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:

_____ _____

Wenruo Hu Date

Genome-wide gene-smoking interaction of Systolic and Diastolic blood pressure among Caucasians

By

Wenruo Hu

Master of Public Health

Epidemiology Department

_____

Yan Sun, PhD

Committee Chair

Genome-wide gene-smoking interaction of Systolic and Diastolic blood pressure among Caucasians

By

Wenruo Hu

Bachelor of Science

Capital Medical University, Beijing, China

2011

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 department

2013

ABSTRACT

Genome-wide gene-smoking interaction of Systolic and Diastolic blood pressure among Caucasians

By Wenruo Hu

High Blood Pressure (HBP) is a leading cause of death in the United States. The attributed mortality rate of HBP remains at a high level in recent years. Cigarette smoking is a major preventable cause of death in U.S. Our study aims to understand etiology of HBP, including genetic susceptibility of high systolic and diastolic blood pressure among the Caucasian population as well as cigarette smoking modification effect on the association between single nucleotide polymorphisms (SNPs) and blood pressure. Statistical tests were used to investigate difference between hypertensive and normotensive study sample. Marginal linear mixed regression model were conducted to discover association between main effect, SNP-smoke interaction effect and blood pressure trait using a genome-wide approach. None of SNPs were detected significantly associated with SBP from both additive and genotypic model. Eight SNPs were identified associated with DBP significantly (FDR Q value < 0.05 with genomic control applied). Seven SNP-smoking interactions were found to be significantly associated with SBP from the additive interaction model and three such significant associations were uncovered from genotypic model. No smoking modification effect associated with DBP. In conclusion, current smoking status could modify the magnitude of SNP-SBP associations, whereas no evidence supported the similar modify of SNP-DBP associations.

KEYWORDS

Genome-wide gene-environment interaction, High Blood Pressure, SNP, cigarette smoking, Caucasian

Genome-wide gene-smoking interaction of Systolic and Diastolic blood pressure among Caucasians

By

Wenruo Hu

Bachelor of Science

Capital Medical University, Beijing, China

2011

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 department

2013

ACKNOWLEDGEMENTS

I would love to thank my thesis advisor, Dr. Yan Sun, who helped me the most. Not only had he taught me how to conduct this one specific research, from designing to preforming, but also he trained me about the logic, principle and thinking process of genomic epidemiologic studies, which are valuable for my future career. I also want to thank two mentors in Dr. Sun's team, who are Qin Hui and Yuanyuan Fang. They helped me a lot with detailed R programming and useful advice. In addition, I would like to thank Rollins School of public health Epidemiology department, Rollins Tech help office and Woodruff Health Sciences library for giving me such a wonderful academic training, resources and support.

Table of Contents

BACKGROUND & INTRODUCTION

BURDEN OF HIGH BLOOD PRESSURE & CIGARETTE SMOKING

Hypertension is a leading health issue in the United States. The diagnosis of essential hypertension is made when the average of 2 or more diastolic blood pressure (DBP) measurements on at least 2 subsequent visits is ≥90 mm Hg or when the average of multiple systolic blood pressure (SBP) readings on 2 or more subsequent visits is consistently ≥140 mm Hg. Essential, primary, or idiopathic hypertension is defined as high blood pressure (HBP) [1]. According to the latest update of American Heart Association statistics, one in three US adults had high blood pressure in 2012. From 1988 to 1994 through 1999 to 2002, the prevalence of high blood pressure among Caucasian increased from 24.3% to 28.1%. The attributed mortality rate of high blood pressure during this time period was 16.5 per 100,000 population per year for Caucasian males and 14.5 per 100,000 population per year for Caucasian females [2]. The total population of the U.S. was about 310.2 million and 64.7% of them were the Caucasian population based on 2010 U.S census statistics [3]. The total direct cost due to HBP is about 131 billion dollars annually in the US, which included health care service, anti-hypertensive medication, and other relevant direct medical expenses. And approximate cost that related to loss of productivity is 25 billion dollars per year [4].

Tobacco consumption, including direct and indirect cigarette smoking, is the major preventable cause of death in the US. Direct cigarette smoking is a well-documented risk factor for lung cancer [5] and respiratory diseases [6], especially chronic obstructive pulmonary disease (COPD) [7]. Other than the respiratory adverse effect, cigarette

smoking is essential risk factor for cardiovascular system. Hypertension [8] and coronary heart disease [9] are closely associated with cigarette smoking. Hypertensive smokers tend to develop more severe forms of hypertension, malignant and renovascular hypertension, than non-smokers [10]. In addition, direct smoking could also cause preterm birth [11] and rheumatoid arthritis (RA) [12]. Secondhand smoke (SHS) is as important as direct cigarette smoking. SHS has been proven by multiple studies that causes or associates with lung cancer [13], respiratory disease [14], heart disease [15] and preterm birth [16]. The cardiovascular effect of SHS is substantial and rapid and that is as large as smoking [17]. In 2010, about 45.3 million (19.3%) of US adults were current smokers and 21.0% of them were Caucasians [18]. The attributed death for smoking and secondhand smoking (SHS) is about 443,000 annually. Shockingly, it is even more deaths than combined cause of death from HIV, drug abuse, alcohol use, motor vehicle injuries, suicides and murders [19]. From 2000 to 2004, about $96 billion dollars were spent on directed medical costs. Additionally, $97 billion were lost due to loss of productivity.

EPIDEMIOLOGY OF HIGH BLOOD PRESSURE

Blood Pressure (BP) is a complex quantitative trait that could be influenced by multiple factors including genetic factors, environmental factors and social-psychological factors. For instance, obesity, insulin resistance, older age, high alcohol intake, high salt intake, low potassium intake, low calcium intake, sedentary lifestyle, stress are all risk factors of hypertension [1]. Behavioral and socioeconomic factors can modify the genetic effects on blood pressure phenotypes. Distinct genetic effect on DBP was found among ever smokers and never smokers. Current alcohol drinkers had significant higher SBP

and DBP comparing to never drinkers. Moreover, physical exercise and education also have independent effects on BP [20].

There are age, sex and racial/ethnic disparities in high blood pressure incidence in the U.S. A linear rise in SBP from age 30 through 84 years and concurrent increase in DBP was found in Framingham Heart study by Franklin et.al in 1997 [21]. Martins et.al examined data from the Third National Health and Nutrition Examination Survey and discovered that SBP was higher in males than in females among adults whose ages were less than 45 year-old. Among adults who were older than 45 year-old, SBP was higher in females. Diastolic blood pressure was lower among adult females than males across all age categories [22]. Carson et.al reported that African Americans had higher hypertension prevalence compared to Caucasians between 45 and 74 years of age controlling for gender and study sites. The Chinese population and the Hispanic population were not significantly different from the Caucasian population at the same age range [23].

Environmental risk factors, for example, smoking, body mass index (BMI) and alcohol consumption, causes or associated with hypertension. Cigarette smoking was modestly associated with an increased risk of developing hypertension in a larger cohort of women. The age-adjusted hazard ratio (HR) of developing hypertension among current smokers of 15 cigarettes per day was 1.10 (95% CI 1.01 to 1.19). Among women who smoked over 25 cigarettes per day, the HR of hypertension was 1.21 (95%CI 1.06 to 1.39) [24]. Multiple studies suggested that maternal tobacco consumption is associated with hypertension of their children. Babies that were exposed to tobacco smoke in utero had higher risk of obesity, hypertension and gestational diabetes mellitus (DM) after they

grow up as adults. The corresponding ORs and 95% confidence intervals (CI) compared exposed group to non-exposed group for each disease were 1.53 (1.45, 1.61) for obesity, 1.69 (1.19, 2.39) for hypertension and 1.32 (1.10, 1.58) for GDM [25]. The Decoda Study Group conducted a meta-analysis, which summarized 16 cohort studies from the Diabetes Epidemiology: Collaborative Analysis of Diagnostic criteria in Asia (DECODA) study and found out BMI, waist circumference (WC), waist-to-hip ratio (WHR), and waist-to-stature ratio (WSR) were equally strong predictors that associated with hypertension. The corresponding ORs for hypertension were 1.68, 1.66, 1.45 and 1.63 respectively [26]. Taylor et.al combined 12 longitudinal quantitative studies from the United States, Japan and Korea and illustrated that a linear positive relationship existed between alcohol consumption and hypertension risk among both males and females [27].

Psychological influence on blood pressure has been emphasized in recent years. Stress is associated with blood pressure linearly. The risk of high blood pressure increased 1.06 times if the perceived stress increased 5 points on a stress scale [28]. SBP reactions to stress correlated with follow-up SBP positively, while no such an association was found for DBP. BP reactions to mental stress could predict future BP status. The magnitude of the prediction appears to vary with socioeconomic position and sex [29]. A longitudinal study from 1994 to 2004 was identified association between major depression (MD) and new-onset high blood pressure. People who suffered from MD had 60% increased risk of HBP in contrast to HBP risk for people who didn't have MD [30].

GENOMIC EPIDEMIOLOGY STUDIES OF HUMAN DISEASES

Human genome epidemiology involves in gene discovery, development and application of genetic tests for diagnosing and treating various disease, or even predicting

and preventing future disease in asymptomatic person. It denotes an evolving field of inquiry that uses systematic applications of epidemiologic methods and approaches to the human genome to assess the impact of human genetic variation on health and disease [31]. Gene discovery is a traditional domain of genetic epidemiology applied to discover novel genes, in which study type includes linkage analysis and family-based association studies [32]. Gene characterization is aiming at investigating population prevalence, genotype-disease association, gene-gene association and gene-environment interaction, which belongs to molecular epidemiology field [33]. In terms of evaluating and assessment health impact, evaluation studies for validity and utility of genetic information in clinical epidemiology and public health epidemiology gets involved [34].

As the development of genotyping technologies, Genome-wide association studies (GWAS) have rapidly become a standard method for disease gene discovery. Because of no-priori hypotheses characteristic, GWAS is an efficient way and widely used to uncover association between novel locus (main effect) and complex disease and traits among different population. Since 2005, about 100 loci for as many as 40 common diseases and traits have been identified and replicated using GWAS approach [35].

Despite revolutionary position GWAS takes, it has several limitations [36]. Firstly, GWAS results can mostly explain a small proportion of heritability. Complex trait such as blood pressure is known to be affected not only by genetic factors but also by many environmental factors. Secondly, considering interaction between the genetic and environmental factors could offer us a better understanding about how the environmental factors modified the genetic effect to certain disease, which allows us to generate important implication for public health intervention. The hypothesis-based association

study for genes and environment interaction and complex disease trait could be useful complement for GWAS. However, inasmuch as the method is based on pre-study hypotheses, it has less flexibility compared to GWAS. Gene-environment wide association studies (G×E study) can be more comprehensive to explain disease etiology by combining both the two methods using hierarchical modeling strategies [37]. In our study, we constructed hierarchy formulated models using gene-environment wide association method to investigate association between SNP-smoke interaction and blood pressure.

PREVIOUS GENOMIC EPIDEMIOLOGY STUDIES ASSOCIATE WITH BLOOD PRESSURE

GWAS and corresponding meta-analysis studies for blood pressure and hypertension provides better understanding on the discovery efforts and impact on public health and clinical researches.

Fox et.al discovered one SNP for DBP and one for SBP attained genome-wide significance in the meta-analysis of GWAS data. SNP that had the strongest association with DBP was rs10474346 ($P=3.6\times10^{-8}$). And rs2258119 was the most significant SNP associated with SBP ($P=4.7\times10^{-8}$). The location of rs10474346 is close to GPR98 and ARRDC3 gene. And rs2258119 is located within C21orf91gene region [38]. Newton et.al identified association between SBP or DBP and common variants in 8 regions near the CYP17A1 ($P=7\times10^{-24}$), CYP1A2 ($P=1\times10^{-23}$), FGF5 ($P=1\times10^{-21}$), SH2B3 ($P=3\times10^{-18}$), MTHFR ($P=2\times10^{-13}$), c10orf107 ($P=1\times10^{-9}$), ZNF652 ($P=5\times10^{-9}$) and PLCD3 ($P=1\times10^{-8}$) genes. All variants associated with continuous blood pressure were associated with dichotomous hypertension status [39].

In 2012, a published research suggested that DBP significant associated with rs1605685 on chromosome 5 among Chinese population utilizing combined linage and association study [43]. Combining 6 cohort studies in CHARGE consortium by meta-analysis, eight loci (ATP2B1, CYP17A1, PLEKHA7, SH2B3, CACNB2, CSK-ULK3, TBX3-TBX5, ULK4) were found significantly associate with SBP or DBP among US and Europe population using a cross-sectional design, in which ATP2B1 and SH2B3 were associated with both SBP and DBP [44]. For African American population, five variants, PMS1, SLC24A4, YWHA7, IPO7, and CACANA1H, were uncovered reaching genome-wide significance for SBP. Of all five loci, SLC24A4, IPO7, and PMS1 remained significant result in the following replication analysis [45]. In addition, other than measurement of blood pressure, hypertensive status was studied as the outcome variable with case-control study design preforming logistic regression model. A statistically significant association between rs13333226 and hypertension was found using such a study design. SNP rs13333226 is located close to Uromodulin (*UMOD*) gene on chromosome 16. The increase of minor G allele variant was related to decreased risk of hypertension (OR=0.87, 95% CI 0.84 – 0.91) [46]. In 2012, a novel locus, rs3918226, located in the promoter region of the endothelial nitric oxide synthase (eNOS) gene was found associate with hypertension. The overall odds ratio from further meta-analysis was 1.34 per T allele (95% CI 1.25 - 1.44, $p=1.032\times10^{-14}$) [47].

There are gene-environment interaction studies that illustrated association between effect of interaction and blood pressure. In 2007, a significant multiplicative interaction between GNB3 genotype, obesity and physical activity in predicting hypertension status was observed by Grove et.al among African Americans [48]. Interaction between FPR1

C32T SNP and age was discovered significantly associate with increased BP level within 5 years. And no such an interaction association was discovered for gender and BMI among French middle-aged adults [49]. The rs1879282 SNP within the calpain (CAPN) 13 gene areas on chromosome 2 had significant gene-BMI interaction on both systolic blood pressure and diastolic blood pressure among African American female offspring, which suggested that association between rs1879282 and BP id depend on BMI level. Four SNPs, rs4035540, rs10499859, rs3771452, and rs12508955, have interaction with BMI that associate with DBP [50]. Alcohol consumption could also modify gene-hypertension association. Interaction between LDH2 genetic polymorphism and alcohol intake was found associated with development of hypertension in a prospective cohort among Asian population. The risk associated with the rs2238152 T allele was stronger in heavy/moderate alcohol drinkers and was reduced in non-drinkers [51].

Furthermore, cigarette smoking served as environmental factor, was studied in many G×E interaction studies. The interactions between lipid-related SNPs of ABCA-1, ACAT-1 and PCSK9; ACAT-1, LDL-R, MTHFR and PCSK9; and ABCA-1, LIPC, PCSK9 and PPARD, and cigarette smoking were detected by factorial regression analysis that could influence SBP and DBP levels among the Chinese population in 2012 by Rui-Xing Yin et al [52]. In 2009, Karlson et.al reported a strong gene-environment interaction was observed between HLA-SE and cigarette smoking that associated with rheumatoid arthritis (RA) among Caucasian population. A strong additive interaction (attributable proportion due to interaction (AP)=0.50, $p<0.001$) and significant multiplicative interaction ($p=0.05$) were found between heavy smoking (>10 pack-years) and any HLA-SE in seropositive RA risk [53] Helbig et.al observed significant negative gene-smoking

interaction related to Crohn's disease (CD) among Germany, essentially Caucasians, which suggested that the risk increase for CD conferred simultaneously by cigarette smoking. The risk of CD among ever-smokers who carried at least one of the NOD2 risk alleles was about 29% lower than non-carriers, while the risk was 32% lower among never-smokers who were also the NOD2 carriers [54].

METHODS

STUDY DESIGN

A cross-sectional study design was used to conduct Genome-wide association study of gene-smoking interaction. Each SNP was first assessed for main effect by following no-interaction regression analysis. And further interaction regression models were used to evaluate all SNP-smoke interactions.

HYPOTHESIS

The study hypothesis is that cigarette smoking modifies genetic susceptibility of blood pressures among Caucasians.

STUDY SAMPLE

The study was approved by the Institutional Review Board (IRB) at Emory University. The dataset that used in this study is the Caucasian subset from Rochester, MN as part of the Genetic Epidemiology Network of Arteriopathy (GENOA), including 1371 individuals whose genotype were performed on the Affymetrix® Genome-Wide Human SNP Array 6.0. GENOA is one of four research networks that form the National Heart Lung and Blood Institute (NHLBI) Family Blood Pressure Program (FBPP) containing two sibling cohorts. The GENOA data already had the quality control performed. Study sample that have SNP call rate and individual call rate lower than 95% was removed, as well as duplicated samples and sex mismatch data. SNPs with minor allele frequency (MAF) less than 0.01 and significant Hardy-Weinberg Equilibrium (HWE) result (HWE P value $<10^{-6}$) were removed.

Data collection consisted of demographic information, medical history, clinical characteristics, lifestyle factors, and blood samples for genotyping and biomarker assays [55]. Blood pressure was measured with random zero sphygmomanometers and cuffs appropriate for arm size. BMI was obtained by the standard calculation of weight (kg) divided by height squared (m^2). Smoking status was documented as current smoker, previous smoker and never-smoker. Current smoker is defined as people who smoked cigarette within one year prior to the GENOA data collection. Previous smoker is defined as people who have smoking history but didn't smoke within one year prior to the data collection. And never-smoker is defined as people who never smoke cigarette. In order to simplify analysis and construct efficient model, smoking status was converted into dichotomized variable representing current smoking status. All current smokers were in current smoking group that coded as 1. Both previous smokers and never smokers are account in non-current smoking group that coded as 0.

Hypertensive status was determined using 140 mm Hg and 90 mm Hg as cut-point. Individuals with SBP≥140 mm Hg and/or DBP≥90 mm Hg were identified in hypertensive group. Otherwise, they were identified in normotensive group. Because majority of individuals in the GENOA dataset were consuming antihypertensive medication, their blood pressure cannot directly be used in the analysis. For people who were taking antihypertensive treatment, 10 mm Hg and 5 mm Hg was added for SBP and DBP respectively [56].

STATISTICAL ANALYSIS

Statistical description was conducted using SAS 9.3 and R 2.15.2. For statistical description, mean and standard deviation are used to describe continuous variables and

count for each category and its percentage are used for categorical variables. Naïve statistical analysis was conducted to depict hypertensive and normotensive group of sample and to determine possible influence factors for blood pressure. Two-sample T Tests were conducted to investigate difference between two groups for SBP, DBP, age, BMI, cholesterol level, high density lipoprotein and TG. And Chi-square tests were used for current smoking status, gender, medication, medical history of parents and alcohol drinking. If the tested variables have homogeneity of variance between hypertensive and normotensive group, Pooled T test result should be accepted. If the tested variables have heterogeneity of variance between the two groups, Satterthwaite T-test result should be accepted.

Regression analysis is conducted in order to depict relationship between the predictors and the outcome controlling for confounding effect. Due to the unbalanced distribution of the binary hypertensive variable, we decided to use the continuous SBP and DBP value as the outcome. In addition to correlated data structure, linear mixed effect model was chosen for the further analysis to take into account family aggregation effect [55]. Since we are not interested in discovering G×E effects for each family, marginal linear mixed model with exchangeable correlation structure for random errors is a proper type to use. To simplify the analysis, only single SNP effect was considered. No SNP-SNP interaction is considered in the model. G×E interactions other than gene-smoking interaction are not allowed in the model as well.

Based on the literature and adjusted by previous analysis [57 - 58], age, gender and BMI were considered as potential confounders that should be controlled in the model. Main effect, SNPs, was coded as (0, 1, 2) for additive models, which represented number

of mutated alleles for each SNP. In genotypic models, two dummy variables, SNP_1 and SNP_2, were created to represent three types of genotypes. If the given SNP has no mutated allele, both SNP1 and SNP2 are coded as 0. If the SNP has one mutated alleles, SNP_1 should be 1 and SNP_2 should be 0. And if both alleles on the given SNP is mutated, SNP_2 is 1 and SNP_1 is 0. Models are stated below.

Model 1. Additive Main effect model:

$$Adjusted\ SBP = \alpha + \beta_1 SNP + \beta_2 Smoke + \beta_3 Age + \beta_4 Gender + \beta_5 BMI + \varepsilon,$$

$$Adjusted\ DBP = \alpha + \beta_1 SNP + \beta_2 Smoke + \beta_3 Age + \beta_4 Gender + \beta_5 BMI + \varepsilon,$$

where α denotes intersection term; βs are correlation coefficients for predictor variables;

Model 2. Additive Interaction model:

$$Adjusted\ SBP = \alpha + \beta_1 SNP + \beta_2 Smoke + \beta_3 Age + \beta_4 Gender + \beta_5 BMI + \delta (SNP \times Smoke) + \varepsilon,$$

$$Adjusted\ DBP = \alpha + \beta_1 SNP + \beta_2 Smoke + \beta_3 Age + \beta_4 Gender + \beta_5 BMI + \delta (SNP \times Smoke) + \varepsilon,$$

where α denotes intersection term; βs are correlation coefficients for predictor variables; δ stand for correlation coefficients of interaction term;

Model 3. Genotypic Main effect model:

$$Adjusted\ SBP = \alpha + \beta_1 SNP_1 + \beta_2 SNP_2 + \beta_3 Smoke + \beta_4 Age + \beta_5 Gender + \beta_6 BMI + \varepsilon,$$

$$Adjusted\ DBP = \alpha + \beta_1 SNP_1 + \beta_2 SNP_2 + \beta_3 Smoke + \beta_4 Age + \beta_5 Gender + \beta_6 BMI + \varepsilon,$$

where α denotes intersection term; βs are correlation coefficients for predictor variables; SNP_1 and SNP_2 are two dummy variables representing three genotypes;

Model 4. Genotypic Interaction model:

$$Adjusted\ SBP = \alpha + \beta_1 SNP_1 + \beta_2 SNP_2 + \beta_3 Smoke + \beta_4 Age + \beta_5 Gender + \beta_6 BMI + \delta_1(SNP_1 \times Smoke) + \delta_2(SNP_2 \times Smoke) + \varepsilon,$$

$$Adjusted\ DBP = \alpha + \beta_1 SNP_1 + \beta_2 SNP_2 + \beta_3 Smoke + \beta_4 Age + \beta_5 Gender + \beta_6 BMI + \delta_1(SNP_1 \times Smoke) + \delta_2(SNP_2 \times Smoke) + \varepsilon,$$

where α denotes intersection term; βs are correlation coefficients for predictor variables; δs stand for correlation coefficients of interaction terms;SNP_1 and SNP_2 are two dummy variables representing three genotypes;

T statistics, standard deviation and corresponding p values were obtained from the output of the model in additive models. Likelihood ratio statistics were calculated as Likelihood Ratio (LR) = $(-2\log(L_{reduced})) - (-2\log(L_{full}))$, where L stands for likelihood for full and reduced model, model 4 is full model and model 3 is a reduced model. P values were calculated using the 1-df Chi-square distribution. Quantile-Quantile Plot (Q-Q plot) was generated to compare observed and expected probability distributions.

GENOMIC CONTROL

In order to correct artificial difference in allele frequencies, it is necessary to calculate inflation factor (λ), which is useful to detect and correct inflation [59 - 60]. Since the sample size is relatively large, Y^2/λ is approximately Chi-square distribution under the null hypothesis. Due to the test statistic in the additive model is T-statistics rather than chi-square statistic. A transformation was performed. T-statistic can be approximated into Z-statistic result from the large sample size. And by square the Z statistic, we can

calculate 1-df chi-square statistics, which is Y^2 for additive models. For the genotypic model, the likelihood ratio statistic is under the 1-df chi-square distribution, so LR= Y^2 for genotypic models. The corresponding p value for additive model (model 1, model 2) can be obtained from chi-square distribution with 1 degree of freedom. And for genotypic model (model3, model4), the corresponding p value can be calculated from 2 degree of freedom chi-square distribution.

MULTIPLE TESTING ADJUSTMENTS:

False Discovery Rate (FDR Q value) adjustment [61] was conducted because of the large number of SNPs examined in this study. The FDR Q value significant threshold is 0.05.

RESULTS

DESCRIPTIVE STATISTICS AND STATISTICAL TESTS

1371 Caucasians, from 555 sibships, enrolled in the study with both phenotypic and genotypic data. The sibship size ranges from 1 to 14. The population characteristics were shown in table 1. Among 1371 individuals, 756 of them were female and 615 of them were male. The youngest person was 25 year-old and the oldest one was 90 year-old. Their average of unadjusted blood pressure for systolic blood pressure was 133 mmHg and for diastolic blood pressure was 78 mmHg. Inasmuch as 996 (73%) of study population were taken anti-hypertension medication, adjusted BP was calculated in order to take into account influence of medication. The average of adjusted SBP and DBP measurement was 139.5 mm Hg and 81.7 mm Hg respectively. The percentage of current smoking status was 14.3% and the percentage of previous smokers and never smokers was 85.7%.

The stratified characteristics and unadjusted statistical analysis results shown in the table 2. Based on hypertensive classification, the study population can be divided into two groups, hypertensive and normotensive group. The average age of hypertensive groups was about 58 year-old, while normotensives was 49 year-old showing significantly difference between two groups. Hypertensive group had more current smokers than normotensive group, but its percentage was lower than normotensives. BMI and waist-hip ratio of normotensives was significantly lower. Correspondingly, some blood index regard of hypertension and cardiovascular disease indicated comprehensible consistent result with previous studies. The cholesterol and Triglyceride level for hypertensive group was significantly greater than normotensive group. High density

lipoprotein (HDL), as a protective factor for cardiovascular system, was relatively low in hypertensive people compared to normotensive people. Majority individuals in the hypertensive group were taking medication as well as a few normotensives. Furthermore, parent's hypertension history and alcohol significantly associated with hypertensive status. The crude analysis result supported choice of confounding variables that suggested in the literatures.

REGRESSION ANALYSIS OF MAIN GENETIC EFFECT

Both the additive model and genotypic model did not identify significant genetic susceptibility for SBP based on FDR Q value after genomic control (GC). The Q-Q plot of additive model (Figure 1) did not indicate inflation of low p-values. The inflation factor (λ) for result of additive genetic effect was very close to 1 (i.e. overall p-value distribution of null hypothesis) . The top two significant SNP, rs8497104 and rs1855589, had positive estimated beta coefficients, which were 3.6 and 3.2 respectively. Their p values were 1.85×10^{-6} and 4.16×10^{-6} respectively. After FDR Q value, none of them were significant (P>0.05). The original model-based p-values from genotypic model, on the contrary, identified over 50 significant SNP-SBP associations but with inflation factor of 1.27. After adjustment of genomic control and multiples testing, none of the associations were statistically significant (Figure 2). The most significant SNP in genotypic model was rs4210869 with FDR Q value of 0.07.

The additive model without interaction did not identify any significant SNP-DBP interactions using an FDR Q value cut-off of 0.05. The inflation was negligible in the additive model (λ=1.01). However, rs2181608 was identified distinctively more significant compared to all the other SNPs (Figure 3). It had an original p-value of

1.29×10^{-7} and FDR Q value as 0.0862 compared to original p-value of 2.67×10^{-6} and FDR Q value 0.53 for the second significant SNP, rs2090418. The estimated beta coefficient for rs2181608 was -3.78 with standard deviation of 0.7. Using FDR 0.05 as the cut-point, we identified 8 SNPs significantly associated with DBP in genotypic model after controlling for inflation factor (λ=1.26).

REGRESSION ANALYSIS FOR G×E INTERACTION

Among all significant SNP-smoking interaction in the additive model (table 4, graph 5), rs2212885, rs8526790, rs4230236, rs8694580, rs2083807, rs2213442, rs8596677 were significantly associated with SBP. The absolute value of estimated beta coefficient for smoking status and SNP-smoking interaction terms were significantly larger than the absolute value of estimated beta coefficient for SNP's. Interaction terms between rs2212885, rs8526790, rs4230236, rs8694580, rs2213442, rs8596677 and smoking status had negative beta values, whereas rs2083807-smoking interaction had positive beta value.

rs4269045, rs8502813 and rs937307 were identified having significant SNP-smoking interaction using the genotypic model (Figure 5, Figure 6). The SNP that had the most significant interaction was rs4269045, which has GC adjusted P value of 1.35×10^{-8} and corresponding FDR Q value of 0.0091. The second and the third most significant SNP-smoking interaction associated with SBP were rs8502813, rs937307, which had the same GC adjusted FDR Q value, 0.0202.

No SNP-smoking interaction was identified to be significantly associated with DBP from either the additive model or the genotypic model. The most significant FDR Q value after genomic control was 0.6 in the additive model and was 0.2 in the genotypic models

(Figure 7, Figure 8). The genome-wide results of the genotypic model were also inflated (λ=1.29).

DISCUSSION

Combining results from main effect model and interaction model, we can conclude that current smoking status was a significant effect modifier of genetic susceptibility for SBP. Overall 10 SNPs, rs2212885, rs8526790, rs4230236, rs8694580, rs2083807, rs2213442, rs8596677, rs4269045, rs8502813 and rs937307, were identified from two genetic models (additive and genotypic), . They had significant interactions with cigarette smoking that associated with SBP.

Searching the top ten most significant SNPs in NCBI dbSNP database, only four of them, rs2212885, rs2083807, rs2213442 and rs2294553, are located in the genic regions. SNP rs2083807 is located in the intron region of LOC100652856, which is currently uncharacterized. It could relate to start of sequence. And SNP rs2294553, the 9th significant SNP, is located in the intron region of *PLCB1* gene, which codes the protein that catalyzes the formation of inositol 1,4,5-trisphosphate and diacylglycerol from phosphatidylinositol 4,5-bisphosphate. This reaction uses calcium as a cofactor and plays an important role in the intracellular transduction of many extracellular signals. *PLCB1* gene was detected in in orbito-frontal cortex samples of schizophrenia-affected patients and in bipolar disorder affected patients [62 - 63]. There is no previous study that reported association between *PLCB1* and blood pressure.

8 out of 10 most significant loci using the genotypic model, rs4269045, rs937307, rs2219771, rs1906984, rs2236815, rs1816847, rs1871152, rs2286468, were validated in dbSNP from NCBI. The SNP with the most significant SNP-smoking interaction was rs4269045, which is located in *RBMS3* gene. *RBMS3* encodes an RNA-binding protein that belongs to the c-myc gene single-strand binding protein family. It may be involved in

a cytoplasmic function, such as controlling RNA metabolism. It is expressed in activated hepatic stellate cells and liver fibrosis indicating novel mechanism of liver fibrosis [64]. The third most significant SNP rs937307 relates to *ZDHHC17* gene, which is a protein coding gene also known as HIP3, HYPH, HIP14, HSPC294. ZDHHC17 has been implicated in genetic neurological disorders by regulating protein palmitoylation. Down-regulation of ZDHHC17 by siRNA in vitro results in increased cell death in neurons, which relates to Huntington's disease [65]. Neither those genes nor those 8 significant SNPs were reported associated with blood pressure. Further study is needed to verify association or causal relationship between BP and SNP-smoking interaction.

The result of the study has several clinical implications. It can be used to identify subjects with genetic susceptibility and provide them prevention strategy, such as monitoring their BP, providing necessary early-stage medication, and educating them with healthy lifestyle advice. In addition, unraveling the role that functional gene polymorphisms play in determining risk and in determining the levels of intermediate phenotypes is crucial to our understanding of the key metabolic pathways and physiology not only in the diseased, but also in the disease-free state [66].

There are some limitations in this study. Most of the sample population was hypertensive. Both their adjusted and unadjusted average SBP and DBP were in high normal category of BP classification [6]. The cross-sectional study design has less power to investigate causal relationship than case-control or cohort study. As a phase I research, however, it was enough to discover novel SNP-smoking interactions for BP to generate hypotheses for further studies. In next stage of research, we could conduct replication studies in different Caucasian population to verify SNPs with significant SNP-smoke

interaction associate to BP from the first stage using independent samples, particularly the untreated samples to eliminate the influence from the antihypertensive medication.

Interestingly, hypertensive group had less current smokers than normotensive groups. Because smoking is a risk factor of high blood pressure, hypertensive group was expected to have more smokers than normotensive group. It could be explained by the following reasons. First, people who got hypertension are more likely to quit smoking in order to control their BP. Second, there could be misclassification bias because of binary coded environmental influence factor.

Exposure assessment is a big challenge when we engage in environmental influence factors. Firstly, cigarette has complex compound and each one of the chemical has different effect for human being. We do not know the effective constituent in cigarette that could adjust association between genetic susceptibility and blood pressure. Secondly, the effective dosage and period of cigarette smoking was not considered, which leads to ignoring variability of environment effect. In addition, it is segmentary that only direct cigarette smoking status was considered as the environmental risk factor in the study models. As previous literature suggested, the effect of secondhand smoking on cardiovascular system is almost as influential as direct smoking [17]. Biomarkers can be useful to monitoring individual smoking effect. Oral fluid cotinine, nicotine, OH-cotinine and norcotinine, which can be obtained from oral fluid, are common biomarker indicating smoking status that is better substitute than just a binary smoking documentation. OH-cotinine performed efficiently to separate smoker from non-smokers. Norcotinine is a better choice to differentiate severity of cigarette smoking [67]. We should consider using

continuous biomarkers level as a more accurate environmental exposure variable in the future G×E studies.

REFERENCE

[1] Oscar A. Carretero, Suzanne Oparil, Essential Hypertension Part I: Definition and Etiology [J]. Circulation, 2000, 101(3): 329-335

[2] Ve´ronique L. Roger, Alan S. Go, Donald M. Lloyd-Jones, et.al. Heart Disease and Stroke Statistics—2012 Update: A Report from the American Heart Association [J]. Circulation, 2012 125(1): 188-197

[3] U.S. Census Bureau, Statistical Abstract of the United States: 2012

[4].Heidenreich PA, Trogdon JG, Khavjou OA, et al. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association [J]. Circulation, 2011, 123(8):933-944

[5] Sadik A. Khuder, Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis [J]. Lung Cancer, 2001, 31(2–3):139–148

[6] Gadek JE, Fells GA, Crystal RG, Cigarette smoking induces functional antiprotease deficiency in the lower respiratory tract of humans [J]. Science, 1979, 206(4424):1315-1316

[7] Sopori M. Effects of cigarette smoke on the immune system [J]. Nature Reviews Immunology, 2002, 2(5): 372-377

[8] Dochi M, Sakata K, Oishi M, et al. Smoking as an independent risk factor for hypertension: a 14-year longitudinal study in male Japanese workers [J]. The Tohoku journal of experimental medicine, 2009, 217(1): 37-43

[9] Huxley R R, Woodward M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies [J]. The Lancet, 2011, 378(9799): 1297-1305

[10] Virdis A, Giannarelli C, Fritsch Neves M, et al. Cigarette smoking and hypertension [J]. Current pharmaceutical design, 2010, 16(23): 2518-2525

[11] Shah N R, Bracken M B. A systematic review and meta-analysis of prospective studies on the association between maternal cigarette smoking and preterm delivery [J]. American journal of obstetrics and gynecology, 2000, 182(2): 465-472

[12] Silman A J, Newman J, Macgregor A J. Cigarette smoking increases the risk of rheumatoid arthritis: results from a nationwide study of disease-discordant twins [J]. Arthritis & Rheumatism, 1996, 39(5): 732-735

[13] Vineis P, Airoldi L, Veglia F, et al. Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study[J]. Bmj, 2005, 330(7486): 277

[14] Coultas D B. Passive smoking and risk of adult asthma and COPD: an update [J]. Thorax, 1998, 53(5): 381-387

[15] Kaur S, Cohen A, Dolor R, et al. The impact of environmental tobacco smoke on women's risk of dying from heart disease: a meta-analysis [J]. Journal of Women's Health, 2004, 13(8): 888-897

[16] Windham G C, Hopkins B, Fenster L, et al. Prenatal active or passive tobacco smoke exposure and the risk of preterm delivery or low birth weight [J]. Epidemiology, 2000, 11(4): 427-433

[17] Barnoya J, Glantz S A. Cardiovascular effects of secondhand smoke nearly as large as smoking [J]. Circulation, 2005, 111(20): 2684-2698

[18] King B, Dube S, Kaufmann R, et al. Vital Signs: Current Cigarette Smoking Among Adults Aged greater than or equal to 18 Years- United States, 2005-2010 [J]. Morbidity and Mortality Weekly Report, 2011, 60(35)

[19] Centers for Disease Control and Prevention. Smoking-Attributable Mortality, Years of Potential Life Lost, and Productivity Losses—United States, 2000–2004. Morbidity and Mortality Weekly Report, 2008, 57(45):1226–1228

[20] Franceschini N, Rose K M, Storti K L, et al. Social-and Behavioral-Specific Genetic Effects on Blood Pressure Traits CLINICAL PERSPECTIVE The Strong Heart Family Study[J]. Circulation: Cardiovascular Genetics, 2009, 2(4): 396-401

[21] Franklin S S, Gustin W, Wong N D, et al. Hemodynamic patterns of age-related changes in blood pressure The Framingham Heart Study [J]. Circulation, 1997, 96(1): 308-315

[22] Martins D, Nelson K, Pan D, et al. The effect of gender on age-related blood pressure changes and the prevalence of isolated systolic hypertension among older adults: data from NHANES III [J]. The journal of gender-specific medicine: JGSM: the official journal of the Partnership for Women's Health at Columbia, 2001, 4(3): 10

[23] Carson A P, Howard G, Burke G L, et al. Ethnic Differences in Hypertension Incidence Among Middle-Aged and Older Adults The Multi-Ethnic Study of Atherosclerosis [J]. Hypertension, 2011, 57(6): 1101-1107

[24] Bowman T S, Gaziano J M, Buring J E, et al. A prospective study of cigarette smoking and risk of incident hypertension in women [J]. Journal of the American College of Cardiology, 2007, 50(21): 2085-2092

[25] Cupul-Uicab L A, Skjaerven R, Haug K, et al. In utero exposure to maternal tobacco smoke and subsequent obesity, hypertension, and gestational diabetes among women in the MoBa Cohort [J]. Environmental health perspectives, 2012, 120(3): 355-360

[26] Nyamdorj R. BMI compared with central obesity indicators in relation to diabetes and hypertension in Asians [J]. Obesity, 2008, 16(7): 1622-1635

[27] Taylor B, Irving H M, Baliunas D, et al. Alcohol and hypertension: gender differences in dose–response relationships determined through systematic review and meta-analysis [J]. Addiction, 2009, 104(12): 1981-1990

[28] Wiernik E, Pannier B, Czernichow S, et al. Occupational Status Moderates the Association Between Current Perceived Stress and High Blood Pressure [J]. 2013, 61:571-577

[29] Carroll D, Ring C, Hunt K, et al. Blood pressure reactions to stress and the prediction of future blood pressure: effects of sex, age, and socioeconomic position [J]. Psychosomatic Medicine, 2003, 65(6): 1058-1064

[30] Patten S B, Williams J V A, Lavorato D H, et al. Major depression as a risk factor for high blood pressure: epidemiologic evidence from a national longitudinal study [J]. Psychosomatic medicine, 2009, 71(3): 273-279

[31] Human Genome Epidemiology: Building the evidence for using genetic information to improve health and prevent disease: Building the evidence for using genetic information to improve health and prevent disease [M]. Oxford University Press, USA, 2010

[32] Khoury M J, Beaty T H, Cohen B H. Fundamentals of genetic epidemiology [M]. Oxford University Press, 1993

[33] Perera F P. Molecular epidemiology: principles and practice [M]. Academic Press, 1993.

[34] Brownson R C, Petitti D B. Applied epidemiology: theory to practice [M]. Oxford University Press on Demand, 1998

[35] Pearson T A, Manolio T A. How to interpret a genome-wide association study [J]. JAMA: the journal of the American Medical Association, 2008, 299(11): 1335-1344

[36] Thomas D. Gene–environment-wide association studies: emerging approaches [J]. Nature Reviews Genetics, 2010, 11(4): 259-272

[37] Thomas D. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies [J]. Annual review of public health, 2010, 31: 21–36

[38] Fox E R, Young J H, Li Y, et al. Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study [J]. Human molecular genetics, 2011, 20(11): 2273-2284

[39] Newton-Cheh C, Johnson T, Gateva V, et al. Eight blood pressure loci identified by genome-wide association study of 34,433 people of European ancestry [J]. Nature genetics, 2009, 41(6): 666

[40] Guo Y, Tomlinson B, Chu T, et al. A genome-wide linkage and association scan reveals novel loci for hypertension and blood pressure traits [J]. PloS one, 2012, 7(2): e31489

[41] Levy D, Ehret G B, Rice K, et al. Genome-wide association study of blood pressure and hypertension [J]. Nature genetics, 2009, 41(6): 677-687

[42] Adeyemo A, Gerry N, Chen G, et al. A genome-wide association study of hypertension and blood pressure in African Americans [J]. PLoS genetics, 2009, 5(7): e1000564

[43] Padmanabhan S, Melander O, Johnson T, et al. Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension [J]. PLoS genetics, 2010, 6(10): e1001177

[44] Salvi E, Kutalik Z, Glorioso N, et al. Genome-wide association study using a high-density SNP-array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of eNOS [J]. Hypertension, 2012, 59(2): 248-255

[45] Grove M L, Morrison A, Folsom A R, et al. Gene–environment interaction and the GNB3 gene in the Atherosclerosis Risk in Communities study [J]. International journal of obesity, 2007, 31(6): 919-926

[46] El Shamieh S, Herbeth B, Azimi-Nezhad M, et al. Human formyl peptide receptor 1 C32T SNP interacts with age and is associated with blood pressure levels [J]. Clinica Chimica Acta, 2012, 413(1): 34-38

[47] Taylor J Y, Sun Y V, Hunt S C, et al. Gene-environment interaction for hypertension among African American women across generations [J]. Biological research for nursing, 2010, 12(2): 149-155

[48] Chang Y C, Chiu Y F, Lee I T, et al. Common ALDH2 genetic variants predict development of hypertension in the SAPPHIRe prospective cohort: Gene-environmental interaction with alcohol consumption [J]. BMC Cardiovascular Disorders, 2012, 12(1): 58

[49] Yin R X, Wu D F, Wu J Z, et al. Interactions of several lipid-related gene polymorphisms and cigarette smoking on blood pressure levels [J]. International journal of biological sciences, 2012, 8(5): 685-696

[50] Karlson E W, Chang S C, Cui J, et al. Gene–environment interaction between HLA-DRB1 shared epitope and heavy cigarette smoking in predicting incident rheumatoid arthritis [J]. Annals of the rheumatic diseases, 2010, 69(01): 54-60

[51] Helbig K L, Nothnagel M, Hampe J, et al. A case-only study of gene-environment interaction between genetic susceptibility variants in NOD2 and cigarette smoking in Crohn's disease aetiology [J]. BMC medical genetics, 2012, 13(1): 14

[52] Meyers K, Chu J, Mosley T, et al. SNP-SNP interactions dominate the genetic architecture of candidate genes associated with left ventricular mass in african-americans of the GENOA study [J]. BMC medical genetics, 2010, 11(1): 160.

[53] Tobin M.D., Sheehan N.A., Scurrah K.J. et.al. Adjusting for treatment effects in studies of quantitative trains: antihypertensive therapy and systolic blood pressure, Stat.Med., 2005, 24(19): 2911-2935

[54] Kidambi S, Ghosh S, Kotchen J M, et al. Non-replication study of a genome-wide association study for hypertension and blood pressure in African Americans [J]. BMC medical genetics, 2012, 13(1): 27

[55] Guo Y, Tomlinson B, Chu T, et al. A genome-wide linkage and association scan reveals novel loci for hypertension and blood pressure traits [J]. PloS one, 2012, 7(2): e31489

[56] Yang J, Weedon M N, Purcell S, et al. Genomic inflation factors under polygenic inheritance [J]. European Journal of Human Genetics, 2011, 19(7): 807-812

[57] Devlin B, Roeder K. Genomic control for association studies [J]. Biometrics, 1999, 55(4): 997-1004

[58] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing [J]. Journal of the Royal Statistical Society. Series B (Methodological), 1995: 289-300

[59] Vasco V R L, Cardinale G, Polonia P. Deletion of PLCB1 gene in schizophrenia-affected patients [J]. Journal of cellular and molecular medicine, 2012, 16(4): 844-851

[60] Vasco V R L, Longo L, Polonia P. Phosphoinositide-specific Phospholipase C β1 gene deletion in bipolar disorder affected patient [J]. Journal of cell communication and signaling, 2013, 7(1): 25–29

[61] Fritz D, Stefanovic B. RNA-binding protein RBMS3 is expressed in activated hepatic stellate cells and liver fibrosis and increases expression of transcription factor Prx1 [J]. Journal of molecular biology, 2007, 371(3): 585-595

[62] Berchtold L A, Størling Z M, Ortis F, et al. Huntingtin-interacting protein 14 is a type 1 diabetes candidate protein regulating insulin secretion and β-cell apoptosis [J]. Proceedings of the National Academy of Sciences, 2011, 108(37): E681-E688

[63] Stephens J W, Humphries S E. The molecular genetics of cardiovascular disease: clinical implications [J]. Journal of internal medicine, 2003, 253(2): 120-127

[64] Scheidweiler K B, Marrone G F, Shakleya D M, et al. Oral fluid nicotine markers to assess smoking status and recency of use [J]. Therapeutic drug monitoring, 2011, 33(5): 609-618

TABLES

Table 1. Study population Characteristics

Continuous variables		**Mean (std)**
	SBP (mm Hg)	133.06 (17.04)
	DBP (mm Hg)	78.50 (9.53)
	Adjusted SBP (mm Hg)*	139.52 (18.65)
	Adjusted DBP (mm Hg)**	81.73 (9.92)
	Age (years)	55.33 (10.87)
	BMI (kg/m^2)	30.42 (6.32)
	Waist-hip ratio	0.91 (0.09)
	Cholesterol level (mg/dl)	209.88 (38.32)
	High density lipoprotein (mg/dl)	51.45 (16.17)
	Triglycerides (mg/dl)	191.11 (105.23)
Discrete variables (N=1371)		**N (%)**
Smoke		
	Current smokers	196 (14.30)
	Non-current smokers	1175 (85.70)
Hypertensive***		
	Yes	996 (72.65)
	No	375 (27.35)
Gender		
	Male	615 (44.86)
	Female	756 (55.14)
Hypertension treatment		
	Taking meds	885 (64.55)
	Not taking meds	486 (35.45)
Alcohol		
	Sometimes	985 (71.85)
	Never	386 (28.15)
Parents Hypertension history		
	Yes	1091 (79.58)
	No	280 (20.42)

Notes: * Adjusted SBP is the variable to describe average systolic blood pressure adjusted for hypertension medication, which was added 10 mm Hg to the original SBP; ** Adjusted DBP is the variable to describe average diastolic blood pressure adjusted for hypertension medication, which was added 5 mm Hg to the original DBP; *** Hypertensive represents whether or not having hypertension (SBP≥140 mm Hg and/or DBP≥90 mm Hg)

Table 2. Statistical tests for influence factors between hypertensive and non-hypertensive group

Continuous variables				
Age (years)[1]	49.20 (10.44)	57.64 (10.11)	-13.66	<0.0001*
BMI (kg/m^2)[2]	28.45 (5.53)	31.16 (6.44)	-7.72	<0.0001*
Waist-Hip ratio[1]	0.88 (0.09)	0.92 (0.09)	-7.33	<0.0001*
Cholesterol level (mg/dl)[1]	206.60 (38.18)	211.10 (38.31)	-1.96	0.0500*
High density lipoprotein (mg/dl)[1]	52.82 (15.58)	50.94 (16.36)	1.92	0.0549
Triglyceride (mg/dl)[2]	161.80 (87.57)	202.10 (109.20)	-7.06	<0.0001*
Discrete variables				
Current Smoking status[3]				
Yes	86 (22.93)	110 (11.04)	31.43	<0.0001*
No	289 (77.07)	886 (88.96)		
Gender[3]				
Male	166 (44.27)	449 (45.08)	0.07	0.7871
Female	209 (55.73)	547 (54.92)		
Hypertension treatment[3]				
Taking meds	22 (5.87)	863 (86.65)	776.88	<0.0001*
Not taking meds	353 (94.13)	133 (13.35)		
Parents Hypertension history[3]				
Yes	285 (76.00)	806 (80.92)	4.06	0.0438*
No	90 (24.00)	190 (19.08)		
Alcohol[3]				
Sometimes	293 (78.13)	692 (69.48)	10.09	0.0015*
Never	82 (21.87)	304 (30.52)		

Note: Mean (std) is used to represent continuous variables; N(%) is used to describe discrete variables; 1. Pooled T-Test was used to test statistical significance for age, cholesterol level and high density lipoprotein between hypertensive and non-hypertensive; 2. Satterthwaite T-test was used to test statistical significance for BMI, Hip-waist ratio and Triglycerides between hypertensive and non-hypertensive; 3. Chi-square test was conducted to test statistical significance for smoking status, gender, hypertension treatment, alcohol use and parents hypertension history between hypertensive and non-hypertensive; * Statistical significance (alpha=0.05);

Table 3. Result of genotypic SNP-DBP association no interaction model (model 3)

SNP	LR	LR P	GC adjusted P	GC adjusted FDR Q
rs4210869*	46.04	1.00×10^{-10}	1.11×10^{-08}	0.0075
rs8676976*	40.58	1.54×10^{-09}	9.80×10^{-08}	0.0266
rs4250742*	40.08	1.98×10^{-09}	1.19×10^{-07}	0.0266
rs8367087*	38.14	5.23×10^{-09}	2.59×10^{-07}	0.0320
rs2089892*	37.92	5.84×10^{-09}	2.82×10^{-07}	0.0320
rs2184255*	37.88	5.96×10^{-09}	2.87×10^{-07}	0.0320
rs8512007*	37.06	8.96×10^{-09}	3.97×10^{-07}	0.0361
rs2116392*	36.84	9.99×10^{-09}	4.33×10^{-07}	0.0361
rs8629316	35.38	2.08×10^{-08}	7.74×10^{-07}	0.0575
rs8572093	32.43	9.08×10^{-08}	2.51×10^{-06}	0.1614

Note: * denote significant SNPs GC adjusted FDR Q value is lower than 0.05.

Table 4 Top 10 significant SNP-Smoke interaction that associate with SBP form additive interaction model (model2)

SNP	SNP_E[1]	Smk_E[2]	G×E_E[3]	G×E_GC_P	G×E_GC_FDR_Q
rs2212885*	1.54	78.72	-27.77	1.0×10^{-08}	0.0067
rs8526790*	0.13	51.93	-28.19	2.9×10^{-08}	0.0098
rs4230236*	2.74	24.68	-29.08	1.9×10^{-07}	0.0423
rs8694580*	1.53	49.45	-18.17	3.3×10^{-07}	0.0423
rs2083807*	-0.97	-3.92	28.85	4.3×10^{-07}	0.0423
rs2213442*	-0.35	166.63	-56.59	4.4×10^{-07}	0.0423
rs8596677*	-0.35	166.63	-56.59	4.4×10^{-07}	0.0423
rs8657661	0.93	52.14	-27.92	2.3×10^{-06}	0.1748
rs2294553	-1.82	-5.01	16.55	2.6×10^{-06}	0.1748
rs8435490	-2.51	-19.98	15.07	2.6×10^{-06}	0.1748

Note: * indicate significant result which FDR Q value after genomic control is lower than 0.05; 1. estimated beta value for SNP; 2. estimated beta value for smoking status; 3. estimated beta value for SNP-smoke interaction

Table 5 Top 10 significant SNP-Smoke interaction that associate with SBP form genotypic interaction model (model4)

SNP	LR statistic	Model_P[1]	GC_P[2]	GC_FDR_Q[3]
rs4269045*	33.93	2.35×10^{-11}	1.35×10^{-08}	0.0091
rs8502813*	33.75	1.89×10^{-10}	6.35×10^{-08}	0.02019
rs937307*	33.75	3.06×10^{-10}	9.06×10^{-08}	0.02019
rs2219771	33.39	2.22×10^{-09}	3.93×10^{-07}	0.0569
rs1906984	33.18	2.48×10^{-09}	4.26×10^{-07}	0.0569
rs2236815	32.86	1.06×10^{-08}	1.25×10^{-06}	0.1368
rs1816847	32.71	1.60×10^{-08}	1.70×10^{-06}	0.1368
rs8389683	32.71	1.70×10^{-08}	1.78×10^{-06}	0.1368
rs1871152	32.46	1.79×10^{-08}	1.84×10^{-06}	0.1368
rs2286468	32.20	2.15×10-08	2.11×10^{-06}	0.1408

Note: * indicate significant result which FDR Q value after genomic control is lower than 0.05; 1. p value obtained directed from the model; 2. GC P value obtained from GC adjusted chi-square statistics, which under chi-square distribution with 2 df; 3. FDR corrected GC adjusted Q values

FIGURES AND FIGURE LEGENDS

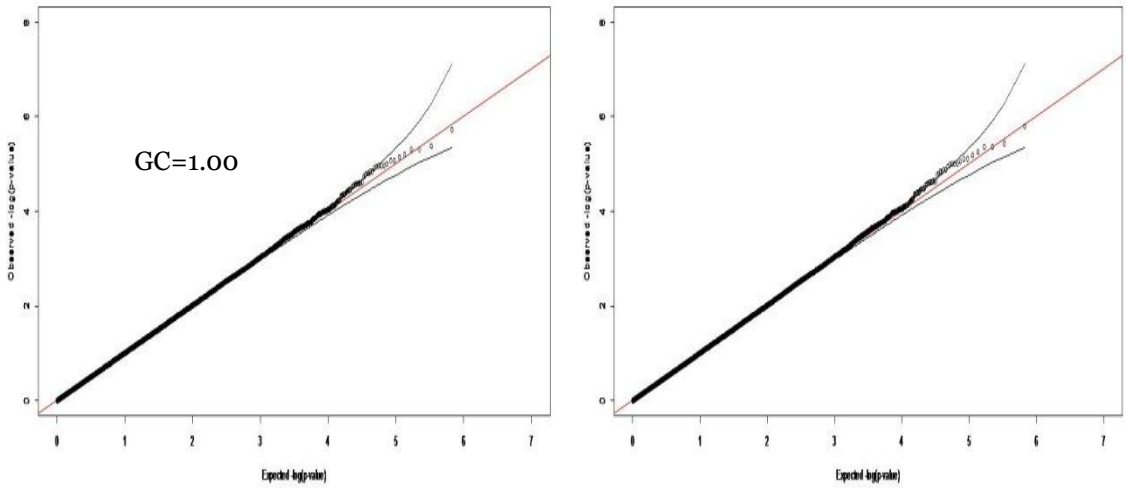

Figure 1. Q-Q plot for Additive No-Interaction Mode (model 1)

The left graph is Q-Q Plot using crude p-value; The Right graph is Q-Q plot after genomic control;

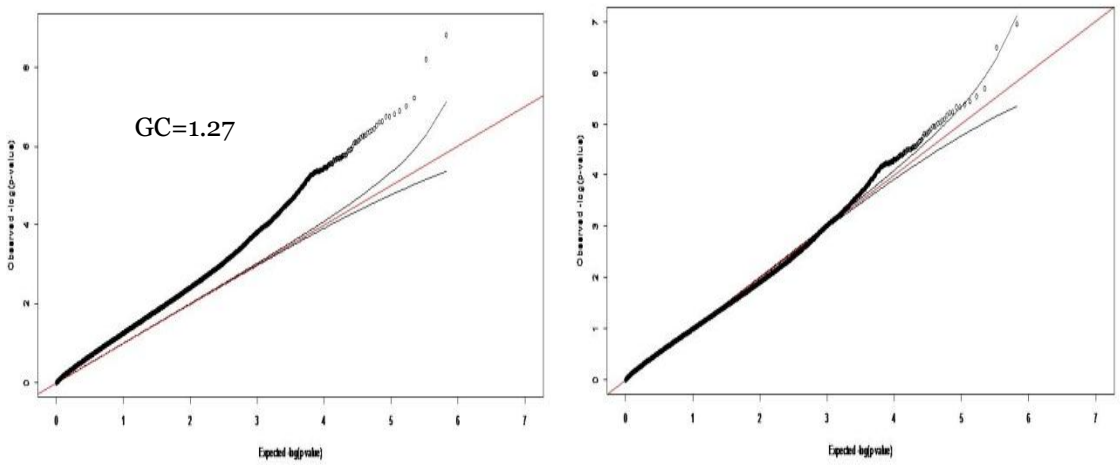

Figure 2. Q-Q Plot for Genotypic No-Interaction Model (model 1)

The left graph is Q-Q Plot using crude p-value; The Right graph is Q-Q plot after genomic control;

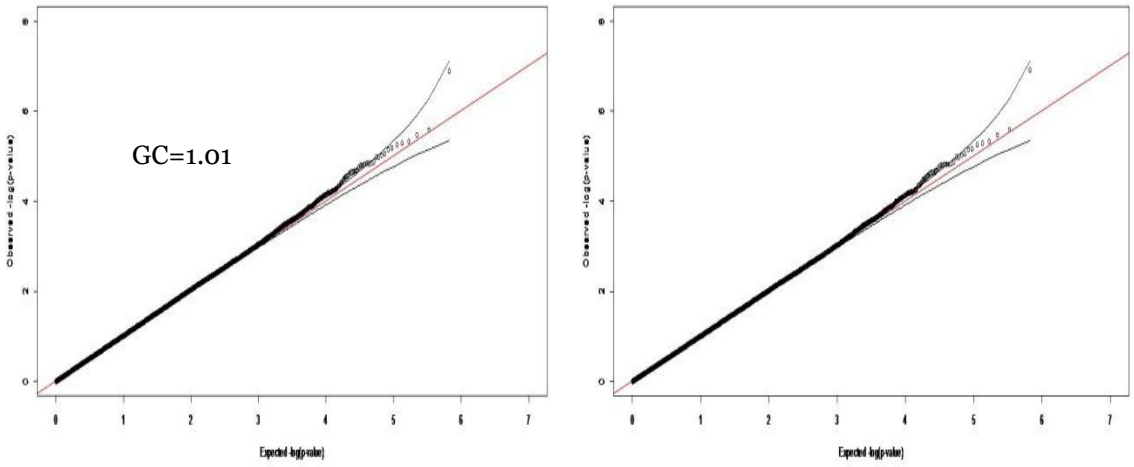

Figure 3. Q-Q plot for Additive No-Interaction Model (model 2)

The left graph is Q-Q Plot using crude p-value; The Right graph is Q-Q plot after genomic control;

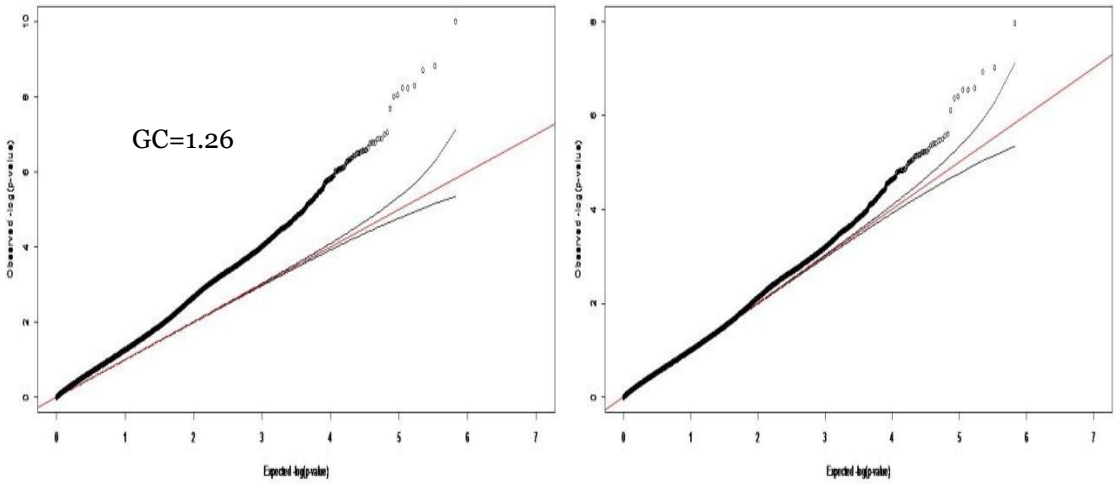

Figure 4. Q-Q plot for Genotypic No-Interaction Model (model 2)

The left graph is Q-Q Plot using crude p-value; The Right graph is Q-Q plot after genomic control;

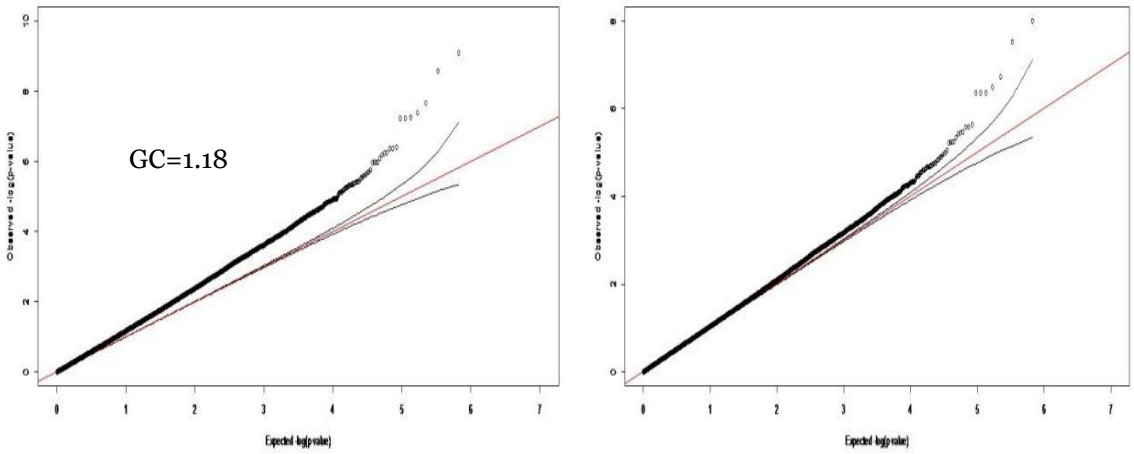

Figure 5. Q-Q plot for Additive Interaction Model (model 3)

The left graph is Q-Q Plot using crude p-value; The Right graph is Q-Q plot after genomic control;

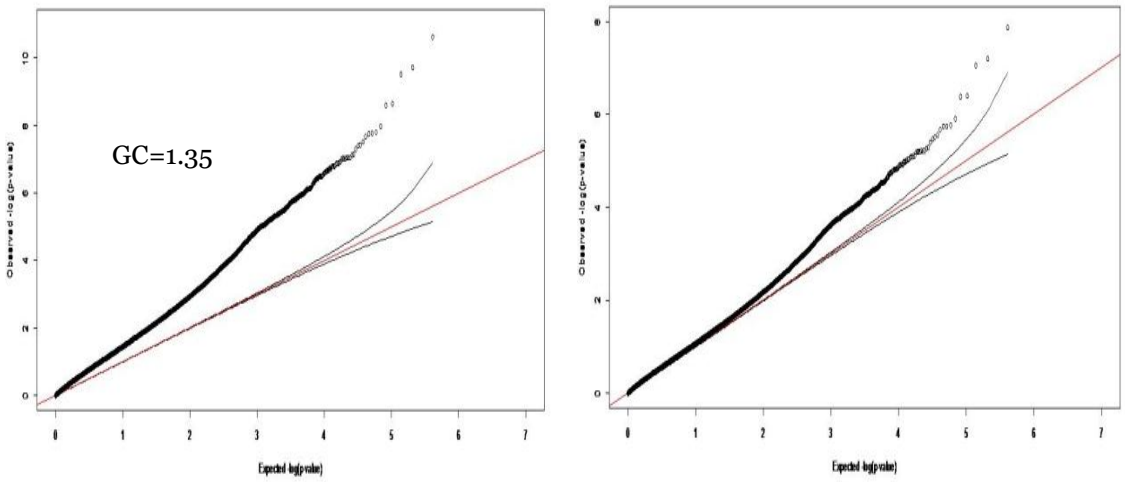

Figure 6. Q-Q plot for Genotypic Interaction Model (model 3)

The left graph is Q-Q Plot using crude p-value; The Right graph is Q-Q plot after genomic control;

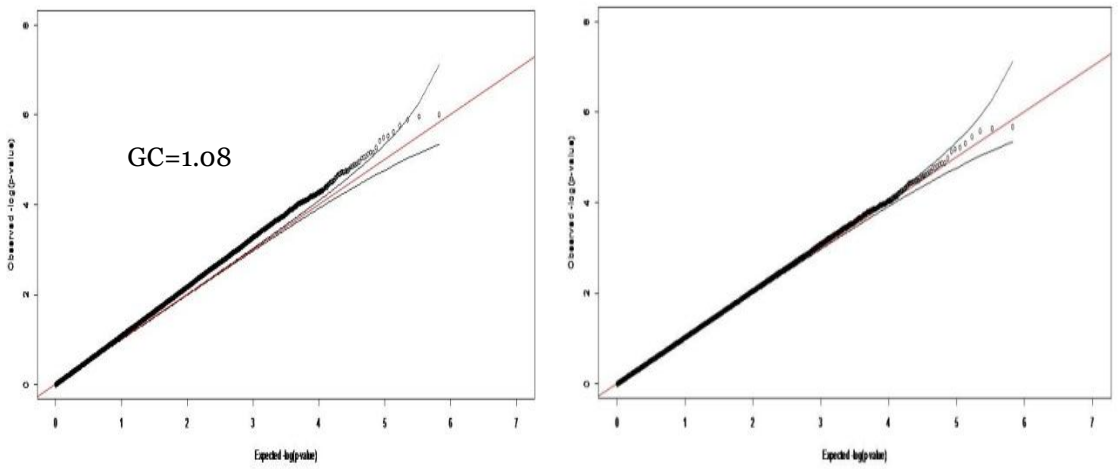

Figure 7. Q-Q plot for Additive Interaction Model (model 4)

The left graph is Q-Q Plot using crude p-value; The Right one is Q-Q plot after genomic control;

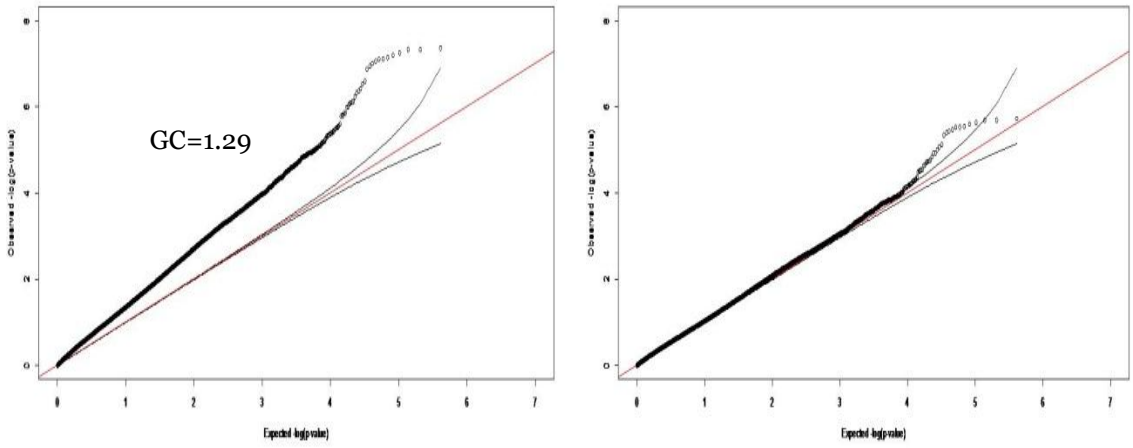

Figure 8. Q-Q plot for Genotypic Interaction Model (model 4)

The left graph is Q-Q Plot using crude p-value; The Right graph is Q-Q plot after genomic control;

APPENDICES

Table 1 Result of association between main effect (SNP) and SBP using Additive Main effect model (model 1)

SNP	SNP_E	smk_E	SNP_SE	smk_SE	SNP_t	smk_t	SNP_p	smk_p	snp_GC_adjP.fdr
SNP_A_8497104	3.6	-2.5	0.7	1.4	4.8	-1.8	1.85E-06	6.71E-02	0.53
SNP_A_1855589	3.2	-2.2	0.7	1.4	4.6	-1.6	4.16E-06	1.04E-01	0.53
SNP_A_1783167	-6.4	-2.0	1.4	1.4	-4.6	-1.4	4.91E-06	1.54E-01	0.53
SNP_A_2134890	4.9	-2.4	1.1	1.4	4.6	-1.7	4.94E-06	8.46E-02	0.53
SNP_A_2251583	-3.2	-2.5	0.7	1.4	-4.5	-1.8	6.33E-06	6.63E-02	0.53
SNP_A_1799276	4.5	-2.1	1.0	1.4	4.5	-1.6	7.31E-06	1.18E-01	0.53
SNP_A_8673840	-3.2	-2.5	0.7	1.4	-4.5	-1.8	8.54E-06	7.30E-02	0.53
SNP_A_8571606	-6.1	-2.7	1.4	1.4	-4.5	-2.0	8.58E-06	4.86E-02	0.53
SNP_A_2031312	-3.6	-2.5	0.8	1.4	-4.4	-1.8	9.91E-06	6.78E-02	0.53
SNP_A_8400743	2.9	-2.2	0.7	1.4	4.4	-1.6	1.09E-05	1.09E-01	0.53

Table 2 Result of association between main effect (SNP) and DBP using Additive Main effect model (model 1)

SNP	SNP_E	smk_E	SNP_SE	smk_SE	SNP_t	smk_t	SNP_p	smk_p	snp_GC_P.fdr
SNP_A_2181608	-3.78	-2.44	0.71	0.91	-5.33	-2.69	1.29E-07	7.19E-03	0.08
SNP_A_2090418	2.79	-2.46	0.59	0.91	4.73	-2.71	2.67E-06	6.92E-03	0.54
SNP_A_2030887	2.72	-2.49	0.58	0.91	4.68	-2.74	3.37E-06	6.28E-03	0.54
SNP_A_8391053	2.43	-2.33	0.53	0.91	4.61	-2.57	4.72E-06	1.03E-02	0.54
SNP_A_1905740	-2.08	-2.42	0.45	0.91	-4.59	-2.67	5.10E-06	7.73E-03	0.54
SNP_A_8661178	3.42	-2.58	0.75	0.91	4.57	-2.84	5.64E-06	4.65E-03	0.54
SNP_A_8557003	2.05	-2.40	0.45	0.91	4.53	-2.64	6.70E-06	8.37E-03	0.54
SNP_A_2108859	-6.39	-2.53	1.41	0.91	-4.52	-2.78	7.05E-06	5.52E-03	0.54
SNP_A_1935698	-3.72	-2.56	0.83	0.91	-4.47	-2.82	8.80E-06	4.92E-03	0.54
SNP_A_8635724	1.98	-2.39	0.44	0.91	4.46	-2.64	9.50E-06	8.55E-03	0.54

Table 3 Result of association between SNP-smoking interaction and SBP using Additive interaction model (model 2)

SNP	SNP_E	smk_E	snpxSmk_E	SNP_p	snpxSmk_p	SNPxSmk_GC_P.fdr
SNP_A_2212885	1.54	78.72	-27.77	3.25E-01	8.30E-10	0.01
SNP_A_8526790	0.13	51.93	-28.19	9.46E-01	2.77E-09	0.01
SNP_A_4230236	2.74	24.68	-29.08	1.58E-01	2.25E-08	0.04
SNP_A_8694580	1.53	49.45	-18.17	2.50E-01	4.18E-08	0.04
SNP_A_2083807	-0.97	-3.92	28.85	6.21E-01	5.66E-08	0.04
SNP_A_2213442	-0.35	166.63	-56.59	8.93E-01	5.83E-08	0.04
SNP_A_8596677	-0.35	166.63	-56.59	8.93E-01	5.83E-08	0.04
SNP_A_8657661	0.93	52.14	-27.92	6.37E-01	3.84E-07	0.17
SNP_A_2294553	-1.82	-5.01	16.55	1.98E-01	4.31E-07	0.17
SNP_A_8435490	-2.51	-19.98	15.07	2.36E-02	4.34E-07	0.17

Table 4 Result of association between SNP-smoking interaction and DBP using Additive interaction model (model 2)

SNP	SNP_E	smk_E	snpxSmk_E	SNP_p	snpxSmk_p	SNPxSmk_GC_P.fdr
SNP_A_2035880	2.12	8.66	-8.59	2.60E-03	1.00E-06	0.58
SNP_A_8569688	0.89	9.29	-11.15	2.97E-01	1.09E-06	0.58
SNP_A_2234877	-2.03	-16.97	8.51	3.94E-03	1.28E-06	0.58
SNP_A_8486216	-0.78	-23.13	7.76	1.83E-01	1.70E-06	0.58
SNP_A_8711920	-0.66	24.60	-26.43	7.37E-01	2.43E-06	0.63
SNP_A_8571519	0.88	7.58	-5.55	6.02E-02	3.03E-06	0.63
SNP_A_8573449	0.61	8.58	-6.02	2.21E-01	3.29E-06	0.63
SNP_A_8620912	-1.15	-13.61	5.60	2.04E-02	3.81E-06	0.63
SNP_A_8322538	0.76	10.70	-6.00	1.13E-01	5.28E-06	0.76
SNP_A_8509818	-0.84	-16.99	6.00	9.78E-02	6.89E-06	0.76

Table 5 Result of association between main effect (SNP) and SBP using genotypic Main effect model (model 3)

SNP	snp_LR	snp_LR_P	SNP_GC_LR	SNP_GC_LR_P	snp_LR_P.fdr	SNP_GC_LR_P.fdr
SNP_A_8458510	40.54	1.57E-09	32.04	1.10E-07	0.00	0.07
SNP_A_4252068	37.78	6.25E-09	29.86	3.29E-07	0.00	0.11
SNP_A_8612116	33.19	6.20E-08	26.23	2.01E-06	0.01	0.39
SNP_A_8574586	32.29	9.71E-08	25.52	2.87E-06	0.02	0.39
SNP_A_2282891	31.73	1.29E-07	25.07	3.59E-06	0.02	0.39
SNP_A_8431619	31.33	1.57E-07	24.76	4.21E-06	0.02	0.39
SNP_A_1977443	31.09	1.77E-07	24.57	4.62E-06	0.02	0.39
SNP_A_8288089	31.04	1.81E-07	24.53	4.71E-06	0.02	0.39
SNP_A_4206714	30.48	2.41E-07	24.09	5.88E-06	0.02	0.41
SNP_A_1857848	30.39	2.52E-07	24.02	6.09E-06	0.02	0.41

Table 6 Result of association between main effect (SNP) and DBP using genotypic Main effect model (model 3)

SNP	snp_LR	snp_LR_P	SNP_GC_LR	SNP_GC_LR_P	snp_LR_P.fdr	SNP_GC_LR_P.fdr
SNP_A_4210869	46.04	1.00E-10	36.62	1.11E-08	0.00	0.01
SNP_A_8676976	40.58	1.54E-09	32.28	9.80E-08	0.00	0.03
SNP_A_4250742	40.08	1.98E-09	31.88	1.19E-07	0.00	0.03
SNP_A_8367087	38.14	5.23E-09	30.33	2.59E-07	0.00	0.03
SNP_A_2089892	37.92	5.84E-09	30.16	2.82E-07	0.00	0.03
SNP_A_2184255	37.88	5.96E-09	30.13	2.87E-07	0.00	0.03
SNP_A_8512007	37.06	8.96E-09	29.48	3.97E-07	0.00	0.04
SNP_A_2116392	36.84	9.99E-09	29.31	4.33E-07	0.00	0.04
SNP_A_8629316	35.38	2.08E-08	28.14	7.74E-07	0.00	0.06
SNP_A_8572093	32.43	9.08E-08	25.79	2.51E-06	0.01	0.16

Table 7 Result of association between SNP-smoking interaction and SBP using genotypic Main effect model (model 4)

SNP	snp_Csmk_LR	snp_Csmk_LR_P	SNPxCsmk_GC_LR	SNPxCsmk_GC_LR_P	snp_Csmk_LR_P.fdr	SNPxCsmk_GC_LR_P.fdr
SNP_A_4269045	48.95	2.35E-11	36.24	1.35E-08	0.00	0.01
SNP_A_8502813	44.78	1.89E-10	33.15	6.35E-08	0.00	0.02
SNP_A_1937307	43.81	3.06E-10	32.43	9.06E-08	0.00	0.02
SNP_A_2219771	39.85	2.22E-09	29.50	3.93E-07	0.00	0.06
SNP_A_1906984	39.63	2.48E-09	29.34	4.26E-07	0.00	0.06
SNP_A_2236815	36.73	1.06E-08	27.19	1.25E-06	0.00	0.14
SNP_A_1816847	35.90	1.60E-08	26.57	1.70E-06	0.00	0.14
SNP_A_8389683	35.78	1.70E-08	26.48	1.78E-06	0.00	0.14
SNP_A_1871152	35.68	1.79E-08	26.41	1.84E-06	0.00	0.14
SNP_A_2286468	35.31	2.15E-08	26.14	2.11E-06	0.00	0.14

Table 8 Result of association between SNP-smoking interaction and DBP using genotypic Main effect model (model 4)

SNP	snp_Csmk_LR	snp_Csmk_LR_P	SNPxCsmk_GC_LR	SNPxCsmk_GC_LR_P	snp_Csmk_LR_P.fdr	SNPxCsmk_GC_LR_P.fdr
SNP_A_8336088	33.93	4.28E-08	26.36	1.89E-06	0.01	0.24
SNP_A_2010915	33.75	4.68E-08	26.22	2.03E-06	0.01	0.24
SNP_A_1834070	33.75	4.68E-08	26.22	2.03E-06	0.01	0.24
SNP_A_8646546	33.39	5.63E-08	25.93	2.34E-06	0.01	0.24
SNP_A_2077495	33.18	6.23E-08	25.77	2.53E-06	0.01	0.24
SNP_A_1889494	32.86	7.32E-08	25.52	2.87E-06	0.01	0.24
SNP_A_1934887	32.71	7.89E-08	25.41	3.04E-06	0.01	0.24
SNP_A_4300438	32.71	7.89E-08	25.41	3.04E-06	0.01	0.24
SNP_A_4249329	32.46	8.95E-08	25.21	3.35E-06	0.01	0.24
SNP_A_1938984	32.21	1.01E-07	25.02	3.70E-06	0.01	0.24