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:

_____ _____

Alexander K. Zaharoff Date

Genome-wide identification of DNA methylation associated with Systolic and Diastolic Blood Pressures among African Americans

By

Alexander K. Zaharoff
Master of Public Health

Epidemiology Department

_____

Yan Sun, PhD
Committee Chair

Genome-wide identification of DNA methylation associated with Systolic and Diastolic Blood Pressures among African Americans

By

Alexander K. Zaharoff

Bachelor of Science
University of California, Santa Barbara
2009

Thesis Committee Chair: Yan Sun, PhD

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
2013

ABSTRACT

Genome-wide identification of DNA methylation associated with Systolic and Diastolic Blood Pressures among African Americans
By Alexander K. Zaharoff

Hypertension is a major risk factor for numerous chronic health diseases. This study aims to further the understanding of the gene-environment interaction related to hypertension. The underlying molecular mechanisms that affect blood pressure measurements will be evaluated. By conducting an epigenome-wide association study (EWAS) of 972 African Americans from Jackson, Mississippi we identified novel methylation sites that were associated with systolic and diastolic blood pressures. After correcting for confounders and multiple testing, two methylation sites were found to be statistically significantly associated with systolic blood pressure. These sites were located on the *CCDC25* gene (p-value = 1.5×10^{-7}, FDR = 0.004) and *COX7A2L* gene (p-value = 1.5×10^{-6}, FDR = 0.020). Findings from this epigenomic study of African Americans may lead to a more comprehensive understanding of hypertension.

KEYWORDS

Epigenome-Wide Association Study (EWAS), DNA methylation (DNAm), epigenetics, hypertension, African American, systolic blood pressure, diastolic blood pressure, *CCDC25*, *COX7A2L*

Genome-wide identification of DNA methylation associated with Systolic and Diastolic Blood Pressures among African Americans

By

Alexander K. Zaharoff

Bachelor of Science
University of California, Santa Barbara
2009

Thesis Committee Chair: Yan Sun, PhD

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
2013

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my main two editors. The first is my advisor, Dr. Yan Sun. He was crucial to locating these novel methylation associations and helped guide me through the manuscript process. I would also like to thank my mother who discussed sentence structure and the need to limit the number of ideas in any given sentence. I would also like to mention Qin Hui and Yuanyuan Fang for their expertise in programming and the support I received from them. Furthermore, I would like to acknowledge the assistance of the informaticists at the Woodruff Health Sciences library. When I needed to locate one specific source for my thesis, they were always there to assist. Lastly, I would like to thank the Rollins School of Public Health Epidemiology Department. Being able to complete a thesis demonstrates all of the skills I have developed during my Masters of Public Health training.

BACKGROUND

A leading chronic health issue in the United States for many Americans is that of high blood pressure. Through measurement of a person's blood pressure (BP), one is able to determine the presence of high blood pressure or hypertension. Hypertension is defined as a systolic blood pressure (SBP) greater than 140 mmHg and/or a diastolic blood pressure (DBP) greater than 90 mmHg without any antihypertensive treatment. High blood pressure is highly prevalent in the United States and leads to other co-morbidities and chronic disease processes. Hypertension can also set the stage for cardiovascular diseases such as heart failure and stroke. In fact, one in three adults in the United States has hypertension (1). A clearer understanding of the causes of hypertension and its resultant disease progression need to be investigated through a compilation of past research and then supplemented by complex dataset analysis.

Based on findings from the 2011 American College of Cardiology Foundation/American Heart Association (AHA) guidelines, hypertension in the United States affects 72 million or 35% of all Americans (2). According to the Centers for Disease Control and Prevention (CDC), the National Health and Nutrition Examination Survey (NHANES) data from 2003 to 2010 estimated that the prevalence of hypertension affected 66.9 million or 30.4% of Americans over 18 years of age (3). Comparatively, an even higher prevalence of hypertension was noted in the 2011 Behavioral Risk Factor Surveillance System (BRFSS)/CDC study conducted in Jackson, Mississippi where 39.3% of adults older than 18 years of age have been told they have high BP (4). From 2007-2009, the age-adjusted death rate due to the effects of high BP and its compounding

factors in the United States was 10.3 per 100,000 people for those 18 years of age and older (5). Undoubtedly, hypertension is a serious contributor to mortality in the United States.

Specifically, the prevalence of hypertension in African Americans continues to rise, creating a significant public health issue. For example, high death rates exist for hypertensive African Americans at 22.5 per 100,000 people (5). Documented prevalence of high BP increased among African Americans from 35.8% to 41.4% during the time periods of 1988-1994 to 1999-2002 (6). In spite of a range of rates across studies, the upward trend of more African Americans having high BP continues to climb (6, 7). Hypertension is a disease that affects African Americans more than any other race (2, 8). The complexity of hypertension disease progression has led to needed research for many races and ethnicities, though African Americans still remain under-represented (9, 10).

There are additional important factors that must be addressed when studying hypertension. Older age, being African American and having a higher body mass index (BMI) are associated with having a high BP (11). Smoking has also been shown to have an effect on hypertension. Hypertensive smokers are at higher risk of developing more serious forms of hypertension (12). Chronic smoking has been found to increase arterial stiffness in hypertensive individuals for up to a decade after cessation (13). Hypertension is a complex disease; therefore direct causal relationships have not been established. Despite this, one prospective study found that cigarette smoking was associated with increased hypertension among women that smoked at least 15 cigarettes a day (14).

Moreover, hypertension has also been demonstrated to be a strong predictor of strokes in multiple populations (15). Research data has shown that compared to whites,

African Americans with hypertension "have a 1.3-times greater rate of nonfatal stroke, a 1.8-times greater rate of fatal stroke, a 1.5-times greater rate of death due to [heart disease], and a 4.2-times greater rate of end-stage kidney disease" (16). The widespread prevalence of hypertension in African Americans and its resultant disease progression indicates the particularly heavy burden on public health systems caused by undiagnosed and under-treated hypertension.

INTRODUCTION

Epigenetics can be modified by both environmental and genetic factors and can mediate these risks for human diseases by heritable changes in epigenetic markers. Epigenetics is most commonly defined as a change in the gene expression that occurs without a change in the actual genome (17). Given the current technology, the most common way of measuring the "epigenome" is through DNA methylation (17).

DNA methylation occurs by the addition of a methyl group ($-CH_3$). Significantly, almost all methylation takes place in CG dinucleotides in DNA, which is why these sites are designated by cytosine-phosphate-guanine bonds and often called "CpG sites". According to Bruce Richardson, "approximately 70-80% of CG pairs are methylated" (18). DNA methylation is critical because the methylation of DNA blocks transcription and results in the blocking of gene expression. Alternatively, de-methylation can also occur and lead to over expression of certain genes, as discovered in the progression of some cancers. Ultimately, DNA methylation is an important mechanism that can alter gene expression (18).

Age is an important modifier of DNA methylation. In some instances, DNA methylation decreases with age, resulting in increased gene expression. This relationship

often follows a linear trend, but increasing or decreasing methylation is gene-specific (19). One study applied this particular attribute specifically to monozygotic twin data. The study noted that the older the twins were, the more dissimilar their DNA methylation locations and levels appeared. Compared to younger twins, the monozygotic twins were epigenetically "indistinguishable" (20). This illuminates that environmental factors are a vital influence on one's DNA methylation status. Particular DNA methylation sites have also been shown to be strongly associated with age (21). Also of importance is the predictive power of DNA methylation data to assess an individual's "epigenetic bio-age". One research group stated that this capability could be used as "a tool in routine medical screening to predict the risk of age-related diseases and to tailor interventions based on the epigenetic bio-age instead of the chronological age" (19). Furthermore, a more recent investigation has shown through a longitudinal study that DNA methylation increases with time, but at different rates for different individuals (22). DNA methylation is a dynamic process in humans (22, 23).

The genome-wide study of DNA methylation is a relatively new strategy for assessing high BP and essential hypertension (24, 25). A new research project evaluated the role of the "methylome", a term to designate an epigenome, using specific DNA methylation sequencing sites as predictors of hypertension. The researchers found that epigenetic factors may very well play a part in the development of hypertension (9). This case-control study suggests an important connection between environmental exposures and disease processes. Clearly, further research is needed so that data findings can be substantiated and thereby validated.

Hypertension is a disease that is affected by many conditions, both genetic and environmental. One review illustrated that studies have shown low birth weight increases the risk of hypertension as an adult (26). Some academic arguments have even described pathways of social stress factors describing the relationship from maternal stress during pregnancy, to perinatal health disparities, to adult cardiovascular diseases (27). "Similarly to type 2 diabetes, hypertension, atherosclerosis and other metabolic disorders, obesity predisposition and weight loss outcomes have been repeatedly associated to changes in epigenetic patterns. Different non-nutritional risk factors that usually accompany obesity seem also to be involved in these epigenetic modifications, especially hyperglycemia, inflammation, hypoxia and oxidative stress." (28). Heritability of hypertension is about 30-50%, which indicates the important role of genetics (29, 30). Moreover, it has been well established that age, sex, and environmental factors, in general, play crucial parts in the pathogenesis of hypertension. Molecular mechanisms or cellular physiology, such as chemical transport or cellular communication, have also been shown in the scientific literature to be associated with hypertension (31). Other suggested risk factors for hypertension include the evolution of the modern diet compared to the humans' Paleolithic genome (32).

Hypertension is a condition that is becoming ever-present among Americans and specifically African Americans. This is a public health concern since hypertension is a precursor of many diseases including, but not limited to, heart disease, heart attack, stroke, aneurysms, heart failure, kidney disease, metabolic syndromes and memory loss. Furthermore, the risk factors of hypertension are just as numerous. Risk factors include age, race, family history, obesity, physical inactivity, tobacco use, elevated sodium

intake, insufficient potassium, insufficient vitamin D, excessive alcohol, stress and other chronic disease conditions. Evaluating hypertension by including all of these risk factors and the resultant measurements in an analysis would be a daunting task. Using DNA methylation data is particularly intriguing because it can incorporate all these environmental exposures into one measurement, as in the case of CpG sites. It would still be essential in one's analysis to control for risk factors that have been proven to have a definitive association with hypertension, such as age, BMI, sex and smoking status.

It was recently shown in a genome-wide association study (GWAS) that several single nucleotide polymorphism (SNP) sites were associated with elevations in both systolic and diastolic BP in African Americans (33). This is crucial information since this study concluded "that blood pressure among [African Americans] is a trait with genetic underpinnings but also with significant complexity" (33). Additionally, the latest genome-wide methylation study found hypertension in young African American males (9). Based on this case-control study, the group suggested that DNA methylation may play a role in the pathogenesis of hypertension (9).

Further analysis will use DNA methylation data to utilize similar epigenetic modeling methods. Understanding the pathway from epigenetic exposures to hypertension as a disease outcome is a novel approach within hypertension research. Research will allow a better comprehension of exposures and disease processes in older, hypertensive African Americans in the United States. This study investigated the role epigenetics studies play in understanding hypertension. In the past, medical treatment has focused on high risk populations with determinants such as race, age, sex, body mass index, and smoking status. A more reliable model would include both environmental and

genetic factors. This investigation focused on the role of epigenetics which encompassed the effects of both environmental exposures and the genetics factors, in the outcome of disease processes as related to hypertension.

METHODS

STUDY SAMPLE

Data used in this analysis comes from the Genetic Epidemiology Network of Arteriopathy (GENOA) study (34). The original GENOA study recruited multiple ethnic groups, though this analysis focused on the African Americans from Jackson, Mississippi, who had been epityped for genome-wide DNA methylation. The inclusion criterion for the GENOA study was the essential diagnosis of hypertension prior to the age of 60. The requirement was met if one or more of the following conditions existed: average systolic BP of the last 2 out of 3 readings was hypertensive, average diastolic BP in 2 out of the last 3 readings were hypertensive, or a previous diagnosis of hypertension and antihypertensive medication prescribed by a physician to be taken daily within the previous month of the visit.

The other inclusion criteria of the GENOA study was that at least two hypertensive siblings were invited into the study and their non-hypertensive siblings. The resulting data of a sibling cohort arranged in this fashion is often called a sibship. Exclusion factors from the study included pregnancy, breast feeding, Type I diabetes, and secondary causes of hypertension such as renal vascular disease and current alcohol abuse. The cohort was originally ascertained from 1995-2000. Approval of the GENOA study was given by all the participating Institutional Review Boards (IRB) (35). The datasets used in this study were de-identified and did not require Emory IRB approval.

PHENOTYPE DATA

The subjects used in this analysis were African Americans from Jackson, Mississippi. The first examination of this cohort had 1854 individuals, while the second

examination that occurred from 2000-2005 had 1482 individuals. For the purpose of this analysis, only 972 individuals are included because these were the only individuals in which high-quality DNA methylation data were available.

MEASUREMENT OF DNA METHYLATION

DNA methylation was measured from peripheral blood leukocytes. This genetic data contained components that reflect genetic make-up as well as environmental exposures. The DNA methylation sites were measured by fluorescent signals and given a beta-value to quantify the level of methylation. A beta-value close to one represents high methylation, while a beta-value close to zero represents very low methylation for a specific site. Further description of the processing and methylation quantitation are described by Sun et al (36).

DNA METHYLATION DATASET

The genotypic methylation data was processed as outlined in Sun et al (35). In summary 1,008 participants were epityped for their methylation profile. Control probes were present on each sequencing chip. Based on the control probes, 49 samples were excluded for low and poor intensity measurements. DNA methylation sites were also excluded from the analysis based on control probe measurements, non-specific probe binding, sex chromosome location, multimodality, and probe overlap (35).

STATISTICAL METHODS

Based on past research criteria outlined in the prior sections of this paper, the final model will include the variables: age, sex, BMI, and smoking status. Since this is a sibship (case individual and their related siblings) dataset, relatedness must also be

considered to adjust for the potential correlation between siblings. The linear mixed effect (LME) model was applied with a random effect based on the sibship relationship.

Age was included as a covariate because it is associated with DNA methylation and has been associated with blood pressure measurements (22, 23). Furthermore, age in this dataset has been shown to have 1385 associated DNA methylation sites (37). This study also controlled for sex, making sure the findings would not be influenced by the unequal distribution of males and females (Table 1). Additionally, body mass index (BMI) was included to control for weight factors as well as the correlation of BMI with type II diabetes (28). Lastly, smoking status was also included in the model.

The primary study variables are the quality controlled 22,927 CpG sites recorded from the 27K Illumina DNA methylation platform. Each CpG site was analyzed within this study cohort to determine the association with hypertension. The covariates included in this analysis were sex, body mass index (BMI), age, and smoking status. Body mass index was defined as kilograms in weight per meter^2 in height on a continuous scale. Age was also on a continuous scale, measured in years. Smoking status was evaluated as a dichotomous variable comparing current smokers with the combined category of never smokers and past smokers.

In this study, we assessed systolic blood pressure (SBP) and diastolic blood pressure (DBP) as the outcome variables to study the epigenetic association with hypertension. SBP measures the pressure in the arteries when the heart beats, which is the biological process when the heart muscle contracts. DBP is the measurement of the pressure between heartbeats, which can also be thought of as the pressure when the heart muscle is resting between beats and is refilling with blood (38). Hypertension is often

designated as a SBP greater than or equal to 140 mmHg or a DBP greater than or equal to 90 mmHg. Moreover, SBP and DBP were adjusted for those on hypertension medications in which 10 mm Hg and 5 mm Hg were added, respectively, to the individuals average blood pressure readings.

The phenotypic and genotypic data were merged using R (39). Further analysis also utilized the R program and the "nlme" package (40). Within this package, a linear mixed model was fit, adjusting for relatedness within the sibships. The raw p-values were adjusted for multiple testing using a Bonferroni correction. Additionally, the raw p-values were also converted to false discovery rates (FDR) as a less conservative adjustment for multiple testing than the Bonferroni correction. Hence an FDR value of 0.05 was set to establish significance. Secondary sites of interest beyond the highly significant sites using the Bonferroni corrected p-value were assessed at an FDR value of 0.2 to suggest an association. Lastly, we checked our results of DNA methylation sites in the *PRCP* and *SULF1* genes to replicate the findings reported by Wang et al (9).

PRINCIPAL COMPONENT ANALYSIS

After the first model was run for all the CpG sites, resulting QQ (quantile-quantile) plots and inflation calculations of the p-values, led to a largely inflated DBP analysis and a fairly largely inflated SBP analysis (Appendix Figures 4 and 5). These issues were likely due to "batch effects" or slight variations because of the nature of genome sequencing. These effects were adjusted for by including principal components in the model. The top 10 principal components were included in the model to adjust for these "batch effects". Principal components are artificial variables that represent most of the variation in data (41). The top principal components of 22,927 autosomal CpG sites

were calculated to adjust for potential confounders, and to control for the inflation of low p-values in the EWAS. The beta-values of each CpG site were first standardized by subtracting the mean beta-value. Using the standardized beta-value, the top principal components and the corresponding eigen-vectors were calculated for each individual using the *princomp* package in R (39). Several studies have shown that by including the main principal components, one can adjust for variation across samples, tissues and for many more variables than are possible to measure (41-45). The principal component analysis technique resulted in a final model that will evaluate the relationship of a high systolic and/or diastolic blood pressure, defined as essential hypertension, based on DNA methylation levels and the original possible confounders.

RESULTS

The analysis itself began with previously cleaned phenotype and DNA methylome data. The phenotypic data from the 1,482 African American individuals was collected during the second visit of the GENOA study. The epigenetic (DNA methylation) data was measured using the stored peripheral leukocytes as previously described (35, 36). The final dataset had 972 individuals with both epigenetic data and phenotype data, among which there were 493 sibships.

Characteristics of the study population by gender group are presented in Table 1. Approximately 66% of the individuals in the study, for both males and females, provided blood for genetic testing. The majority of both males and females were hypertensive based on average adjusted systolic blood pressure, but not their average adjusted diastolic blood pressure. The mean age, BMI, SBP and DBP for this cohort were generally higher for hypertensive individuals than those that were normotensive (Table 1). Additionally, a slightly higher percentage of females were hypertensive (71.6%) than normotensive (66.5%). And overall, about 12% of the participants had been smokers in the last year (Table 1).

The sibships were used to insert random effects based on relatedness into the analysis. Batch effects were also adjusted for by including the top 10 principal components of the DNA methylation data as covariates. The final model included several covariates based on the literature in addition to the principal components, as shown in Appendix Figure 1. The resulting linear mixed effects models found several sites that were associated with high SBP. Using a suggestive false discovery rate (FDR) cutoff of 0.2 led to seven methylation sites of interest for SBP and none for DBP. Moreover, using

a Bonferroni corrected p-value resulted in two specific significant sites relating to SBP. The sites that were found to be significantly associated with the adjusted SBP levels were cg08773844 on the *CCDC25* gene and cg02278165 on the *COX7A2L* gene. The other sites that were found to be possible sites of explanation, based on the FDR cutoff, of the adjusted SBP were on several genes (*C9orf48* [or *KIF24*], *CDK10, KLHL25, APOLD1*, and *PRKACA*) and are shown in Table 2.

The site with the smallest p-value for SBP was also the site that was the second most associated site, based on the corrected p-values, with adjusted diastolic blood pressure (Appendix Table 1). Despite this, the site was found not to be significant with DBP. Another site, cg04481779, on the *IL20RA* gene was found to have the smallest p-value associated with adjusted diastolic blood pressure (Table 3). The analysis took precautions to exclude as many false positives as possible, in order to prevent false negative results.

DISCUSSION

Using data from the hypertensive African Americans in the GENOA study, we identified two CpG sites that were significantly associated with SBP. These two sites were associated with the adjusted systolic blood pressure (Table 2). No significant sites were found to be associated with DBP (Table 3).

The top associated site for SBP was on the coiled-coil domain containing 25 (*CCDC25*) gene. It has been shown to be associated with cholesterol and triglyceride levels in the Family Heart Study (46, 47). In one particular study, the deletion of the *CCDC25* gene was associated with the decreased survival of those with hepatocellular carcinoma (48). This is particularly intriguing because, generally, DNA methylation leads to decreased expression of the methylated gene. The *CCDC25* gene could be one that is principally associated with poor health, or an expedited progression of disease. Interestingly, another DNA methylation site within the *CCDC25* gene was nearing significance as shown by being the ninth ranked site for SBP (Table 2), which may further indicate the relationship between the DNA methylation of *CCDC25* gene and SBP.

The second site that was associated was the cytochrome c oxidase subunit VIIa polypeptide 2 like (*COX7A2L*) gene (49). This gene has been shown to be expressed in all tissues. Furthermore, in a laboratory study, it was up-regulated in a breast cancer cell line (50). The *COX7A2L* gene has been shown to be a CpG island of intermediate enrichment for p53 binding (51). This particular gene appears to have many functions and could make it important in the pathogenesis of hypertension. Hypertension is an

extremely complex disease and this gene, as well as the DNA methylation site should be further studied.

The top site for DBP was not significant based on the Bonferroni correction and FDR for multiple testing (Table 3). Regardless, the site with the smallest p-value is located on the interleukin 20 receptor, alpha (*IL20RA*) gene (52). This gene and its related beta gene (*IL20RB*) are highly expressed in skin. *IL20RB* is expressed almost everywhere, when the *IL20RA* gene has been found to have a restricted tissue distribution (53). They are also significantly up-regulated in psoriasis, a common skin condition that causes skin redness and irritation (52, 54). This gene has also been associated with the CD40 ligand, HDL and LDL cholesterol, and multiple sclerosis (46, 55).

The genes associated with SBP based on the FDR cutoff of 0.2 had a range of roles. One was *KIF24*, a kinesin that is known to act in cellular division and intracellular vesicle transport (56). This shows that the process leading to high SBP is an extremely dynamic process, even when age, sex, BMI and smoking status are incorporated.

Of the next five associated sites (Table 2) that were nearing significance, three had interesting phenotypic expressions. The *C9orf48* gene was found to be a likely risk factor for sporatic Frontotemporal Lobar Degeneration (56). One group found that the expression of *CDK10* is associated with hepatocellular carcinoma and biliary tract cancer (57, 58). *APOLD1* was found to be associated with testicular germ cell tumors among differentially methylated regions (59). It is also believed to be important for vascular function and endothelial cell signaling (60). Based on this analysis, these genes were not statistically significant, but could be important for future studies investigating epigenetic factors associated with hypertension.

Traditional DNA sequence-based approaches are complemented by DNA methylation data. Epigenetics in general has the potential to explain additional molecular mechanisms and the pathogenesis of hypertension (9). Additionally, peripheral blood leukocytes have been hypothesized to be a major measurement tool for the change in DNA methylation (36). The resulting data obtained from genetic sample collection can be used to characterize hypertension (9). Because of the small sample size, the Wang study used cases and controls. They found two CpG sites, which were located in the *SULF1* and *PRCP* genes, and attempted to validate the findings (9). The data located fundamental genes that could be further validated to explain hypertension in African Americans in the Southern part of the United States.

In addition to the findings presented in this analysis, we were able to separately validate the findings of the Wang group study, without specifically including their same sites. Similar to their study, we also were looking at hypertension among African Americans. The Wang group used younger cases and controls of an African American cohort, where as we used hypertensive sibships. We too found the same specific CpG site (Illumina ID cg04845579) within the *SULF1* gene as associated with elevated DBP when controlling for age, sex, BMI, smoking status and the top 10 principal components. This site was significant (nominal p-value = 2.6×10^{-4}), but based on our predetermined FDR cutoff (FDR corrected p-value = 0.999) or the conservative Bonferroni p-value (p-value = 1) as shown in Table 3, the site was insignificant after the multiple test corrections. Furthermore, the Wang study, using the same DNA methylation platform, found associations with the *PRCP* gene (9). Despite these findings, the other CpG site related to *SULF1* and the two sites related to *PRCP* showed minimal probability of association

(Table xx). In spite of these other discoveries, the top *SULF1* related CpG site was still among the most significant sites relating to DBP (Appendix Table 3). In addition, the same study did a Gene Ontology (GO) analysis and found the interleukin-13 gene to be "among the top GO categories" (9). In our epigenetic association study, the top CpG site for DBP was within the interleukin receptor 20-alpha gene (Table 3), which belongs to the same gene family of *IL13* and shares similar biological function. These findings show that these genes respond to biotic stimulus by inflammation, such as *IL20RA*. The implication is that hypertension is likely a disease that responds to epigenetic influences.

Awareness of environment and genetic components in diagnosing hypertension is crucial. Future directions for this study would be to include more African Americans from other parts of the United States. Epigenetics analysis can be further utilized to study other chronic diseases where risk factors and genetic relationships are not well defined. This same analysis using a different chronic disease could help establish the key role epigenetics could play in a better understanding of chronic diseases and illnesses that affect the health of Americans.

STRENGTHS AND LIMITATIONS

The first strength of this study is the large sample size compared to previous epigenomic studies. This allowed for statistically significant results when corrections for multiple testing were included. Further strengths of this analysis include a robust modeling technique. The linear mixed effects model was crucial for adjusting the relatedness of this sibship. A key strength of this study is the fact that the dataset has a family structure of relatedness. This means the siblings that were normotensive were excellent controls for their hypertensive counterparts. Furthermore, this study was able to

control for a huge inflation of the p-values by including principal components in the model. This additional statistical analysis was likely a strong regulation technique in reducing the bias based on the processing of the genetic data in the laboratory. The control for inflation was also important for the reduction of type-I errors.

Furthermore, hypertension in practice does not have a strict cutoff value. It is a continuous measurement that varies by individual and fluctuates based on almost any exposure. This means a linear model was appropriate to relating CpG sites to SBP and DBP levels. This analysis also found two associated DNA methylation sites that are strongly associated with hypertension. These sites can be further validated as genes of interest in hypertension epigenetic studies.

The limitations of my study include that fact that most of the individuals were hypertensive. Because blood pressures are on a continuous scale, having participants that cover more of the range of values may help with the strength of an epigenetic association. Additionally, to prove that these sites are essential in explaining hypertension, additional studies on other cohorts must be done. These cohorts could be ones where other medical backgrounds are analyzed using related techniques and attempting to replicate these same findings. Furthermore, there was also a lack of replication samples and the results could not be further validated.

FUTURE DIRECTIONS

The role of genome sequencing in public health and personal health care are constantly discussed in research and policy alike. Genome sequencing is a relatively new field that can play a crucial role in the preventative health of Americans. Knowing more about the risks imposed by genetics and environmental exposures allows for a targeted approach for better individual health.

Genetic analysis and applied applications along with technology advancements and the drastically decreasing cost of sequencing will likely speed up a rapid increase in the role DNA methylation and other individually targeted approaches. Tailoring prevention towards groups and later individuals can help with the overall health of all Americans. Individuals are exposed to many non-genetic factors such as aging, diet, exercise and environmental chemicals. These exposures may function via epigenetic mechanisms (i.e. DNA methylation) to cause chronic diseases. This knowledge can be acted upon to better the health of people through prevention and targeted treatment. Targeting particular groups or types of people at risk through education is essential. This could lead to an early and accurate diagnosis of chronic diseases. Furthermore, knowing particular sites related to a chronic disease could allow for studies that would focus on a novel intervention or drug for therapy.

Of those in the most recent NHANES dataset, about 5.7 million or 15.8% with hypertension were not receiving pharmacologic treatment (3). High blood pressure could be explained more comprehensively by a combination of the known environmental risk factors and genomic data. Hypertension is a modifiable risk factor when compared to other diseases, particularly chronic diseases. Untreated hypertension can lead to highly

debilitating and difficult to treat disease states.(15). Educating an individual as to the best way to deal with hypertension could lead to better results in behavior change. Examples could be a better diet, reducing chemical or pollutant exposures, and the cessation of smoking. An individual's actions could be monitored through their DNA methylation profile. A correct diagnosis of hypertension is crucial in preventing further disease development. The time has come to effectively deal with hypertension as greater health care coverage seems imminent in the United States. Preventing chronic disease is a feasible and effective way to keep health care costs down.

REFERENCES

1. Fields LE**.** The Burden of Adult Hypertension in the United States 1999 to 2000: A Rising Tide. Hypertension. 2004;44(4):398-404.
2. Bertoia ML, Waring ME, Gupta PS, Roberts MB, Eaton CB**.** Implications of new hypertension guidelines in the United States. Hypertension. 2012;60(3):639-44.
3. CDC**.** Vital Signs: Awareness and Treatment of Uncontrolled Hypertension Among Adults — United States, 2003–2010. Morbidity and Mortality Weekly Report. 2012;61(35):7.
4. BRFSS: turning information into health. Centers for Disease Control and Prevention.
5. Health Data Interactive File, 1981–2009. National Center for Health Statistics.
6. Hertz RP, Unger AN, Cornell JA, Saunders E**.** Racial Disparities in Hypertension Prevalence, Awareness, and Management. Arch Intern Med. 2005;165:2098-104.
7. Collins R**.** African American Women and Men at High and Low Risk for Hypertension: A Signal Detection Analysis of NHANES III, 1988–1994. Preventive Medicine. 2002;35(4):303-12.
8. Hajjar I, Kotchen TA**.** Trends in Prevalence, Awareness, Treatment, and Control of Hypertension in the United States, 1988-2000. Journal of the American Medical Association. 2003;290(2):8.
9. Wang X, Falkner B, Zhu H, Shi H, Su S, Xu X, et al. A Genome-Wide Methylation Study on Essential Hypertension in Young African American Males. PLoS ONE. 2013;8(1):e53938.
10. Hicken M, Gragg R, Hu H**.** How cumulative risks warrant a shift in our approach to racial health disparities: the case of lead, stress, and hypertension. Health Aff (Millwood). 2011;30(10):1895-901.

11. Banerjee D, Chung S, Wong EC, Wang EJ, Stafford RS, Palaniappan LP**.** Underdiagnosis of hypertension using electronic health records. Am J Hypertens. 2012;25(1):97-102.

12. Virdis A, Giannarelli C, Neves M, Taddei S, Ghiadoni L**.** Cigarette smoking and hypertension. Current Pharmaceutical Design. 2010;23(16):8.

13. Jatoi NA, Jerrard-Dunne P, Feely J, Mahmud A**.** Impact of smoking and smoking cessation on arterial stiffness and aortic wave reflection in hypertension. Hypertension. 2007;49(5):981-5.

14. Bowman TS, Gaziano JM, Buring JE, Sesso HD**.** A prospective study of cigarette smoking and risk of incident hypertension in women. J Am Coll Cardiol. 2007;50(21):2085-92.

15. Elkind MSV**.** Epidemiology and Risk Factors. Continuum. 2011;17(6).

16. Lloyd-Jones D, Adams RJ, Brown TM, Carnethon M, Dai S, De Simone G, et al. Heart disease and stroke statistics--2010 update: a report from the American Heart Association. Circulation. 2010;121(7):e46-e215.

17. Meyers RA**.** *Epigenetic Regulation and Epidgenomics*: Wiley-Blackwell; 2012.

18. Richardson B**.** Impact of aging on DNA methylation. Ageing Research Reviews. 2003;2(3):245-61.

19. Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S, et al. Epigenetic Predictor of Age. PLoS ONE. 2011;6(6).

20. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A. 2005;102(30):10604-9.

21. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Research. 2010;20(4):440-6.

22. Madrigano J, Baccarelli A, Mittleman MA, Sparrow D, Vokonas PS, Tarantini L, et al. Aging and epigenetics: longitudinal changes in gene-specific DNA methylation. Epigenetics. 2012;7(1):63-70.

23. Numata S, Ye T, Hyde TM, Guitart-Navarro X, Tao R, Wininger M, et al. DNA methylation signatures in development and aging of the human prefrontal cortex. Am J Hum Genet. 2012;90(2):260-72.

24. Wang X, Prins BP, Sõber S, Laan M, Snieder H**.** Beyond Genome-Wide Association Studies: New Strategies for Identifying Genetic Determinants of Hypertension. Current Hypertension Reports. 2011;13(6):442-51.

25. Millis RM**.** Epigenetics and Hypertension. Current Hypertension Reports. 2010;13(1):21-8.

26. Adair L, Dahly D**.** Developmental Determinants of Blood Pressure in Adults. Annual Review of Nutrition. 2005;25(1):407-34.

27. Kuzawa CW, Sweet E**.** Epigenetics and the embodiment of race: Developmental origins of US racial disparities in cardiovascular health. American Journal of Human Biology. 2009;21(1):2-15.

28. Milagro FI, Mansego ML, De Miguel C, Martínez JA**.** Dietary factors, epigenetic modifications and obesity outcomes: Progresses and perspectives. Molecular Aspects of Medicine. 2012.

29. Cowley AW, Nadeau JH, Baccarelli A, Berecek K, Fornage M, Gibbons GH, et al. Report of the National Heart, Lung, and Blood Institute Working Group on Epigenetics and Hypertension. Hypertension. 2012;59(5):899-905.

30. Ward R**.** Familiar Aggregation and Genetic Epidemiology of Blood Pressure. In: Laragh JH, Brenner BM, eds. Hypertension: Pathophysiology, Diagnosis, and Management. New York: Raven Press; 1990:81-100.

31. Mateos-Cáceres PJ, Zamorano-León JJ, Rodríguez-Sierra P, Macaya C, López-Farré AJ**.** New and Old Mechanisms Associated with Hypertension in the Elderly. International Journal of Hypertension. 2012;2012:1-10.

32. Kunes J, Zicha J**.** The Interaction of Genetic and Environmental Factors in the Etiology of Hypertension. Physiological Research. 2009;58(Supplemental 2).

33. Fox ER, Young JH, Li Y, Dreisbach AW, Keating BJ, Musani SK, et al. Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study. Hum Mol Genet. 2011;20(11):2273-84.

34. NCBI**.** dbGaP : Genotypes and Phenotypes. Genetic Epidemiology Network of Arteriopathy (GENOA) http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000379.v1.p1.

35. Sun Y, Smith A, Conneely K, Chang Q, Li W, Lazarus A, et al. Epigenomic Association Analysis Identifies Smoking-related DNA Methylation Sites in African Americans. Human Genetics. 2013.

36. Sun YV, Turner ST, Smith JA, Hammond PI, Lazarus A, Rostyne JL, et al. Comparison of the DNA methylation profiles of human peripheral blood cells and transformed B-lymphocytes. Human Genetics. 2010;127(6):651-8.

37. Abstracts from the Annual Meeting of the International Genetic Epidemiology Society. Genetic Epidemiology. 2012;36(7):720-77.

38. AHA**.** In: Association AH, ed. Understanding Blood Pressure Readings. http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/AboutHighBloodPressure/Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp.

39. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: R Core Team; 2012.

40. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team atRDC. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-106. 2012.

41. Cheryl C. Y. Li JEC, Mark J. Cowley, Thomas Preiss, David I. K. Martin, Catherine M. Suter. A Sustained Dietary Change Increases Epigenetic Variation in Isogenic Mice. PLoS ONE. 2011;7(4):10.

42. Arend Voorman TL, Barbara McKnight, Kenneth Rice. Behavior of QQ-Plots and Genomic Control in Studies of Gene-Environment Interaction. PLoS ONE. 2011;6(5):7.

43. Laurila K, Oster B, Andersen CL, Lamy P, Orntoft T, Yli-Harja O, et al. A beta-mixture model for dimensionality reduction, sample classification and analysis. BMC Bioinformatics. 2011;12:215.

44. Zheng H, Wu H, Li J, Jiang SW. CpGIMethPred: computational model for predicting methylation status of CpG islands in human genome. BMC Med Genomics. 2013;6 Suppl 1:S13.

45. Wolstenholme JT, Warner JA, Capparuccini MI, Archer KJ, Shelton KL, Miles MF. Genomic analysis of individual differences in ethanol drinking: evidence for non-genetic factors in C57BL/6 mice. PLoS One. 2011;6(6):e21100.

46. NCBI. Association Results. Phenotype-Genotype Integrator. http://www.ncbi.nlm.nih.gov/gap/phegeni?tab=1&gene=55246.

47. NCBI. NHLBI Family Heart Study. dbGaP: Genotypes and Phenotypes. http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000221.v1.p1.

48. Roessler S, Long EL, Budhu A, Chen Y, Zhao X, Ji J, et al. Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. Gastroenterology. 2012;142(4):957-66 e12.

49. NCBI**.** COX7A2L cytochrome c oxidase subunit VIIa polypeptide 2 like [Homo sapiens (human)]. Gene. http://www.ncbi.nlm.nih.gov/gene/9167.

50. Watanabe T, Inoue S, Hiroi H, Orimo A, Kawashima H, Muramatsu M**.** Isolation of Estrogen-Responsive Genes with a CpG Island Library. Molecular and Cellular Biology. 1998;18(1):8.

51. Krieg AJ, Hammond EM, Giaccia AJ**.** Functional analysis of p53 binding under differential stresses. Molecular and cellular biology. 2006;26(19):7030-45.

52. NCBI**.** IL20RA interleukin 20 receptor, alpha [Homo sapiens (human)] Gene. http://www.ncbi.nlm.nih.gov/gene/53832.

53. Sheikh F, Baurin VV, Lewis-Antes A, Shah NK, Smirnov SV, Anantha S, et al. Cutting Edge: IL-26 Signals through a Novel Receptor Complex Composed of IL-20 Receptor 1 and IL-10 Receptor 2. The Journal of Immunology. 2004;172(4):5.

54. Encyclopedia ADAMM**.** Psoriasis. PubMed Health. http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0001470/.

55. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9.

56. Venturelli E, Villa C, Fenoglio C, Clerici F, Marcone A, Benussi L, et al. Is KIF24 a genetic risk factor for Frontotemporal Lobar Degeneration? Neurosci Lett. 2010;482(3):240-4.

57. Cui Y-F**.** CDK10 functions as a tumor suppressor gene and regulates survivability of biliary tract cancer cells. Oncology Reports. 2011.

58. Zhong X-y, Xu X-x, Yu J-h, Jiang G-x, Yu Y, Tai S, et al. Clinical and biological significance of Cdk10 in hepatocellular carcinoma. Gene. 2012;498(1):68-74.

59. Cheung HH, Lee TL, Davis AJ, Taft DH, Rennert OM, Chan WY**.** Genome-wide DNA methylation profiling reveals novel epigenetically regulated genes and non-coding RNAs in human testicular cancer. British Journal of Cancer. 2010;102(2):419-27.

60. Regard JB**.** Verge: A Novel Vascular Early Response Gene. Journal of Neuroscience. 2004;24(16):4092-103.

TABLES

Table 1. Stratified by Gender Characteristics of African Americans from Jackson, MS in the Genetic Epidemiology Network of Arteriopathy (GENOA) Study (n=972)

	Hypertensive (n=799)		Normotensive (n=173)	
	Mean±SD	**Range**	**Mean±SD**	**Range**
Age, years	67.0±7.4	[40.1–94.7]	63.2±7.8	[39.3–91.6]
BMI, kg/m^2	31.6±6.3	[16.4–60.4][a]	29.2±5.7	[17.7–50.4]
SBP, mmHg	152.2±21.2	[89.0–253.0]	125.2±10.4	[95.0–163.0]
DBP, mmHg	83.4±11.4	[50.0–126.0]	75.7±6.9	[54.0–91.0]
	No.	**%**	**No.**	**%**
Gender				
Male	227	28.4	58	33.5
Female	572	71.6	115	66.5
Smoker				
Yes, within last year	100	12.5	22	12.7
Other[b]	699	87.5	151	87.3
On Hypertensive Medications				
Yes	699	87.5	24	13.9
No	100	12.5	149	86.1

[a] There are four missing observations for BMI among hypertensive individuals

[b] Combined "Not in the last year" and "Never" responses

Table 2. Linear Mixed Effects Model Evaluating the Adjusted Systolic Blood Pressure by DNA Methylation Site, adjusting for Age, Sex, BMI, Smoking Status, 10 Principal Components, and the Cohort Relatedness

Methylation Site	Chromosome	Gene	Beta Estimate	Standard Error	Uncorrected P-value	Bonferroni Corrected P-value	FDR Corrected P-value	
cg08773844	8	*CCDC25*	210.3	39.4	1.5×10^{-7}	0.004	0.004	*
cg02278165	2	*COX7A2L*	201.4	41.4	1.5×10^{-6}	0.041	0.020	*
cg19947621	9	*C9orf48*	121.5	27.8	1.5×10^{-5}	0.402	0.102	
cg12119029	16	*CDK10*	-50.1	11.5	1.6×10^{-5}	0.417	0.102	
cg24345138	15	*KLHL25*	424.1	98.2	1.9×10^{-5}	0.510	0.102	
cg02813863	12	*APOLD1*	166.2	39.4	2.9×10^{-5}	0.778	0.130	
cg03856723	19	*PRKACA*	203.8	49.1	3.9×10^{-5}	1	0.148	
cg05942970	10	*C10orf39*	-119.3	29.6	6.6×10^{-5}	1	0.218	
cg13467649	8	*CCDC25*	122.5	31.1	9.3×10^{-5}	1	0.272	
cg20609368	11	*HBD*	-29.9	7.7	1.1×10^{-4}	1	0.303	
cg19278809	4	*KIAA0232*	-224.4	58.9	1.6×10^{-4}	1	0.377	
cg09766383	6	*GPR63*	424.6	112.1	1.7×10^{-4}	1	0.380	
cg21505886	4	*TMEM129*	-66.8	17.8	2.0×10^{-4}	1	0.397	
cg02164442	16	*ITGAD*	112.7	30.2	2.1×10^{-4}	1	0.397	

These are all the sites with an FDR <0.4

Table 3. Linear Mixed Effects Model Evaluating the Adjusted Diastolic Blood Pressure by DNA Methylation Site, adjusting for Age, Sex, BMI, Smoking Status, 10 Principal Components, and the Cohort Relatedness

Methylation Site	Chromosome	Gene	Beta Estimate	Standard Error	Uncorrected P-value	Bonferroni Corrected P-value	FDR Corrected P-value
cg04481779	6	*IL20RA*	43.3	10.0	2.0×10^{-5}	0.531	0.531
cg08773844	8	*CCDC25*	80.9	20.0	6.3×10^{-5}	1	0.609
cg01602416	16	*GCSH*	-51.8	12.9	6.9×10^{-5}	1	0.609
cg04845579	8	*SULF1*	19.1	5.2	2.6×10^{-4}	1	0.999
cg08451957	2	*FOXD4L1*	37.5	10.2	2.8×10^{-4}	1	0.999
cg20657383	19	*CEACAM1*	-37.9	10.4	3.0×10^{-4}	1	0.999
cg00646492	10	*GSTO1*	-54.6	15.3	4.1×10^{-4}	1	0.999
cg10691387	3	*IQCF2*	59.5	17.2	5.7×10^{-4}	1	0.999
cg17132967	19	*ZNF83*	-19.5	5.6	5.9×10^{-4}	1	0.999
cg22477971	1	*C1QB*	29.8	8.6	6.1×10^{-4}	1	0.999

These are the top 10 associated sites

FIGURES AND FIGURE LEGENDS

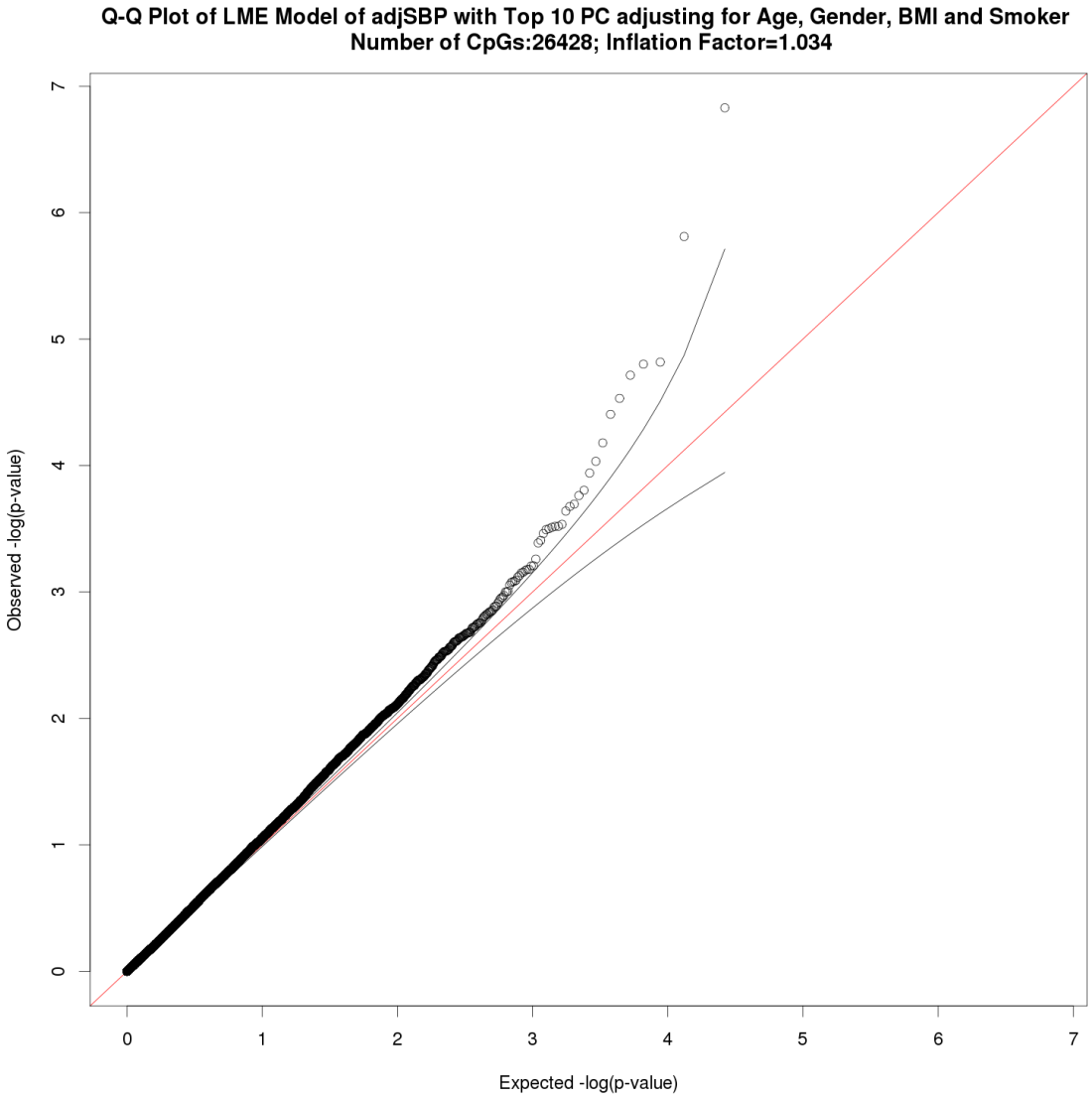


Figure 1. Log quantile-quantile plot of the observed and expected p-values for adjusted SBP using the final model

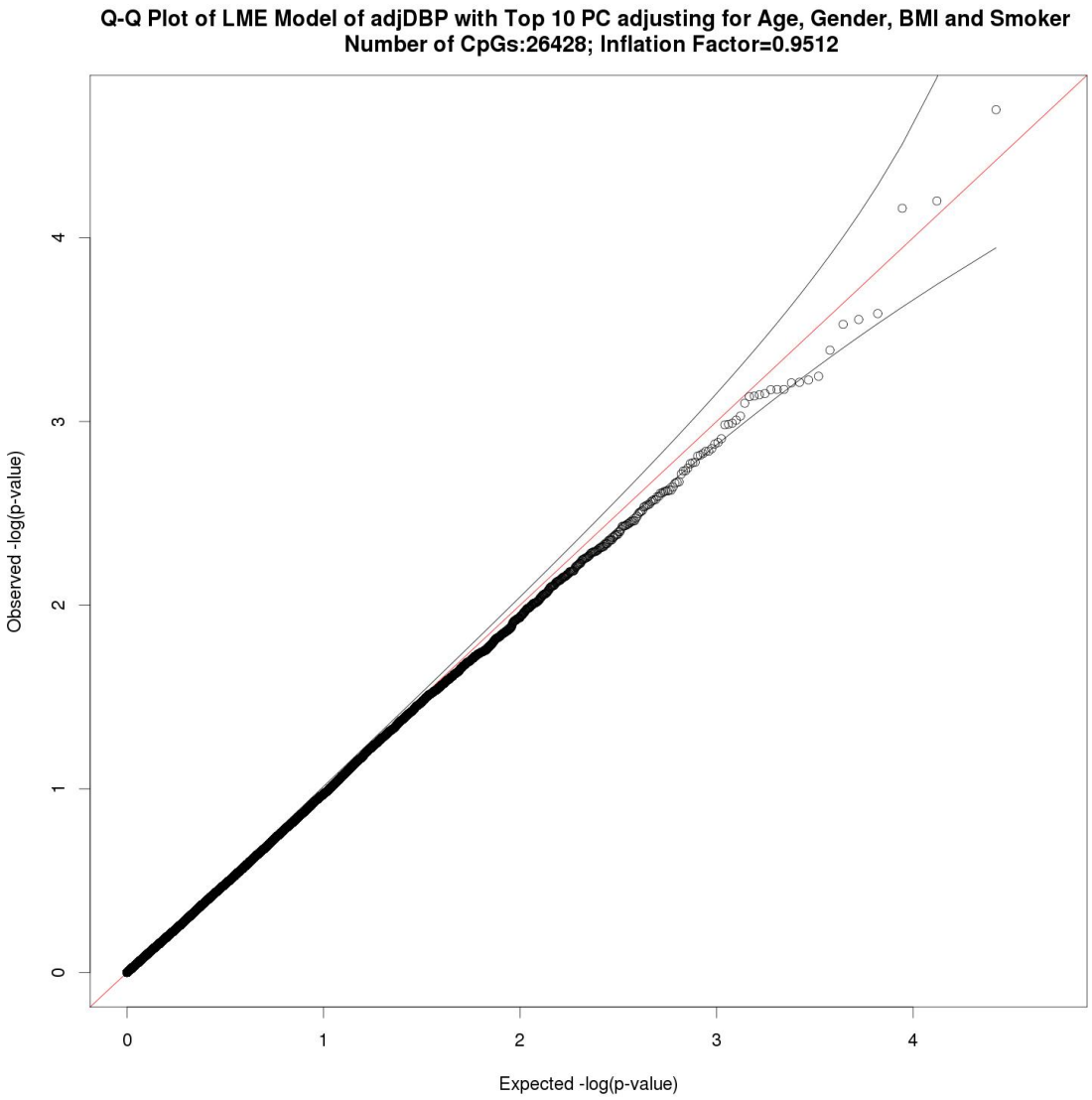


Figure 2. Log quantile-quantile plot of the observed and expected p-values for adjusted DBP using the final model

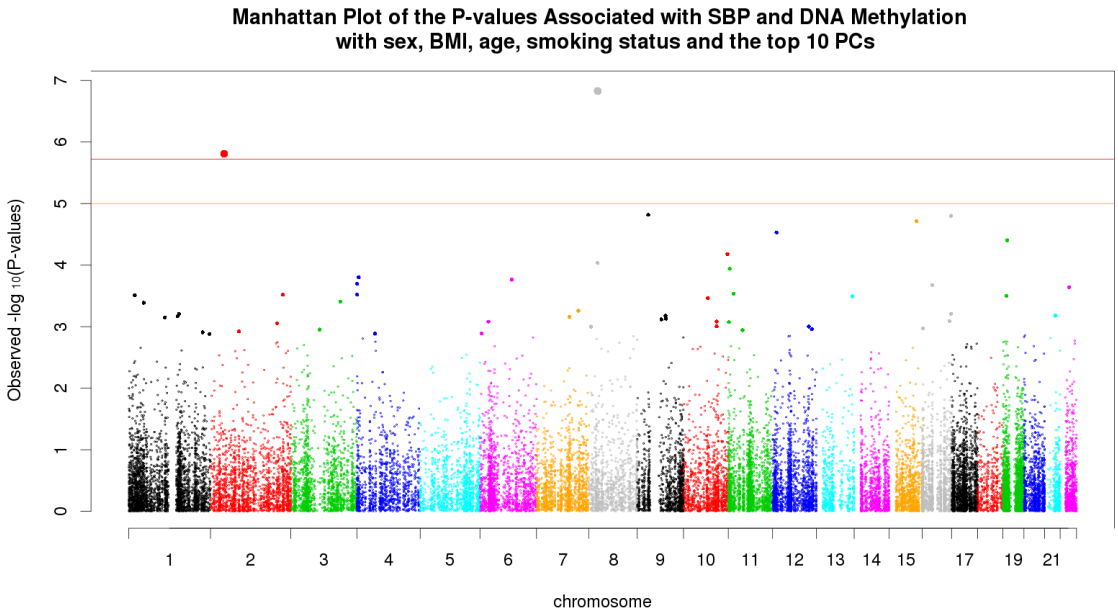


Figure 3. Manhattan plot of the observed p-values for adjusted SBP using the final model. The red-line represents the Bonferroni corrected p-value significance level. The yellow-line represents the False Discovery Rate significance level (q-value=0.05).

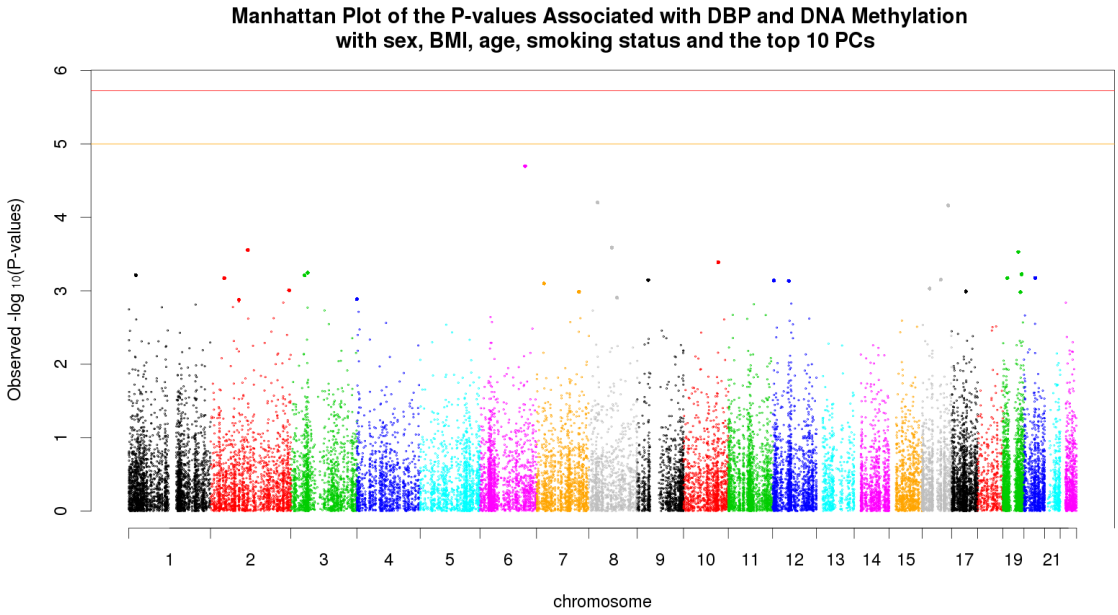
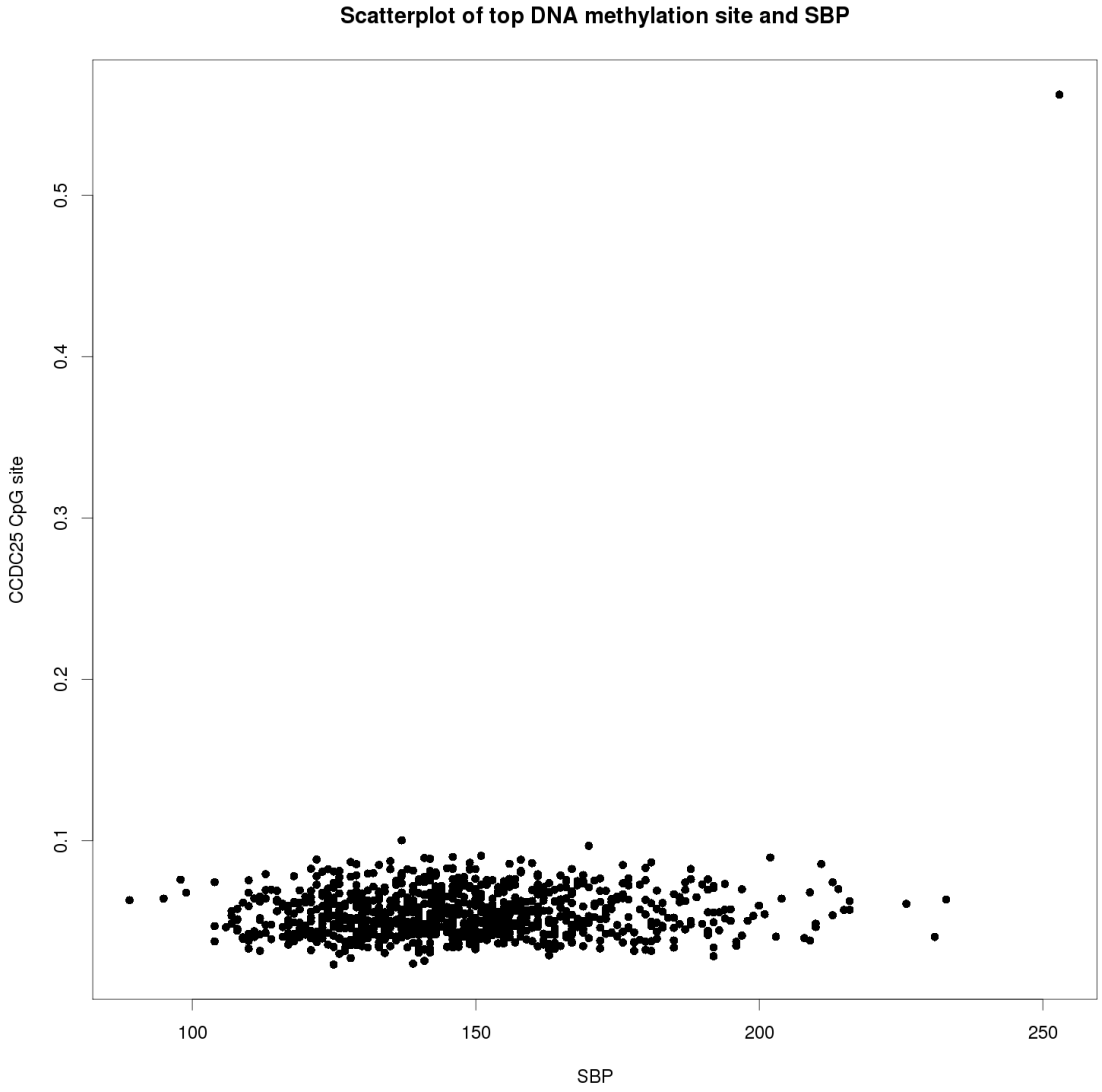


Figure 4. Manhattan plot of the observed p-values for adjusted DBP using the final model. The red-line represents the Bonferroni corrected p-value significance level. The yellow-line represents the False Discovery Rate significance level (q-value=0.05).

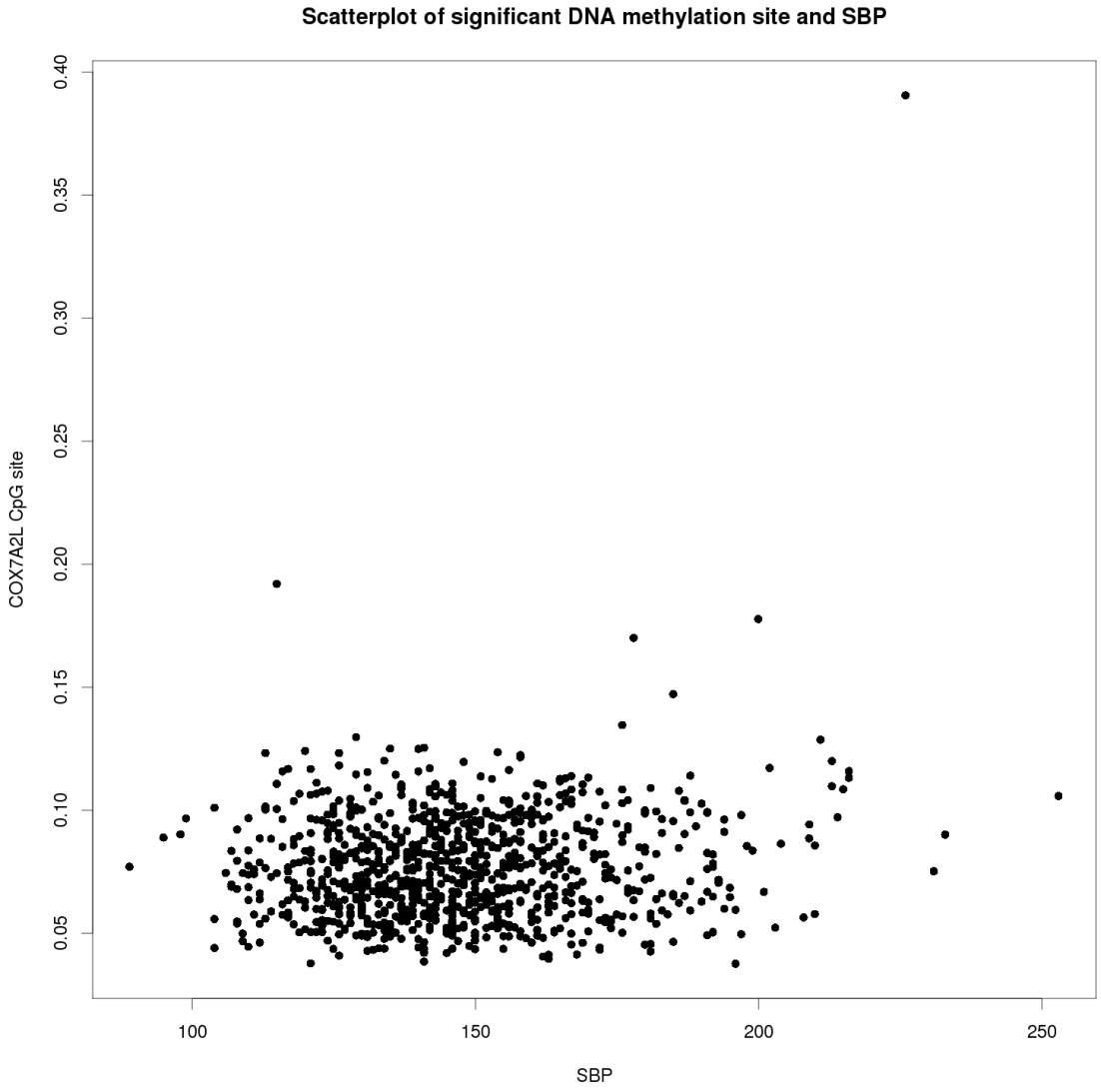
APPENDICES

$$SBP \text{ or } DBP = \beta_1 \text{CpG} + \beta_2 \text{age} + \beta_3 \text{sex} + \beta_4 \text{BMI} + \beta_5 \text{smoker} + B_6 \text{PC1} + \beta_7 \text{PC2} + \beta_8 \text{PC3} + \beta_9 \text{PC4} + \beta_{10} \text{PC5} + \beta_{11} \text{PC6} + \beta_{12} \text{PC7} + \beta_{13} \text{PC8} + \beta_{14} \text{PC9} + \beta_{15} \text{PC10} + \text{random effects for sibship}$$

Appendix Figure 1. The final model fit for this analysis. Each DNA methylation site was run (n=24,618) for SBP and DBP and controlled for covariates described in the literature and ten principal components. This model was adjusted using related sibships as a mixed effect in the model.



Appendix Figure 2. Scatterplot of the participants adjusted SBP and the corresponding beta-value for a DNA methylation site in the CCDC25 gene (cg08773844)



Appendix Figure 3. Scatterplot of the participants adjusted SBP and the corresponding beta-value for a DNA methylation site in the COX7A2L gene (cg02278165)

Appendix Table 1. Relationship of the most significant SBP sites for the final model and the related values for DBP

			SBP			DBP		
Methylation Site	Chromosome	Gene	Uncorrected P-value	Bonferroni Corrected P-value	FDR Corrected P-value	Uncorrected P-value	Bonferroni Corrected P-value	FDR Corrected P-value
cg08773844	8	*CCDC25*	1.5E-07	0.004	0.004	6.30E-05	1	0.609
cg02278165	2	*COX7A2L*	1.5E-06	0.041	0.020	6.71E-04	1	0.999
cg19947621	9	*C9orf48*	1.5E-05	0.402	0.102	7.15E-04	1	0.999
cg12119029	16	*CDK10*	1.6E-05	0.417	0.102	1.10E-02	1	0.999
cg24345138	15	*KLHL25*	1.9E-05	0.510	0.102	3.09E-03	1	0.999
cg02813863	12	*APOLD1*	2.9E-05	0.778	0.130	3.17E-03	1	0.999
cg03856723	19	*PRKACA*	3.9E-05	1	0.148	2.49E-02	1	0.999
cg05942970	10	*C10orf39*	6.6E-05	1	0.218	2.12E-01	1	0.999
cg13467649	8	*CCDC25*	9.3E-05	1	0.272	1.09E-02	1	0.999
cg20609368	11	*HBD*	1.1E-04	1	0.303	2.26E-01	1	0.999
cg19278809	4	*KIAA0232*	1.6E-04	1	0.377	1.08E-01	1	0.999
cg09766383	6	*GPR63*	1.7E-04	1	0.380	3.89E-01	1	0.999
cg21505886	4	*TMEM129*	2.0E-04	1	0.397	1.95E-02	1	0.999
cg02164442	16	*ITGAD*	2.1E-04	1	0.397	1.53E-01	1	0.999

These are all the SBP associated sites with an FDR <0.4

Appendix Table 2. Relationship of the most significant DBP sites for the final model and the related values for SBP

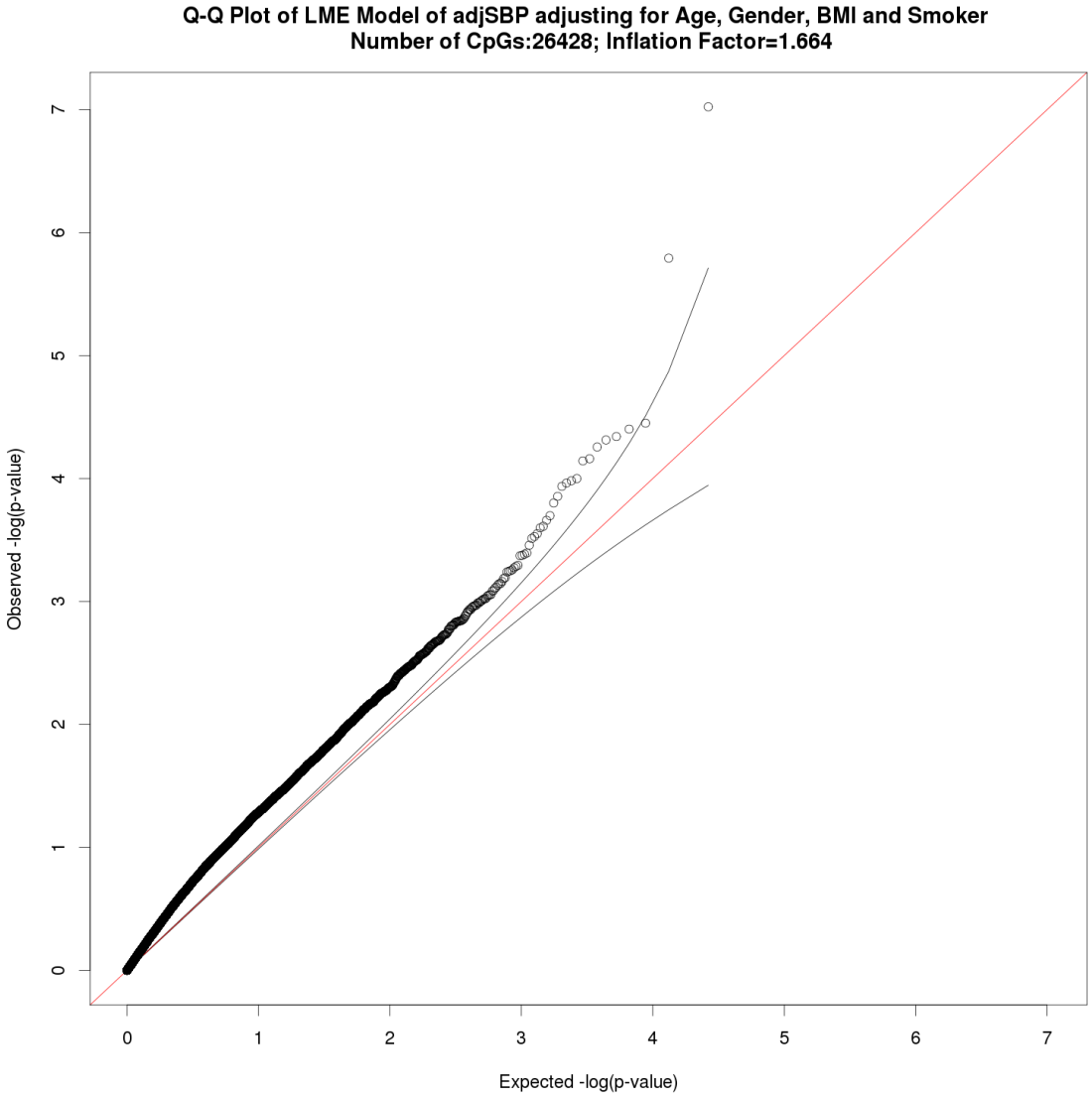
			DBP			SBP		
Methylation Site	Chromosome	Gene	Uncorrected P-value	Bonferroni Corrected P-value	FDR Corrected P-value	Uncorrected P-value	Bonferroni Corrected P-value	FDR Corrected P-value
cg04481779	6	*IL20RA*	2.0E-05	0.531	0.531	0.013106	1	0.746
cg08773844	8	*CCDC25*	6.3E-05	1	0.609	1.48E-07	0.004	0.004
cg01602416	16	*GCSH*	6.9E-05	1	0.609	0.00385	1	0.695
cg04845579	8	*SULF1*	2.6E-04	1	0.999	0.008525	1	0.742
cg08451957	2	*FOXD4L1*	2.8E-04	1	0.999	0.024057	1	0.780
cg20657383	19	*CEACAM1*	3.0E-04	1	0.999	0.025991	1	0.791
cg00646492	10	*GSTO1*	4.1E-04	1	0.999	0.004954	1	0.708
cg10691387	3	*IQCF2*	5.7E-04	1	0.999	0.077956	1	0.876
cg17132967	19	*ZNF83*	5.9E-04	1	0.999	0.014599	1	0.759
cg22477971	1	*C1QB*	6.1E-04	1	0.999	0.137927	1	0.907

These are all the DBP associated sites with an FDR <0.4

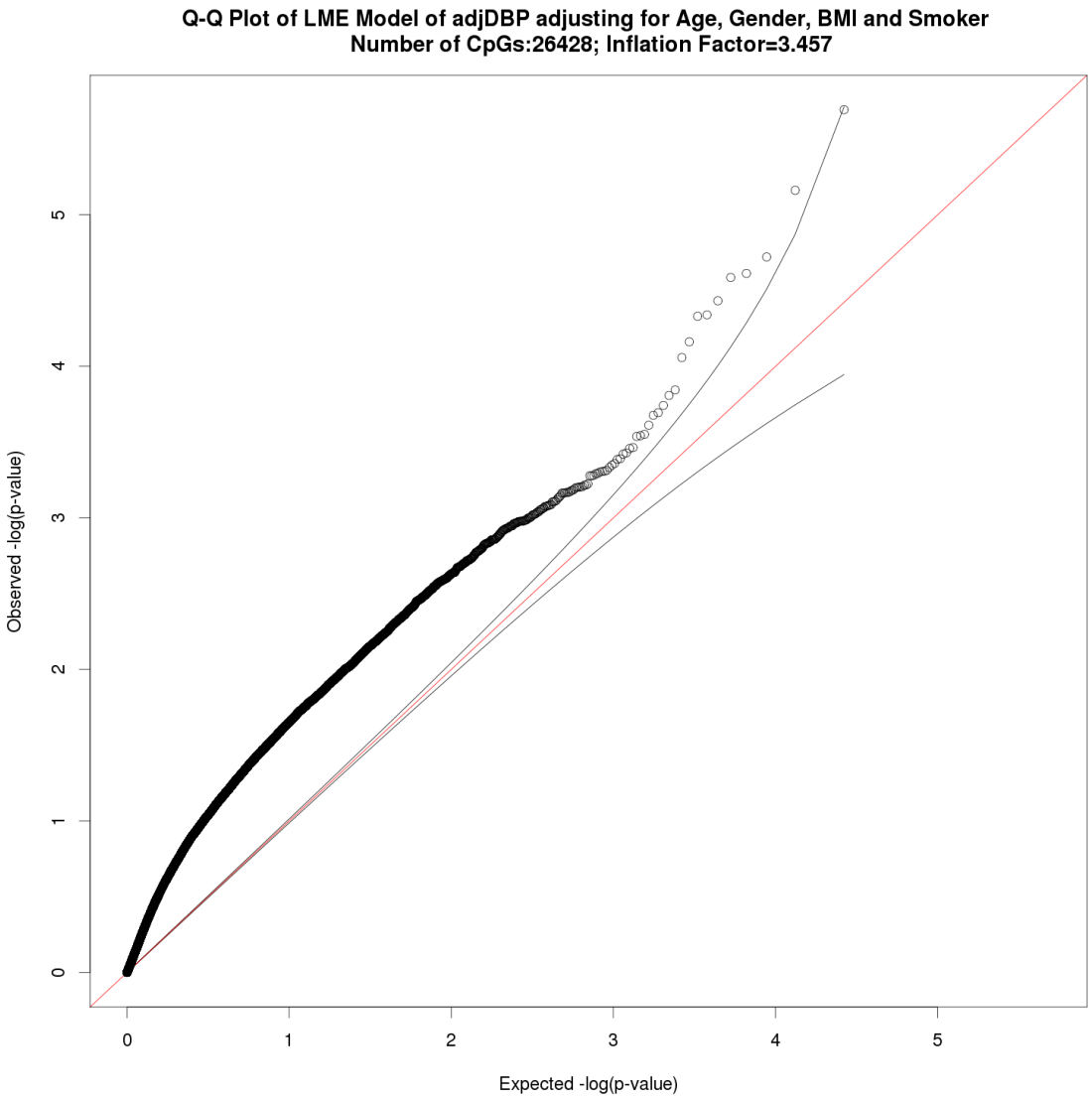
Appendix Table 3. Relationship of the most significant sites for the Wang Study compared to this analysis

			SBP			DBP		
Methylation Site	**Chromosome**	**Gene**	**Uncorrected P-value**	**Bonferroni Corrected P-value**	**FDR Corrected P-value**	**Uncorrected P-value**	**Bonferroni Corrected P-value**	**FDR Corrected P-value**
cg04845579	8	*SULF1*	0.009	1	0.742	0.0003	1	0.999
cg02283643	8	*SULF1*	0.011	1	0.746	0.040	1	0.999
cg09772827	11	*PRCP*	0.123	1	0.896	0.281	1	0.999
cg27561006	11	*PRCP*	0.739	1	0.991	0.492	1	0.999

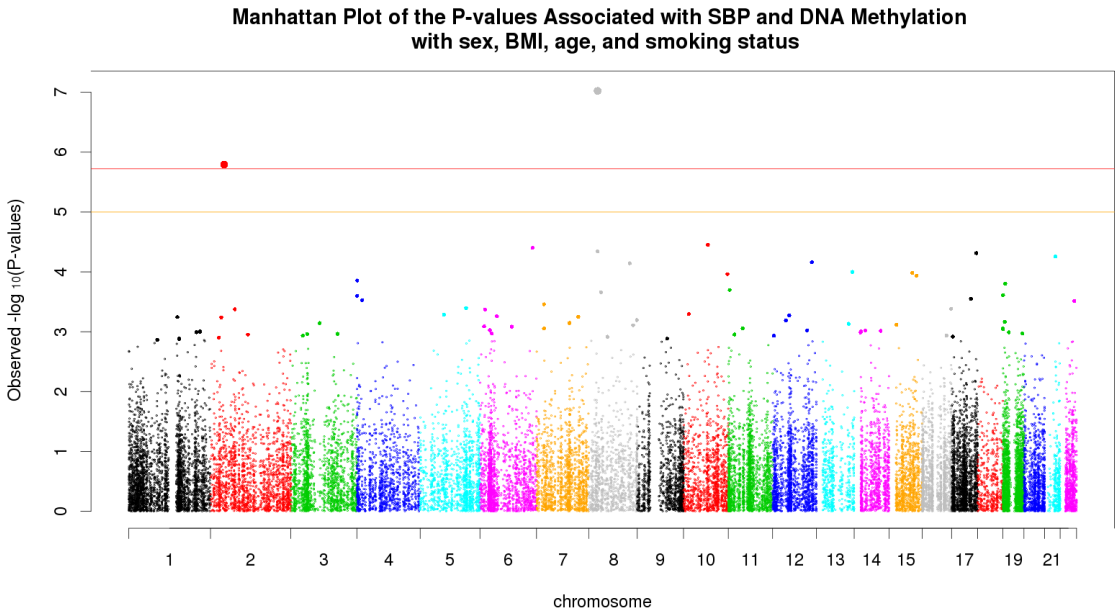
These are all the associated *PRCP* and *SULF1* sites



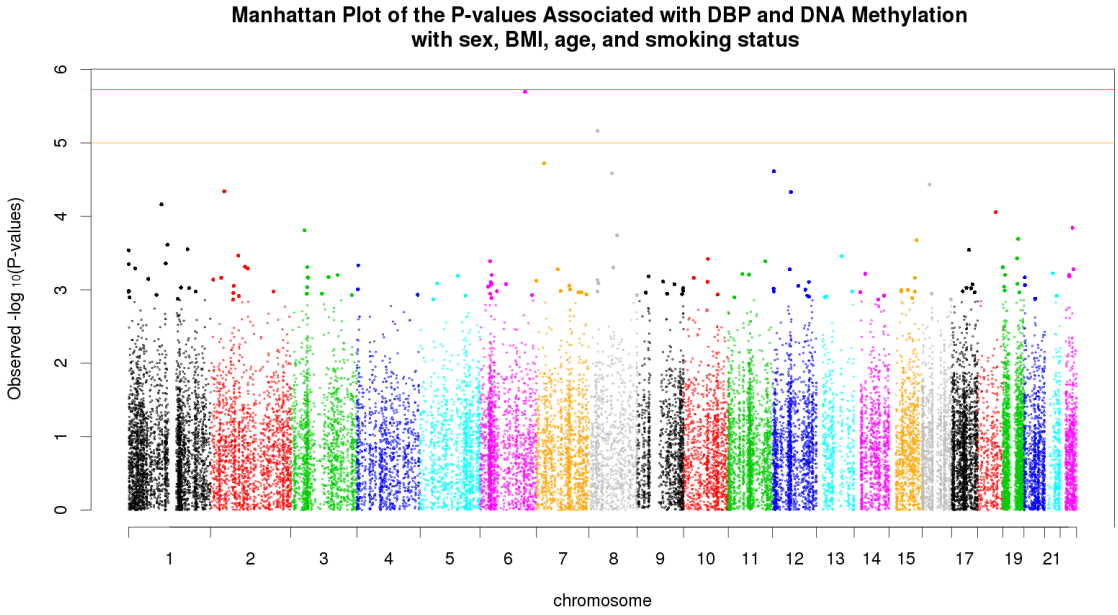
Appendix Figure 4. Log quantile-quantile plot of the observed and expected p-values for adjusted SBP using a linear mixed effects model without principal components



Appendix Figure 5. Log quantile-quantile plot of the observed and expected p-values for adjusted DBP using a linear mixed effects model without principal components



Appendix Figure 6. Manhattan plot of the observed p-values for adjusted SBP using a linear mixed effects model without principal components. The red-line represents the Bonferroni corrected p-value significance level. The yellow-line represents the False Discovery Rate significance level.



Appendix Figure 7. Manhattan plot of the observed p-values for adjusted DBP using a linear mixed effects model without principal components. The red-line represents the Bonferroni corrected p-value significance level. The yellow-line represents the False Discovery Rate significance level.