	0.486	4878	0.002 (-0.02,0.02)	0.8242	4675
rs1169287	*HNF1A*	T	-1.14 (-2.63,0.34)	0.13	7740	-1.25 (-2.81,0.30)	0.11	7127	-0.002 (-0.06,0.06)	0.9493	7741	0.0003 (-0.06,0.06)	0.993	7127
rs1169288	*HNF1A*	C	0.27 (-0.15,0.69)	0.20	10139	0.19 (-0.21,0.59)	0.36	9487	-0.002 (-0.02,0.01)	0.8122	10144	0.0007 (-0.01,0.02)	0.9235	9491

[1]Model 1 adjusted for age, sex, study center, population structure (cohort-specific metric);

[2]Model 2 adjusted for covariates in Model 1 plus BMI (continuous, kg/m2), smoking (2 categories: never/past vs. current smokers), physical activity (continuous, based on study-specific metric), drinking (2 categories: current vs. former/never), estrogen therapy usage (2 categories: yes/no), lipid-lowering medication usage (2 categories: yes/no), education level (categorical: cohort-specific metric), total energy intake (continuous, kcal/day), dietary total fat intake (continuous, % total energy intake/day), dietary total folate intake (continuous, mcg/day), dietary VitB12 intake (continuous, mcg/day), glycemic load (quintile of g/day) (or glycemic index (quintile) or whole grain intake (quintile of g/day));

[3]β represents the regression coefficient for the expected changes in blood HDL (mg/dl) or ln(TG) (mg/dl) per copy of coded allele.

Table 5.3. Meta-analysis of interactions between SNPs and circulation fatty acids in association with blood HDL and TG

Fatty acids (% total fatty acids)	SNP	Gene	Coded allele	HDL (mg/dl)						ln(TG) (mg/dl)					
				Model 1[1]			Model 2[2]			Model 1[1]			Model 2[2]		
				β[3] (95%CI)	*P*	N	β[3] (95%CI)	*P*	N	β[3] (95%CI)	*P*	N	β[3] (95%CI)	*P*	N
PA (16:0)	rs405509	*APOE*	A	-0.04 (-0.20,0.13)	0.66	6342	-0.04 (-0.23,0.14)	0.65	5888	-0.004 (-0.01,0.0007)	0.091	6347	-0.01 (-0.01,0.0003)	0.06	5892
	rs2246293	*ABCA1*	G	-0.05 (-0.23,0.12)	0.56	6709	-0.04 (-0.22,0.14)	0.68	6096	-0.003 (-0.01,0.01)	0.925	6710	-0.001 (-0.01,0.005)	0.73	6101
	rs3761740	*HMGCR*	A	0.13 (-0.11,0.38)	0.28	10087	0.12 (-0.10,0.34)	0.30	9436	0.002 (-0.01,0.01)	0.655	10092	0.01 (-0.003,0.02)	0.17	9458
	rs662799	*APOA5*	G	-0.23 (-0.63,0.17)	0.26	7999	-0.33 (-0.67,0.02)	0.06	7384	0.01 (-0.01,0.02)	0.441	8000	0.001 (-0.01,0.01)	0.85	7384
	rs2479409	*PCSK9*	C	-0.15 (-0.56,0.26)	0.47	4850	-0.18 (-0.57,0.22)	0.38	4647	0.003 (-0.01,0.02)	0.669	4850	0.01 (-0.01,0.02)	0.37	4643
	rs1169287	*HNF1A*	T	-0.95 (-1.55,-0.36)	0.002	7712	-0.64 (-1.20,-0.08)	0.03	7099	0.01 (-0.01,0.02)	0.383	7713	0.01 (-0.0009,0.02)	0.07	7098
	rs1169288	*HNF1A*	C	-0.01 (-0.18,0.17)	0.94	10111	0.02 (-0.16,0.20)	0.85	9459	-0.001 (-0.01,0.005)	0.645	10116	0.001 (-0.01,0.01)	0.83	9445
OA (18:1)	rs405509	*APOE*	A	0.26 (0.06,0.46)	0.01	6342	0.18 (-0.03,0.38)	0.10	5888	-0.01 (-0.01,-0.001)	0.022	6347	-0.01 (-0.01,0.0006)	0.07	5892
	rs2246293	*ABCA1*	G	-0.06 (-0.32,0.21)	0.68	6709	-0.03 (-0.31,0.26)	0.86	6096	-0.01 (-0.01,0.002)	0.125	6710	-0.01 (-0.02,0.002)	0.15	6101
	rs3761740	*HMGCR*	A	0.07 (-0.24,0.38)	0.65	10087	0.05 (-0.27,0.38)	0.75	9436	0.01 (-0.003,0.02)	0.178	10092	0.005 (-0.01,0.02)	0.42	9458
	rs662799	*APOA5*	G	-0.38 (-0.80,0.03)	0.07	7999	-0.26 (-0.68,0.16)	0.22	7384	0.003 (-0.01,0.02)	0.66	8000	0.005 (-0.01,0.02)	0.41	7384
	rs2479409	*PCSK9*	C	-0.12 (-0.71,0.47)	0.69	4850	-0.07 (-0.63,0.48)	0.80	4647	-0.004 (-0.02,0.02)	0.70	4850	-0.003 (-0.02,0.02)	0.74	4643
	rs1169287	*HNF1A*	T	-0.59 (-1.50,0.32)	0.20	7712	-0.24 (-1.16,0.68)	0.61	7099	0.01 (-0.01,0.04)	0.366	7713	-0.002 (-0.02,0.01)	0.81	7098
	rs1169288	*HNF1A*	C	0.12 (-0.09,0.32)	0.26	10111	0.09-0.12,0.29)	0.41	9459	-0.002 (-0.01,0.004)	0.507	10116	-0.004 (-0.01,0.003)	0.29	9445
LA (18:2n6)	rs405509	*APOE*	A	-0.04 (-0.18,0.11)	0.62	6342	-0.02 (-0.17,0.14)	0.83	5888	0.004 (-0.0003,0.01)	0.07	6347	0.004 (-0.0005,0.01)	0.09	5892
	rs2246293	*ABCA1*	G	0.02 (-0.14,0.18)	0.83	6709	0.05 (-0.11,0.22)	0.52	6096	-0.001 (-0.01,0.004)	0.622	6710	-0.002 (-0.01,0.004)	0.59	6101
	rs3761740	*HMGCR*	A	-0.05 (-0.26,0.15)	0.61	10087	-0.08 (-0.28,0.12)	0.44	9436	-0.003 (-0.01,0.003)	0.338	10092	-0.004 (-0.01,0.003)	0.29	9458
	rs662799	*APOA5*	G	0.08 (-0.23,0.39)	0.60	7999	0.16 (-0.12,0.45)	0.25	7384	-0.003 (-0.01,0.01)	0.581	8000	-0.002 (-0.009,0.004)	0.50	7384
	rs2479409	*PCSK9*	C	0.03 (-0.22,0.28)	0.83	4850	-0.02 (-0.26,0.22)	0.85	4647	0.0004 (-0.01,0.01)	0.925	4850	0.002 (-0.01,0.01)	0.62	4643
	rs1169287	*HNF1A*	T	0.59 (0.09,1.09)	0.02	7712	0.36 (-0.14,0.85)	0.16	7099	-0.01 (-0.03,-0.0008)	0.038	7713	-0.002 (-0.01,0.01)	0.62	7098
	rs1169288	*HNF1A*	C	-0.09 (-0.22,0.05)	0.20	10111	-0.11 (-0.24,0.03)	0.11	9459	0.001 (-0.004,0.01)	0.775	10116	0.001 (-0.004,0.01)	0.74	9445
AA (20:4n6)	rs405509	*APOE*	A	-0.11 (-0.37,0.14)	0.37	6342	0.02 (-0.23,0.26)	0.90	5888	0.005 (-0.004,0.01)	0.307	6347	0.001 (-0.01,0.01)	0.88	5892
	rs2246293	*ABCA1*	G	0.12 (-0.14,0.39)	0.35	6709	0.19 (-0.07,0.44)	0.15	6096	0.01 (-0.004,0.01)	0.271	6710	0.002 (-0.01,0.01)	0.71	6101
	rs3761740	*HMGCR*	A	-0.35 (-0.70,-0.004)	0.05	10087	-0.20 (-0.52,0.13)	0.24	9436	0.002 (-0.01,0.01)	0.80	10092	0.0003 (-0.01,0.01)	0.97	9458

	rs662799	*APOA5*	G	0.35 (-0.12,0.82)	0.15	7999	0.59 (0.13,1.05)	0.01	7384	-0.01 (-0.03,0.004)	0.122	8000	-0.01 (-0.02,-0.0002)	0.04	7384
	rs2479409	*PCSK9*	C	0.06 (-0.28,0.40)	0.74	4850	0.03 (-0.29,0.34)	0.87	4647	0.004 (-0.008,0.02)	0.517	4850	-0.003 (-0.01,0.01)	0.59	4643
	rs1169287	*HNF1A*	T	0.43 (-0.43,1.29)	0.33	7712	-0.06 (-0.96,0.85)	0.90	7099	-0.005 (-0.03,0.02)	0.76	7713	-0.01 (-0.02,0.01)	0.33	7098
	rs1169288	*HNF1A*	C	0.05 (-0.17,0.26)	0.66	10111	0.08 (-0.13,0.29)	0.45	9459	-0.002 (-0.009,0.01)	0.649	10116	0.002 (-0.01,0.01)	0.74	9445
ALA (18:3n3)	rs405509	*APOE*	A	-1.77 (-5.70,2.16)	0.38	6342	-2.76 (-6.57,1.05)	0.16	5888	0.12 (-0.0002,0.24)	0.051	6347	0.16 (0.04,0.27)	0.01	5892
	rs2246293	*ABCA1*	G	1.62 (-4.29,7.53)	0.59	6709	2.25 (-3.80,8.30)	0.47	6096	-0.14 (-0.31,0.03)	0.114	6710	-0.19 (-0.39,-0.003)	0.05	6101
	rs3761740	*HMGCR*	A	1.99 (-4.56,8.53)	0.55	10087	2.85 (-3.60,9.30)	0.39	9436	0.12 (-0.09,0.32)	0.261	10092	0.08 (-0.13,0.28)	0.48	9458
	rs662799	*APOA5*	G	-3.16 (-11.32,5.00)	0.45	7999	-3.40 (-11.69,4.88)	0.42	7384	0.10 (-0.16,0.36)	0.464	8000	0.03 (-0.2,0.27)	0.78	7384
	rs2479409	*PCSK9*	C	-3.24 (-16.01,9.53)	0.62	4850	1.13 (-10.29,12.56)	0.85	4647	-0.04 (-0.45,0.38)	0.862	4850	-0.16 (-0.54,0.23)	0.42	4643
	rs1169287	*HNF1A*	T	15.30 (2.01,28.60)	0.02	7712	17.97 (4.71,31.22)	0.008	7099	-0.06 (-0.48,0.36)	0.789	7713	-0.10 (-0.42,0.23)	0.56	7098
	rs1169288	*HNF1A*	C	-4.54 (-8.40,-0.69)	0.02	10111	-3.15 (-6.95,0.65)	0.10	9459	0.11 (-0.01,0.24)	0.072	10116	0.06 (-0.06,0.19)	0.33	9445
EPA (20:5n3)	rs405509	*APOE*	A	-0.80 (-2.15,0.56)	0.25	6342	-1.01 (-2.39,0.36)	0.15	5888	-0.004 (-0.04,0.04)	0.863	6347	0.005 (-0.04,0.05)	0.83	5892
	rs2246293	*ABCA1*	G	1.69 (0.49,2.88)	0.006	6709	1.15 (-0.01,2.32)	0.05	6096	0.01 (-0.03,0.05)	0.636	6710	0.03 (-0.02,0.08)	0.25	6101
	rs3761740	*HMGCR*	A	-0.67 (-2.45,1.10)	0.46	10087	-1.32 (-2.89,0.25)	0.10	9436	0.01 (-0.05,0.06)	0.858	10092	0.05 (-0.01,0.11)	0.09	9458
	rs662799	*APOA5*	G	0.46 (-2.23,3.15)	0.74	7999	-1.24 (-3.37,0.88)	0.25	7384	0.0002 (-0.08,0.08)	1.00	8000	0.003 (-0.06,0.06)	0.93	7384
	rs2479409	*PCSK9*	C	0.35 (-1.26,1.97)	0.67	4850	0.47 (-1.05,2.00)	0.54	4647	0.01 (-0.05,0.07)	0.795	4850	-0.01 (-0.06,0.04)	0.80	4643
	rs1169287	*HNF1A*	T	5.45 (-0.39,11.29)	0.07	7712	6.77 (0.98,12.56)	0.02	7099	-0.02 (-0.17,0.13)	0.754	7713	0.03 (-0.04,0.1)	0.37	7098
	rs1169288	*HNF1A*	C	0.07 (-1.05,1.19)	0.90	10111	-0.14 (-1.29,1.01)	0.81	9459	-0.02 (-0.06,0.01)	0.226	10116	-0.02 (-0.06,0.02)	0.34	9445
DHA (22:6n3)	rs405509	*APOE*	A	-0.28 (-0.83,0.26)	0.31	6342	-0.35 (-0.87,0.17)	0.18	5888	0.002 (-0.02,0.02)	0.80	6347	0.01 (-0.01,0.02)	0.57	5892
	rs2246293	*ABCA1*	G	0.17 (-0.35,0.69)	0.52	6709	0.18 (-0.32,0.68)	0.47	6096	-0.0002 (-0.02,0.02)	0.98	6710	0.01 (-0.01,0.03)	0.45	6101
	rs3761740	*HMGCR*	A	0.06 (-0.62,0.73)	0.87	10087	-0.01 (-0.63,0.60)	0.96	9436	0.01 (-0.02,0.03)	0.555	10092	0.02 (-0.01,0.05)	0.15	9458
	rs662799	*APOA5*	G	0.22 (-0.73,1.16)	0.65	7999	-0.14 (-1.05,0.77)	0.76	7384	-0.02 (-0.06,0.01)	0.183	8000	-0.02 (-0.04,-0.002)	0.03	7384
	rs2479409	*PCSK9*	C	-0.14 (-0.79,0.50)	0.67	4850	0.06 (-0.54,0.66)	0.85	4647	0.01 (-0.01,0.03)	0.279	4850	0.02 (-0.0004,0.04)	0.06	4643
	rs1169287	*HNF1A*	T	0.11 (-1.46,1.69)	0.89	7712	1.08 (-0.38,2.54)	0.15	7099	0.004 (-0.06,0.07)	0.90	7713	0.002 (-0.02,0.03)	0.89	7098
	rs1169288	*HNF1A*	C	0.03 (-0.41,0.46)	0.91	10111	-0.06 (-0.47,0.35)	0.76	9459	-0.001 (-0.02,0.01)	0.911	10116	0.02 (0.001,0.03)	0.04	9445

[1]Model 1 adjusted for age, sex, study center, population structure (cohort-specific metric);
[2]Model 2 adjusted for covariates in Model 1 plus BMI (continuous, kg/m2), smoking (2 categories: never/past vs. current smokers), physical activity (continuous, based on study-specific metric), drinking (2 categories: current vs. former/never), estrogen therapy usage (2 categories: yes/no), lipid-lowering medication usage (2 categories: yes/no), education level (categorical: cohort-specific metric), total energy intake (continuous, kcal/day), dietary total fat intake (continuous, % total energy intake/day), dietary total folate intake (continuous, mcg/day), dietary VitB12 intake (continuous, mcg/day), glycemic load (quintile of g/day) (or glycemic index (quintile) or whole grain intake (quintile of g/day));
[3]β represents the regression coefficient for the expected difference in size of the dietary fatty acids association with blood HDL (mg/dl) or ln(TG) (mg/dl) per copy of coded allele.

Supplementary Table 5.1. Association between SNPs and DNA methylation of CpG sites.

Gene	SNP	Coded allele	N of CpG tested	CpG sites							
				N	name	Model 1[1]			Model 2[2]		
						β[3]	sem[3]	*P*	β[3]	sem[3]	*P*
APOE	rs405509	A	13	1	cg04406254	0.46	0.17	0.01	0.46	0.17	0.009
				2	cg06750524	0.49	0.22	0.03	0.48	0.22	0.03
				3	cg16471933	0.46	0.20	0.02	0.52	0.20	0.01
ABCA1	rs2246293	G	19	1	cg02945674	18.24	1.16	1.52E-13	18.48	1.19	7.51E-14
				2	cg14019050	8.84	0.28	3.50E-18	8.82	0.28	5.64E-18
HMGCR	rs3761740	A	17	1	cg11003009	-0.19	0.06	0.003	-0.21	0.06	0.002
APOA5	rs662799	G	8	1	cg02157083	-0.72	0.26	0.01	-0.70	0.27	0.01
				2	cg03481039	0.85	0.31	0.01	0.86	0.32	0.01
				3	cg25454270	0.99	0.38	0.03	1.06	0.40	0.02
PCSK9	rs2479409	C	23	1	cg00045070	33.11	0.67	3.76E-18	33.11	0.68	1.35E-18
				2	cg04554817	-0.52	0.21	0.02	-0.51	0.21	0.02
				3	cg05118916	-1.21	0.16	2.56E-09	-1.26	0.17	2.74E-09
				4	cg06197377	-0.98	0.15	1.68E-07	-0.97	0.16	3.27E-07
				5	cg13462158	-2.31	0.73	0.002	-2.46	0.74	0.002
				6	cg14993491	-0.71	0.19	3.81E-04	-0.77	0.21	5.17E-04
				7	cg17826594	-0.27	0.11	0.02	-0.28	0.11	0.01
				8	cg20245116	-2.02	0.73	0.007	-2.16	0.75	0.006
				9	cg25957967	-0.16	0.05	0.003	-0.13	0.05	0.01
				10	cg26666107	-0.40	0.10	1.65E-04	-0.39	0.11	3.64E-04
HNF1A	rs1169287	T	17	1	cg01394199	-0.38	0.12	0.03	-0.35	0.10	0.02
				2	cg03495030	2.64	0.71	0.05	2.57	0.73	0.04
				3	cg16175725	-1.03	0.33	0.04	-1.06	0.34	0.03
HNF1A	rs1169288	C	17	1	cg01341572	-0.24	0.09	0.01	-0.24	0.09	0.01
				2	cg01394199	-0.13	0.06	0.03	-0.13	0.05	0.02
				3	cg02153339	0.69	0.13	1.01E-05	0.67	0.13	1.46E-05
				4	cg10573521	0.32	0.12	0.01	0.28	0.12	0.02
				5	cg14101638	1.54	0.15	6.03E-11	1.57	0.16	7.53E-11
				6	cg23661013	0.55	0.14	2.24E-04	0.56	0.14	2.07E-04

[1]Model 1 adjusted for age, sex, study center, pedigree, and the first 4 principle components of cellulary purity and population structure;
[2]Model 2 adjusted for covariates in Model 1 plus BMI (continuous, kg/m2), smoking (2 categories: never/past vs. current smokers), physical activity (continuous, based on study-specific metric), drinking (2 categories: current vs. former/never), estrogen therapy usage (2 categories: yes/no), lipid-lowering medication usage (2 categories: yes/no), education level (categorical), total energy intake (continuous, kcal/day), dietary total fat intake (continuous, % total energy intake/day), dietary total folate intake (continuous, mcg/day), dietary VitB12 intake (continuous, mcg/day), glycemic load (quintile of g/day);
[3]β and sem represent the regression coefficient and standard error, respectively, for the expected changes in methylation per copy of coded allele.

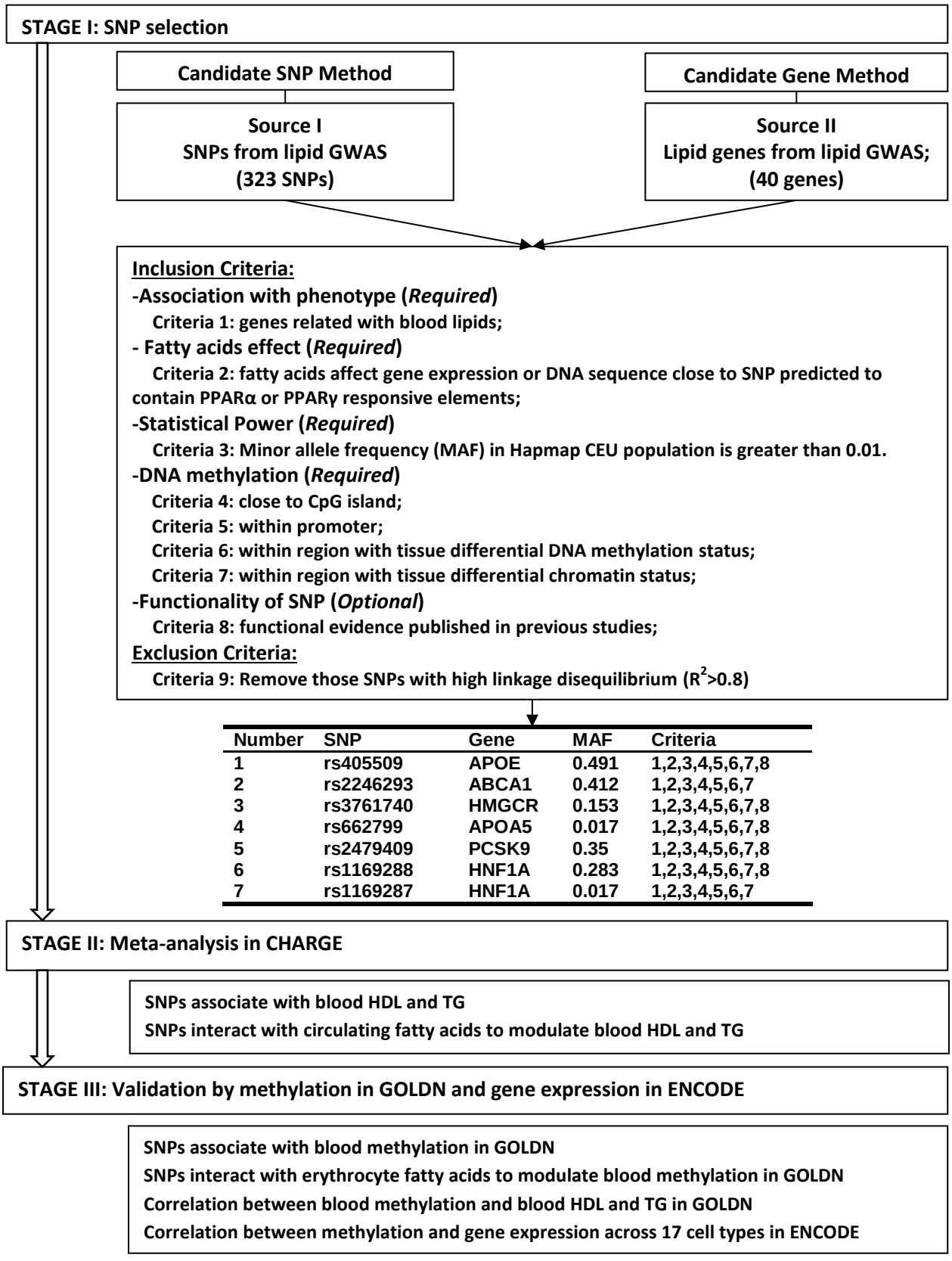


Figure 5.1. Study design of Chapter 5

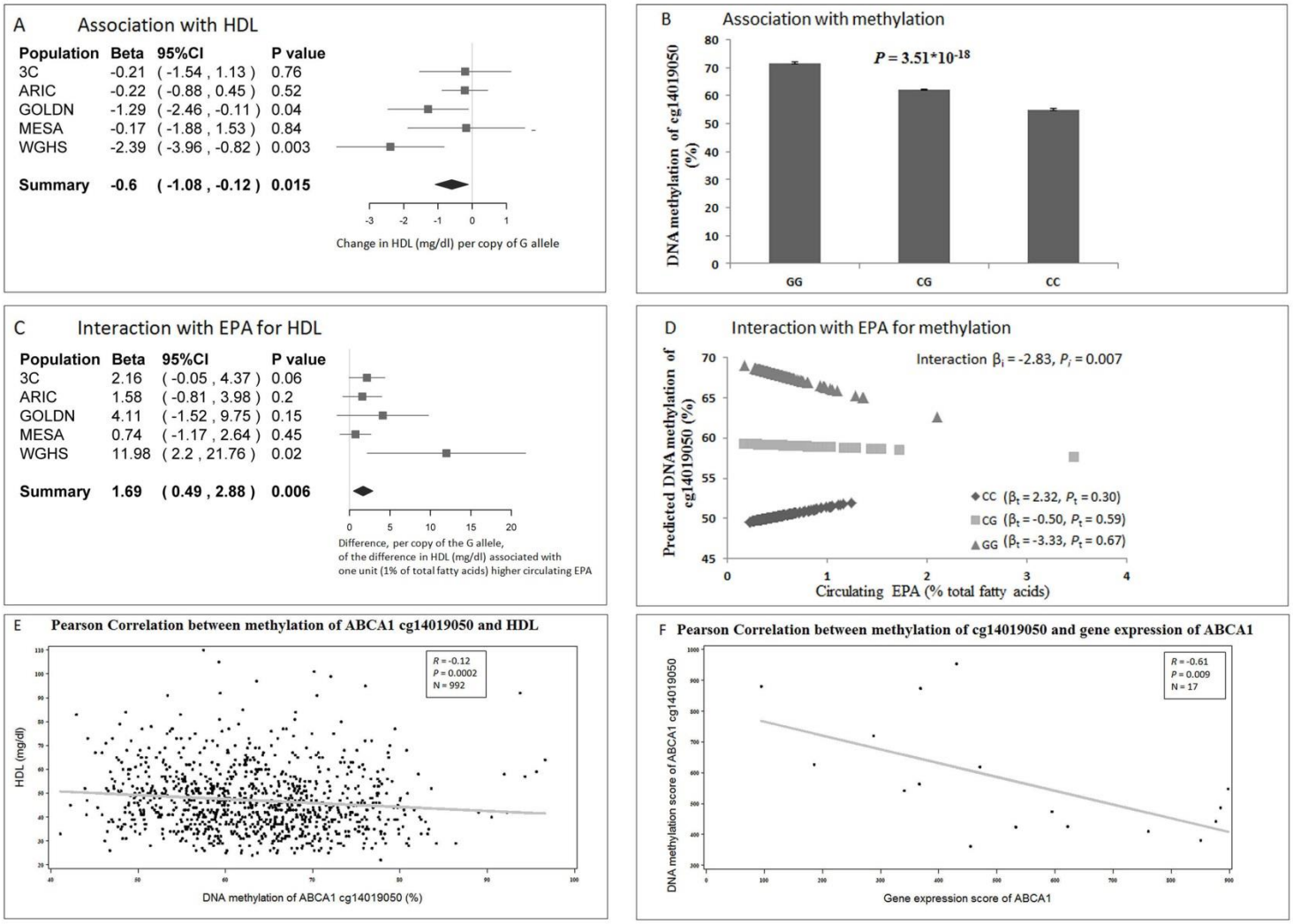


Figure 5.2. *ABCA1* rs2246293 modifies the association of EPA with HDL through genotype-dependent changes in DNA methylation of cg14019050. Genetic associations with HDL (Model 2) and methylation of *ABCA1* cg14019050 (Model 1) were presented by forest plot with meta-analysis in CHARGE consortium (A) and bar-charts by genotype in GOLDN study (B), respectively. Interactions between rs2246293 and EPA in associations with HDL (Model 1) and methylation of *ABCA1* cg14019050 (Model 1) were presented by forest plot with meta-analysis in CHARGE consortium (C) and plot with genotype-dependent regression lines in GOLDN study (D), respectively. In forest plots, estimated beta and its 95% CI, illustrating the changes in HDL per copy of the rs2246293 G allele in panel A and the association with each one percent higher level of circulating EPA with HDL per copy of the rs2246293 G allele in panel C, respectively, were represented by the filled square and horizontal line for each population, or filled diamond for the summary. In panel D, βi and *P*i represented regression coefficient and the corresponding statistical significance for the interaction term, which was interpreted as the association with each one percent higher level of circulating EPA with methylation of cg14019050 per copy of the rs2246293 G allele, while βt and *P*t represented regression coefficient and statistical significance for trend, which were interpreted as the association of circulating EPA with cg14019050 methylation according to three genotype groups of rs2246293 (GG, CG, and CC). Panel E and F illustrated the Pearson correlation of cg14019050 methylation with HDL in GOLDN (E) and with gene expression of *ABCA1* in ENCODE consortium (F), respectively, in which each black dot represented one observation, the gray line represented the regression line, and the correlation coefficient (R), *P* value and sample size (N) were displayed in the box. Model 1 adjusted for age, sex, study center, population structure/pedigree, principle components for cellular purity (only for methylation analysis), and Model 2 further adjusted for BMI, smoking, physical activity, drinking, estrogen therapy usage, lipid-lowering medication usage, education level, total energy intake, dietary intake of total fat, total folate, and vitamin B12, glycemic load (or glycemic index or whole grain intake).

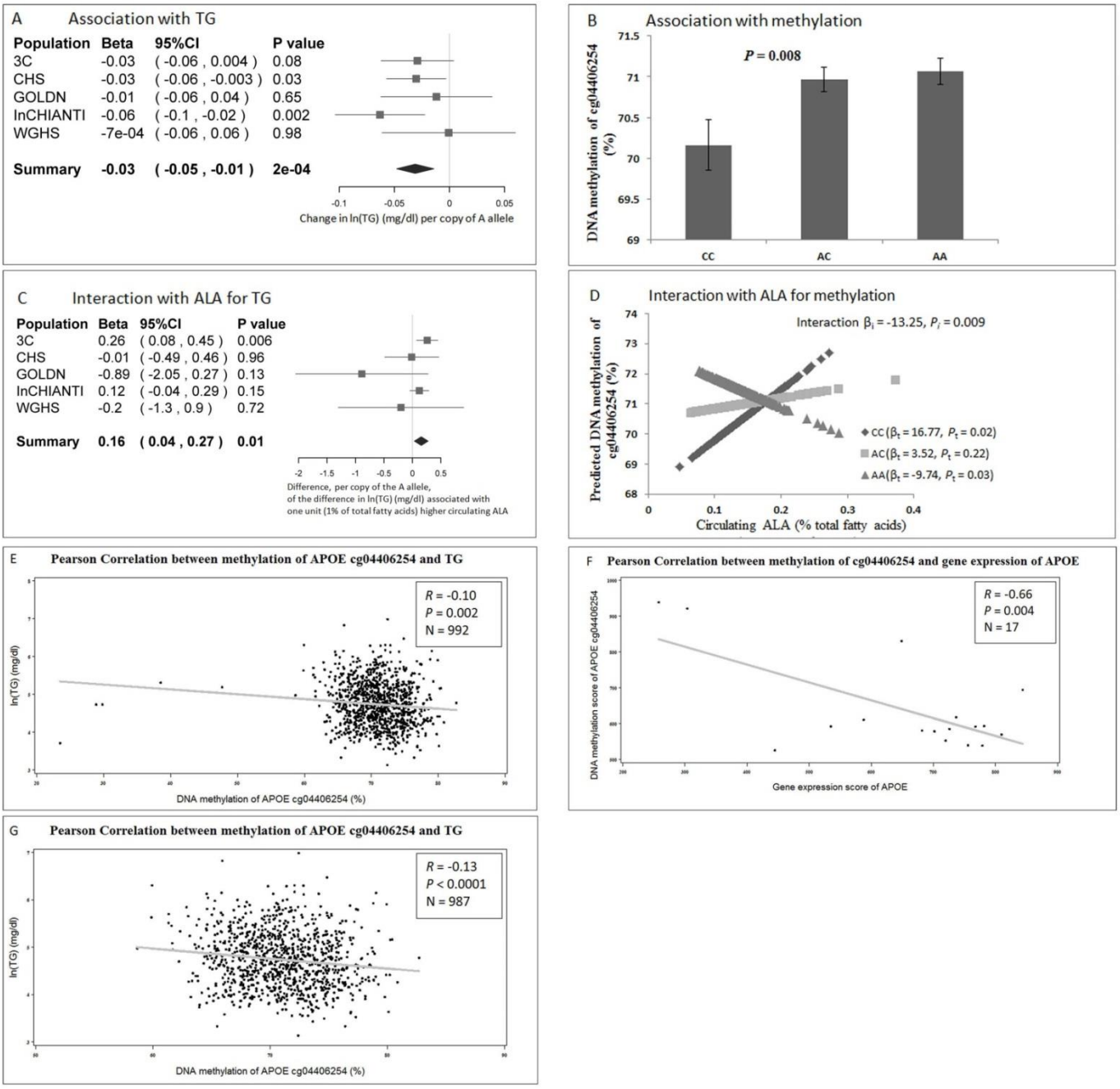


Figure 5.3. *APOE* rs405509 modifies the association of ALA with TG through genotype-dependent changes in DNA methylation of cg04406254. Genetic associations with natural logged TG (Model 1) and methylation of *APOE* cg405509 (Model 1) were presented by forest plot with meta-analysis in CHARGE consortium (A) and bar-charts by genotype in GOLDN study (B), respectively. Interactions between rs405509 and ALA in associations with natural logged TG (Model 2) and methylation of *APOE* cg04406254 (Model 1) were presented by forest plot with meta-analysis in CHARGE consortium (C) and plot with genotype-dependent regression lines in GOLDN study (D), respectively. In forest plots, estimated beta and its 95% CI, illustrating the changes in ln(TG) per copy of the rs405509 A allele in panel A and the association with each one percent higher level of circulating ALA with ln(TG) per copy of the rs405509 A allele in panel C, respectively, were represented by the filled square and horizontal line for each population, or filled diamond for the summary. In panel D, βi and *P*i represented regression coefficient and the corresponding statistical significance for the interaction term, which was interpreted as the association with each one percent higher level of circulating ALA with methylation of cg04406254 per copy of the rs405509 A allele, while βt and *P*t represented regression coefficient and statistical significance for trend, which were interpreted as the association of circulating ALA with cg04406254 methylation according to three genotype groups of rs405509 (CC, AC, and AA). Panel E, F, and G illustrated the Pearson correlation of cg04406254 methylation with ln(TG) in GOLDN (E and G) and with gene expression of *APOE* in ENCODE consortium (F), respectively, in which each black dot represented one observation, the gray line represented the regression line, and the correlation coefficient (*R*), *P* value and sample size (N) were displayed in the box. Model 1 adjusted for age, sex, study center, population structure/pedigree, principle components for cellular purity (only for methylation analysis), and Model 2 further adjusted for BMI, smoking, physical activity, drinking, estrogen therapy usage, lipid-lowering medication usage, education level, total energy intake, dietary intake of total fat, total folate, and vitamin B12, glycemic load (or glycemic index or whole grain intake).

5.7 Supplementary methods of Chapter 5

1. Study cohort description

Three-City (3C) Study: The 3C Study is a population-based prospective cohort designed to investigate the vascular risk factors for dementia. In 1999-2000, 9294 participants, aged ≥ 65 years, were recruited from communities in Bordeaux (n = 2104), Dijon (n = 4931), and Montpellier (n = 2259) in France [1]. There were 1240 participants in the current study. The protocol was approved by the Consultative Committee for the Protection of Persons participating in Biomedical Research of the Kremlin-Bicetre University Hospital (Paris). All participants signed an informed consent form.

The Atherosclerosis Risk in Communities (ARIC) Study: The ARIC study is a multi-center community-based prospective study with the purpose to study the risk factors related to atherosclerosis [2] (http://www2.cscc.unc.edu/aric/). Launched in 1987, ARIC recruited 15,792 men and women aged 45-64 years from 4 communities: Forsyth County, NC; Jackson, MI; suburban areas of Minneapolis, MN; and Washington County, MD. 3385 White participants were included in the current study. This study was approved by the Institutional Review Board at each field center. Informed consent was received by all participants or their representatives.

The Cardiovascular Health Study (CHS): The CHS is a population-based prospective cohort study of risk factors of coronary heart disease and stroke [3] (https://chs-nhlbi.org/). From a random sample of people on Medicare eligibility lists, 5201 predominantly White individuals aged ≥ 65 years were recruited in 1989-1990 and an additional 687 African-Americans were enrolled in 1992-93 from 4 centers: Forsyth County, NC; Sacramento County, CA; Washington County, MD; and Pittsburgh, PA. For the current study, a total of 2399 White participants were included. All institutional Review Board requirements were met prior to the initiation of the study at all sites. All participants signed the informed consent form.

Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) Study: The GOLDN study is a population-based intervention study to evaluate the genetic factors of triglyceride response to a diet intervention with a high fat meal and fenofibrate treatment [4] (https://dsgweb.wustl.edu/goldn/). Overall 1120 participants with Northern European origin were re-recruited from the National Heart, Lung, and Blood Institute Family Heart Study [5] from 2 study centers in Minneapolis, MN and Salt Lake City, UT, and all of them were included in the analysis. The protocol of this study was approved by the Human Studies Committee of institutional review board at the University of Minnesota, University of Utah, and Tufts University/New England Medical Center. Written informed consent was obtained from all participants.

Invescchiare in Chianti (InCHIANTI): The InCHIANTI study is a prospective population-based study of factors influencing walking ability in the elderly [6] (http://inchiantistudy.net/wp/). 1453 eligible subjects, with age ranges from 20 to 102 years, were randomly selected from the population registry of Greve in Chianti and Bagno a Ripoli in Italy and agreed to participate in the study. For the current study, 1002 participants were included. The protocol of this study was approved and informed consent from participants were obtained.

Multi-Ethnic Study of Atherosclerosis (MESA): The MESA study is a population-based cohort designed to evaluate factors related with subclinical cardiovascular disease. 6500 subjects, aged from 45 to 84 years, were enrolled from 6 regions in USA, including Baltimore City and Baltimore County, MD; Chicago, IL; Forsyth County, NC; Los Angeles County, CA; New York, NY; and St. Paul, MN [7] (http://www.mesa-nhlbi.org/).

The proportions of White, African-American, Hispanics, and Asians are 38%, 28%, 23%, and 11%, respectively. The current study included 674 eligible Whites. The study was approved by the institutional review boards of the participating centers. All participants provided written informed consent.

Women's Genome Health Study (WGHS): The WGHS study is a prospective cohort designed for genome-wide association study of the population based on the NIH-funded Women's Health Study (WHS) [8] initiated in 1992. More than 25,000 healthy women were included in the study and followed for more than 12 years for the incidence of common diseases. For the current study, 652 subjects were included. The study was approved by the institutional review board of Brigham and Women's Hospital, Boston, MA, and monitored by an external data and safety monitoring board. All participants provided written informed consent.

2. Biochemical measurements

3C: Fasting blood samples were collected at baseline to measure HDL cholesterol, triglycerides, and fatty acids. Details of the assessment of plasma fatty acids were described previously [9]. Briefly, plasma of fasting blood samples was used to extract total lipids, from which the composition of fatty acids were determined [10]. Gas chromatograph (Trace, Thermoelectron, Cergy-Pontoise, France) with a 25-m Carbowax capillary column was applied to measure the level of plasma fatty acids, expressed as a percentage of total fatty acids.

ARIC: Fasing blood samples were collected. TG and HDL cholesterol were assayed using enzymatic methods and dextran-magnesium precipitation respectively [2].Plasma fatty acids from frozen fasting blood samples were measured at the Collaborative Studies Clinical Laboratory at Fairview-University Medical Center (Minneapolis, MN, USA), using an HP-5890 gas chromatograph (Hewlett-Packard, Palo Alto, CA, USA) with a 100-meter capillary Varian CP7420 column. Details were described previously [11].

CHS: Fasting blood samples were used to measure plasma lipids and fatty acids. Plasma TG was measured by enzymatic methods on an Olympus Demand System (Olympus corp., Lake Success, NY., USA), and HDL cholesterol was measured by enzymatic method after precipitation of apolipoprotein B-containing lipoproteins with dextran sulfate/magnesium sulfate [12]. Methods of Folch was used to extract total lipids from plasma, and a one-dimentional TLC was used to separate the phospholipids from neutral lipids. Gas chromatography (Agilent5890) was used to measure the levels of plasma fatty acids, expressed as the percentage of total fatty acids. All the measurements were conducted at the Fred Hutchinson Cancer Research Center, and further confirmed by GC-MS at USDA (Peoria, IL, USA).

GOLDN: TG was measured by glycerol-blanked enzymatic method on the Roche COBAS FARA centrifugal analyzer (Roche Diagnostics Corporation, USA). HDL cholesterol was measured by cholesterol oxidase reaction (Chol R1; Roche Diagnostics, USA) after precipitation of non-HDL cholesterol with magnesium/dextran. Erythrocyte membrane levels of fatty acids were measured. Lipids were extracted from the erythrocyte membrane with a mixture of chloroform:methanol (2:1, v/v) and then collected in heptanes. Fatty acids were separated and measured using a capillary Varian CP7420 100-meter column with a Hewlett Packard 5890 gas chromatograph equipped with a HP6890A autosampler [13]. The levels of fatty acids were expressed as percentage of total fatty acids.

InCHIANTI: TG and HDL cholesterol were determined by commercial assays (Roche Diagnostics, Mannheim, Germany). Plasma fatty acids were separated from fatty acid methyl esters (FAME), which were prepared through transesterification using modified Lepage and Roy's method. HP6890 gas chromatograph (Hewlett-Packard, Palo Alto, CA, USA) with a 30-meter fused silica column (HP-225; Hewlett-Packard) was applied for FAME separation. Calibration curve with the addition of six increasing amounts of individual FAME standards (NU Chek Prep, Inc., Elysian, MA, USA) to the same internal standard (C17:0; 50xg) was used for quantitative analysis of the fatty acids, yielding the levels as percentage of total fatty acids.

MESA: TG was measured using a TG GB reagent (Roche Diagnostics, Indianapolis, IN, USA). HDL cholesterol was measured in EDTA plasma on a Vitros analyzer (Johnson & Johnson Clinical Diagnostic, Inc.) using a standard cholesterol oxidase method. Lipids were extracted from the plasma using a chloroform/methanol extraction method and the cholesterol esters, triglyceride, phospholipids and free fatty acids were separated by thin layer chromatography. Phospholipids were used to obtain the fatty acids methyl

esters, which were further detected by gas chromatography flame ionization. The level of fatty acids was expressed as the percentage of total fatty acids.

WGHS: TG and HDL cholesterol were measured by direct assay and had low coefficeints of variation [8]. Erythrocyte membrane fatty acids were extracted with a mixture of chloroform:methanol (2:1, v/v) dissolved in heptane, and separated by a capillary Varian CP7420 100-meter column with a HP5890 gas chromatograph flame ionization. The results were expressed as the percentage of total fatty acids.

3. Genotyping methodology in each cohort:

Supplementary Table 5.2. Genotyping, imputation, and quality control by each cohort

	Genotyping platform and calling method	Sample and filter criteria filter criteria	Imputation
3C	Illumina Human610-Quad BeadChips and Illuminus software	SNPs were removed if one or more following conditions are met: Hardy-Weinberg $P < 1\times10^{-6}$; minor allele frequency < 1%; and missingness > 2%.	MACH/MINIMAC using the EUR population reference haplotype data from the 1000 Genomes Project (2010 interim release based on sequence data freeze from 04 Aug 2010 and phased haplotypes from Dec 2010).
ARIC	Affymetrix 6.0 and Birdseed	Sample filter criteria: Call rate<95%; discordant with previous genotype data; genotypic and phenotypic sex mismatch; suspected first-degree relative of an included individual based on genotype data; genetic outlier as assessed by Identity by State (IBS) using PLINK and >8SD along any of the first 10 principal components in EIGENSTRAT with 5 iteractions. SNP filter criteria: Call rate<95%; HWE $P<5\times10^{-5}$; MAF<1%.	MACH with NCBI Build 36 as reference
CHS	Illumina 370CNV and BeadStudio	Sample filter criteria: Call rate<95%; 2 duplicate errors or Mendelian inconsistencies (for reference CEPH trios). SNP filter criteria: Call rate <97%; HWE $P<5\times10^{-5}$;	BIMBAM10v0.91 with reference to HapMap CEU relsease 21A, Build 35 using one round of imputations and the default expectation-maximization warm-ups and runs. SNPs were excluded for variance on the allele dosage≤0.01.
GOLDN	Affymetrix 6.0 and Birdseed	SNP filter criteria: Monomorphic SNPs; Call rate <96%; Number of families with Mendelian error by MAF of SNPs: >3 families for MAF≥20%, >2 families for MAF≥10%, >1 family for MAF≥5%, any family for MAF<5%, SNPs were set for missing in those families remained errors; HWE $P<5\times10^{-6}$. Rs405509 and rs662799 were genotyped with Applied Biosystems TaqMan SNP genotyping system.	MACH v1.0.16 with reference to Human Genome Build 36
InCHIANTI	Illumina 550K and BeadStudio	Sample filter criteria: Call rate<98.5%; Sex misspecification. SNP filter criteria: Call rate <99%; HWE $P<5\times10^{-4}$; MAF<1%.	MACH
MESA	Affymetrix 6.0 and Birdseed	Sample filter criteria: Heterozygosity>53%; individual-level genotyping call rate<95%. SNP filter criteria: Monomorphic SNP; call rate <95%.	IMPUTE v2.1.0 using HapMap Phase I and II-CEU as the reference panel (release 24-NCBI Build 36(dbSNP b126)).
WGHS	Illumina Infinium II and BeadStuido	Sample filter criteria: Call rate<95%. SNP filter criteria: Call rate <95%; MAF≤1%; HWE $P<5\times10^{-6}$.	MACH

4. Measurement of covariates in each cohort:

3C: Demographic and social-economic information were recorded at baseline through inerview. Educational levels were grouped into 4 classes, which were no education or primary school only, secondary (middle) school, high school or vocational school, and university. Dietary information was collected through a food frequency questionnaire (FFQ) and a 24h dietary recall administered by a specifically trained dietician [14, 15]. Practice of physical activity was defined as regular when doing sport regularly or having at least 1 hour of leisure or household activity per day [16].

ARIC Study: Dietary intake was obtained by a interviewer-administered semi-quantitative FFQ with 66 items, which was modified from the validated Willett 61-item FFQ [17, 18]. Education was categorized into three groups: less than high school (<12 years); completion of high school or vocational school (12 years); or more than high school (any college/professional school, >12 years) [19]. Physical activity was assessed by the modified Baecke questionnaire which defines 3 semi-continuous indices ranging from 1 (low) to 5 (high) for physical activity in sports, during leisure time, and at work [20, 21].

CHS: A self-administered 99-item picture-sort version of the National Cancer Institute FFQ and a validated 131-item Willett FFQ were used to collect the usual dietary intakes in 1989-90 and 1992-93, respectively [22, 23]. Education was grouped into <high school, high school graduate or some college, and ≥ college graduate. Physical activity was measured using a validated modified Minnesota Leisure-Time Activities Questionnaire [24, 25].

GOLDN: An interviewer-administered validated dietary history questionnaire (DHQ) was developed by the National Cancer Institute to estimate the habitual dietary intakes [26]. Education was categorized into 5 levels: grade 8 or less, 1 or more years of high school, 1 or more years of trade school, 1 or more years of college, or 1 or more years of post college. Physical activity was measured by a questionnaire with questions of the number of hours per day dedicated to different levels (heavy, slight, and sedentary) of activity [27].

InCHIANTI: The usual dietary intakes were collected by a interviewer-administered 236-item FFQ. Nutrient data for specific foods were obtained from the Food Composition Database for Epidemiological Studies in Italy [28, 29]. Education was estimated by the years of school attendance. Physical activity was categorized into 3 groups: sedentary (completely inactive or with less than 1 h/wk for light-intensity activity, light physical activity (with at least 5 h/wk or more of light activity or at least

1-2 h/wk moderate activity), and moderate to high physical activity (with at least 5 h/wk or more of light activity or at least 1-2 h/wk moderate activity).

MESA: Dietary intakes were collected by the self-administered 120-item Block FFQ, which was further modified and validated in non-Hispanic white, non-Hispanic black, Hispanic individuals [30], and Asians [31, 32]. Education was classified into 3 groups: less than high school, complete high school or equivalent certification, and complete college or more. Physical activity was collected by The MESA Typical Week Physical Activity Survey (TWPAS), which was adapted from the Cross cultural Activity Participation Study [33]. The sum of minutes spent in all activity types was multiplied by the metabolic equivalent (MET) level assigned to each activity [34].

WGHS: Between 1992 and 1994, dietary intakes were estimated using the 131-item Willett semi-quantitative FFQ derived from the National Health Service [35]. Physical activity was obtained by multiplying the average time spent per week with the metabolic equivalent (MET) values [34].

5.8 Supplementary references:

1. Group, C.S., *Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population.* Neuroepidemiology, 2003. **22**(6): p. 316-25.
2. *The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators.* Am J Epidemiol, 1989. **129**(4): p. 687-702.
3. Fried, L.P., et al., *The Cardiovascular Health Study: design and rationale.* Ann Epidemiol, 1991. **1**(3): p. 263-76.
4. Corella, D., et al., *The -256T>C polymorphism in the apolipoprotein A-II gene promoter is associated with body mass index and food intake in the genetics of lipid lowering drugs and diet network study.* Clin Chem, 2007. **53**(6): p. 1144-52.
5. Higgins, M., et al., *NHLBI Family Heart Study: objectives and design.* Am J Epidemiol, 1996. **143**(12): p. 1219-28.
6. Ferrucci, L., et al., *Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study.* J Am Geriatr Soc, 2000. **48**(12): p. 1618-25.
7. Bild, D.E., et al., *Multi-ethnic study of atherosclerosis: objectives and design.* Am J Epidemiol, 2002. **156**(9): p. 871-81.
8. Ridker, P.M., et al., *Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women.* Clin Chem, 2008. **54**(2): p. 249-55.

9. Samieri, C., et al., *Low plasma eicosapentaenoic acid and depressive symptomatology are independent predictors of dementia risk.* Am J Clin Nutr, 2008. **88**(3): p. 714-21.
10. Peuchant, E., et al., *One-step extraction of human erythrocyte lipids allowing rapid determination of fatty acid composition.* Anal Biochem, 1989. **181**(2): p. 341-4.
11. Ma, J., et al., *Plasma fatty acid composition as an indicator of habitual dietary fat intake in middle-aged adults. The Atherosclerosis Risk in Communities (ARIC) Study Investigators.* Am J Clin Nutr, 1995. **62**(3): p. 564-71.
12. Ettinger, W.H., et al., *Lipoprotein lipids in older people. Results from the Cardiovascular Health Study. The CHS Collaborative Research Group.* Circulation, 1992. **86**(3): p. 858-69.
13. Cao, J., et al., *Incorporation and clearance of omega-3 fatty acids in erythrocyte membranes and plasma phospholipids.* Clin Chem, 2006. **52**(12): p. 2265-72.
14. Feart, C., et al., *Energy, macronutrient and fatty acid intake of French elderly community dwellers and association with socio-demographic characteristics: data from the Bordeaux sample of the Three-City Study.* Br J Nutr, 2007. **98**(5): p. 1046-57.
15. Samieri, C., et al., *Dietary patterns derived by hybrid clustering method in older people: association with cognition, mood, and self-rated health.* J Am Diet Assoc, 2008. **108**(9): p. 1461-71.
16. Samieri, C., et al., *Olive oil consumption, plasma oleic acid, and stroke incidence: the Three-City Study.* Neurology, 2011. **77**(5): p. 418-25.
17. Willett, W.C., et al., *Reproducibility and validity of a semiquantitative food frequency questionnaire.* Am J Epidemiol, 1985. **122**(1): p. 51-65.
18. June Stevens, P.A.M., Barbara H. Dennis, Grethe S. Tell, Tomoko Shimakawa, Aaron R. Folsom, *Reliability of a food frequency questionnaire by ethnicity, gender, age and education.* Nutrition Research, 1996. **16**(5): p. 735-745.
19. Gottesman, R.F., et al., *Impact of differential attrition on the association of education with cognitive change over 20 years of follow-up: the ARIC neurocognitive study.* Am J Epidemiol, 2014. **179**(8): p. 956-66.
20. Baecke, J.A., J. Burema, and J.E. Frijters, *A short questionnaire for the measurement of habitual physical activity in epidemiological studies.* Am J Clin Nutr, 1982. **36**(5): p. 936-42.
21. Folsom, A.R., et al., *Physical activity and incidence of coronary heart disease in middle-aged women and men.* Med Sci Sports Exerc, 1997. **29**(7): p. 901-9.
22. Kumanyika, S., et al., *Picture-sort method for administering a food frequency questionnaire to older adults.* J Am Diet Assoc, 1996. **96**(2): p. 137-44.
23. Rimm, E.B., et al., *Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals.* Am J Epidemiol, 1992. **135**(10): p. 1114-26; discussion 1127-36.
24. Taylor, H.L., et al., *A questionnaire for the assessment of leisure time physical activities.* J Chronic Dis, 1978. **31**(12): p. 741-55.
25. Folsom, A.R., et al., *Test-retest reliability of the Minnesota Leisure Time Physical Activity Questionnaire.* J Chronic Dis, 1986. **39**(7): p. 505-11.

26. Subar, A.F., et al., *Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires : the Eating at America's Table Study.* Am J Epidemiol, 2001. **154**(12): p. 1089-99.
27. Corella, D., et al., *A high intake of saturated fatty acids strengthens the association between the fat mass and obesity-associated gene and BMI.* J Nutr, 2011. **141**(12): p. 2219-25.
28. Bartali, B., et al., *Dietary intake estimated using different methods in two Italian older populations.* Arch Gerontol Geriatr, 2004. **38**(1): p. 51-60.
29. Pisani, P., et al., *Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S152-60.
30. Mayer-Davis, E.J., et al., *Validity and reproducibility of a food frequency interview in a Multi-Cultural Epidemiology Study.* Ann Epidemiol, 1999. **9**(5): p. 314-24.
31. Nettleton, J.A., et al., *Dietary patterns are associated with biochemical markers of inflammation and endothelial activation in the Multi-Ethnic Study of Atherosclerosis (MESA).* Am J Clin Nutr, 2006. **83**(6): p. 1369-79.
32. Nettleton, J.A., et al., *Associations between dietary macronutrient intake and plasma lipids demonstrate criterion performance of the Multi-Ethnic Study of Atherosclerosis (MESA) food-frequency questionnaire.* Br J Nutr, 2009. **102**(8): p. 1220-7.
33. Ainsworth, B.E., et al., *Moderate physical activity patterns of minority women: the Cross-Cultural Activity Participation Study.* J Womens Health Gend Based Med, 1999. **8**(6): p. 805-13.
34. Ainsworth, B.E., et al., *Compendium of physical activities: an update of activity codes and MET intensities.* Med Sci Sports Exerc, 2000. **32**(9 Suppl): p. S498-504.
35. Liu, S., et al., *Fruit and vegetable intake and risk of cardiovascular disease: the Women's Health Study.* Am J Clin Nutr, 2000. **72**(4): p. 922-8.

CHAPTER 6

Discussion

6.1 Summary of the studies:

This thesis provides innovative evidence for the interplay among genetic variants, DNA methylation, and environmental factors for cardiovascular disease (CVD) related risk factors. On a genome-wide scale, we found that DNA methylation is not only related with genetic variants, but also associated with haplotypes of multiple CpG-related SNPs (CGSs). Using a candidate gene-based approach, we identified significant interactions between genetic variants and environmental factors to modulate CVD related traits, most likely mediated by the observed genotype-dependent methylation effects. Specifically, we showed that promoter SNPs interacted with age to modulate the DNA methylation of *APOE*, a candidate gene for aging, CVD, and Alzheimer's Disease as well. We also found that the genetic variants within the promoter region of *IL6*, an inflammatory cytokine, are associated with methylation of a potentially functional CpG site within that region, which was also shown to have a significant association with erythrocyte membrane level of N3 PUFAs. Finally, the methylation-related SNPs within *ABCA1* and *APOE* locus were observed to interact with circulating fatty acids to modulate blood lipids through their genotype-dependent methylation mechanisms. Our studies provide evidence that epigenetic mechanisms may play a novel role in the observed phenomenon of gene-by-environment (G×E) interactions, which may have important implications for the eventual development of personalized medicine.

6.2 Potential role of DNA methylation as one of the molecular mechanisms for G×E interactions

Despite the extensive evidence for G×E interactions, the underlying biological mechanisms are still unclear. Our results suggested that DNA methylation is one possible mechanisms for G×E interactions, which is consistent with the established regulatory role of DNA methylation as the interface between "nature" and "nurture".

DNA methylation has been demonstrated to be determined by the local nucleotide sequence. Although 25% of methylation occurs outside of CpG sites in embryonic stem cells, most of these non-CG methylation events disappear when the stem cells undergo differentiation, leading to the result that almost all of the methylation (99.98%) in differentiated mammalian cells occur on the CpG dinucleotides [1]. Manipulated insertions of the promoter constructs from genes with various DNA methylation patterns into a uniform epigenetically inert locus can recapitulate their originally patterns of DNA methylation, demonstrating the deterministic role of the genetic sequence and the proteins binding to it in controlling the regulation of DNA methylation [2]. Furthermore, the phenomenon of allele-specific methylation (ASM), suggested by observed associations between genetic variants and DNA methylation, is widespread across the human genome. For example, according to analysis of twin pairs and their parents, >35,000 CpG sites were shown to have ASM events [3]. Across 16 human pluripotent and adult cell lines, approximately 37% of heterozygous SNPs were shown to demonstrate ASM events [4]. With 153 human adult cerebellum samples, 736 out of 8590 tested CpG sites were shown to have significant cis

associations with 2878 SNPs, while these numbers in trans associations were 12 CpG sites with 38 SNPs [5]. The analysis with a three-generation family showed that 8.1% of heterozygous SNPs had cis associations with ASM events and the predicted methylation levels of CpG sites based on the genotype of those ASM-associated SNPs were highly correlated with the actual measured methylation levels in two unrelated individuals (R^2= 0.82 and 0.83) [6].

Evidence has been accumulating in support of changes to DNA methylation in response to different types of environmental factors. Studies with monozygotic and dizygotic twins suggested the potential role of environmental factors play in the regulation of DNA methylation [7, 8]. It has been shown that DNA methylation can be altered by irritant chemicals such as air pollutants and hair dyes [9-11]. Sun exposure was reported to have a tendency to lower DNA methylation on a genome-wide scale [12]. Smoking has been shown to change DNA methylation of the coagulation factor II receptor-like 3 gene (*F2RL3*) [13] and tumor suppressor genes, such as *p16* and death-associated protein *(DAP)* kinase [14]. Famine in humans caused by weather or historical events has provided us with the valuable tool of natural experiments to explore the effects of environmental factors, particularly nutrition, on DNA methylation. The rainy season in rural Gambia presents substantial challenges to the nutritional status of the Gambian individuals. Compared to those conceived during the nutrition rich season, those individuals who were conceived during the rainy season were shown to have higher methylation of genetic regions identified as “metastable epialleles”, which were defined as the regions contributing

to the dramatic and systemic inter-individual variations in epigenetic regulation [15]. Also, the Dutch Hunger Winter at the end of the Second World War provided us with unique evidence of nutritionally regulated changes in DNA methylation. Individuals who experienced famine prenatally due to *in utero* exposure to the Dutch Hunger Winter demonstrated less DNA methylation of the imprinted gene insulin-like growth factor 2 (*IGF2*) and 15 other cardio-metabolic disease genes compared with their unexposed siblings [16, 17]. Longitudinal epidemiology studies found that self-reported folic acid supplementation during pregnancy was reported to increase methylation level of imprinted gene insulin-like growth factor 2 (*IGF2*) in infants [18, 19], which was further shown to have inverse associations with birth weight [19]. Interventional studies with animals demonstrate the cause-effect relationship between different types of nutrients and DNA methylation. Dietary intake of folate and other B vitamins during gestational will affect methylation of the agouti viable yellow (A^{vy}) allele in the mouse offspring [20]. Feeding with a low protein diet during gestational in the rat was reported to increase the incidence of type 2 diabetes in the offspring by changing the DNA methylation levels in pancreatic islets of hepatocyte nuclear factor 4 a (*Hnf4a*), a transcription factor associated with the etiology of type 2 diabetes [21], and also to reduce the methylation of peroxisomal proliferator-activated receptor alpha (*PPAR-α*) gene and glucocorticoid receptor (*GR*) gene in liver [22]. A high-fat diet early in gestation was shown to result in global DNA hypomethylation in the placenta of female mice [23].

Our results consistently indicated that DNA methylation might represent one potential regulatory mechanism for the observed G×E interactions. We confirmed the genetic contribution to DNA methylation by obtaining at least one SNP-CpG pair with significant associations for each of the 7 studied genes in this thesis, including *APOE*, *IL6*, *ABCA1*, *APOA5*, *PCSK9*, *HMGCR* and *HNF1A*. Our genome-wide study further suggested that the haplotype of multiple highly linked CpG related SNPs (CGSs) were significantly associated with DNA methylation patterns in that linkage disequilibrium (LD) block. Additionally, we observed significant interactions between these methylation-related SNPs and environmental factors of interest to the current study, which are age and circulating fatty acids. For example, we found significant interactions for the promoter SNP of *APOE*, which interacted with age and α-linolenic acid (ALA), the promoter SNP of *IL6* with eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), and the promoter SNP of *ABCA1* with EPA. These interactions were not only observed for the CVD traits but also for the DNA methylation measurements of the corresponding genes. Furthermore, the results from the correlations between methylation and CVD traits and gene expression were in the same direction of the observed genetic associations and interactions. The consistent findings for the interplay among genetic variants, DNA methylation, environmental factors, CVD traits, and gene expression shown by our study suggest a potential but partial mechanistic role of DNA methylation to explain the observed G×E interactions.

6.3 Implications:

This thesis provides a model system for how to use observational studies to evaluate the hypothesis that epigenetic may represent one of the underlying molecular mechanisms for observed gene-by-environment (G×E) interactions. Our approach used several methods including bioinformatics prediction, meta-analysis of G×E interactions across multiple cohorts, association tests between genetic variants and DNA methylation (mQTL), and association analysis between genetic variants and gene expression (eQTL).

With respect to clinical implications, our study provides evidence that may eventually facilitate the implementation of genotyping testing to facilitate primary and secondary prevention of diseases. Currently, it is virtually impossible to change the genotype status of the risk variant one individual carries. In addition, informing the patient of his/her genotyping status may in some cases cause anxiety and other distress, and the desirability of the routine, clinical use of genotypes for common SNPs in the clinical setting of primary or secondary prevention remains controversial. *APOE* is one example, in that the ε4 variant was demonstrated to have a dosage effect on the incidence of and on the age of onset of the late-onset Alzheimer's disease (AD) [24]. However, debates persist over whether the genotyping test for *APOE* ε4 is necessary or desirable, because there are no medications or clinical strategies to counter the deleterious effect of the ε4 isoforms [25-27]. However, our finding that ε4 is associated with *APOE* methylation and expression suggest that the deleterious

effects of ε4 might be mitigated by applying appropriate lifestyle-based modifiers that reduce the difference in methylation across different APOE isoforms.

6.4 Limitations:

This thesis does have some limitations which need to be addressed. Firstly, this study is based on an observational design, so that a cause-effect relationship cannot be established. Although we have observed relationships between SNPs and DNA methylation, we can not determine which is the cause and which is the effect. On one hand, it is possible that the CpG site formed by the SNPs leads to methylation of this site. On the other hand, it is also possible for the DNA methylation event to create or disrupt a CpG site by causing mutation at that site. The cause-and-effect relationship between DNA methylation and blood lipids is also unclear. However, the effect of N3 PUFAs on DNA methylation and then on gene expression might be plausible based on the fact that most N3 PUFAs comes from dietary/supplement intake and the main effects of DNA methylation is changes in gene expression. Secondly, the tissue for DNA methylation in our study is the immune cells, which are ideal for the measurement of inflammatory markers but not of blood lipids. However, the fact plasma lipids are affected, to some degree, by immune cells, i.e. the reverse cholesterol pathway by macrophages, may reduce the limitations associated with using immune cells rather than hepatic cells. Also, the consistent methylation pattern

across different cell types, exemplified by the analysis of *APOE* locus, mitigates limitations related to tissue specificity.

6.5 Future direction:

The results from this thesis illustrate the need for future mechanistic experiments, such as interventional studies and molecular studies, to increase the level of evidence supporting our hypothesis that genotype-dependent epigenetic changes are the underlying molecular mechanism for the observed G×E interactions. With mechanistic experiments, cause-and-effect relationships will be explored and our hypothesis of the genotype-specific effects of fatty acids on DNA methylation-gene expression-blood lipids can be more conclusively demonstrated.

6.6 Conclusion:

We found that DNA methylation is related to genetic variants by showing haplotypes of multiple CpG related SNPs (CGSs) that were associated with the DNA methylation patterns on a genome-wide scale. Our studies with candidate loci related to CVD risk factors suggested that genotype-dependent effects on DNA methylation may contribute, in part, to the underlying molecular mechanisms for observed G×E interactions. Our studies call for further demonstrations with interventional studies and molecular mechanistic experiments, with the ultimate goal of providing reliable

evidence to advance the development of more personalized approaches to nutrition and medical care.

6.7 References:

1. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-22.
2. Lienert, F., et al., *Identification of genetic elements that autonomously determine DNA methylation states.* Nat Genet, 2011. **43**(11): p. 1091-7.
3. Schalkwyk, L.C., et al., *Allelic skewing of DNA methylation is widespread across the genome.* Am J Hum Genet, 2010. **86**(2): p. 196-212.
4. Shoemaker, R., et al., *Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome.* Genome Res, 2010. **20**(7): p. 883-9.
5. Zhang, D., et al., *Genetic control of individual differences in gene-specific methylation in human brain.* Am J Hum Genet, 2010. **86**(3): p. 411-9.
6. Gertz, J., et al., *Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation.* PLoS Genet, 2011. **7**(8): p. e1002228.
7. Fraga, M.F., et al., *Epigenetic differences arise during the lifetime of monozygotic twins.* Proc Natl Acad Sci U S A, 2005. **102**(30): p. 10604-9.
8. Wong, C.C., et al., *A longitudinal study of epigenetic variation in twins.* Epigenetics, 2010. **5**(6): p. 516-26.
9. Baccarelli, A., et al., *Rapid DNA methylation changes after exposure to traffic particles.* Am J Respir Crit Care Med, 2009. **179**(7): p. 572-8.
10. Bollati, V., et al., *Changes in DNA methylation patterns in subjects exposed to low-dose benzene.* Cancer Res, 2007. **67**(3): p. 876-80.
11. Langevin, S.M., et al., *The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood.* Epigenetics, 2011. **6**(7): p. 908-19.
12. Gronniger, E., et al., *Aging and chronic sun exposure cause distinct epigenetic changes in human skin.* PLoS Genet, 2010. **6**(5): p. e1000971.
13. Breitling, L.P., et al., *Tobacco-smoking-related differential DNA methylation: 27K discovery and replication.* Am J Hum Genet, 2011. **88**(4): p. 450-7.
14. Belinsky, S.A., et al., *Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers.* Cancer Res, 2002. **62**(8): p. 2370-7.
15. Waterland, R.A., et al., *Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles.* PLoS Genet, 2010. **6**(12): p. e1001252.
16. Heijmans, B.T., et al., *Persistent epigenetic differences associated with prenatal exposure to famine in humans.* Proc Natl Acad Sci U S A, 2008. **105**(44): p. 17046-9.

17. Tobi, E.W., et al., *DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific.* Hum Mol Genet, 2009. **18**(21): p. 4046-53.
18. Hoyo, C., et al., *Methylation variation at IGF2 differentially methylated regions and maternal folic acid use before and during pregnancy.* Epigenetics, 2011. **6**(7): p. 928-36.
19. Steegers-Theunissen, R.P., et al., *Periconceptional maternal folic acid use of 400 microg per day is related to increased methylation of the IGF2 gene in the very young child.* PLoS One, 2009. **4**(11): p. e7845.
20. Waterland, R.A. and R.L. Jirtle, *Transposable elements: targets for early nutritional effects on epigenetic gene regulation.* Mol Cell Biol, 2003. **23**(15): p. 5293-300.
21. Sandovici, I., et al., *Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the Hnf4a gene in rat pancreatic islets.* Proc Natl Acad Sci U S A, 2011. **108**(13): p. 5449-54.
22. Lillycrop, K.A., et al., *Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring.* J Nutr, 2005. **135**(6): p. 1382-6.
23. Gallou-Kabani, C., et al., *Sex- and diet-specific changes of imprinted gene expression and DNA methylation in mouse placenta under a high-fat diet.* PLoS One, 2010. **5**(12): p. e14398.
24. Corder, E.H., et al., *Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families.* Science, 1993. **261**(5123): p. 921-3.
25. Gordon, S.C. and D. Landa, *Disclosure of the genetic risk of Alzheimer's disease.* N Engl J Med, 2010. **362**(2): p. 181-2; author reply 182.
26. Kane, R.A. and R.L. Kane, *Effect of genetic testing for risk of Alzheimer's disease.* N Engl J Med, 2009. **361**(3): p. 298-9.
27. Green, R.C., et al., *Disclosure of APOE genotype for risk of Alzheimer's disease.* N Engl J Med, 2009. **361**(3): p. 245-54.

APPENDIX

Lipoprotein lipase variants interact with polyunsaturated fatty acids for obesity traits in women: replication in two populations

Y. Ma[a], K.L. Tucker[b], C.E. Smith[a], Y.C. Lee[a], T. Huang[c], K. Richardson[a], L.D. Parnell[a], C.Q. Lai[a], K.L. Young[d], A.E. Justice[e], Y. Shao[e], K.E. North[f], J.M. Ordovás[a,g,h*]

[a] Jean Mayer US Department of Agriculture Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA

[b] Clinical Laboratory and Nutritional Sciences, University of Massachusetts, Lowell, MA, USA

[c] Department of Nutrition, Harvard School of Public Health, Boston, MA, USA

[d] Department of Epidemiology and Carolina Population Center, University of North Carolina, Chapel Hill, NC, USA

[e] Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA

[f] Department of Epidemiology and Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC, USA

[g] Department of Epidemiology, Centro Nacional Investigaciones Cardiovasculares (CNIC), Madrid, Spain.

[h] Instituto Madrileño de Estudios Avanzados en Alimentación (IMDEA-FOOD), Madrid, Spain.

Nutrition, Metabolism and Cardiovascular Diseases, 2014,
http://dx.doi.org/10.1016/j.numecd.2014.07.003

Abstract:

Background and aims

Lipoprotein lipase (*LPL)* is a candidate gene for obesity based on its role in triglyceride hydrolysis and the partitioning of fatty acids towards storage or oxidation. Whether dietary fatty acids modify *LPL* associated obesity risk is unknown.

Methods and results

We examined five single nucleotide polymorphisms (SNPs) (rs320, rs2083637, rs17411031, rs13702, rs2197089) for potential interaction with dietary fatty acids for obesity traits in 1171 participants (333 men and 838 women, aged 45-75 y) of the Boston Puerto Rican Health Study (BPRHS). In women, SNP rs320 interacted with dietary polyunsaturated fatty acids (PUFA) for body mass index (BMI) (*P*=0.002) and waist circumference (WC) (*P*=0.001) respectively. Higher intake of PUFA was associated with lower BMI and WC in homozygotes of the major allele (TT) (*P*=0.01 and 0.005) but not in minor allele carriers (TG and GG). These interactions were replicated in an independent population, African American women of the Atherosclerosis Risk in Communities (ARIC) study (n=1334).

Conclusion

Dietary PUFA modulated the association of *LPL* rs320 with obesity traits in two independent populations. These interactions may be relevant to the dietary management of obesity, particularly in women.

Key words:

gene-diet interaction; lipoprotein lipase; polyunsaturated fatty acids; obesity**.**

Running title: *LPL* interacts with PUFA for obesity in women

Introduction

Obesity in the US has reached an overall prevalence of nearly 34% [1], with greater prevalence in some ethnic minorities [2], which might be related to differences in genetic background and behavioral factors [3-5]. The investigation of genetic variants for obesity in conjunction with behavioral factors, especially diet, may benefit development of more specific strategies to ameliorate susceptibility to weight gain.

Lipoprotein lipase (*LPL*) is a candidate gene for obesity, based on its encoded function to absorb fatty acids across tissues[6, 7]. *LPL* contributes to fat storage in adipocytes [8], regulation of thermogenesis in skeletal muscle [9]. However, in spite of LPL's demonstrated role in obesity, relevant association studies with *LPL* single nucleotide polymorphism (SNP) have inconsistent findings and show sex-specific differences [10, 11].

One hypothesis that may account for the inconsistency is that unexamined factors may modulate *LPL*-associated obesity risk. Dietary fat type (e.g., saturated fatty acids (SFA), monounsaturated fatty acids (MUFA), and polyunsaturated fatty acids (PUFA)) have been evaluated for obesity risk independently of genotype for decades [4]. However, it remains to be explored whether intakes of different fatty acids alter obesity-related traits in the context of *LPL* genotype.

Therefore, we aimed to determine whether dietary fatty acids interact with *LPL* variants for obesity traits in a population of multiple ancestries, stratified by sex. We also aimed to replicate our findings in an independent population.

Methods

Study Populations

Discovery population: The Boston Puerto Rican Health Study (BPRHS)

In the BPRHS, there were 1171 participants of Puerto Rican origin, aged 45-75 years, living in the Greater Boston, MA metropolitan area, after excluding those with missing data and implausible energy intake, defined as <2512 kJ (600 kcal) per day or >20093 kJ (4800 kcal) per day. Details for the study have been described previously [12]. Fasting blood were collected for biochemical and genetic analyses. Anthropometric methods were consistent with techniques used by the National Health and Nutrition Examination Surveys. The study protocol was approved by the Institutional Review Board at Tufts Medical Center and Tufts University Health Sciences Campus. Informed consent was received by all participants or their representatives.

Replication population: Atherosclerosis Risk in Communities Study (ARIC)

Participants of replication study in ARIC included 2186 African American (AA) and 8689 European American (EA), considering the multiple ancestry nature of the BPRHS. ARIC was a multi-center study with participants aged 44-66 years from Forsyth County, North Carolina; Jackson, Mississippi; suburban areas of Minneapolis, Minnesota; and Washington County, Maryland [13]. Individuals with implausible energy intakes, defined as in the BPRHS, were excluded from analysis. Body weight was measured using a calibrated scale with subjects in scrub suits without shoes and height was measured using a ruler. Waist circumference (WC) at the umbilicus was measured using a tape measure. Fasting blood was collected from an antecubital vein into a vacuum tube with ethylenediamine tetraacetic acid. Triglycerides and high-density lipoprotein (HDL) were assayed using enzymatic methods and dextran-magnesium precipitation respectively [14]. This study was approved by the Institutional Review Board at each field center, and the University of North Carolina at Chapel Hill. Informed consent was received by all participants or their representatives.

SNP Selection, Genotyping and Linkage disequilibrium (LD) analysis

Four lipids related SNPs (rs2083637, rs17411031, rs13702, and rs2197089) [15-17], and rs320 (common name as *HindIII*) with inconsistent associations with obesity [10, 11] were selected as discovery panel tested in BPRHS. Genotyping in BPRHS was performed using the ABI TaqMan SNP genotyping system 7900HT (Applied Biosystems, Foster City, CA). Hardy-Weinberg equilibrium (HWE) was evaluated by Chi-square tests. LD and haplotype was analyzed by HaploView4.2 [18] according to 1000 Genomes Project. Genotypes of replication SNP rs327 in ARIC was imputed by MACH (v1.0.16) [19] with HapMap r22 reference populations, Utah residents with Northern and Western European Ancestry (CEU) and Yoruba in Ibadan, Nigeria (YRI) based on the genome-wide SNP data obtained by the Affymetrix 6.0 chip (Affymetrix, Santa Clara, CA).

Dietary assessment

The BPRHS used a semi-quantitative food frequency questionnaire (FFQ) [20]. The ARIC study used a modified 66-item interviewer-administered FFQ [21]. Dietary fatty acids intake were expressed as a percentage of total energy intake.

Population ancestry admixture

The population admixture of participants in the BPRHS was estimated with reference to three ancestral populations including Native American (15%), Southern European (57%), and West African (27%), and the major principal component estimated by EIGENSTRAT was adjusted in the analysis [22]. The first 10 principal components, estimated using Eigensoft, represent admixture for ARIC EAs. Percentage of European ancestry for ARIC AAs was estimated based on the reference population of CEU using 1350 ancestry informative markers by ANCESTRYMAP [23].

Statistical Analysis

Analysis of covariance and general linear models were applied to test genetic associations and interactions between SNPs and different types of dietary fat intake (SFA, MUFA and PUFA) to modulate body mass index (BMI) and WC, assuming an additive genetic model and adjusting age, admixture, smoking, drinking, total fat intake (interaction test), total energy intake, antilipemic medication, diabetes status, hormone replacement therapy (in women), physical activity, and education. Dietary

fat intake was analyzed as both continuous and categorical variables, dichotomized by the population median, for validation. Log transformation was performed when necessary. Data were analyzed using SAS 9.2 (SAS Institute, Inc. Cary, NC). A two-tailed *P*-value of <0.02 was considered statistically significant, adjusting for three multiple tests introduced by three SNPs reaching HWE. Successful replication was considered when the analysis with the same additive model and covariates resulted in the significant interactions in the same direction across different populations. Inverse-variance weighted fixed-effects meta-analyses were conducted using METAL (University of Michigan; www.sph.umich.edu/csg/abecasis/metal/) to summarize the interactions across populations.

Bioinformatics analysis

Sequences surrounding the candidate SNP with significant interactions were screened via NUBIScan [24], a software evaluated peroxisome proliferator-activated receptors response elements (PPREs), for which fatty acids (especially PUFAs) are the primary ligand.

Results

Population characteristics

Demographic data, anthropometrics, lipids, nutrient intakes, disease status and medication use for both the BPRHS and ARIC are presented by sex (**Table A.1**). Women in the BPRHS and in ARIC AAs were characterized for obesity status, using the cutoff of BMI ≥ 30 kg/m^2 for obesity and cutoff of WC ≥ 88 cm in women for abdominal obesity. Dietary PUFA contributed 8.5% of energy intake in the BPRHS and 5% in ARIC, and it was similar for men and women within each cohort.

Discovery in BPRHS

Minor allele frequencies (MAF) were greater than 0.25 (**Table A.2**). Rs2083637 and rs17411031 deviated from HWE (*P*=0.03 and *P*=0.02) and thus were excluded in the following analysis.

Rs320, rs13702, and rs2197089 were not associated with BMI and WC (**Supplementary Table A.1**). Both categorical PUFA (**Supplementary Table A.2)** and continuous PUFA (**Figure A.1**) showed significant interaction with rs320 for BMI (*P*=0.02 and *P*=0.002, respectively) and WC (*P*=0.02 and *P*=0.001, respectively). In homozygotes of the major allele (TT), higher PUFA intake associated with lower BMI (*P*=0.01) and smaller WC (*P*=0.005) (Figure A.1). However, in carriers of minor allele (TG and GG), there were no associations (P>0.1) (Figure A.1). Significance and interaction pattern were similar for N6- and N3-PUFA (Supplementary Table A.2). Results remained the same after excluding women (n=22) using hormone replacement (Data not shown). No significant interactions were observed in men (Figure A.1 and Supplementary Table A.2), for SFA or MUFA, or with rs13702 or rs2197089 (data not shown). Data reported below apply to the single SNP rs320.

Replication in ARIC

Rs327 was selected for replication in ARIC because of its close distance to rs320 (459 bp), the high imputation quality (R^2=1), and its strong LD with rs320 in CEU, African ancestry in the Southwest USA (ASW), and Mexican ancestry in Los Angeles (MXL) populations in the 1000 Genomes Project (R^2>0.8) (**Figure A.2**).

Interaction with continuous PUFA in AA women was significant for WC (P=0.01) and approached significance for BMI (P=0.07). Interactions were not significant in AA men or EAs (**Figure A.2**), and similar results were with categorical PUFA (**Supplementary Figure A.1**).

In a meta-analysis of the women from all three populations, rs320/rs327 showed significant interaction with continuous PUFA intake for both BMI (P=0.009) and WC (P=0.003) (**Figure A.2**). However, interaction P values were even lower when the meta-analysis included only two women population, BPRHS and ARIC AAs (P=0.001 for BMI and $P=4\times10^{-5}$ for WC). The betas for the interaction term were consistently negative in three populations. There were no interactions in men (Figure A.2). Analysis with categorical PUFA exhibited similar results (**Supplementary FigureA. 1**).

Bioinformatics analysis

There was a PPRE approximately 14 kb downstream of rs320 (**Figure A.3**).

Discussion

In the current study, we identified and replicated significant interactions between dietary PUFA intake and *LPL* rs320 for obesity traits in women. In the homozygosity for the major allele (TT), increase in PUFA intake associated with lower BMI and WC. Our observed gene-diet interaction may have improved our ability to detect dietary effects that are apparent only under certain genotype.

Genetic modulation of dietary associations has been suggested by the studies for plasma lipids [25, 26]. TT group showed greater improvement in plasma lipids with calorie restriction intervention [25]. Our group also found that TT genotype had a greater postprandial lipidemia response following an oral fat load [26]. These may suggest that T-carriers are more responsive to diet than G-carriers, which is consistent with our findings for PUFA intake. Considering the existence of the PPRE near rs320, we hypothesize that the apparently greater responsiveness of T allele to PUFA may involve the higher binding affinity to the transcription factor of this allele compared to G allele [27]. The anti-obesity effects of PUFA found are consistent with previous studies [28], which may be related to decreased adipocyte LPL activity [29], reduced triglyceride storage and less fat deposition.

Our sex-specific finding is consistent with previous study [7], and may be due to the sex-specific *LPL* response. *LPL* expression in muscle was 160% higher in women than men [6]. Our finding of null genetic association with obesity is consistent with a US study [11] not a French one [10], which may further support our hypothesis of gene-diet interaction due to potential differences in dietary pattern.

Limitations include the cross-sectional design and the absence of adipose LPL expression measurement. The variable ancestral admixtures and high prevalence of obesity in both minority populations (BPRHS and ARIC AAs) restricted our study from drawing conclusions about the relative contributions of genetic background, behavioral and socio-economic factors to the observed interactions. As our results are preliminary, additional studies are desirable.

In summary, we detected a significant interaction between *LPL* rs320 and dietary PUFA for obesity-related outcomes in women, and these findings were strengthened by replication in a second, independent population. Our observations may be relevant to the nutritional management of obesity in women.

References

1. Flegal, K.M., et al., *Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010.* JAMA, 2012. **307**(5): p. 491-7.
2. Zhang, H. and R. Rodriguez-Monguio, *Racial disparities in the risk of developing obesity-related diseases: a cross-sectional study.* Ethn Dis, 2012. **22**(3): p. 308-16.
3. Cheng, C.Y., et al., *Admixture mapping of obesity-related traits in African Americans: the Atherosclerosis Risk in Communities (ARIC) Study.* Obesity (Silver Spring), 2010. **18**(3): p. 563-72.
4. Moussavi, N., V. Gavino, and O. Receveur, *Could the quality of dietary fat, and not just its quantity, be related to risk of obesity?* Obesity (Silver Spring), 2008. **16**(1): p. 7-15.
5. Kirby, J.B., et al., *Race, place, and obesity: the complex relationships among community racial/ethnic composition, individual race/ethnicity, and obesity in the United States.* Am J Public Health, 2012. **102**(8): p. 1572-8.
6. Kiens, B., et al., *Lipid-binding proteins and lipoprotein lipase activity in human skeletal muscle: influence of physical activity and gender.* J Appl Physiol, 2004. **97**(4): p. 1209-18.
7. Corella, D., et al., *Gender specific associations of the Trp64Arg mutation in the beta3-adrenergic receptor gene with obesity-related phenotypes in a Mediterranean population: interaction with a common lipoprotein lipase gene variation.* J Intern Med, 2001. **250**(4): p. 348-60.
8. Gonzales, A.M. and R.A. Orlando, *Role of adipocyte-derived lipoprotein lipase in adipocyte hypertrophy.* Nutr Metab (Lond), 2007. **4**: p. 22.
9. Kusunoki, M., et al., *NO-1886 (ibrolipim), a lipoprotein lipase activator, increases the expression of uncoupling protein 3 in skeletal muscle and suppresses fat accumulation in high-fat diet-induced obesity in rats.* Metabolism, 2005. **54**(12): p. 1587-92.
10. Jemaa, R., et al., *Lipoprotein lipase gene polymorphisms: associations with hypertriglyceridemia and body mass index in obese people.* Int J Obes Relat Metab Disord, 1995. **19**(4): p. 270-4.
11. Nicklas, B.J., et al., *Lipoprotein lipase gene variation is associated with adipose tissue lipoprotein lipase activity, and lipoprotein lipid and glucose concentrations in overweight postmenopausal women.* Hum Genet, 2000. **106**(4): p. 420-4.
12. Tucker, K.L., et al., *The Boston Puerto Rican Health Study, a longitudinal cohort study on health disparities in Puerto Rican adults: challenges and opportunities.* BMC Public Health, 2010. **10**: p. 107.

13. Monda, K.L., C.M. Ballantyne, and K.E. North, *Longitudinal impact of physical activity on lipid profiles in middle-aged adults: the Atherosclerosis Risk in Communities Study.* J Lipid Res, 2009. **50**(8): p. 1685-91.
14. *The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators.* Am J Epidemiol, 1989. **129**(4): p. 687-702.
15. Aulchenko, Y.S., et al., *Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts.* Nat Genet, 2009. **41**(1): p. 47-55.
16. Wallace, C., et al., *Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia.* Am J Hum Genet, 2008. **82**(1): p. 139-49.
17. Deo, R.C., et al., *Genetic differences between the determinants of lipid profile phenotypes in African and European Americans: the Jackson Heart Study.* PLoS Genet, 2009. **5**(1): p. e1000342.
18. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**(2): p. 263-5.
19. Li, Y., et al., *MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes.* Genet Epidemiol. **34**(8): p. 816-34.
20. Tucker, K.L., et al., *Adaptation of a food frequency questionnaire to assess diets of Puerto Rican and non-Hispanic adults.* Am J Epidemiol, 1998. **148**(5): p. 507-18.
21. Hu, F.B., et al., *Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire.* Am J Clin Nutr, 1999. **69**(2): p. 243-9.
22. Lai, C.Q., et al., *Population admixture associated with disease prevalence in the Boston Puerto Rican health study.* Hum Genet, 2009. **125**(2): p. 199-209.
23. Patterson, N., et al., *Methods for high-density admixture mapping of disease genes.* Am J Hum Genet, 2004. **74**(5): p. 979-1000.
24. Podvinec, M., et al., *NUBIScan, an in silico approach for prediction of nuclear receptor response elements.* Mol Endocrinol, 2002. **16**(6): p. 1269-79.
25. Jemaa, R., et al., *Hind III polymorphism of the lipoprotein lipase gene and plasma lipid response to low calorie diet.* Int J Obes Relat Metab Disord, 1997. **21**(4): p. 280-3.
26. Lopez-Miranda, J., et al., *The influence of lipoprotein lipase gene variation on postprandial lipoprotein metabolism.* J Clin Endocrinol Metab, 2004. **89**(9): p. 4721-8.
27. Chen, Q., et al., *Functional significance of lipoprotein lipase HindIII polymorphism associated with the risk of coronary artery disease.* Atherosclerosis, 2008. **200**(1): p. 102-8.
28. Whigham, L.D., A.C. Watras, and D.A. Schoeller, *Efficacy of conjugated linoleic acid for reducing fat mass: a meta-analysis in humans.* Am J Clin Nutr, 2007. **85**(5): p. 1203-11.
29. Baltzell, J.K., J.T. Wooten, and D.A. Otto, *Lipoprotein lipase in rats fed fish oil: apparent relationship to plasma insulin levels.* Lipids, 1991. **26**(4): p. 289-94.

Tables

Table A.1. Participant characteristics in the BPRHS and ARIC, by sex[a].

	Discovery		Replication			
	BPRHS		ARIC African American		ARIC European Americans	
	Men (n=333)	Women (n=838)	Men (n=852)	Women (n=1334)	Men (n=4170)	Women (n=4519)
General Characteristics						
Age, year	57 (8)	58 (7)	53.92(6)	53.2(6)	54.66(6)	53.9(6)
Current smoker, n	111 (33%)	172 (21%)	310(36%)	335(25%)	1009(24%)	1142(25%)
Current drinker, n	158 (47%)	290 (35%)	423(50%)	280(21%)	2961(71%)	2818(62%)
Anthropometric measures						
BMI, kg/m^2	29.6 (5.1)	32.9 (7.0)	28(4.8)	30.7(6.6)	27.4(4.0)	26.5(5.4)
WC, cm	102 (15)	102 (16)	98(13)	100(16)	100(10)	93(15)
Lipids						
Triglycerides, mmol/l	1.66 (1.77)	1.55 (1.63)	1.4(1.25)	1.24(0.74)	1.66(1.13)	1.45(0.93)
Total cholesterol, mmol/l	4.47 (1.14)	4.87 (1.05)	5.47(1.11)	5.62(1.16)	5.45(0.99)	5.64(1.1)
HDL cholesterol, mmol/l	1.03 (0.31)	1.21 (0.31)	1.29(0.42)	1.5(0.44)	1.11(0.32)	1.49(0.44)
LDL cholesterol, mmol/l	2.57 (0.91)	2.86 (0.89)	3.57(1.06)	3.57(1.13)	3.61(0.91)	3.49(1.02)
Dietary Intake						
Total energy, kcal/d	2483 (883)	2044 (875)	1732(657)	1514(569)	1799(652)	1515(522)
Carbohydrate, % total energy/d	48.9 (7.5)	51.5 (7.7)	48.5(9.1)	50.5(9.6)	47.3(9.1)	49.5(9.3)
Protein, % total energy/d	16.8 (2.9)	17.2 (3.5)	17.2(3.9)	18.7(4.5)	17.1(3.7)	18.5(4.2)
Total fat, % total energy/d	32.9 (5.7)	31.7 (5.5)	32.0(6.3)	32.1(6.5)	33.6(6.8)	32.8(6.8)
SFA, % total energy/d	10.1 (2.5)	9.6 (2.3)	11.4(2.7)	11.4(2.7)	12.4(3.0)	12.1(3.1)
MUFA, % total energy/d	11.4 (2.1)	11.0 (2.1)	12.6(2.8)	12.5(2.9)	13.0(3.0)	12.4(3.0)
PUFA, % total energy/d	8.6 (2.0)	8.5 (2.0)	4.7(1.2)	4.9(1.3)	5.1(1.5)	5.1(1.5)
Cholesterol, mg/d	377 (191)	283 (162)	312(162)	250(124)	269(139)	223(104)
Disease and medication						
Have diabetes, n	137 (41%)	331 (40%)	156(18%)	256(19%)	405(10%)	334(7%)
Take antilipemic medication, n	133 (40%)	343 (41%)	6(1%)	7(1%)	84(2%)	87(2%)
Take anti-depressants, n	84 (25%)	323 (39%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Take hormone replacement therapy in women, n	0 (0%)	22 (3%)	0 (0%)	211(16%)	0 (0%)	984(22%)

[a]Data are means (standard deviation) or n (%)

Table A.2. Minor allele frequency (MAF) of *LPL* SNPs in the BPRHS[a].

		Men	Women	*P*[b]
rs320	MAF	0.29	0.27	0.08
	TT	154	418	
	TG	135	332	
	GG	22	50	
rs2083637	MAF	0.28	0.24	0.03
	AA	166	454	
	AG	131	334	
	GG	25	32	
rs17411031	MAF	0.28	0.24	0.02
	CC	162	446	
	CG	130	332	
	GG	24	31	
rs13702	MAF	0.36	0.35	0.70
	TT	125	334	
	TC	142	370	
	CC	41	93	
rs2197089	MAF	0.39	0.41	0.25
	AA	123	286	
	AG	149	390	
	GG	50	140	

[a]Data are n;
[b]*P*: *P* value for Hardy-Weinberg Equilibrium testing.

Supplementary Table A.1. Obesity traits by genotype and gender in BPRHS[a].

Anthropometric measures	SNP	Genotype			P^b
BMI, kg/m^2	rs320	TT	TG	GG	
	men	29.6(0.4)	29.7(0.4)	29.8(1.0)	0.99
	women	32.9(0.3)	32.7(0.4)	34.1(0.9)	0.4
	rs13702	TT	TC	CC	
	men	29.7(0.4)	29.5(0.4)	30.1(0.8)	0.81
	women	32.8(0.4)	32.8(0.3)	32.9(0.7)	0.99
	rs2197089	AA	AG	GG	
	men	30.3(0.4)	29.1(0.4)	29.4(0.7)	0.12
	women	32.9(0.4)	32.8(0.3)	33.0(0.6)	0.99
Waist circumference, cm	rs320	TT	TG	GG	
	men	101.8(1.1)	102.8(1.2)	100.2(3.0)	0.64
	women	101.3(0.7)	101.8(0.8)	103.8(2.1)	0.5
	rs13702	TT	TC	CC	
	men	101.5(1.2)	102.7(1.1)	101.8(2.1)	0.74
	women	101.1(0.8)	101.8(0.7)	100.7(1.5)	0.73
	rs2197089	AA	AG	GG	
	men	102.9(1.2)	101.4(1.1)	101.2(1.9)	0.62
	women	101.3(0.9)	101.6(0.7)	102.1(1.3)	0.85

[a]Data are means (standard error);
[b]*P*: *P* value for genetic association between SNPs and anthropometric traits adjusting age, admixture, smoking, drinking, total energy intake, diabetes status, antilipemic medication, hormone replacement therapy in women, physical activity, and education.

Supplementary Table A.2. Discovery of categorical interaction between rs320 and dietary PUFAs for obesity traits in BPRHS by sex[a].

			Women				Men			
***LPL* rs320**			TT	TG	GG		TT	TG	GG	
			(n=418)	(n=332)	(n=50)	P1[b]	(n=154)	(n=135)	(n=22)	*P*1[b]
PUFA Intake, % of total energy intake	<8.6	BMI, kg/m^2	33.6(0.5)	32.3(0.6)	32.3(1.3)	0.02	30.4(0.6)	29.9(0.6)	30.8(2.1)	0.86
	≥8.6		32.2(0.5)	33.3(0.6)	35.0(1.6)		29.0(0.6)	29.3(0.6)	29.3(2.1)	
			*P*2[c] 0.06	0.26	0.27		0.13	0.57	0.68	
	<8.6	WC, cm	103.3(1.1)	102.0(1.3)	101.2(2.9)	0.02	103.9(1.6)	102.7(1.9)	100.6(4.6)	0.52
	≥8.6		99.6(1.0)	101.5(1.4)	104.3(3.6)		100.1(1.5)	102.5(1.96)	102.1(4.6)	
			*P*2[c] 0.02	0.82	0.56		0.11	0.93	0.85	
N6-PUFA Intake, % of total energy intake	<7.9	BMI, kg/m^2	33.5(0.5)	32.3(0.6)	32.7(1.3)	0.04	30.3(0.6)	29.7(0.6)	30.8(2.1)	0.81
	≥7.9		32.3(0.5)	33.3(0.6)	34.3(1.5)		29.2(0.6)	29.5(0.6)	29.3(2.1)	
			*P*2[c] 0.10	0.26	0.47		0.22	0.85	0.68	
	<7.9	WC, cm	102.8(1.0)	101.9(1.3)	101.9(2.8)	0.06	103.7(1.6)\	102.3(1.9)	100.6(4.6)	0.46
	≥7.9		100.0(1.1)	101.7(1.4)	103.2(3.3)		100.3(1.5)	103.0(1.9)	102.1(4.6)	
			*P*2[c] 0.08	0.9	0.8		0.16	0.81	0.85	
N3-PUFA Intake, % of total energy intake	<0.7	BMI, kg/m^2	32.9(0.5)	33.0(0.5)	30.4(1.6)	0.05	30.1(0.6)	29.7(0.6)	30.1(1.6)	0.83
	≥0.7		32.9(0.5)	32.4(0.5)	35.4(1.2)		29.1(0.6)	29.5(0.6)	30.0(2.2)	
			*P*2[c] 0.99	0.43	0.03		0.29	0.78	0.97	
	<0.7	WC, cm	102.5(1.1)	102.6(1.2)	97.9(3.5)	0.03	102.6(1.5)	101.9(1.9)	100.3(3.4)	0.43
	≥0.7		100.5(1.0)	101.0(1.3)	105.6(2.7)		100.9(1.6)	103.3(1.9)	103.2(4.7)	
			*P*2[c] 0.19	0.40	0.13		0.48	0.62	0.65	

[a]Data are means (standard error);
[b]*P*1: P for interaction between diet and SNP adjusting age, sex, physical activity, admixture, smoking, drinking, total energy intake, total fat intake, diabetes status, antilipemic medication, hormone replacement therapy in women, and education;
[c]*P*2: P for trend for dietary intake in each genotype group adjusting same covariates;

Figures

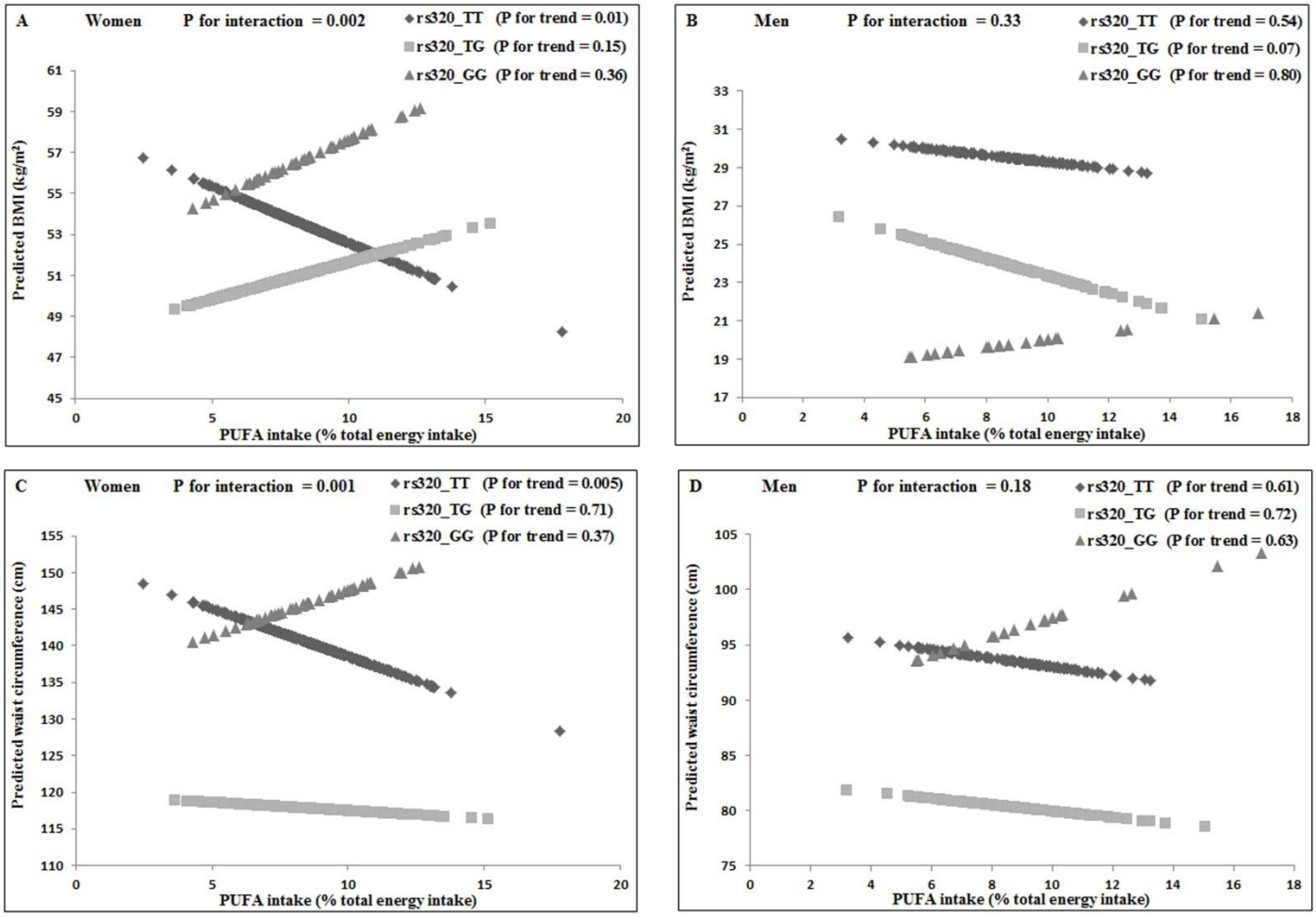


Figure A.1 Discovery of continuous interaction between rs320 and dietary PUFA intake for obesity traits in the BPRHS, by sex. Predicted BMI (A for women, B for men) and WC (C for women, D for men) by rs320 genotype were plotted against continuous dietary PUFA adjusted for age, admixture, smoking, drinking, total energy intake, total fat intake, diabetes status, antilipemic medication, hormone replacement therapy in women, physical activity, and education. *P* values indicate the statistical significance of the adjusted interaction term and adjusted regression coefficients in the regression line corresponding to three genotype groups of rs320 (TT, TG, and GG).

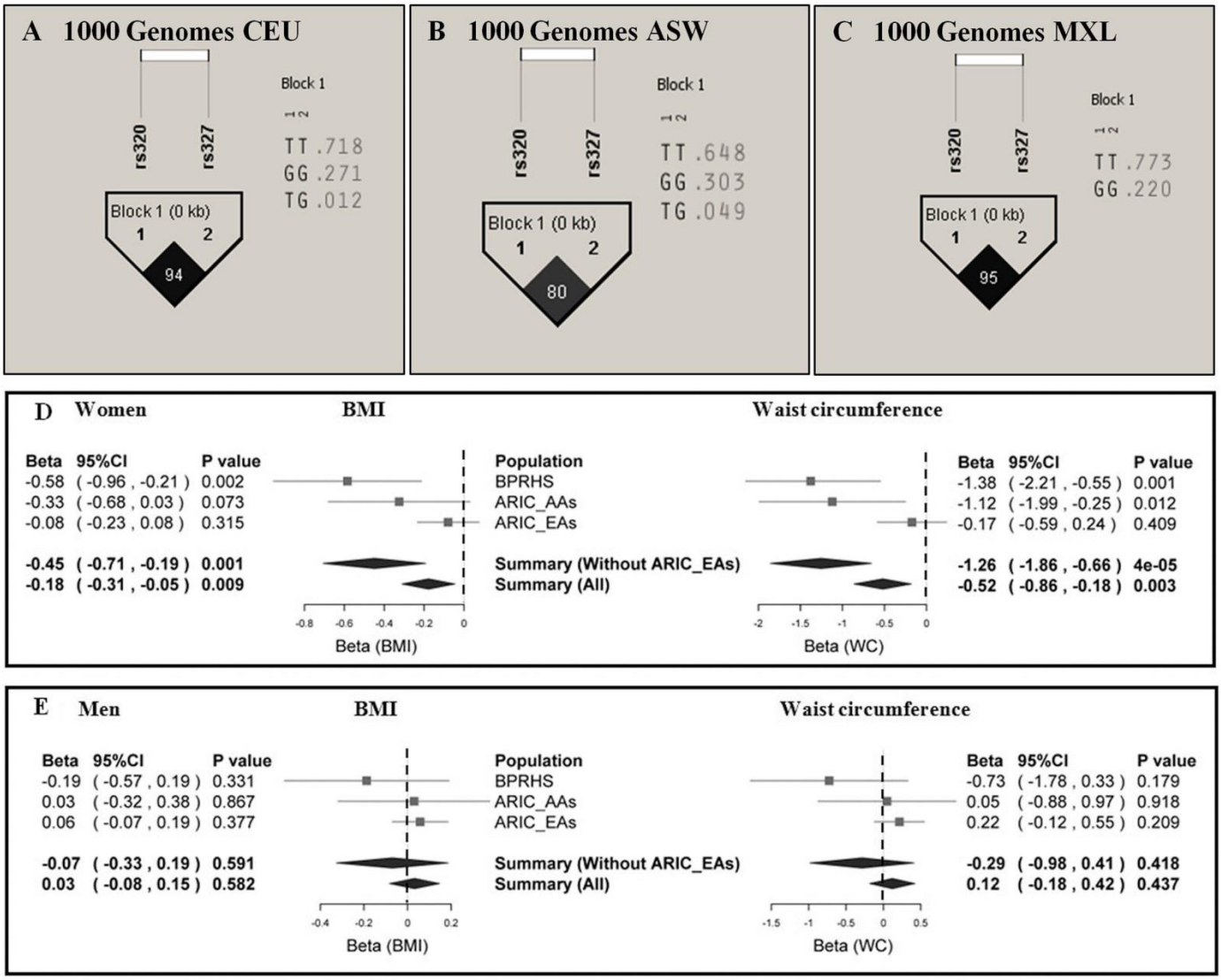


Figure A.2 Replication and meta-analysis of continuous interaction between rs320/rs327 and dietary PUFA intake for obesity traits by sex. LD correlation ($100\times R^2$ value in diamond), haplotypes and corresponding frequencies of rs320 and replication SNP rs327 in 1000 Genomes CEU (A), ASW (B) and MXL (C) are presented. Forest plots of this continuous interactions modulating BMI (left) and WC (right) in women (D) and men (E) are presented. The estimates represent the difference in the magnitude of the PUFA association (per +1-unit intake of PUFA) with BMI or WC per copy of the T allele of rs320/rs327, adjusting for age, admixture, smoking, drinking, total energy intake, total fat intake, diabetes status, antilipemic medication, hormone replacement therapy in women, physical activity, and education. Filled square and horizontal line represent the estimate and 95% CI for each population, and the filled diamond represent the summary estimate and its 95% CI.

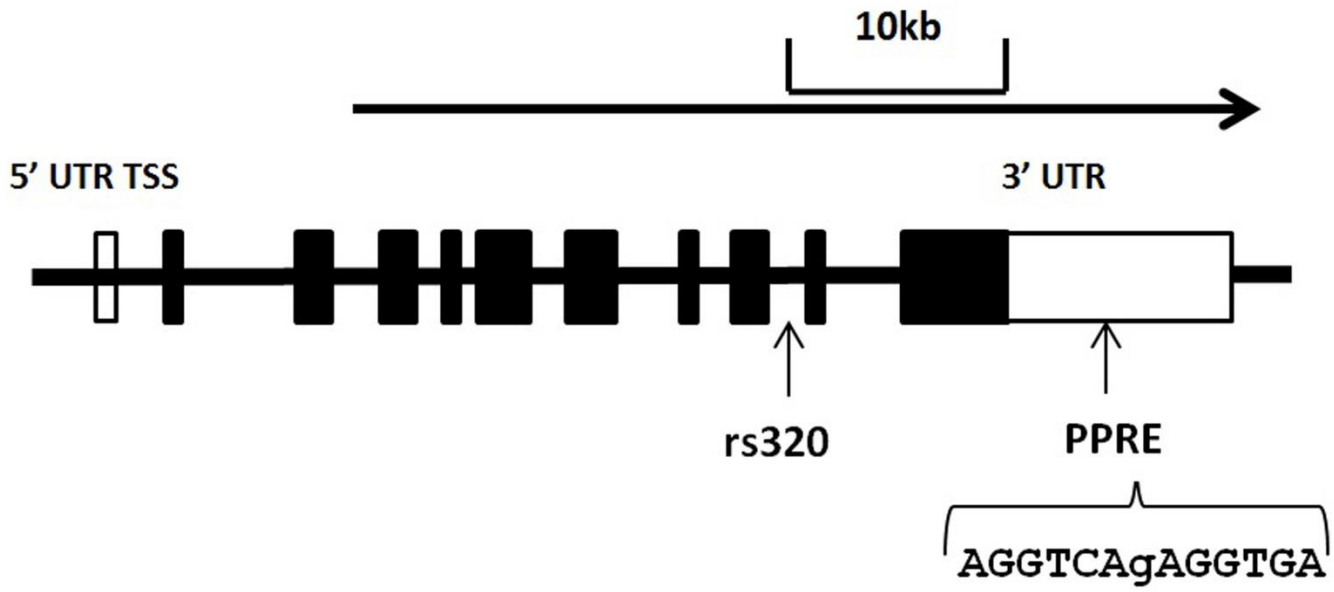
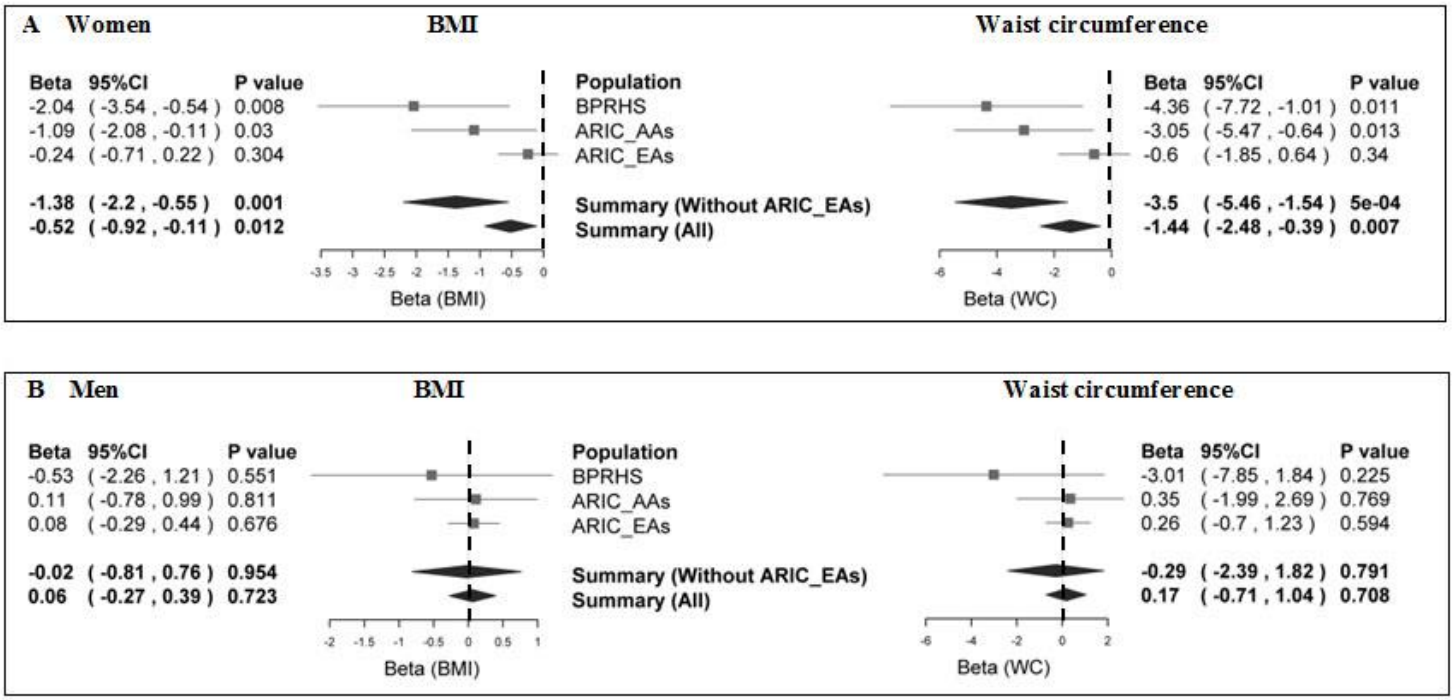


Figure A.3 Genomic structure of *LPL* locus. The gene is transcribed as indicated by the large horizontal arrow. Exons are the larger rectangular black boxes.



Supplementary Figure A.1 Forest plots of categorical interactions between rs320/rs327 and dietary PUFA intake for obesity traits, by women (A) and men (B). The estimates represent the difference in the magnitude of the PUFA association (higher than population median vs. lower than population median) with BMI or WC per copy of the T allele of rs320/rs327, adjusting for age, admixture, smoking, drinking, total energy intake, total fat intake, diabetes status, antilipemic medication, hormone replacement therapy in women, physical activity, and education. Filled square and horizontal line represent the estimate and 95% CI for each population, and the filled diamond represent the summary estimate and its 95% CI.