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:

_____ _____

Chang Liu Date

Effects of Smoking-related DNA Methylation on Coronary Flow Reserve

By

Chang Liu

Master of Public Health

Epidemiology

Yan Sun, PhD

Committee Chair

Effects of Smoking-related DNA Methylation on Coronary Flow Reserve

By

Chang Liu

Master of Medicine
Liaoning University of Traditional Chinese Medicine
2015

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
2017

ABSTRACT

Effects of Smoking-related DNA Methylation on Coronary Flow Reserve

By Chang Liu

Background: Smoking is a major risk factor for cardiovascular disease, which is a leading cause of death in the United States. Smoking has been shown to affect human health through changes of epigenetics such as DNA methylation. Previous publications suggested that smoking-related DNA methylation might mediate the effect of smoking on cardiovascular function, and the smoking-related DNA methylation markers could potentially improve the prediction of cardiovascular diseases.

Methods: Genome-wide DNA methylation data of peripheral blood (218 male twins) were utilized to investigate the associations between smoking, DNA methylation and coronary flow reserve (CFR), which is a parameter that reflects the coronary vasodilator capacity by showing the status of pericardial coronary arteries and the microvascular coronary circulation. An epigenome-wide association analysis was conducted to identify the smoking-related DNA methylation markers, which were further investigated in a linear mixed effect model, adjusting for age, body mass index, cell types (granulocytes, monocytes, natural killer cells, B cells, CD4+ and CD8+ T cells) and treating Beadchip and pair as random effects. Then, smoking was additionally controlled in the model to study the role of DNA methylation as a mediator in the effect of smoking on cardiovascular function. Furthermore, twin-specific models were performed to assess within-twin and between-twin effects.

Results: We successfully replicated 50 smoking-related CpG sites, which were reported 3 or more times in previous publications. 15 of the 50 markers were significantly associated with CFR ($p < 0.05$) in linear mixed effect model, adjusting for age, body mass index, cell types (granulocytes, monocytes, natural killer cells, B cells, CD4+ and CD8+ T cells) and treating Beadchip and pair as random effects. After additional adjustment for smoking, cg21121843 in huntingtin (HTT) gene on Chromosome 4 remained associated with CFR ($p=0.046$).

Conclusions: Our findings suggest that smoking-related DNA methylation sites of peripheral blood cells are associated with CFR but may not directly mediate the effects of smoking on cardiovascular function.

Effects of Smoking-related DNA Methylation on Coronary Flow Reserve

By

Chang Liu

Master of Medicine
Liaoning University of Traditional Chinese Medicine
2015

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
2017

TABLE OF CONTENTS

BACKGROUND

Smoking has been one of the major public health concerns in the US. In 2015, 15.1% (36.5 million) of all adults aged 18 or older were current smokers. The prevalence among males (16.7%) was higher than that in females (13.6%). 21.9% of non-Hispanic American Indians/Alaska Natives and 7.0% of non-Hispanic Asians were current smokers, which denote the highest and lowest percentage among racial subgroups respectively (1). According to the data from Surgeon General's Report, through years of 1965-2014, 20,830,000 premature deaths occurred due to exposure to smoking including secondhand smoke (2).

Many studies have demonstrated that smoking is harmful to health through multiple pathways. Blakely et al. found evidence of active smoking being strongly associated with lung, larynx (including ear and nasosinus), and bladder cancers with RR estimates of 9.28 (95 % CI: 8.31-10.4), 6.14 (95 % CI: 4.55-8.30), and 2.22 (95 % CI: 1.94-2.55), respectively; cervical (RR: 1.82; 95 % CI: 1.51-2.20), kidney (RR: 1.29; 95 % CI: 1.07-1.56), esophageal (RR: 2.14; 95 % CI: 1.73-2.65), oropharyngeal (RR: 2.30; 95 % CI: 1.94-2.72), pancreatic (RR: 1.68; 95 % CI: 1.44-1.96), and stomach (RR: 1.42; 95 % CI: 1.22-1.66) cancers were found in moderate associations with active smoking (3). Cunningham et al. confirmed the previous findings of associations between smoking and COPD by estimating that among current smokers, the likelihood of having COPD was almost four times higher (PR: 3.9, 95% CI: 3.7, 4.1) compared to never smokers (4). Weitzman et al. found a cotinine-confirmed relationship between tobacco smoke and metabolic syndrome among adolescents, with environmental tobacco smoke exposure odds ratio of 4.7 (95% CI: 1.7-12.9) and active smoking odds ratio of 6.1 (95% CI: 2.8-

13.4) (5). A large cross-sectional study was conducted by Kurmi et al., estimating that the odds ratios of having airflow obstruction comparing regular smokers to never-smokers, were 1.42 (95% CI: 1.34-1.50) and 1.53 (95% CI: 1.43-1.65) among males and females respectively (6). In addition, passive smoking is also proved associated with cancers, cardiovascular diseases and lung damage (7-11).

In particular, smoking is a well-known major risk factor for coronary artery disease and sudden cardiac death (12). Based on a 16-year follow-up study, Shields et al. found that current daily smokers had a 60% higher risk of incident heart disease compared to those who never smoked daily during the follow-up (13). Previous studies have confirmed the effects of smoking on coronary flow reserve (CFR) (14-16). CFR is a parameter that reflects the coronary vasodilator capacity by showing the status of pericardial coronary arteries and the microvascular coronary circulation (17). An impaired CFR can be due to narrowed coronary arteries. In the absence of angiographically demonstrable atherosclerotic disease, impaired CFR can reflect coronary microcirculation dysfunction, which can be caused by structural or functional changes that involve neurohumoral factors or endothelial dysfunction. In addition, an abnormal CFR may also indicate changes in coronary or systemic hemodynamics as well as extravascular coronary resistance (18). Park et al., Tanaka et al. and Czernin et al. found that smoking reduces CFR (14-16), and Tanaka's group also found that the deteriorating effect on the coronary flow reserve by smoking can be corrected after cessation (19).

Furthermore, cigarette smoking has been shown to affect human health though changes of epigenetics such as DNA methylation (20). DNA methylation is a key mechanism in controlling gene expression involving the covalent transfer of a methyl group to the C-5

position of the cytosine ring of DNA by DNA methyltransferases (21). DNA methylation is crucial player in epigenetic silencing gene expression, therefore diseases will be induced when not regulated (22). Epigenome-wide association studies (EWAS) have been conducted multiple times, repeatedly verifying the association between smoking and DNA methylation changes, with many loci consistently identified across different populations (23-34). Based on literature review, twelve EWASs were conducted in recent three years (Table 1). In these studies, many strong signals were identified in Gene Aryl-Hydrocarbon Receptor Repressor (AHRR) on Chromosome 5 (23-34). Eleven of the twelve EWASs reported cg05575921 in AHRR gene as a highly significant CpG site, and Zhang et al. (23), Su et al. (24), Lee et al. (25), Ambatipudi et al. (26), Sayols-Baixeras et al. (27), Zaghlool et al. (28), Guida et al. (29), Besingi et al. (30), Elliott et al. (33) reported that it was the most significant smoking-related marker in the study. Cg21161138 in AHRR was also highly significant, and was reported among the top ten associations by Zhang et al. (23), Su et al. (24), Ambatipudi et al. (26), Sayols-Baixeras et al. (27), Zaghlool et al. (28), Guida et al. (29), Dogan et al. (32), and Tsaprouni et al. (34). Two markers in Alkaline Phosphatase, Placental Like 2 (ALPPL2), Chromosome 2 were revealed highly significant, with cg05951221 among the top ten hits by Zhang et al. (23), Su et al. (24), Lee et al. (25), Ambatipudi et al. (26), Sayols-Baixeras et al. (27), Guida et al. (29), Elliott et al. (33), Tsaprouni et al. (34) and cg2156642 reported by Zhang et al. (23), Su et al. (24), Ambatipudi et al. (26), Sayols-Baixeras et al. (27), Guida et al. (29), Besingi et al. (30), Elliott et al. (33), Tsaprouni et al. (34). In addition, cg03636183 in Coagulation Factor II (thrombin) Receptor-Like 3 (F2PL3) Gene, Chromosome 19 was identified as a marker among top ten associations by Zhang et al.

(23), Lee et al. (25), Ambatipudi et al. (26), Sayols-Baixeras et al. (27), Zaghlool et al. (28), Guida et al. (29), Besingi et al. (30), Harlid et al. (31), Elliott et al. (33), Tsaprouni et al. (34). Furthermore, Gao et al. (35) conducted a systematic review of DNA methylation studies in 2015, focusing on DNA methylation changes of whole blood cells in response to active smoking in adults. This systematic review identified 17 studies addressing the association of active smoking exposure with methylation modifications in blood DNA, including 14 epigenome-wide association studies (EWASs) and 3 gene-specific methylation studies (GSMSs) on the gene regions identified by EWASs. Overall, 1460 smoking-associated CpG sites were identified in the EWASs, of which 62 sites were detected in multiple (≥3) studies. The three most frequently reported CpG sites (genes) were cg05575921 (AHRR, Chr. 5), cg03636183 (F2RL3, Chr. 19), and cg19859270 (GPR15, Chr.3).

Reynolds et al. identified novel associations between cg05575921 (AHRR, Chr. 5) methylation and carotid plaque scores, which remained significant in current and former smokers even after adjusting for self-reported smoking habits, urinary cotinine, and well-known cardiovascular risk factors (36). Yixiang Li studied the associations between the smoking-related CpGs and atherosclerosis, and identified cg03991871 (AHRR, Chr. 5) and cg06126421 (6p21.33, Chr. 6) significant after adjusting for smoking status (37). These findings suggest that DNA methylation might mediate the effect of cigarette smoking on cardiovascular function, and we might use potential CpG sites to independently improve the prediction of cardiovascular function including CFR.

METHODS

Data source

The Vietnam Era Twin (VET) Registry is one of the largest twin registries in the US. The VET registry is composed of 7,369 middle-aged male-male twin pairs both of whom served in the military during the time of the Vietnam conflict (1965–1975). The registry was initially aiming at addressing long-term health effects of military service in Vietnam, then it has evolved into a data resource for genetic epidemiologic studies of mental and physical health conditions (38). Subjects in this study were drawn from the Emory Twin Study (ETS), which is a subcohort of Vietnam Era Twin (VET) Registry. The ETS consists of 307 middle-aged male Caucasian monozygotic and dizygotic twin pairs who were born between 1946 and 1956 (39, 40). All twins were examined in pairs at the Emory University General Clinical Research Center between 2002 and 2010. DNA samples of 218, including 164 monozygotic and 54 dizygotic twins were epityped and used for this study. The ETS was approved by the Emory Institutional Review Board, and informed consent were obtained from all subjects.

DNA methylation methods

DNA samples of 218 twins were epityped. Venous blood samples were drawn and stored at −80 °C before biomedical assay. Zygosity was determined by DNA analysis. We used Illumina HumanMethylation450 (450K) Beadchip to profile genome-wide DNA methylation sites. This array interrogates more than 485,000 methylation sites per sample at single nucleotide resolution. It covers 99% of RefSeq genes, with an average of 17 CpG sites per gene region distributed across the promoter, 5'UTR, first exon, gene body, and 3'UTR. It covers 96% of CpG islands, with additional coverage in island shores and

the regions flanking them. 0.5 μg of genomic DNA per sample from peripheral blood leukocytes was bisulfite-converted using the EZ DNA Methylation Kit (Zymo Research, Orange CA). Samples were whole-genome amplified, enzymatically fragmented, purified, then hybridized in batches of 12 to the BeadChip containing locus-specific DNA oligomers. The arrays were fluorescently stained, scanned, and fluorescence intensities were assessed at each bead site. GenomeStudio software was used to quantify each methylation site as β-values:

$$\beta - value = \frac{\max(I_{i\ methylated}, 0)}{|I_{i\ methylated}| + |I_{i\ unmethylated}| + \alpha}$$

where $I_{i\ methylated}$ and $I_{i\ unmethylated}$ denote the methylated and unmethylated intensities. β-values are continuous values represent the ratio of fluorescence intensity of the methylated and unmethylated sites, ranging from 0 to 1. The β-values were used for data pre-processing and quality control. CpG sites were excluded from analyses if they had missing rate above 5 %, overlapped with single nucleotide polymorphisms base on Illumina's 450K annotation, or were not uniquely mapped to the reference genome. After quality control, 409,806 autosomal sites for 218 subjects (164 monozygotic and 54 dizygotic twins) were eligible for analysis (41). In addition, DNA methylation patterns vary between different cell types in the blood sample and need to be adjusted. Therefore, cell type proportions of granulocytes, monocytes, natural killer cells (NK cells), B cells, CD4+ and CD8+ T cells were calculated using cell-type specific DNAm sites, and then adjusted as covariates in association analyses (42-44).

Phenotypic measurements

Twins were given the same low-fat diet the night before the tests and refrained from smoking and alcoholic or caffeinated drinks. All medications were held for about 24 hours before testing. After an overnight fasting, all measurements were conducted in the morning, and both twins in pairs were tested at the same time. Biochemical assays were processed in the same analytical run. Medical history for twins were obtained and physical exams were conducted. Body mass index (BMI) was calculated based on weight and height measured. Cigarette smoking was categorized as current smokers (any number of cigarettes) versus non-current smokers.

Positron emission tomography (PET) scanning was performed in 2-dimensional mode using a CTI ECAT Exact 47 (921) camera (5-mm resolution) (Siemens, Knoxville, Tennessee). During a single imaging session, subjects underwent myocardial blood flow (MBF) imaging with PET nitrogen 13 [^{13}N] ammonia at rest, and the following 4-min pharmacological (adenosine infusion, 0.14 mg/kg/min) stress. To measure MBF at rest and during adenosine hyperemia using established methodology (45, 46), the left ventricle was sampled radially from 40 different angles, and 40 samples of flow were obtained for each short-axis slice, yielding hundreds of samples, which were grouped into 20 segments. CFR of the entire myocardium was calculated as the mean ratio of maximum MBF during stress to MBF at rest across all 20 segments.

Statistical analysis

First, a linear mixed effect model was conducted to identify the effect of smoking on CFR, treating co-twin as a random effect:

$$CFR = \beta_0 + \beta_1 * SMK + \beta_2 * AGE + \beta_3 * BMI + E \qquad (1)$$

Here, CFR is the ratio mean of maximum MBF during stress to MBF at rest across all 20 segments; SMK denotes the smoking status (current smoker vs non-current smoker); AGE indicates age in years; BMI is the body mass index, and E suggests the error.

Second, EWAS was performed to find out the significant associations between smoking and DNA methylation based on each CpG site. A linear mixed effect model was used, adjusting for Beadchips and pairs as random effects:

$$\beta - value = \beta_0 + \beta_1 * SMK + \beta_2 * BMI + \beta_3 * AGE + \beta_4 * GRAN + \beta_5 * MONO + \beta_6 * NK + \beta_7 * B + \beta_8 * CD4 + \beta_9 * CD8 + E \qquad (2)$$

In this model, GRAN, MONO, NK, B, CD4 and CD8 denotes the cell type proportions of granulocytes, monocytes, natural killer (NK) cells, B cells, CD4+ and CD8+ T cells.

Then, smoking-associated CpG sites were chosen for the following analyses. Within the list of 62 markers, which were replicated 3 or more times based on the systematic review by Gao et al. (35), markers replicated in our EWAS were selected.

For the selected smoking-associated CpG sites, a linear mixed effect model was conducted to identify the associations between these markers and CFR, adjusting for Beadchips and twin pairs as random effects in model (3):

$$CFR = \beta_0 + \beta_1 * \beta - value + \beta_2 * BMI + \beta_3 * AGE + \beta_4 * GRAN + \beta_5 * MONO + \beta_6 * NK + \beta_7 * B + \beta_8 * CD4 + \beta_9 * CD8 + E \qquad (3)$$

Finally, to identify the independent epigenetic effect of the smoking-associated CpG sites on CFR, a similar linear mixed effect model conditional on smoking was used in addition to covariates in model (3), adjusting for Beadchips and twin pairs as random effects:

$$CFR = \beta_0 + \beta_1 * \beta - value + \beta_2 * SMK + \beta_3 * BMI + \beta_4 * AGE + \beta_5 * GRAN + \beta_6 * MONO + \beta_7 * NK + \beta_8 * B + \beta_9 * CD4 + \beta_{10} * CD8 + E \quad (4)$$

Finally, model (5) and model (6) were implemented to assess within-twin and between-twin effects (47):

$$CFR = \beta_0 + beta_W * \beta - diff + beta_B * \beta - mean + \beta_1 * BMI + \beta_2 * AGE + \beta_3 * GRAN + \beta_4 * MONO + \beta_5 * NK + \beta_6 * B + \beta_7 * CD4 + \beta_8 * CD8 + E \quad (5)$$

$$CFR = \beta_0 + beta_W * \beta - diff + beta_B * \beta - mean + \beta_1 * SMK + \beta_2 * BMI + \beta_3 * AGE + \beta_4 * GRAN + \beta_5 * MONO + \beta_6 * NK + \beta_7 * B + \beta_8 * CD4 + \beta_9 * CD8 + E \quad (6)$$

where β-mean represents the mean value of DNA methylation within each twin pair and β-diff denotes twin-pair methylation difference, which is the value of each individual methylation β-value subtracted by β-mean of the pair. The within-pair coefficient $beta_W$ gives the expected change in CFR for a one unit change in the difference between the individual methylation and the twin pair average methylation, while holding the latter constant. The between-pair coefficient $beta_B$ gives the expected change in CFR for a one-unit change in the twin pair average DNA methylation level, while holding the individual deviation from the average constant. These two models mainly focus on explaining the association between DNA methylation and CFR, by exploring the shared genetic or environmental effects within twin pairs and other unshared environmental effects (47).

Statistical analyses were performed in R (version 3.3.2, released 2016-10-31. https://www.r-project.org). Linear mixed effect models were implemented using R package nlme.

RESULTS

Among the total of 218 subjects, 64 (29%) were current smokers and 154 (71%) were non-current smokers. The mean age of 218 subjects were 55.6 in years, with the youngest of 48 and the oldest of 63. One subject had missing BMI value, yielding the total sample size of 217 in model (1) and model (2). The average BMI was 29.3 kg/m^2, with the minimum of 16.5 kg/m^2 and maximum of 55.6 kg/m^2. Fifty-three subjects had missing CFR values, yielding total sample size of 164 in model (3), model (4), model (5) and model (6). The mean CFR was 2.45; the minimum and maximum values were 1.11 and 4.56, respectively. Results of model (1) indicated that the current smoking status was significantly associated with CFR, and current smokers had CFR 0.513 units lower than non-current smokers ($p<0.001$).

Using model (2), a total of 1,052 CpG sites revealed significant association with current smoking status (FDR adjusted p-values < 0.05). The distribution of observed vs. expected p-values for all 409,806 CpG sites were shown in the QQ plot (Figure 1). The inflation factor of 1.37 is relatively close to 1, indicating that the EWAS results were not inflated by population stratification or relatedness. Figure 2 summarizes p-values from the EWAS of current smoking ordered by chromosomal position (i.e., Manhattan plot). The effect sizes, standard errors, t statistics, unadjusted p-values and FDR adjusted p-values of the 62 smoking-related markers replicated 3 or more times (35), were listed in Table 2. Four out of the 62 markers were not included in the DNA methylation data, thus excluded in further analyses. For the remaining 58 markers, 50 were successfully replicated with FDR adjusted p-values less than 0.05. Four of the top five CpGs were in AHRR gene, including cg05575921 ($p=1.92\times10^{-19}$), cg26703534 ($p=2.68\times10^{-14}$), cg21161138

($p=7.54\times10^{-13}$), cg25648203 ($p=5.79\times10^{-12}$). Cg01940273 (q37.1, Chr. 2) was found the second most significant with p-value of 3.40×10^{15}.

From the 50 smoking-related CpGs in our study (Table 3), 15 markers were significantly associated with CFR ($p< 0.05$) using model (3), with cg21121843 in huntingtin (HTT) gene, Chromosome 4 having the lowest p-value of 0.001. Under a less stringent alpha threshold of 0.1, 22 markers were significantly associated with CFR. Notably, cg21121843 (HTT, Chr. 4) was the only marker remained positively associated with CFR ($p=0.046$) after controlling for smoking status in model (4), even when alpha level of 0.1 was applied.

At alpha level of 0.05, 5 CpG sites had significant p-values for solely within-twin effect, while 7 CpG sites had solely between-twin effect using twin-specific model (5). cg21121843 was the only marker had both within-twin ($p=0.037$) and between-twin ($p=0.008$) effects significant. However, for the twin-specific model (6) conditioning on smoking status, none of the markers remained significant (Table 4).

DISCUSSION

Smoking was identified significantly associated with DNA methylation change using model in our study. Top consistently reported associations were successfully replicated using our data. Therefore, the smoking-associated markers selected for analysis in relation to CFR were well confirmed. Except for cg22132788 and cg12803068 in myosin IG (MYO1G), Chromosome 7, all the smoking-related DNA methylation sites had negative effect size, suggesting that current smokers had lower methylation levels on these sites, compared to non-current smokers.

Cg21121843 in HTT gene had the most significant association with CFR and remained significant (p=0.046) when adjusting for smoking status. HTT is a disease gene linked to Huntington's disease, which is a neurodegenerative disorder characterized by loss of striatal neurons. DNA sequence of HTT was identified and the precise association between of HTT mutation and Huntington's disease was determined. the HTT gene contains a region where the triplet nucleotide CAG is repeated several times, and the number of repeats within the gene determines the occurrence of Huntington's disease (48, 49). The mutant HTT gene encodes a protein huntingtin with an expanded polyglutamine domain, but the mechanism of how the mutant protein leads to the degeneration of neurons and neuropathological manifestations remains unclear (50). More importantly, several studies have identified changes in DNA methylation associated with expression of huntingtin protein in different Huntington's disease model systems and human brain (50). As a disease of the central nervous system, mortality surveys showed that heart disease is a leading cause of death among Huntington's disease patients. Clinical findings indicated that cardiac dysautonomia may be a result of the central autonomic network,

since automatic dysfunction often accompanies Huntington's disease (51). Therefore, it may be reasonable to hypothesize that in addition to the mediator of smoking effect in the association between cg21121843 methylation and CFR, cg21121843 methylation also had effect on CFR through the central autonomic network. However, Huntington's disease is a relatively rare disease and our sample size may not be sufficient to capture the true positive associations. Furthermore, since we tested 50 markers in model (3) and model (4), making conclusion based on a marginal significant p-value (p=0.046) can lead to false positives because no correction for multiple tests was conducted. Using model (5), both within-twin effect (p=0.037) and between-twin effect (p=0.008) were significant, which suggests that both shared genetic effect and unshared environmental effect contribute to CFR in the unconditional model. However, neither of both remained significant after adjusting for smoking status, which may be due to the relatively small sample size and insufficient statistical power.

Reynolds et al. found that AHRR methylation is associated with atherosclerosis, which can lead to heart dysfunction including abnormal CFR, considering cg05575921 and cg21161138 in AHRR remained significant in current and former smokers even after adjusting for self-reported smoking habits, urinary cotinine, and well-known cardiovascular disease risk factors (36). In our study, cg05575921, cg21161138 and cg25648203 in AHRR had significant p-values in model (3), but did not remain significant in model (4) controlling for smoking. Findings in our study did not replicate the associations reported. The relatively comparative sample size of current smokers and non-current smokers strengthened the validity of this study, yet the total sample size might lead to less detection power.

Yixiang Li (37) studied the impact of smoking on atherosclerosis through DNA methylation in 2015, using a subgroup of 134 subjects from our study cohort. His study consisted of 47 (34%) current smokers and 87 (66%) non-current smokers. Even with a smaller total sample size, his analyses identified that smoking status was marginally significantly associated with carotid intima-media thickness (p=0.095), which is an important quantitative assessment of atherosclerosis. After controlling for smoking status in the model, cg03991871 (AHRR, Chr. 5, p=0.027) and cg06126421 (6p21.33, Chr. 6, p=0.009) remained significant in his study. To further investigate using larger sample size, analyses were performed for these CpGs and found that cg03991871 was significantly associated with smoking (FDR adjusted p-value=0.007), but no significant association was found using model (3) (p=0.671) or model (4) (p=0.355); cg06126421 was also identified as a significant smoking-related CpG site (FDR adjusted p-value=8.00×10^{-7}), significant using model (3) (p=0.036), but after controlling for smoking, the p-value increased to 0.655. As a result of the increased sample size in our study, the statistical power was enhanced and the probability of retaining incorrect false negatives (type II error) was reduced. A possible explanation for the reverse results might be the analyses being more subjected to potential outliers due to smaller sample size.

STRENGTHS & WEAKNESSES

Since various sources of tissue collected for DNA methylation measurement can have various methylation profile, it is critical that we collected blood for measurement, which is consistent with the publications from literature review. Therefore, an impressive proportion of the previous reported markers could be replicated using our data. When selecting CpGs, we considered all 62 CpG sites that were previously reported for 3 or more times, so that it was well confirmed that these markers were smoking-associated. On the other hand, proportions of leukocyte subtypes were computed and added in models as covariates in order to avoid the confounding effect due to the shift of leukocyte populations. Furthermore, linear mixed effect model was implemented to correct the co-twin effect due to the shared genetic profile in twin pairs.

One of the major drawbacks of this study is the relatively small sample size, which may result in insufficient statistical power to detect the significant loci, and the sample could not represent the general population since all the subjects were middle-aged male-male twins. The cross-sectional design of this study also narrowed down the data to only the time point when the measurements were performed, instead of investigating under the progression of DNA methylation and CFR changes.

FUTURE DIRECTIONS

More studies utilizing larger sample size that can better represent the general population need to be conducted in the future, not only to replicate the findings from this study, but also to investigate the associations among a larger pool of smoking-related loci as more CpGs being revealed significant. Instead of cross-sectional design, longitudinal data can allow us to draw a more concrete conclusion by exploring the changes of DNA methylation and CFR over time. Based on our findings, the identified associated CpG site cg21121843 in HTT, Chromosome 4 should be further studied to understand the role of the coded protein and its function on the cardiovascular disease pathology. If the findings from this study can be replicated, we can use cg21121843 as a quantitative predictor to assess abnormal CFR and cardiovascular disease risks.

REFERENCES

1. Centers for Disease Control and Prevention. Current Cigarette Smoking Among Adults—United States, 2005-2015. Morbidity and Mortality Weekly Report 2016;65(44):1205-1211.

2. U.S. Department of Health and Human Services. The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General 2014.

3. Blakely T, Barendregt JJ, Foster RH, et al. The association of active smoking with multiple cancers: national census-cancer registry cohorts with quantitative bias analysis. Cancer Causes & Control 2013;24(6):1243-1255.

4. Cunningham TJ, Ford ES, Rolle IV, et al. Associations of Self-Reported Cigarette Smoking with Chronic Obstructive Pulmonary Disease and Co-Morbid Chronic Conditions in the United States. COPD: Journal of Chronic Obstructive Pulmonary Disease 2015;12(3):281-291.

5. Weitzman M. Tobacco Smoke Exposure Is Associated with the Metabolic Syndrome in Adolescents. Circulation 2005;112(6):862-869.

6. Kurmi OP, Chen Z, Wang J, et al. COPD and its association with smoking in the Mainland China: a cross-sectional analysis of 0.5 million men and women from ten diverse areas. International Journal of Chronic Obstructive Pulmonary Disease 2015;10:655-65.

7. Li B, Wang L, Lu M, et al. Passive Smoking and Breast Cancer Risk among Non-Smoking Women: A Case-Control Study in China. Plos One 2015;10(4): e0125894.

8. Kim CH, Lee YC, Hung RJ, et al. Exposure to secondhand tobacco smoke and lung cancer by histological type: a pooled analysis of the International Lung Cancer Consortium (ILCCO). International Journal of Cancer 2014;135(8):1918-30.

9. Chen W, Yun M, Fernandez C, et al. Secondhand smoke exposure is associated with increased carotid artery intima-media thickness: The Bogalusa Heart Study. Atherosclerosis 2015;240(2):374-379.

10. Ding D, Fung JW, Zhang Q, et al. Effect of household passive smoking exposure on the risk of ischaemic heart disease in never-smoke female patients in Hong Kong. Tobacco Control 2009;18(5):354-357.

11. Wang C, Mugler JP, Lange EE, et al. Lung injury induced by secondhand smoke exposure detected with hyperpolarized helium-3 diffusion MR. Journal of Magnetic Resonance Imaging 2013;39(1):77-84.

12. U.S. Department of Health and Human Services. How Tobacco Smoke Causes Disease: What It Means to You. A Report of the Surgeon General 2010.

13. Shields M, Wilkins K. Smoking, smoking cessation and heart disease risk: A 16-year follow-up study. Health Rep. 2013;24(2):12-22.

14. Park SM, Shim WJ, Song WH, et al. Effects of smoking on coronary blood flow velocity and coronary flow reserve assessed by Transthoracic Doppler Echocardiography. Echocardiography 2006;23(6):465-470.

15. Tanaka T, Oka Y, Tawara I, et al. Acute effects of nicotine content in cigarettes on coronary flow velocity and coronary flow reserve in men. The American Journal of Cardiology 1998;82(10):1275-1278.

16. Czernin J, Sun K, Brunken R, et al. Effect of acute and long-term smoking on myocardial blood flow and flow reserve. Circulation 1995;91(12):2891-2897.

17. Bradley AJ, Alpert JS. Coronary flow reserve. American Heart Journal 1991;122(4):1116-1128.

18. Kaufmann PA, Camici PG. Myocardial blood flow measurement by PET: technical aspects and clinical applications. J Nucl Med. 2005;46:75-88.

19. Tanaka T, Oka Y, Tawara I, et al. Impaired coronary flow reserve due to long-term smoking recovers after quitting. Journal of Cardiology 1998;31(6):337-41.

20. Breitling LP, Yang R, Korn B, et al. Tobacco-smoking-related differential DNA Methylation: 27K discovery and replication. The American Journal of Human Genetics 2011;88(4):450-457.

21. Robertson KD. DNA methylation and human disease. Nature Reviews Genetics 2005;6(8):597-610.

22. Jin B, Li Y, Robertson KD. DNA Methylation: Superior or subordinate in the epigenetic hierarchy? Genes & Cancer 2011;2(6):607-617.

23. Zhang Y, Florath I, Saum KU, et al. Self-reported smoking, serum cotinine, and blood DNA methylation. Environmental Research 2016;146:395-403.

24. Su D, Wang X, Campbell MR, et al. Distinct epigenetic effects of tobacco smoking in whole blood and among Leukocyte subtypes. Plos One 2016;11(12):e0166486.

25. Lee MK, Hong Y, Kim SY, et al. DNA methylation and smoking in Korean adults: Epigenome-wide association study. Clinical Epigenetics 2016:8(1).

26. Ambatipudi S, Cuenin C, Hernandez-Vargas H, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics 2016;8(5):599-618.

27. Sayols-Baixeras S, Lluís-Ganella C, Subirana I, et al. Identification of a new locus and validation of previously reported loci showing differential methylation associated with smoking. The REGICOR study. Epigenetics 2015;10(12):1156-1165.

28. Zaghlool SB, Al-Shafai M, Al Muftah WA, et al. Association of DNA methylation with age, gender, and smoking in an Arab population. Clinical Epigenetics 2015;7(1):6.

29. Guida F, Sandanger TM, Castagne R, et al. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. Hum Mol Genet. 2015;24(8):2349–59.

30. Besingi W, Johansson A. Smoke-related DNA methylation changes in the etiology of human disease. Human Molecular Genetics 2014;23(9):2290-2297.

31. Harlid S, Xu Z, Panduri V, et al. CpG sites associated with cigarette smoking: analysis of epigenome-wide data from the sister study. Environ Health Perspect. 2014;122(7):673–8.

32. Dogan MV, Shields B, Cutrona C, et al. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. BMC Genomics 2014;15(1):151.

33. Elliott HR, Tillin T, McArdle WL, et al. Differences in smoking associated DNA methylation patterns in south Asians and Europeans. Clinical Epigenetics 2014;6(1):4.

34. Tsaprouni, LG, Yang TP, Bell J, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. Epigenetics 2014;9(10):1382-1396.

35. Gao X, Jia M, Zhang Y, et al. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clinical Epigenetics 2015;7(1):113.

36. Reynolds LM, Wan M, Ding J, et al. DNA methylation of the Aryl Hydrocarbon Receptor Repressor associations with cigarette smoking and subclinical atherosclerosis. Circulation: Cardiovascular Genetics 2015;8(5):707-716.

37. Li Y. Exploration of the Impact of Smoking on Atherosclerosis through DNA Methylation. Atlanta, GA: Emory University; 2015.

38. Goldberg J, Curran B, Vitek ME, et al. The Vietnam era twin registry. Twin Res. 2002;5:476-481.

39. Vaccarino V, Brennan ML, Miller AH, et al. Association of major depressive disorder with serum myeloperoxidase and other markers of inflammation: a twin study. Biol Psychiatry 2008;64(6):476-83.

40. Vaccarino V, Lampert R, Bremner JD, et al. Depressive symptoms and heart rate variability: evidence for a shared genetic substrate in a study of twins. Psychosom Med. 2008;70(6):628-36.

41. Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 2013;8(2):203-9.

42. Houseman E, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012;13(1):86.

43. Rahmani E, Zaitlen N, Baran Y, et al. Correcting for cell-type heterogeneity in DNA methylation: A comprehensive evaluation. Nature Methods 2017;14(3):218-219.

44. McGregor K, Bernatsky S, Colmegna I, et al. An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. Genome Biology 2016;17(1).

45. Hutchins GD, Schwaiger M, Rosenspire KC, et al. Noninvasive quantification of regional blood flow in the human heart using N-13 ammonia and dynamic positron emission tomographic imaging. J Am Coll Cardiol. 1990;15:1032-42.

46. El Fakhri G, Sitek A, Guerin B, et al. Quantitative dynamic cardiac 82Rb PET using generalized factor and compartment analyses. J Nucl Med. 2005;46:1264-71.

47. Carlin JB, Gurrin LC, Sterne JA, et al. Regression models for twin studies: a critical review. International Journal of Epidemiology 2005;34(5):1089-99.

48. Macdonald M. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 1993;72(6):971-983.

49. Chial H. Huntington's disease: The discovery of the Huntingtin gene. Nature Education 2008;1(1):71.

50. Thomas EA. DNA methylation in Huntington’s disease: Implications for transgenerational effects. Neuroscience Letters 2016;625:34-39.

51. Abildtrup M, Shattock M. Cardiac dysautonomia in Huntington’s disease. Journal of Huntington’s Disease 2013;2:251-261.

TABLES

Table 1. Characteristics Summary of EWASs in 2014-2016

First author (year)	Country	Source of DNA	Sample size	Measure-ment	Correction of multiple tests	Identified number of CpGs
Zhang, (2016) (23)	Germany	Whole blood	500	450K	Bonferroni	40
Su, (2016) (24)	US	Whole blood	253	450K	Bonferroni	738
Lee, (2016) (25)	Korean	Blood	100	450K	FDR	108
Ambatipudi, (2016) (26)	European countries	Peripheral blood	910	450K	FDR	748
Sayols-Baixeras, (2015) (27)	Spain	Peripheral blood	645	450K	FDR	66
Zaghlool, (2015) (28)	Qatar	Whole blood	123	450K	FDR	8
Guida, (2015) (29)	UK	Whole blood	745	450K	FWER	461
Besingi, (2014) (30)	Sweden	Peripheral blood	432	450K	FDR	95
Harlid, (2014) (31)	US	Whole blood	908	27/450K	FDR	8
Dogan, (2014) (32)	US	Peripheral blood	111	450K	FDR	910
Elliott, (2014) (33)	UK	Peripheral blood	192	450K	FWER	29
Tsaprouni, (2014) (34)	UK	Peripheral blood	464	450K	Bonferroni	30

Table 2. Summary of Smoking-associated CpG Sites from Systematic Review

CpG	Chr.	Gene	SMK_E	SMK_SE	SMK_t	SMK_P[b]	SMK_P[c]
cg05575921	5	AHRR	-0.1919	0.0138	-13.9250	4.69E-25	1.92E-19
cg01940273	2	X[a]	-0.0981	0.0083	-11.7615	1.66E-20	3.40E-15
cg26703534	5	AHRR	-0.0575	0.0051	-11.2639	1.96E-19	2.68E-14
cg21161138	5	AHRR	-0.0680	0.0065	-10.5379	7.36E-18	7.54E-13
cg25648203	5	AHRR	-0.0520	0.0052	-10.0867	7.08E-17	5.79E-12
cg03636183	19	F2RL3	-0.1016	0.0103	-9.8460	2.37E-16	1.62E-11
cg21566642	2	X[a]	-0.0999	0.0108	-9.2038	5.96E-15	3.49E-10
cg15342087	6	X[a]	-0.0377	0.0041	-9.1699	7.07E-15	3.62E-10
cg24859433	6	X[a]	-0.0398	0.0047	-8.5464	1.60E-13	7.27E-09
cg09935388	1	GFI1	-0.0790	0.0099	-8.0136	2.24E-12	9.19E-08
cg20295214	1	AVPR1B	-0.0294	0.0038	-7.6927	1.09E-11	3.71E-07
cg19572487	17	RARA	-0.0493	0.0065	-7.6041	1.67E-11	4.90E-07
cg14817490	5	AHRR	-0.0473	0.0063	-7.4969	2.82E-11	7.70E-07
cg05951221	2	X[a]	-0.0787	0.0105	-7.4634	3.32E-11	8.00E-07
cg06126421	6	X[a]	-0.0787	0.0105	-7.4738	3.16E-11	8.00E-07
cg23576855	5	AHRR	-0.1527	0.0205	-7.4385	3.75E-11	8.08E-07
cg02451831	7	KIAA0087	-0.0342	0.0047	-7.3420	5.98E-11	1.22E-06
cg03329539	2	X[a]	-0.0404	0.0057	-7.1266	1.69E-10	2.88E-06
cg21121843	4	HTT	-0.0476	0.0068	-6.9873	3.29E-10	5.39E-06
cg27241845	2	X[a]	-0.0431	0.0063	-6.8516	6.27E-10	9.87E-06
cg22851561	14	C14orf43	-0.0295	0.0045	-6.5007	3.26E-09	3.82E-05
cg14753356	6	X[a]	-0.0328	0.0051	-6.4161	4.83E-09	5.11E-05
cg12876356	1	GFI1	-0.0716	0.0113	-6.3505	6.55E-09	6.70E-05
cg14580211	5	C5orf62	-0.0379	0.0060	-6.2959	8.42E-09	7.80E-05
cg19859270	3	GPR15	-0.0196	0.0031	-6.2984	8.33E-09	7.80E-05
cg23161492	15	ANPEP	-0.0403	0.0064	-6.2642	9.75E-09	8.67E-05
cg21611682	11	LRP5	-0.0348	0.0056	-6.2371	1.10E-08	9.61E-05
cg24090911	5	AHRR	-0.0312	0.0050	-6.2224	1.18E-08	0.0001
cg18316974	1	GFI1	-0.0570	0.0092	-6.1694	1.50E-08	0.0001
cg05284742	14	ITPK1	-0.0279	0.0046	-6.1199	1.89E-08	0.0001
cg07123182	11	KCNQ1OT1	-0.0222	0.0038	-5.7773	8.80E-08	0.0006
cg26963277	11	KCNQ1OT1	-0.0326	0.0057	-5.6985	1.25E-07	0.0008
cg01899089	5	AHRR	-0.0378	0.0067	-5.6272	1.71E-07	0.0009
cg00310412	15	SEMA7A	-0.0290	0.0053	-5.5101	2.85E-07	0.0013
cg08709672	1	AVPR1B	-0.0240	0.0044	-5.4535	3.64E-07	0.0015
cg24996979	14	C14orf43	-0.0152	0.0028	-5.3866	4.85E-07	0.0019
cg23916896	5	AHRR	-0.0452	0.0086	-5.2645	8.18E-07	0.0026
cg26271591	2	NFE2L2	-0.0415	0.0079	-5.2593	8.36E-07	0.0027
cg02657160	3	CPOX	-0.0169	0.0032	-5.2438	8.93E-07	0.0028
cg22132788	7	MYO1G	0.0481	0.0094	5.1052	1.60E-06	0.0042
cg04885881	1	X[a]	-0.0369	0.0074	-4.9833	2.66E-06	0.0057
cg03991871	5	AHRR	-0.0310	0.0063	-4.9274	3.35E-06	0.0066
cg13976502	14	C14orf43	-0.0232	0.0049	-4.7445	7.04E-06	0.0106
cg21913886	1	TMEM51	-0.0239	0.0053	-4.5337	1.63E-05	0.0170
cg12806681	5	AHRR	-0.0166	0.0037	-4.4354	2.38E-05	0.0209
cg11660018	11	PRSS23	-0.0300	0.0068	-4.4028	2.70E-05	0.0222
cg25189904	1	GNG12	-0.0463	0.0106	-4.3733	3.03E-05	0.0234
cg23079012	2	X[a]	-0.0327	0.0076	-4.3084	3.88E-05	0.0267
cg12803068	7	MYO1G	0.0462	0.0109	4.2344	5.14E-05	0.0311
cg01901332	11	ARRB1	-0.0255	0.0062	-4.1272	7.67E-05	0.0389
cg25949550	7	CNTNAP2	-0.0108	0.0028	-3.8220	0.0002	0.0649
cg01731783	14	C14orf43	-0.0145	0.0038	-3.8018	0.0002	0.0667
cg13193840	2	X[a]	-0.0095	0.0027	-3.5671	0.0006	0.0930
cg12075928	8	PTK2	-0.0223	0.0069	-3.2093	0.0018	0.1499
cg03547355	1	X[a]	-0.0167	0.0055	-3.0606	0.0028	0.1783
cg23771366	11	PRSS23	-0.0198	0.0069	-2.8696	0.0050	0.2219
cg11314684	1	AKT3	-0.0157	0.0061	-2.5814	0.0113	0.2982
cg06644428	2	X[a]	-0.0146	0.0069	-2.1305	0.0356	0.4448
cg01692968[d]	9	X[a]	-	-	-	-	-
cg06060868[d]	5	SDHA	-	-	-	-	-
cg11207515[d]	7	CNTNAP2	-	-	-	-	-
cg11231349[d]	1	NOS1AP	-	-	-	-	-

a. According to UCSC Genome Browser, no annotated transcripts are associated with these CpG sites.
b. Unadjusted p-values.
c. FDR adjusted p-values.
d. After quality control, these CpGs were excluded in the following analyses.

Table 3. Summary of Smoking-associated CpGs in Relation to CFR

CpG	Chr.	Gene	SMK_P[b]	Unconditional				Conditional on SMK			
				cg_E	cg_SE	cg_t	cg_P	cg_E	cg_SE	cg_t	cg_P
cg05575921	5	AHRR	1.92E-19	1.3472	0.4718	2.8558	***0.0057***	0.3060	0.7144	0.4283	0.6698
cg01940273	2	X[a]	3.40E-15	2.1547	0.8397	2.5659	***0.0125***	0.4767	1.1134	0.4282	0.6699
cg26703534	5	AHRR	2.68E-14	1.1105	1.5275	0.7270	0.4698	-2.4556	1.7984	-1.3655	0.1767
cg21161138	5	AHRR	7.54E-13	2.4122	1.1557	2.0873	***0.0407***	-0.0347	1.4797	-0.0235	0.9813
cg25648203	5	AHRR	5.79E-12	3.6383	1.5439	2.3565	***0.0214***	0.7603	1.9299	0.3939	0.6949
cg03636183	19	F2RL3	1.62E-11	1.1926	0.7478	1.5948	0.1155	-0.7196	0.9266	-0.7766	0.4402
cg21566642	2	X[a]	3.49E-10	1.4785	0.7009	2.1095	***0.0386***	0.1384	0.8628	0.1604	0.8730
cg15342087	6	X[a]	3.62E-10	5.0560	1.8908	2.6740	***0.0094***	2.4135	2.0817	1.1594	0.2505
cg24859433	6	X[a]	7.27E-09	4.2808	1.7293	2.4755	***0.0158***	1.7822	1.9188	0.9288	0.3564
cg09935388	1	GFI1	9.19E-08	1.5482	0.6875	2.2520	***0.0276***	0.5926	0.7546	0.7853	0.4351
cg20295214	1	AVPR1B	3.71E-07	1.7251	2.1990	0.7845	0.4355	-0.7224	2.2496	-0.3211	0.7491
cg19572487	17	RARA	4.90E-07	1.9361	1.3188	1.4681	0.1468	-0.5812	1.5048	-0.3862	0.7006
cg14817490	5	AHRR	7.70E-07	2.2073	1.2955	1.7039	0.0930	0.3058	1.4364	0.2129	0.8320
cg05951221	2	X[a]	8.00E-07	1.2302	0.7856	1.5660	0.1221	-0.1506	0.9023	-0.1669	0.8679
cg06126421	6	X[a]	8.00E-07	1.5987	0.7477	2.1382	***0.0362***	0.3780	0.8414	0.4493	0.6547
cg23576855	5	AHRR	8.08E-07	0.2359	0.3599	0.6555	0.5144	-0.2675	0.3759	-0.7117	0.4792
cg02451831	7	KIAA0087	1.22E-06	2.9343	1.8286	1.6047	0.1133	0.1461	1.9715	0.0741	0.9411
cg03329539	2	X[a]	2.88E-06	1.3820	1.5435	0.8954	0.3738	-1.5917	1.7579	-0.9054	0.3685
cg21121843	4	HTT	5.39E-06	4.1671	1.2197	3.4165	***0.0011***	2.7713	1.3636	2.0324	***0.0461***
cg27241845	2	X[a]	9.87E-06	0.6063	1.3677	0.4433	0.6590	-1.4044	1.4495	-0.9689	0.3361
cg22851561	14	C14orf43	3.82E-05	3.7539	1.9039	1.9718	0.0528	1.5346	2.0137	0.7621	0.4487
cg14753356	6	X[a]	5.11E-05	3.4574	1.7655	1.9583	0.0544	0.9585	1.9421	0.4935	0.6233
cg12876356	1	GFI1	6.70E-05	1.5928	0.6346	2.5100	***0.0145***	0.8928	0.6533	1.3667	0.1764
cg14580211	5	C5orf62	7.80E-05	1.5194	1.3744	1.1055	0.2729	-1.0139	1.4850	-0.6828	0.4972
cg19859270	3	GPR15	7.80E-05	4.6306	2.8503	1.6246	0.1089	1.0666	2.9994	0.3556	0.7233
cg23161492	15	ANPEP	8.67E-05	2.3756	1.2628	1.8812	0.0643	0.5665	1.3863	0.4087	0.6841
cg21611682	11	LRP5	9.61E-05	2.5054	1.4958	1.6750	0.0986	0.4736	1.6248	0.2915	0.7716
cg24090911	5	AHRR	0.0001	1.9295	1.8340	1.0521	0.2965	-0.6193	1.9320	-0.3206	0.7495
cg18316974	1	GFI1	0.0001	2.0563	0.8313	2.4735	***0.0159***	1.1207	0.8538	1.3126	0.1939
cg05284742	14	ITPK1	0.0001	2.8899	1.9063	1.5159	0.1342	0.3024	2.0312	0.1489	0.8821
cg07123182	11	KCNQ1OT1	0.0006	2.9822	2.3129	1.2893	0.2017	-0.1466	2.4446	-0.0600	0.9524
cg26963277	11	KCNQ1OT1	0.0008	3.3667	1.5044	2.2379	***0.0286***	1.5064	1.5845	0.9507	0.3452
cg01899089	5	AHRR	0.0009	1.7948	1.2801	1.4021	0.1655	0.0621	1.3603	0.0457	0.9637
cg00310412	15	SEMA7A	0.0013	1.4618	1.5480	0.9443	0.3484	-0.5193	1.5684	-0.3311	0.7416
cg08709672	1	AVPR1B	0.0015	1.8081	1.9418	0.9311	0.3551	-0.4571	2.0132	-0.2270	0.8211
cg24996979	14	C14orf43	0.0019	6.6972	2.9407	2.2774	***0.0260***	3.5731	3.0364	1.1767	0.2435
cg23916896	5	AHRR	0.0026	0.7471	0.9587	0.7793	0.4386	-0.2733	0.9744	-0.2805	0.7800
cg26271591	2	NFE2L2	0.0027	0.4165	1.0909	0.3818	0.7038	-0.6276	1.0942	-0.5736	0.5682
cg02657160	3	CPOX	0.0028	4.5415	2.8306	1.6044	0.1133	1.4226	2.7917	0.5096	0.6120
cg22132788	7	MYO1G	0.0042	-1.6351	0.7934	-2.0609	***0.0432***	-0.7174	0.8384	-0.8557	0.3953
cg04885881	1	X[a]	0.0057	0.8911	1.1552	0.7714	0.4432	-0.6968	1.1923	-0.5845	0.5609
cg03991871	5	AHRR	0.0066	0.5481	1.2865	0.4261	0.6714	-1.2087	1.2971	-0.9318	0.3548
cg13976502	14	C14orf43	0.0106	1.2004	1.9201	0.6252	0.5340	-0.9730	1.9660	-0.4949	0.6223
cg21913886	1	TMEM51	0.0170	1.4991	1.5143	0.9899	0.3258	0.1789	1.5059	0.1188	0.9058
cg12806681	5	AHRR	0.0209	-0.2027	2.1865	-0.0927	0.9264	-2.6686	2.1445	-1.2444	0.2178
cg11660018	11	PRSS23	0.0222	0.8676	1.3359	0.6495	0.5182	-0.4660	1.3620	-0.3421	0.7333
cg25189904	1	GNG12	0.0234	-0.0929	0.8461	-0.1098	0.9129	-0.9855	0.8669	-1.1368	0.2597
cg23079012	2	X[a]	0.0267	2.8704	1.4476	1.9829	0.0515	1.1409	1.4994	0.7609	0.4494
cg12803068	7	MYO1G	0.0311	-1.2184	0.6673	-1.8260	0.0723	-0.4753	0.6948	-0.6841	0.4963
cg01901332	11	ARRB1	0.0389	0.7679	1.3676	0.5615	0.5763	-0.6533	1.3652	-0.4785	0.6339

a. According to UCSC Genome Browser, no annotated transcripts are associated with these CpG sites.
b. FDR-adjusted p-values for CpG sites in relation to smoking.

Table 4. Summary of Twin Specific Models

CpG	Chr.	Gene	Unconditional								Conditional on SMK							
			W_E[b]	W_SE[b]	W_t[b]	W_P[b]	B_E[c]	B_SE[c]	B_t[c]	B_P[c]	W_E[b]	W_SE[b]	W_t[b]	W_P[b]	B_E[c]	B_SE[c]	B_t[c]	B_P[c]
cg05575921	5	AHRR	1.1010	0.9032	1.2189	0.2271	1.4332	0.5446	2.6315	***0.0105***	-0.2473	1.1443	-0.2161	0.8296	0.4099	0.7348	0.5578	0.5788
cg01940273	2	X[a]	1.7250	1.5312	1.1266	0.2639	2.3321	0.9715	2.4005	***0.0191***	-0.4470	1.8092	-0.2471	0.8056	0.7059	1.1700	0.6034	0.5483
cg26703534	5	AHRR	1.4908	2.3158	0.6437	0.5219	0.8618	1.8605	0.4632	0.6447	-2.9921	2.6732	-1.1193	0.2671	-2.2316	1.9807	-1.1267	0.2638
cg21161138	5	AHRR	3.2599	1.9386	1.6816	0.0973	1.9691	1.3933	1.4132	0.1622	0.5652	2.2650	0.2495	0.8037	-0.2589	1.6143	-0.1604	0.8731
cg25648203	5	AHRR	5.0040	2.6417	1.8943	0.0625	3.0223	1.8284	1.6530	0.1029	1.5118	3.1237	0.4840	0.6300	0.5400	2.0684	0.2611	0.7948
cg03636183	19	F2RL3	2.4216	1.3843	1.7494	0.0848	0.7478	0.8500	0.8798	0.3821	0.1076	1.6352	0.0658	0.9477	-0.8776	0.9661	-0.9084	0.3669
cg21566642	2	X[a]	0.9742	1.4516	0.6711	0.5045	1.6304	0.7859	2.0746	***0.0418***	-0.7607	1.6109	-0.4722	0.6383	0.3191	0.9060	0.3522	0.7258
cg15342087	6	X[a]	8.0293	3.5523	2.2603	***0.0271***	3.9811	2.1754	1.8300	0.0716	4.3793	4.0786	1.0737	0.2869	2.0099	2.2126	0.9084	0.3669
cg24859433	6	X[a]	6.0333	2.9089	2.0741	***0.0419***	3.3300	2.1416	1.5549	0.1246	2.8559	3.3228	0.8595	0.3932	1.4014	2.1524	0.6511	0.5172
cg09935388	1	GFI1	1.0114	1.4818	0.6825	0.4973	1.6778	0.7580	2.2134	***0.0302***	-0.5670	1.6121	-0.3517	0.7262	0.7910	0.7942	0.9961	0.3228
cg20295214	1	AVPR1B	2.4525	3.6250	0.6765	0.5010	1.3423	2.6796	0.5009	0.6180	-1.5685	3.9036	-0.4018	0.6891	-0.3940	2.5708	-0.1533	0.8786
cg19572487	17	RARA	3.4914	1.8203	1.9180	0.0594	0.6655	1.6215	0.4104	0.6828	1.3539	1.9909	0.6801	0.4988	-1.8914	1.7371	-1.0888	0.2801
cg14817490	5	AHRR	2.6989	2.3202	1.1632	0.2489	2.0327	1.4713	1.3816	0.1716	0.0994	2.5503	0.0390	0.9690	0.3594	1.5407	0.2333	0.8163
cg05951221	2	X[a]	0.1104	1.7355	0.0636	0.9495	1.5303	0.8689	1.7613	0.0827	-1.5201	1.8419	-0.8253	0.4122	0.1196	0.9564	0.1251	0.9008
cg06126421	6	X[a]	2.8877	1.4650	1.9711	0.0528	1.2386	0.8246	1.5021	0.1377	1.3168	1.6335	0.8061	0.4231	0.2105	0.8818	0.2388	0.8120
cg23576855	5	AHRR	0.6334	0.7281	0.8698	0.3875	0.1200	0.4019	0.2987	0.7661	-0.4356	0.8166	-0.5334	0.5955	-0.2411	0.3938	-0.6121	0.5425
cg02451831	7	KIAA0087	3.9575	3.0773	1.2860	0.2029	2.4109	2.2295	1.0814	0.2834	0.9412	3.3278	0.2828	0.7782	-0.1647	2.2389	-0.0736	0.9416
cg03329539	2	X[a]	0.9042	2.6554	0.3405	0.7345	1.6284	1.8722	0.8697	0.3875	-2.0510	2.8568	-0.7179	0.4753	-1.3967	1.9994	-0.6986	0.4872
cg21121843	4	HTT	5.0972	2.3983	2.1253	***0.0372***	3.8473	1.4116	2.7255	***0.0082***	4.0392	2.4933	1.6200	0.1100	2.3116	1.5338	1.5071	0.1364
cg27241845	2	X[a]	0.8540	2.4796	0.3444	0.7316	0.4961	1.6479	0.3010	0.7643	-2.4260	2.6972	-0.8994	0.3717	-1.0758	1.6264	-0.6615	0.5105
cg22851561	14	C14orf43	3.7328	2.7532	1.3558	0.1797	3.7645	2.1681	1.7363	0.0870	1.4150	2.9243	0.4839	0.6301	1.5842	2.2014	0.7196	0.4742
cg14753356	6	X[a]	3.0239	2.2999	1.3148	0.1931	3.5943	1.8215	1.9733	0.0525	0.5401	2.4566	0.2199	0.8267	1.0723	1.9896	0.5390	0.5917
cg12876356	1	GFI1	1.9968	1.5236	1.3106	0.1945	1.5108	0.6896	2.1908	***0.0319***	0.7073	1.6323	0.4333	0.6662	0.9213	0.6866	1.3418	0.1841
cg14580211	5	C5orf62	1.9846	2.3349	0.8500	0.3984	1.3973	1.4633	0.9550	0.3430	-0.8848	2.4985	-0.3541	0.7244	-1.0383	1.5384	-0.6749	0.5020
cg19859270	3	GPR15	5.3321	4.3290	1.2317	0.2224	4.1530	3.6183	1.1478	0.2551	1.4909	4.6264	0.3223	0.7483	0.8373	3.5533	0.2357	0.8144
cg23161492	15	ANPEP	2.6064	2.2830	1.1416	0.2577	2.3013	1.4026	1.6408	0.1055	0.0408	2.4895	0.0164	0.9870	0.6868	1.4694	0.4674	0.6417
cg21611682	11	LRP5	3.2935	2.4403	1.3496	0.1817	2.0986	1.7857	1.1752	0.2440	0.8711	2.6345	0.3306	0.7420	0.3113	1.8358	0.1696	0.8659
cg24090911	5	AHRR	3.4030	2.9086	1.1700	0.2462	1.0906	2.2387	0.4872	0.6277	0.9054	3.0810	0.2939	0.7698	-1.3095	2.2200	-0.5899	0.5572
cg18316974	1	GFI1	2.8217	1.8250	1.5461	0.1268	1.8623	0.9181	2.0283	***0.0464***	1.3141	1.9537	0.6726	0.5035	1.0818	0.9094	1.1896	0.2383
cg05284742	14	ITPK1	1.9772	2.7614	0.7160	0.4765	3.2874	2.0937	1.5701	0.1210	-1.1094	2.9409	-0.3772	0.7072	0.7806	2.1584	0.3617	0.7187
cg07123182	11	KCNQ1OT1	4.6151	3.7543	1.2293	0.2233	1.9909	2.9354	0.6782	0.4999	1.9429	3.9571	0.4910	0.6251	-1.2243	2.9248	-0.4186	0.6768
cg26963277	11	KCNQ1OT1	5.7143	2.4883	2.2964	***0.0248***	2.0445	1.8733	1.0914	0.2790	4.1268	2.6389	1.5638	0.1226	0.2955	1.8613	0.1587	0.8743
cg01899089	5	AHRR	4.0900	2.3996	1.7044	0.0929	0.9055	1.4895	0.6079	0.5453	2.3611	2.5301	0.9332	0.3541	-0.6628	1.5229	-0.4352	0.6648
cg00310412	15	SEMA7A	4.7724	3.2455	1.4705	0.1461	0.5906	1.7209	0.3432	0.7325	1.6544	3.4736	0.4763	0.6355	-0.9064	1.6678	-0.5435	0.5886
cg08709672	1	AVPR1B	1.8001	3.8541	0.4670	0.6420	1.8102	2.1506	0.8417	0.4029	-2.7373	4.1536	-0.6590	0.5122	-0.0351	2.1259	-0.0165	0.9869
cg24996979	14	C14orf43	5.9352	4.9444	1.2004	0.2342	7.0857	3.5267	2.0091	***0.0485***	3.9339	5.0966	0.7719	0.4429	3.4082	3.5402	0.9627	0.3391
cg23916896	5	AHRR	0.7803	2.1874	0.3567	0.7224	0.7393	1.0422	0.7094	0.4805	-0.7358	2.2688	-0.3243	0.7467	-0.1978	1.0325	-0.1916	0.8486
cg26271591	2	NFE2L2	1.9302	2.0544	0.9395	0.3508	-0.1087	1.2263	-0.0886	0.9296	0.6173	2.1396	0.2885	0.7739	-0.9474	1.1967	-0.7917	0.4313
cg02657160	3	CPOX	5.4753	4.5370	1.2068	0.2317	3.9553	3.5386	1.1178	0.2676	3.2147	4.7169	0.6815	0.4979	0.5825	3.3112	0.1759	0.8609
cg22132788	7	MYO1G	-2.8157	1.7570	-1.6025	0.1137	-1.3710	0.8679	-1.5797	0.1188	-1.5783	1.8535	-0.8515	0.3975	-0.5652	0.8884	-0.6362	0.5268
cg04885881	1	X[a]	5.3923	2.2173	2.4319	***0.0177***	-0.5395	1.3025	-0.4142	0.6800	3.3101	2.4026	1.3777	0.1729	-1.5909	1.2829	-1.2401	0.2192
cg03991871	5	AHRR	4.6433	3.1198	1.4883	0.1414	-0.2168	1.3931	-0.1557	0.8768	1.3507	3.3818	0.3994	0.6909	-1.5146	1.3561	-1.1169	0.2680
cg13976502	14	C14orf43	1.2805	3.1310	0.4090	0.6839	1.1569	2.3473	0.4929	0.6237	-0.2041	3.2180	-0.0634	0.9496	-1.3445	2.3236	-0.5786	0.5648
cg21913886	1	TMEM51	0.4818	3.0650	0.1572	0.8756	1.7822	1.6898	1.0547	0.2953	-1.3228	3.1508	-0.4198	0.6760	0.5295	1.6421	0.3224	0.7481
cg12806681	5	AHRR	4.4858	4.4904	0.9990	0.3214	-1.6451	2.4954	-0.6593	0.5119	0.6847	4.7394	0.1445	0.8856	-3.4111	2.3479	-1.4529	0.1509
cg11660018	11	PRSS23	-0.3637	2.1098	-0.1724	0.8637	1.3976	1.5031	0.9298	0.3557	-2.0079	2.1707	-0.9250	0.3583	0.0921	1.4944	0.0616	0.9510
cg25189904	1	GNG12	0.0957	1.5696	0.0609	0.9516	-0.1790	1.0005	-0.1789	0.8585	-1.1055	1.6379	-0.6750	0.5021	-0.9383	0.9823	-0.9552	0.3429
cg23079012	2	X[a]	4.9408	2.1956	2.2503	***0.0277***	1.5207	1.7996	0.8450	0.4011	3.2990	2.3409	1.4093	0.1634	0.0479	1.7559	0.0273	0.9783
cg12803068	7	MYO1G	-2.4007	1.5773	-1.5221	0.1327	-0.9761	0.7281	-1.3406	0.1845	-1.4573	1.6456	-0.8856	0.3791	-0.3101	0.7385	-0.4200	0.6758
cg01901332	11	ARRB1	2.8588	2.5362	1.1272	0.2637	0.2033	1.4828	0.1371	0.8913	0.7470	2.6579	0.2810	0.7796	-0.9259	1.4419	-0.6421	0.5229

a. According to UCSC Genome Browser, no annotated transcripts are associated with these CpG sites.

b. Estimates for with-in twin coefficients.

c. Estimates for between-twin coefficients.

FIGURES

Figure 1.

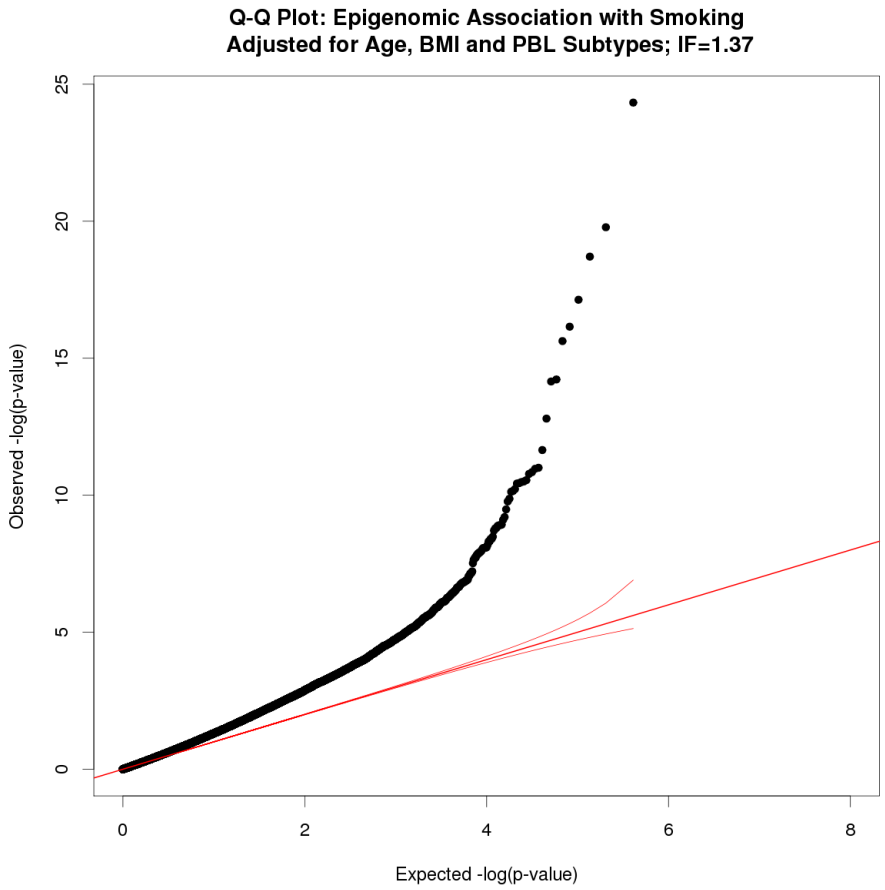
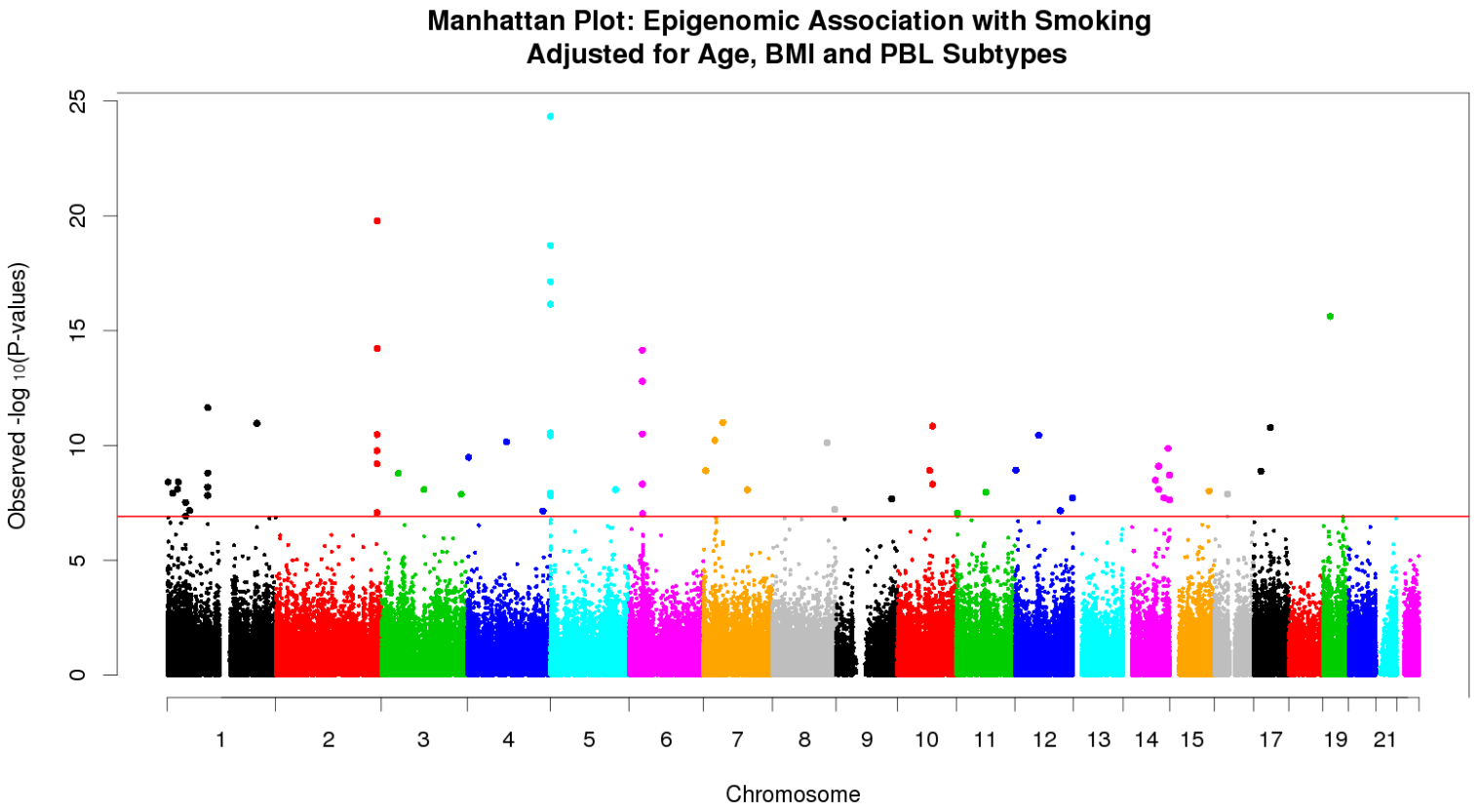


Figure 2.


Manhattan Plot: Epigenomic Association with Smoking
Adjusted for Age, BMI and PBL Subtypes
Observed -log 10(P-values)
0
5
10
15
20
25
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
17
19
21
Chromosome