Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_______________________________ _______________
Yu-Hsuan Chuang Date

Association of DNA Methylation Markers with C-reactive protein in African Americans

By

Yu-Hsuan Chuang
MPH

Department of Epidemiology

____

Yan Sun
Committee Chair

____

Committee Member

____

Committee Member

____

Committee Member

Association of DNA Methylation Markers with C-reactive protein in African Americans

By

Yu-Hsuan Chuang

B.S.
Taipei Medical University
2009

Thesis Committee Chair: Yan Sun, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2012

Abstract

Association of DNA Methylation Markers with C-reactive protein in African Americans
By Yu-Hsuan Chuang

C-reactive protein (CRP), a nonspecific acute-phase protein in response to inflammatory reaction, has been shown highly correlated with increased risks for cardiovascular disease in epidemiological studies. DNA methylation is an epigenetic mechanism regulating gene expression. Investigating the association of DNA methylation sites with serum CRP levels may improve understanding of the etiology of inflammation and cardiovascular disease. A cross sectional study was used to determine the association between serum levels of CRP and gene-specific DNA methylation among African Americans. The study population consisted of 966 African Americans from 492 hypertensive sibships of the Genetic Epidemiology Network of Arteriopathy (GENOA) study. The GENOA data included demographic information and biomarkers of cardiovascular disease. DNA methylation data of 27,578 DNA methylation sites were obtained from Illumina Infinium HumanMethylation27 BeadChip using DNA samples extracted from peripheral blood cell. Linear mixed models were applied to identify gene-specific DNA methylation sites associated with the serum levels of CRP, and adjust for relatedness and potential confounders, including age, gender, BMI, and smoking status. Two hundred and fifty seven (1.12%) DNA methylation sites are significantly associated with serum level of CRP correcting for 22,927 tests using Bonferroni method. Moreover, among 257 significant DNA methylation sites, 80.5% (n=207) of the significant DNA methylation sites were hypomethylated with higher CRP levels. Analysis of the gene ontology showed that the CRP-associated DNA methylation sites were enriched by genes involved in inflammation and immune response. (Fisher Exact p-value= 4.50×10^{-4} and 2.10×10^{-15} respectively)

Association of DNA Methylation Markers with C-reactive protein in African Americans

By

Yu-Hsuan Chuang

B.S.
Taipei Medical University
2009

Thesis Committee Chair: Yan Sun, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2012

Acknowledgements

I would like to thank a number of people who have assisted me with writing this thesis. I thank Dr. Sun for conversations that clarified my thinking, which trained me to be an independent researcher. I am grateful to my colleagues, Rose and Weiyan, for the discussions which helped me develop the ideas. My parents, friends, and family also provided emotional and spiritual support. I would like to express my thanks to them too.

Table of Contents

INTRODUCTION……………………………………………………………………...1

LITERATURE REVIEW………………………………………………………….3

METHODS…………………………………………………………………………...9

RESULTS…………………………………………………………………………….13

DISCUSSION………………………………………………………………………..16

REFERENCES……………………………………………………………………….19

TABLES………………………………………………………………………………...27

INTRODUCTION

C-reactive protein (CRP) is a nonspecific acute-phase protein produced by the liver in response to inflammatory reaction.[1] Recent studies show that high serum CRP levels are correlated with increased risks for cardiovascular disease[2], which has been the leading cause of death in the United States since 1991.[3] Also, CRP is stable enough to measure as a long-term predictor of coronary heart disease (CHD).[4] In Ridker's study, reducing the serum CRP levels by Rosuvastatin among people without hyperlipidemia was observed to significantly lower the incidence of major cardiovascular events, including myocardial infarction, stroke, unstable angina, and sudden cardiac death.[5] Moreover, the association between CRP and cardiovascular disease is not only well established in population-based studies, but the functional involvement of CRP in CHD is demonstrated.[6] CRP plays a role in binding with oxidized phosphocholine on low-density lipoprotein (LDL) to increase the ability of phogocytosis by macrophage.[6] When there is too much LDL in the blood, LDL passes through endothelial cells to enter the intima and undergoes oxidation. Oxidized-LDL is toxic and induces an inflammatory reaction, which in turn increases the serum level of CRP.

There are several environmental factors correlated with serum levels of CRP, such as age, gender, race, socioeconomic status, diet, behavioral factors (e.g. poor physical fitness or low levels of physical activity, smoking), and medical conditions, including hypertension, congestive heart failure, obesity, and metabolic syndrome.[7] Additionally, Pankow's study indicated that as much as 40% of the total variation in the serum CRP level is due to genetic factors.[8] Some genetic association studies have shown significant correlations between single nucleotide polymorphisms (SNP) on metabolic and inflammatory regulation genes (e.g. LEPR, HNF1A, IL6R, and GCKR) and serum levels of CRP[6,9], but the causal role for SNPs in cardiovascular disease remains unclear.[6,10]

DNA methylation sites are measurable epigenetic markers which may contribute to disease risks and can be modified by age and other physical conditions.[11] A significant inverse association was observed between IL-6 methylation and the serum CRP level among people with depression in Uddin's study.[12] In order to better understand the etiology of CHD and inflammatory response, our study aims to identify gene-specific DNA methylation sites associated with serum CRP levels in 966 African Americans.

LITERATURE REVIEW

C-reactive protein (CRP), produced by the liver, is one of major positive acute phase inflammatory markers in humans.[13] The serum level of CRP increases when there is inflammation throughout the body. The elevated CRP level is measureable within 4-5 hours after an inflammatory stimulus, and the elevation ranges widely from 5 times to 1000 times normal CRP values.[13] An elevated level of CRP can be caused by any forms of inflammation, such as bacterial infections (e.g. pneumococcal pneumonia, tuberculosis), inflammatory diseases (e.g. rheumatoid arthritis, lupus, cardiovascular diseases), and cancer.[14] Although serum CRP tests cannot specifically check which locations of the body undergo the inflammatory reaction, a CRP test is included in routine blood examinations in clinics and hospitals to measure the level of systematic inflammation, especially for cardiovascular disease prediction. According to the American Heart Association, a CRP level less than 0.1 mg/dl, between 0.1 and 0.3 mg/dl, and more than 0.3 mg/dl, are three categories to differentiate low risk, average risk, and high risk of developing cardiovascular diseases.[14] People with higher levels of CRP are more likely to develop cardiovascular diseases. The link of serum CRP levels and cardiovascular diseases has been established in several epidemiological studies since 1990's. A population-based prospective cohort study by Arima found that the risk of coronary heart disease in the highest CRP quartile group was 2.98 times higher compared to that in the lowest CRP group among individuals aged 40 years or older after adjusting for other cardiovascular risk factors, including age, gender, body mass index (BMI), smoking habits, alcohol intake and regular exercise.[15] A positive relationship between an elevated CRP level and atherosclerosis has also been found in young adults in the Pathobiological Determinants of Atherosclerosis in Youth study.[16] Moreover, according to the Heart and Soul Study, coronary heart disease patients with elevated CRP levels have

a higher risk of being hospitalized due to heart failure, independent of prior history of heart failure, medication use, and severity of coronary heart disease.[17]

Among all of the biomarkers of cardiovascular disease, CRP has the strongest correlation. CRP is not only stable enough within the circulation system to be a long-term predictor of coronary heart disease, but also has been identified as a predictor of morbidity independent of LDL cholesterol values.[4,18] A long-term observational study by Ridker found that even after adjusting for all major cardiovascular disease risk factors, including LDL cholesterol levels, the positive influence of plasma CRP levels on cardiovascular events remains.[18] Ridker suggested that CRP is a stronger predictor of cardiovascular events relative to the LDL cholesterol values based on the following evidence:[18] The risk of first cardiovascular events among apparently healthy American women with the highest CRP quintile is 2.3 times higher than that among women with the lowest CRP quintile, but the corresponding relative risk of LDL cholesterol level between the highest quintile and the lowest is only 1.5 times higher.[18] Furthermore, a significant decrease of the incidence of major cardiovascular events through lowering blood CRP levels by Rosuvastatin was shown among people without hyperlipidemia in the JUPITER trial (Justification for the Use of Statins in Primary Prevention).[19] These studies emphasize the important role that CRP plays in cardiovascular disease prevention.

However, some scholars in the field believe that CRP is more than just an inflammatory marker that increases the risk of cardiovascular disease.[20] CRP was shown to participate in the atherosclerosis process by involving foam cell formation and increasing LDL cholesterol uptake by macrophage in Zwaka's study.[21] CRP also facilitates the ability of phogocytosis by macrophage through binding with oxidized phosphocholine on LDL cholesterol.[6] In addition, Fay found that high CRP levels are associated with increasing the risk of myocardial infarction by promoting thrombosis, with a focus on the effects of CRP on hemostasis, platelet function, and fibrinolysis.[22]

The evidence of CRP as a mediator of atherothrombotic diseases was demonstrated in a clinical trial. After infusion of CRP into the blood, the coagulation and inflammatory pathways are activated.[23] In conclusion, CRP is not only a predictor of cardiovascular disease but also can be a therapeutic target.

DNA methylation is an epigenetic mechanism that regulates gene function through the addition of methyl groups to cytosine, resulting in the formation of 5-methyl-cytosine (5mC).[24] Epigenetics is the study of inheritable change in gene expression without altering nucleotide base to another.[25] Recent studies have suggested that epigenetic differences may be associated with cardiovascular diseases. Newman has hypothesized that global DNA hypomethylation is associated with atherosclerosis due to elevated plasma homocysteine values.[26] Elevated plasma homocysteine, which is a risk factor for atherosclerosis, causes DNA hypomethylation by inhibiting DNA methyltransferase.[27] An epigenetic study by Cash found that LINE-1 hypomethylation in peripheral blood was associated with higher values of LDL cholesterol, which is also associated with atherosclerosis and high cardiovascular risk.[28] But a correlation between DNA methylation in a LINE-1 repetitive element and plasma CRP levels was not seen in Baccarelli's study.[29] Stenvinkel found that the ratio of HpaII/ MspI is correlated with blood CRP values and increases cardiovascular mortality in chronic kidney disease patients.[30] Although there are quite a few global DNA methylation studies aiming to investigate the etiology of cardiovascular diseases, mechanisms linking DNA methylation to increase cardiovascular risk are still unclear. A DNA methylation study focusing on inflammatory related genes by Kim found that Forkhead box P3 (FOXP3) DNA hypermethylation is related to elevated blood CRP levels.[31] In addition, a relationship of IL-6 hypomethylation and elevated plasma CRP levels has been found among individuals with lifetime depression by Uddin from the Detroit Neighborhood Health Study.[12] These studies support the idea that the epigenetic marker, DNA methylation, is associated with the regulation of serum CRP levels at both the global level and the gene-specific level.

Several risk factors of CRP levels, such as age, gender, BMI, and smoking habits, have been identified. Increased serum levels of CRP are associated with older age in several studies across a large span of age groups. A general adult population study including individuals aged 25-64 years by Hutchinson found that the plasma CRP level in the oldest (55-64 years) was approximately two times higher than the plasma CRP level in the youngest (25-34 years).[32] The increasing linear trend of CRP with age could also be seen among U.S. children and young adults 3-19 years of age in the National Health and Nutrition Examination Survey, 1999-2000.[33] In epidemiological studies targeting the elderly population, the association of CRP and age remain. A 20-year follow-up study of elderly men aged 60-79 years in the British Regional Heart Study showed that 57% of the median CRP increased between the 60-64 year age group and the 75-79 year age group.[34] Also, Caparevic indicated that healthy women over 56 years have higher levels of CRP compared with healthy women younger than 45 years.[35] The influence of age on the CRP level can be observed in different age groups and different gender groups.

The mean of CRP levels varies widely between males and females.[36] Overall, females have higher CRP levels and wider CRP ranges compared with males.[36] Data from multiethnic study of atherosclerosis (MESA) showed that African American females aged 45-85 years have double the median CRP compared to African American males with similar ages.[37] When matching the BMI of males and females, the significant gender differences in CRP levels can also be seen within different BMI groups: BMI<25 kg/m^2 and BMI>30 kg/m^2.[38] In addition, Wener suggests a different upper reference limit of CRP values for males and females 25-70 years old based on the Third National Health and Nutrition Evaluation Survey.[39] The upper reference limit of CRP values for females is 0.6 mg/dl more than that for males of the same age. (CRP = age/50 mg/dl for males, and CRP = age/50 +0.6 mg/dl for females)[39] This gender difference of serum CRP levels is maintained after

adjusting for age, BMI, smoking habits, and other cardiovascular disease risk factors, and is consistent across ethical groups.[37]

A higher level of CRP is positively correlated with obesity and higher BMI. A linear correlation between CRP and BMI after adjustments for age and gender was demonstrated in Pieroni's study.[38] The median CRP levels increased from 0.41 mg/dl for individuals with BMI less than 25 kg/m^2 to 3.00 mg/dl for individuals having BMI more than 30 kg/m^2.[38] The evidence of this correlation can be found in both epidemiological studies focusing on adults and studies on children and adolescents. A case control study by Huffman found that both higher waist circumference and BMI have a strong association with an increasing CRP level among Cuban Americans over 30 years old, controlling for age, gender, and smoking status.[40] Moreover, a study by MacKenzie found that the positive correlation between BMI and serum CRP levels is consistent over time, and the correlation is statistically significant among children with Type 1 Diabetes, but not among healthy children.[41]

There is also epidemiologic evidence showing that smoking is related to an elevated CRP concentration in serum. Current smokers have higher CRP levels compared to past smokers and non-smokers.[42] A positive dose-response association was found by Madsen between CRP levels and the amount of current cigarette smoking per day, and between CRP levels and the number of pack-years of smoking.[42] The dose effect of cigarette smoking on an elevated CRP level has also been shown in adolescent studies and in elderly populations.[43,44] Moreover, Ahonen's study showed that the CRP level is significantly higher in smoking men compared to non-smoking men, but there is no significant difference in CRP between smoking women and non-smoking women.[45] The association is particularly stronger among males relative to females.[45] However, the gender difference of cigarette smoking and CRP concentrations was not seen in Tsai's study focusing on older Taiwanese.[46] Several studies have also found a negative relationship between smoking cessation and

CRP levels. A cohort study by Ohsawa found that male past smokers with long smoking cessation (≧5 year) have a significantly lower mean CRP level compared to those male smokers who have smoking cessation less than 5 years, and their CRP level is similar to the mean CRP in non-smokers.[47]

Other potential risk factors for serum CRP levels include ethnicity, low socioeconomic status, high fat and low fiber consumption, poor physical fitness or low levels of physical activity, and medical conditions, such as hypertension, congestive heart failure, obesity, and metabolic syndrome.[7]

METHODS

Study Purpose:

Analyses were performed to determine if there are any associations between serum levels of CRP and gene-specific DNA methylation among African Americans. The null hypothesis of this study is that there is no significant association between the gene-specific DNA methylation and serum levels of CRP among African Americans after controlling for potential confounders.

Study Design:

A cross-sectional study design was used to test the hypothesis. The data for this analysis was from the Genetic Epidemiology Network of Arteriopathy (GENOA) study. The GENOA study was a longitudinal cohort study of hypertensive sibships that had the goal of identifying genetic predictors of hypertension.[48] The African American cohort of the GENOA study was recruited from Jackson, Mississippi from 1995-2001. Inclusion criteria for the GENOA cohort study were that at least two siblings had been diagnosed with essential hypertension before age 60. They were defined as a sibship. The diagnosis of hypertension was based on blood pressure levels measured at the study visit (systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg), or a prior diagnosis of hypertension by a physician and current use of antihypertensive medications. Exclusion criteria for the GENOA cohort study were potential causes of secondary hypertension, such as previously diagnosed kidney disease. The GENOA study collected data of demographic risk factors (e.g. age, gender, BMI, history of smoking, and history of disease and medication), biomarkers of cardiovascular diseases (e.g. CRP, homocysteine, fibrinogen, cholesterol, high-density lipoprotein cholesterol (HDL-C), triglyceride, systolic and diastolic blood pressure, etc.), and DNA methylation data. Questionnaires were used in the GENOA study to obtain demographic information. The values of biomarkers were collected by physical examination and

laboratory tests using blood samples. DNA methylation data containing 27,578 DNA methylation sites were obtained from Illumina Infinium HumanMethylation27 BeadChip (Illumina, San Diego, CA) using DNA samples extracted from peripheral blood cell.[49]

Study Population:

The African American cohort in the GENOA study had 1,482 participants. Subjects were excluded from analyses if they did not have DNA methylation results (n=510) or had missing values for the serum CRP levels (n=2). Four subjects whose BMI were missing were excluded from the database as well. After exclusion, 966 subjects from 492 sibships remained in these analyses. (Table3) The GENOA study was approved by the Institutional Review Boards of all participating institutions. Informed consent was obtained from all study participants.

Data Analysis:

The outcome variable in these analyses was the serum CRP level, which was measured by a high sensitive immunoturbidimetric assay. CRP was natural log transformed (lnCRP) to obtain normality for the regression analysis. The main predictor variable for these analyses was DNA methylation sites. The DNA methylation data included DNA methylation sites on X and Y chromosomes, as well as 26,485 autosomal DNA methylation sites. Descriptive analyses and histograms were used for data cleaning, identification of missing values, and outliers.

Continuous variables that were considered in these analyses include: age (years); body mass index (BMI) (kg/m^2); serum homocysteine value (umol/l); fibrinogen value (mg/dl); cholesterol average value (mg/dl); high-density lipoprotein cholesterol (HDL-C) value (mg/dl); triglycerides average value (mg/dl); the amounts of cigarette smoking over time (pack-year); and systolic and diastolic blood pressure (mmHg). Mean and standard deviation of each continuous variable were identified overall and stratified by gender (Table1). Two sample t-tests were used to test for differences of the means for each continuous variable in the study cohort, and stratified by males

and females. If the variance of a continuous variable between males and females was not statistically different, the p-value was obtained by using a pooled method[50] Otherwise, SAS calculated the standard errors by Satterthwaite method using the weighted average of the two variances to obtain the p-value.

Categorical variables analyzed in this study included: gender (male/female); ever smoker (yes/no); smoking status at the study visit (never/not in last year/smoker within 1 year); diagnosis of hypertension by a physician (yes/no); and history of antihypertensive medication (yes/no). Ever smoker was defined as having ever smoked more than 100 cigarettes. In the analyses, a binary variable - current smoker (yes/no) - was used to represent the status of cigarette smoking. People who reported smoking within one year were defined as current smokers. Number and frequency of each categorical variable were identified overall and stratified by gender (Table1). Chi-square test was used to test for differences in the frequencies of each categorical variable in the study cohort between males and females.

Linear Regression Analysis:

In the univariate analyses, the estimated slopes (β-coefficients) represented the mean difference of lnCRP per unit of predictors.

Multiple regression models were used to control for potential confounders when estimating the correlation between lnCRP and the main predictor, DNA methylation sites. To adjust for the relatedness due to the sibship structure among participants, nlme (Linear and Nonlinear Mixed Effects Models) -package in R was used, and linear mixed model was performed to adjust for the random effect, sibships. Gender, BMI, and current smoker were identified as significant predictors of lnCRP based on the univariate analyses, and age was forced in the regression models.BMI In addition, to assess if age, gender, BMI, and current smoker modify the effect of DNA methylation sites in the prediction of lnCRP, we also performed analyses to investigate interactions between

DNA methylation sites and age (DNAm*age), DNA methylation sites and gender (DNAm*sex), DNA methylation sites and BMI (DNAm*bmi), and DNA methylation sites and current smoker (DNAm*smk) separately in the linear mixed models, after adjusting for the main effects of age, gender, BMI, and current smoker.

In this study, we excluded DNA methylation sites located on the X and Y chromosomes, and Bonferroni corrected p-values were calculated to adjust for repeated measurements. We also excluded 2,984 DNA methylation sites with non-specific binding probes indicated in Chen's study and 875 DNA methylation sites reported in the database of genotypes and phenotypes because their probes were overlapped with SNPs from the analysis.[51] After exclusion, 22,927 DNA methylation sites remained in the data to perform multivariate analyses. All statistical analyses were performed in SAS version 9.2. or in R 2.13.1 at the 5% significance level. The Bonferroni adjusted p-value threshold was used to identify the number of statistically significant lnCRP-associated DNA methylation sites or the number of significant interactions between DNA methylation sites and one of the covariates in the regression models. The Bonferroni adjusted p-value threshold for 22,927 repeated measurements was 2.18×10^{-6}. (Table 4) Annotation from NCBI build 36.1 was used to identify chromosome locations and genes of DNA methylation sites. (Table 5) In addition, a list of CRP-associated genes identified by multivariate analyses was used for functional annotation analysis. We used an online bioinformatics tool – the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 – to identify the enrichment Gene Ontology (GO) terms among CRP-associated genes. In DAVID, the p-value for gene enrichment analysis was measured by Fisher Exact test. We used 10^{-4} as the Fisher Exact p-value threshold for over-represented annotation terms. (Table 6)

RESULTS

The analyses included 966 African Americans from 492 sibships who participated in the GENOA study and had DNA methylation data. The baseline characteristics of the study population for all continuous and categorical variables considered in these analyses can be seen in Table1. Seventy (70.91%) percent of the subjects were female. Subjects included in the study had an average age of 66.27 ± 7.59 years and an average BMI of 31.16 ± 6.28 kg/m^2. At the 5% significance level, the average BMI among females (32.06 ± 6.58 kg/m^2) was significantly higher than that among males (28.98 ± 4.81 kg/m^2) ($p < 0.0001$). The study contained 797 (82.51%) subjects with self-reported physician-diagnosed hypertension and 719 (74.43%) subjects taking anti-hypertension medicine. The average systolic and diastolic blood pressures were 140.06 ± 21.40 mmHg and 78.36 ± 10.97 mmHg respectively for these subjects. 67% of males and 30% of females were ever smokers (e.g. smoked more than 100 cigarettes during their life time) ($p<0.0001$). About twenty percent of males had smoked within 1 year before the visit, but only nine percent (9.49%) of females had smoked within past 1 year ($p<0.0001$). Among 401 ever smokers, the total amount of cigarette smoking (measured by pack-year) for males (29.64 ± 23.56 pack-year) was significantly higher than that of females (20.60 ± 17.66) with $p<0.0001$. The above three smoking related variables showed that the behavior of cigarette smoking is significantly different between males and females. The serum level of CRP ranged from 0.02 mg/dl to 27.10 mg/dl, and the distribution of CRP was right-skewed with a median of 0.35 (Q1=0.17, Q3=0.71) mg/dl. There were significant differences in serum CRP levels, homocysteine values, fibrinogen values, serum cholesterol average values, and HDL-C values among different gender groups at alpha level of 0.05. Serum levels of CRP in females (0.71 ± 1.31 mg/dl) were significantly higher than that in males (0.54 ± 0.93 mg/dl) ($p=0.025$).

Table 2 shows the distribution of sibships by size among 966 African American subjects from 492 sibships. The sibship size ranged from 1 to 10. The majority (59.76%) of subjects in the study had at least one sibling enrolled. About forty (40.24%) percent of the subjects did not have siblings involved in the analyses. Three hundred and sixty subjects (36.59%) had one sibling involved in the analyses

Table 3 shows the association of age, gender, BMI, biomarkers for hypertension, and smoking status with lnCRP in the study population from univariable linear regression analyses. Gender, BMI, serum fibrinogen values, HDL-C values, serum triglyceride average values, and current smoker were each significantly associated with lnCRP at the 5% significance level. Increased value of BMI, fibrinogen and triglycerides are positively associated with natural log transformed CRP levels. The elevation of lnCRP was also associated with female sex and cigarette smoking (e.g. current smoker). The average lnCRP among females was 0.33 mg/dl higher than the average lnCRP among males ($p<0.0001$). The average lnCRP among current smokers was 0.30 mg/dl ($p=0.0043$) higher than the average lnCRP among non-smokers and former smokers. Based on the univariable analyses in Table3, gender, BMI, and current smoker were considered in the multivariable regression models, because they were identified as significant predictors of lnCRP. Age was forced in the models because of previously reported association with CRP levels, although the correlation between lnCRP and age was not statistically significant in this study.

The results of multivariable linear regression models are summarized in Table 4. Controlling for age, gender, BMI, and current smoker, there were 5,634 (20.43%) DNA methylation sites significantly associated with lnCRP at the 5% significance level. Two hundred and fifty seven (1.12%) DNA methylation sites remained statistically significant, after Bonferroni corrected p-value threshold of 0.05 was implemented to adjust for multiple testing of 22,927 autosomal DNA methylation sites. Furthermore, sex was noted to have a modified effect on 326 (1.18%) DNA

methylation sites in the prediction of lnCRP at the 5% significance level, after adjustment for age, gender, BMI, and current smoker in linear mixed model. Using Bonferroni corrected threshold of 0.05, we did not identify any significant interactions between covariates (Age, BMI, sex and cigarette smoking) and DNA methylation.

The top 30 most significant CRP-associated DNA methylation sites obtained from the linear mixed model, are listed in Table 5. The hypomethylation of all top 30 DNAm sites are associated with higher CRP levels. In addition, among 257 significant DNA methylation sites across 22 autosomes, 80.54% (n=207) of the significant DNA methylation sites were hypomethylated with higher CRP levels.

A bioinformatic analysis of the significant CRP-associated sites identified over-represented biological pathways (Table 6). Two hundred and forty one genes were annotated to 257 CRP-associated DNA methylation sites and were mapped to GO terms in human to determine the enrichment. Through the gene-enrichment analysis, we found 4 “biological process” (BP) GO terms, 3 “cellular component” (CC) GO terms, and 1 “molecular function” (MF) GO terms over-represented among CRP-associated genes. Moreover, “immune response” and “defense response” were identified as the major biological processes that these CRP-associated genes are involved. (Fisher Exact p-value= 2.10×10^{-15} and 6.60×10^{-10} respectively)

DISCUSSION

In this epigenome-wide association study on a population of 966 African Americans, we identified 257 DNA methylation sites on 241 genes across 22 autosomes significantly associated with serum CRP levels after adjusting for age, gender, BMI, and smoking status. The most significant CRP-associated DNA methylation site is cg07073964, which is located on chromosome 19 in kallikrein-related peptidase 10 (KLK10) gene (Bonferroni corrected p-value of 1.34×10^{-7}, Table5). KLK10 has been indicated as a tumor suppressor gene, and the expression of its encoded protein is used as a marker of prognosis for hepatocellular carcinoma, colorectal cancer, breast cancer, and prostate cancer.[52-55] Moreover, Ni et al. reported that the expression of KLK10 in endothelial cells was down regulated by disturbed blood flow, which is associated with hypertension.[56] The result inferred that KLK10 may play a role in the development of vascular diseases, such as atherosclerosis.[56]

Cg09358725, the second significant CRP-associated DNA methylation site in the analysis, is located on chromosome 11 on LIM domain only 2 (LMO2) gene (Bonferroni corrected p-value=3.88×10^{-7}). The LMO2 protein is a transcription factor involved in the early stage of hematopoiesis.[57] Without the participation of LMO2 protein in hematopoietic development, hematopoietic stem cells cannot be differentiated into erythrocytes.[57] It is likely that the expression of LMO2 has a comprehensive effect on cell differentiation, including leukocytes, since hematopoietic stem cells are the origin of all kinds of blood cells. In addition, cg09358725 is not the only significant CRP-associated DNA methylation site on LMO2, cg11822932 has also shown a significant associationwith serum CRP levels (rank 43/257, Bonferroni corrected p-value=1.65×10^{-4}).

The coagulation factor XII (F12) gene, where cg06625767 (Bonferroni corrected p-value=2.38×10^{-5}) is located, is on chromosome 5. Its encoded protein coagulation factor XII has a

role in the initiation of coagulation and fibrinolysis, as well as the release of bradykinin and angiotensin.[58] The association between inflammation and coagulator factor XII can be linked by bradykinin, which results in vasodilation when coagulator factor XII stimulates the kalikrein-kinin system.[59] Since angiotensin has an opposite effect on vascular permeability, the counterbalancing influences of bradykinin and angiotensin on inflammatory reactions were also suggested in Schmaier's study.[59]

The relationship between inflammation and the annotated genes can also be seen from gene-enrichment analysis using GO terms. The most significant GO term was immune response, which contains 41 genes (Fisher Exact p-value of 2.10×10^{-15}, Table6). The annotated genes, such as CCL26, F12, IRF7, and IL9, were defined highly associated with this biological process. CCL26 (chemokine (C-C motif) ligand 26) is an inflammatory chemokine, which act as a chemoattractant to guide leukocytes from the blood to the site of infection or tissue damage, and initiate immune response.[60] Interferon regulatory factor 7 (IRF7) is a regulator of type I interferons against infections.[61] Aberrant production of type I interferons was demonstrated to associate with cancers and autoimmune disorders.[61] Also, interleukin 19 (IL19) is an anti-inflammatory cytokine.[62] IL19 was suggested to have a role in chronic inflammatory disorders because of its function to induce cell-mediated immune response.[62]

Using the same platform for measuring DNA methylation, Breitling et al. reported that cg03636183, located in coagulation factor II (thrombin) receptor-like 3 (F2RL3) gene, was significantly hypomethylated among smokers compared to non-smokers and former smokers.[63] Also, smoking status has been identified as a risk factor for higher serum CRP levels in several studies.[42-44] These results imply that smoking status can be a confounder for studying epigenetic association with CRP levels, and needs to be considered in the analyses. In our analyses, the p-value of cg03636183 is not significant in the linear mixed model adjusted for smoking status (p-

value=$9.94x10^{-2}$), but the p-value is norminally significant in the model without adjusting for smoking status (p-value=$2.11x10^{-4}$)) (model1 vs. model6 in Table3). The stronger association between serum CRP levels and cg03636183 in model6 compared to the association in model1 could be explained by the mixed effect of smoking.

There are several limitations in this study. The study population is composed of only one African American cohort. Without further replication studies, the significant association of serum CRP levels with DNA methylation sites may not be generalizable to other populations or ethnic groups. In addition, DNA methylation was demonstrated to associate with physical conditions, such as essential hypertension, and the GENOA cohort is highly enriched with hypertensive subjects.[64] Therefore, the significant association of serum CRP levels with DNA methylation sites might be confounded by hypertension.

In conclusion, our results on the African American population in the Jackson, Mississippi area demonstrated the epigenetic modifications on serum levels of CRP, a biomarker of systemic inflammation. Our results provide gene-specific evidence about the epigenetic association with serum levels of CRP . Many genes involved in inflammatory process and immune responses are implicated as associated with CRP levels. Although inflammation is a known mechanism underlying chronic diseases such as cardiovascular disease and cancer, further investigations are warranted to understand how epigenetics influence these diseases on the gene-specific level.

REFERENCES

1. Pepys MB, Hirschfield GM. C-reactive protein: a critical update. *J Clin Invest.* Jun 2003;111(12):1805-1812.

2. Bassuk SS, Rifai N, Ridker PM. High-sensitivity C-reactive protein: clinical importance. *Curr Probl Cardiol.* Aug 2004;29(8):439-493.

3. Centers for Disease Control and Prevention. Deaths, Percent of Total Deaths, and Death Rates for the 15 Leading Causes of Death: United States and Each State, 1999-2007. http://www.cdc.gov/nchs/nvss/mortality/lcwk9.htm. Accessed oct 06, 2011.

4. Danesh J, Wheeler JG, Hirschfield GM, et al. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N Engl J Med.* Apr 1 2004;350(14):1387-1397.

5. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med.* Nov 20 2008;359(21):2195-2207.

6. Hage FG, Szalai AJ. C-reactive protein gene polymorphisms, C-reactive protein blood levels, and cardiovascular disease risk. *J Am Coll Cardiol.* Sep 18 2007;50(12):1115-1122.

7. Kushner I, Rzewnicki D, Samols D. What does minor elevation of C-reactive protein signify? *Am J Med.* Feb 2006;119(2):166 e117-128.

8. Pankow JS, Folsom AR, Cushman M, et al. Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis.* Feb 15 2001;154(3):681-689.

9. Ridker PM, Pare G, Parker A, et al. Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. *Am J Hum Genet.* May 2008;82(5):1185-1192.

10. Elliott P, Chambers JC, Zhang W, et al. Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA.* Jul 1 2009;302(1):37-48.

11. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature.* Jun 10 2010;465(7299):721-727.

12. Uddin M, Koenen KC, Aiello AE, Wildman DE, de los Santos R, Galea S. Epigenetic and inflammatory marker profiles associated with depression in a community-based epidemiologic sample. *Psychol Med.* May 2011;41(5):997-1007.

13. Gruys E, Toussaint MJ, Niewold TA, Koopmans SJ. Acute phase reaction and acute phase proteins. *J Zhejiang Univ Sci B.* Nov 2005;6(11):1045-1056.

14. Michael E. Makover, David Zieve. C-reactive protein. http://www.nlm.nih.gov/medlineplus/ency/article/003356.htm.

15. Arima H, Kubo M, Yonemoto K, et al. High-sensitivity C-reactive protein and coronary heart disease in a general population of Japanese: the Hisayama study. *Arterioscler Thromb Vasc Biol.* Jul 2008;28(7):1385-1391.

16. Zieske AW, Tracy RP, McMahan CA, et al. Elevated serum C-reactive protein levels and advanced atherosclerosis in youth. *Arterioscler Thromb Vasc Biol.* Jun 2005;25(6):1237-1243.

17. Williams ES, Shah SJ, Ali S, Na BY, Schiller NB, Whooley MA. C-reactive protein, diastolic dysfunction, and risk of heart failure in patients with coronary disease: Heart and Soul Study. *Eur J Heart Fail.* Jan 2008;10(1):63-69.

18. Ridker PM, Rifai N, Rose L, Buring JE, Cook NR. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. *N Engl J Med.* Nov 14 2002;347(20):1557-1565.

19. Ridker PM. Rosuvastatin in the primary prevention of cardiovascular disease among patients with low levels of low-density lipoprotein cholesterol and elevated high-sensitivity C-reactive

protein: rationale and design of the JUPITER trial. *Circulation.* Nov 11 2003;108(19):2292-2297.

20. Stumpf C, Hilgers KF. C-reactive protein: more than just a marker of inflammation? *J Hypertens.* Sep 2009;27(9):1748-1749.

21. Zwaka TP, Hombach V, Torzewski J. C-reactive protein-mediated low density lipoprotein uptake by macrophages: implications for atherosclerosis. *Circulation.* Mar 6 2001;103(9):1194-1197.

22. Fay WP. Linking inflammation and thrombosis: Role of C-reactive protein. *World J Cardiol.* Nov 26 2010;2(11):365-369.

23. Bisoendial RJ, Kastelein JJ, Levels JH, et al. Activation of inflammation and coagulation after infusion of C-reactive protein in humans. *Circ Res.* Apr 15 2005;96(7):714-716.

24. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature.* May 24 2007;447(7143):433-440.

25. Wierda RJ, Geutskens SB, Jukema JW, Quax PH, van den Elsen PJ. Epigenetics in atherosclerosis and inflammation. *J Cell Mol Med.* Jun 2010;14(6A):1225-1240.

26. Newman PE. Can reduced folic acid and vitamin B12 levels cause deficient DNA methylation producing mutations which initiate atherosclerosis? *Med Hypotheses.* Nov 1999;53(5):421-424.

27. Nehler MR, Taylor LM, Jr., Porter JM. Homocysteinemia as a risk factor for atherosclerosis: a review. *Cardiovasc Surg.* Dec 1997;5(6):559-567.

28. Cash HL, McGarvey ST, Houseman EA, et al. Cardiovascular disease risk factors and DNA methylation at the LINE-1 repeat region in peripheral blood from Samoan Islanders. *Epigenetics.* Oct 1 2011;6(10):1257-1264.

29. Baccarelli A, Tarantini L, Wright RO, et al. Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA normative aging study. *Epigenetics.* Apr 1 2010;5(3).

30. Stenvinkel P, Karimi M, Johansson S, et al. Impact of inflammation on epigenetic DNA methylation - a novel risk factor for cardiovascular disease? *J Intern Med.* May 2007;261(5):488-499.

31. Kim J, Bhattacharjee R, Khalyfa A, et al. DNA Methylation in Inflammatory Genes among Children with Obstructive Sleep Apnea. *Am J Respir Crit Care Med.* Feb 1 2012;185(3):330-338.

32. Hutchinson WL, Koenig W, Frohlich M, Sund M, Lowe GD, Pepys MB. Immunoradiometric assay of circulating C-reactive protein: age-related values in the adult general population. *Clin Chem.* Jul 2000;46(7):934-938.

33. Ford ES, Giles WH, Myers GL, Rifai N, Ridker PM, Mannino DM. C-reactive protein concentration distribution among US children and young adults: findings from the National Health and Nutrition Examination Survey, 1999-2000. *Clin Chem.* Aug 2003;49(8):1353-1357.

34. Rumley A, Emberson JR, Wannamethee SG, Lennon L, Whincup PH, Lowe GD. Effects of older age on fibrin D-dimer, C-reactive protein, and other hemostatic and inflammatory variables in men aged 60-79 years. *J Thromb Haemost.* May 2006;4(5):982-987.

35. Caparevic Z, Kostic N. [The influence of age and the beginning of menopause on the lipid status, LDL oxidation, and CRP in healthy women]. *Srp Arh Celok Lek.* May-Jun 2007;135(5-6):280-285.

36. Garcia-Moll X, Zouridakis E, Cole D, Kaski JC. C-reactive protein in patients with chronic stable angina: differences in baseline serum concentration between women and men. *Eur Heart J.* Oct 2000;21(19):1598-1606.

37. Lakoski SG, Cushman M, Criqui M, et al. Gender and C-reactive protein: data from the Multiethnic Study of Atherosclerosis (MESA) cohort. *Am Heart J.* Sep 2006;152(3):593-598.

38. Pieroni L, Bastard JP, Piton A, Khalil L, Hainque B, Jardel C. Interpretation of circulating C-reactive protein levels in adults: body mass index and gender are a must. *Diabetes Metab.* Apr 2003;29(2 Pt 1):133-138.

39. Wener MH, Daum PR, McQuillan GM. The influence of age, sex, and race on the upper reference limit of serum C-reactive protein concentration. *J Rheumatol.* Oct 2000;27(10):2351-2359.

40. Huffman FG, Whisner S, Zarini GG, Nath S. Waist circumference and BMI in relation to serum high sensitivity C-reactive protein (hs-CRP) in Cuban Americans with and without type 2 diabetes. *Int J Environ Res Public Health.* Mar 2010;7(3):842-852.

41. MacKenzie KE, Wiltshire EJ, Pena AS, et al. Hs-CRP is associated with weight, BMI, and female sex but not with endothelial function in children with type 1 diabetes. *Pediatr Diabetes.* Feb 2009;10(1):44-51.

42. Madsen C, Nafstad P, Eikvar L, Schwarze PE, Ronningen KS, Haaheim LL. Association between tobacco smoke exposure and levels of C-reactive protein in the Oslo II Study. *Eur J Epidemiol.* 2007;22(5):311-317.

43. O'Loughlin J, Lambert M, Karp I, et al. Association between cigarette smoking and C-reactive protein in a representative, population-based sample of adolescents. *Nicotine Tob Res.* Mar 2008;10(3):525-532.

44. Melbye H, Halvorsen DS, Hartz I, et al. Bronchial airflow limitation, smoking, body mass index, and statin use are strongly associated with the C-reactive protein level in the elderly. The Tromso Study 2001. *Respir Med.* Dec 2007;101(12):2541-2549.

45. Ahonen TM, Kautiainen HJ, Keinanen-Kiukaanniemi SM, Kumpusalo EA, Vanhala MJ. Gender difference among smoking, adiponectin, and high-sensitivity C-reactive protein. *Am J Prev Med.* Dec 2008;35(6):598-601.

46. Tsai AC, Tsai HJ. The association of age, gender, body fatness and lifestyle factors with plasma C-reactive protein concentrations in older Taiwanese. *J Nutr Health Aging.* Jun 2010;14(6):412-416.

47. Ohsawa M, Okayama A, Nakamura M, et al. CRP levels are elevated in smokers but unrelated to the number of cigarettes and are decreased by long-term smoking cessation in male smokers. *Prev Med.* Aug 2005;41(2):651-656.

48. The FBPP Investigators, Boerwinkle E, Brown CA, et al. A Multi-Center Genetic Study of Hypertension: The Family Blood Pressure Program (FBPP). *Hypertension.* 2002;39:3-9.

49. Sun YV, Turner ST, Smith JA, et al. Comparison of the DNA methylation profiles of human peripheral blood cells and transformed B-lymphocytes. *Hum Genet.* Jun 2010;127(6):651-658.

50. ANNOTATED OUTPUT: SAS T-TEST. 2006; http://www.bgsu.edu/downloads/cas/file36799.pdf.

51. Chen YA, Choufani S, Ferreira JC, Grafodatskaya D, Butcher DT, Weksberg R. Sequence overlap between autosomal and sex-linked probes on the Illumina HumanMethylation27 microarray. *Genomics.* Apr 2011;97(4):214-222.

52. NCBI, Information NCfB. KLK10 kallikrein-related peptidase 10 [Homo sapiens]. http://www.ncbi.nlm.nih.gov/gene/5655.

53. Petraki C, Youssef YM, Dubinski W, et al. Evaluation and prognostic significance of human tissue kallikrein-related peptidase 10 (KLK10) in colorectal cancer. *Tumour Biol.* Mar 22 2012.

54. Kioulafa M, Kaklamanis L, Stathopoulos E, Mavroudis D, Georgoulias V, Lianidou ES. Kallikrein 10 (KLK10) methylation as a novel prognostic biomarker in early breast cancer. *Ann Oncol.* Jun 2009;20(6):1020-1025.

55. Lu CY, Hsieh SY, Lu YJ, et al. Aberrant DNA methylation profile and frequent methylation of KLK10 and OXGR1 genes in hepatocellular carcinoma. *Genes Chromosomes Cancer.* Dec 2009;48(12):1057-1068.

56. Ni CW, Qiu H, Rezvan A, et al. Discovery of novel mechanosensitive genes in vivo using mouse carotid artery endothelium exposed to disturbed flow. *Blood.* Oct 14 2010;116(15):e66-73.

57. Yamada Y, Warren AJ, Dobson C, Forster A, Pannell R, Rabbitts TH. The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. *Proc Natl Acad Sci U S A.* Mar 31 1998;95(7):3890-3895.

58. NCBI, Information NCfB. F12 coagulation factor XII (Hageman factor) [Homo sapiens]. http://www.ncbi.nlm.nih.gov/gene/2161.

59. Schmaier AH. The kallikrein-kinin and the renin-angiotensin systems have a multilayered interaction. *Am J Physiol Regul Integr Comp Physiol.* Jul 2003;285(1):R1-13.

60. NCBI. CCL26 chemokine (C-C motif) ligand 26 [Homo sapiens]. http://www.ncbi.nlm.nih.gov/gene/10344.

61. Ning S, Pagano JS, Barber GN. IRF7: activation, regulation, modification and function. *Genes Immun.* Sep 2011;12(6):399-414.

62. Gallagher G. Interleukin-19: multiple roles in immune regulation and disease. *Cytokine Growth Factor Rev.* Oct 2010;21(5):345-352.

63. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet.* Apr 8 2011;88(4):450-457.

64. Smolarek I, Wyszko E, Barciszewska AM, et al. Global DNA methylation changes in blood of patients with essential hypertension. *Med Sci Monit.* Mar 2010;16(3):CR149-155.

TABLES

Table1. Characteristics of the Study Participants Stratified by Gender (n=966).

variable	Overall n=966 mean	(±SD)	Male n=281 (29.09%) mean	(±SD)	Female n=685 (70.91%) mean	(±SD)	p-value*
Age, years	66.27	-7.59	66.70	-7.64	66.10	-7.56	0.2651
C reactive protein, mg/dl [a]	0.66	-1.22	0.54	-0.93	0.71	-1.31	0.0247
lnCRP, mg/dl	-1.05	1.09	-1.28	1.09	-0.95	1.08	<0.0001
Body Mass Index, kg/m^2	31.16	-6.28	28.98	-4.81	32.06	-6.58	<0.0001
Homocysteine, umol/l	10.78	-4.57	11.86	-4.78	10.33	-4.41	<0.0001
Fibrinogen, mg/dl	369.43	-82.29	353.54	-86.02	375.95	-79.87	0.0001
Cholesterol, mg/dl	203.87	-42.03	192.59	-40.41	208.50	-41.83	<0.0001
HDL-C, mg/dl	58.34	-18.00	50.20	-14.60	61.67	-18.20	<0.0001
Triglycerides, mg/dl	119.86	-64.34	119.90	-74.07	119.84	-59.96	0.9896
Cigarette, pack-year [b]	24.86	-21.12	29.64	-23.56	20.60	-17.66	<0.0001
Systolic Blood Pressure, mmHg	140.06	-21.40	138.05	-20.75	140.88	-21.62	0.0622
Diastolic Blood Pressure, mmHg	78.36	-10.97	80.39	-11.06	77.52	-10.83	0.0002
	n	(%)	n	(%)	n	(%)	
Ever smoker (≥ 100 cigarettes)	401	-41.51	189	-67.26	212	-30.95	<0.0001
Current smoker	122	-12.63	57	-20.28	65	-9.49	<0.0001
Diagnosis of hypertension	797	-82.51	225	-80.07	572	-83.50	0.2022
Using hypertension medications	719	-74.43	199	-70.82	520	-75.91	0.0993

* P-values are significant at alpha = 0.05, and p-value is for difference between male and female using two sample t-test for continuous variables and chi-square test for categorical variables.

[a] The distribution of CRP is skewed.
(All: median = 0.35; Q1 = 0.17; Q3 = 0.71; Male: 0.27 (0.12, 0.58); Female: 0.38 (0.19, 0.80))

[b] Information is only available for ever smokers, 92 males and 473 females have missing values.

Table 2. Summary of sibships. (n=966)

sibship size	sibships		no. subject
	n	(%)	
1	198	40.24	198
2	180	36.59	360
3	72	14.63	216
4	29	5.89	116
5	7	1.42	35
6	4	0.81	24
7	1	0.2	7
8	0	0.00	0
9	0	0.00	0
10	1	0.20	10
Total (n)	492	-	966

Table 3. Univariable Linear Regression Analyses for Predictors of lnCRP. (n=966)

variable	Beta	SE	P-value	
Age, years	-0.0018	0.0046	0.6909	
Gender, female	0.3280	0.0766	<0.0001	***
Body Mass Index, kg/m^2	0.0569	0.0053	<0.0001	***
Homocystiene, umol/l	0.0046	0.0077	0.5500	
Fibrinogen, mg/dl	0.0066	0.0004	<0.0001	***
Cholesterol, mg/dl	0.0001	0.0008	0.9450	
HDL-C, mg/dl	-0.0059	0.0019	0.0024	**
Triglycerides, mg/dl	0.0016	0.0005	0.0039	**
Ever smoker	0.0855	0.0712	0.2300	
Current smoker	0.3015	0.1053	0.0043	**

Table 4. The number of significant CRP-associated DNA methylation sites and interactions in multivariable linear regression models using different p-value thresholds. (n=966)

No.	model	variable	p-value < 0.05		p-value* < 2.18×10^{-6}	
			n	(%)	n	(%)
1	lnCRP~DNAm+age+sex+bmi+smk	DNAm	5634	20.43	257	1.12
2	lnCRP~DNAm+age+sex+bmi+smk+DNAm*age	DNAm*age	506	1.83	0	0.00
3	lnCRP~DNAm+age+sex+bmi+smk+DNAm*sex	DNAm*sex	326	1.18	0	0.00
4	lnCRP~DNAm+age+sex+bmi+smk+DNAm*bmi	DNAm*bmi	260	0.94	0	0.00
5	lnCRP~DNAm+age+sex+bmi+smk+DNAm*smk	DNAm*smk	801	2.90	0	0.00
6	lnCRP~DNAm+age+sex+bmi	DNAm	6880	24.95	347	1.51

* Bonferroni adjusted p-value threshold for 22,927 tests; smk: current smoker

Table 5. The summary of top 30 significant CRP-associated DNA methylation sites.

DNAm	Gene	Chr.*	Beta	(SE)	P-value	Bonferroni P-value**
cg07073964	KLK10	19	-4.12	(0.58)	5.85×10^{-12}	1.34×10^{-7}
cg09358725	LMO2	11	-3.60	(0.52)	1.69×10^{-11}	3.87×10^{-7}
cg04121771	TM4SF4	3	-4.42	(0.68)	2.05×10^{-10}	4.70×10^{-6}
cg08458487	SFTPD	10	-2.79	(0.43)	2.26×10^{-10}	5.18×10^{-6}
cg09305224	FUT7	9	-3.38	(0.52)	2.48×10^{-10}	5.69×10^{-6}
cg00645579	IRF7	11	-3.80	(0.59)	2.94×10^{-10}	6.74×10^{-6}
cg05556717	CCL26	7	-3.32	(0.52)	3.94×10^{-10}	9.03×10^{-6}
cg17496921	TSPAN16	19	-2.94	(0.46)	4.97×10^{-10}	1.14×10^{-5}
cg03801286	KCNE1	21	-2.62	(0.41)	5.61×10^{-10}	1.29×10^{-5}
cg21969640	GPR84	12	-3.08	(0.49)	6.03×10^{-10}	1.38×10^{-5}
cg05501357	HIPK3	11	-3.35	(0.53)	6.29×10^{-10}	1.44×10^{-5}
cg03600318	SFTPD	10	-3.58	(0.57)	7.01×10^{-10}	1.61×10^{-5}
cg18084554	ARID3A	19	-2.68	(0.43)	7.91×10^{-10}	1.81×10^{-5}
cg06625767	F12	5	-2.80	(0.45)	1.04×10^{-9}	2.38×10^{-5}
cg15248035	CCIN	9	-2.62	(0.42)	1.22×10^{-9}	2.80×10^{-5}
cg05546038	NOL3	16	-3.96	(0.64)	1.40×10^{-9}	3.21×10^{-5}
cg09303642	NFE2	12	-2.70	(0.44)	1.60×10^{-9}	3.67×10^{-5}
cg03330678	SEPT9	17	-2.62	(0.43)	1.69×10^{-9}	3.87×10^{-5}
cg17753124	IER2	19	-3.34	(0.54)	1.72×10^{-9}	3.94×10^{-5}
cg22242539	SERPINF1	17	-3.25	(0.53)	2.08×10^{-9}	4.77×10^{-5}
cg17166812	NDUFS2	1	-3.89	(0.64)	2.28×10^{-9}	5.23×10^{-5}
cg22266967	S100P	4	-2.84	(0.47)	2.29×10^{-9}	5.25×10^{-5}
cg12380764	IL19	1	-2.93	(0.48)	2.35×10^{-9}	5.39×10^{-5}
cg10275770	ICAM2	17	-3.64	(0.60)	2.51×10^{-9}	5.75×10^{-5}
cg21492378	CEP1	9	-3.68	(0.60)	2.53×10^{-9}	5.80×10^{-5}
cg22381196	DHODH	16	-2.24	(0.37)	2.99×10^{-9}	6.86×10^{-5}
cg23140706	NFE2	12	-4.22	(0.70)	2.99×10^{-9}	6.86×10^{-5}
cg20283107	FAM91A1	8	-3.65	(0.60)	3.12×10^{-9}	7.15×10^{-5}
cg27606341	FYB	5	-2.55	(0.42)	3.21×10^{-9}	7.36×10^{-5}
cg26861460	PARVG	22	-2.90	(0.48)	3.27×10^{-9}	7.50×10^{-5}

*Chromosomal location is obtained from NCBI 36.1.

**Bonferroni adjusted p-value for 22,927 tests.

Table 6. GO Terms for CRP-associated genes (p-value threshold=10^{-4})

GO Term ID	GO Term Name	GO Domain*	Number of Genes	p-value
GO:0006955	immune response	BP	41	2.10×10^{-15}
GO:0006952	defense response	BP	31	7.50×10^{-10}
GO:0009611	response to wounding	BP	22	9.20×10^{-6}
GO:0009617	response to bacterium	BP	12	6.00×10^{-5}
GO:0031226	intrinsic to plasma membrane	CC	39	1.20×10^{-6}
GO:0005887	integral to plasma membrane	CC	38	1.90×10^{-6}
GO:0044459	plasma membrane part	CC	54	2.00×10^{-5}
GO:0019955	cytokine binding	MF	9	4.50×10^{-5}

* BP: Biological Process; CC: Cellular Component; MF: Molecular Function