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:

_____ _____

Yixiang Li Date

Exploration of the Impact of Smoking on Atherosclerosis through DNA Methylation

By

Yixiang Li

Master of Science in Public Health

Epidemiology

_____

Yan Sun

Committee Chair

Exploration of the Impact of Smoking on Atherosclerosis through DNA Methylation

By

Yixiang Li

Bachelor of Science
Nankai University
2013

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 Science in Public Health
in Epidemiology
2015

ABSTRACT

Exploration of the Impact of Smoking on Atherosclerosis through DNA Methylation

By Yixiang Li

Atherosclerosis is a cardiovascular disease (CVD) which is the leading cause of death in the United States. Cigarette smoking is a known risk factor for CVD and many other chronic diseases. Led by development of micro-array, we obtained genome-wide DNA methylation data of 67 pairs of monozygotic twins to investigate the relationship between cigarette smoking, DNA methylation and atherosclerosis. We assessed the association between smoking and intima-media thickness, a quantitative measurement of subclinical atherosclerosis. Comparing to smoking-related DNA methylation sites from publications, we successfully replicated 17 DNA methylation sites. 16 of these sites are significantly associated with IMT adjusting age, body mass index and six cell traits (granulocyte, monocyte, natural killer cells (NK), B cell, CD4+ and CD8+ T cells) at alpha level of 0.05. After performing linear mixed effect model which conditioned on smoking status and adjusted for chip effect and co-twin effect, we found two significant loci: cg03991871 in AHRR (β-coefficient = -0.0503, p-value = 0.0267) and cg06126421 in 6p21.33 (β-coefficient= -0.0604, p-value = 0.0088). The result of our study suggested that DNA methylation might mediate the effect of cigarette smoking on atherosclerosis, and the smoking-related DNA methylation markers could potentially improve the prediction of atherosclerosis and CVD.

Exploration of the Impact of Smoking on Atherosclerosis through DNA Methylation

By

Yixiang Li

Bachelor of Science
Nankai University
2013

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 Science in Public Health
in Epidemiology
2015

Table of Contents

BACKGROUND

Tobacco usage is an established risk factor for many human diseases. More than 1 billion people worldwide and up to 1/3 of the adult population are exposed to tobacco use, and the number of deaths due to smoking is still increasing [1, 2]. According to the National Adult Tobacco Survey in the US, as of the year of 2013, the national prevalence of current smoking was 17.8%. Among all ethnicities, the non-Hispanic American Indians/Alaska Natives had the highest rate of current smoker of 21.8% while the non-Hispanic Asians (exclude Native Hawaiians/Pacific Islanders) had the lowest rate of current smoker of 10.7%[3]. In addition, smoking status also differed by education levels and social economic status. People with 12 or fewer years of education (no diploma) have a smoking rate of 24.2% while those with graduate degree have a smoking rate of 5.6%; 29.2% of adults who live below the poverty level were current smokers while a rate of 16.2% for the others[3].

Tobacco smoking affected human health through both active and passive forms. Howard et al. confirmed the strong relationship between active smoking and increased carotid wall thickness [4]; Glantz et al. found evidences that passive smoking increases risk of death from heart disease: increased resting heart rate led to reduced oxygencarrying capacity and components of environmental tobacco smoking such as benzo(a)pyrene, together with other biochemical evidences, triggered the development of heart diseases[5]. Also, research has shown that tobacco smoking affects multiple tissues such as brain, lung, and heart. Tobacco smoking is a major contributor to brain

aging[6], is associated with lung malfunction such as lung cancer and respiratory infections[7-9], and significantly increases the risk of heart diseases[10]. Moreover, smoking has been proved a major cardiac artery disease risk factor that affects endothelial dysfunction, an early key event in atherogenesis [11]. In addition, active smoking causes harmful and chronic effect with structural lesions that led to coronary atherosclerosis in a long term. In the meantime, passive smoking injures cardiac health for people who established ischemic heart disease[12].

Specifically, tobacco smoking increases the risk for developing atherosclerosis. Atherosclerosis is a disease that sticky substance of plasma lipids (especially cholesterol), cells, and connective tissue matrix, accumulates in artery and causes artery wall thickness. In addition to older age and being male [13], the arterial hypertension, cigarette smoking, high serum levels of cholesterol, and chronic inflammation are potential risk factors [14]. In the developed countries, cardiovascular diseases, especially the atherosclerotic diseases and their complications are among the top causes of mortality [15]. Carotid intima-media thickness (IMT) is an important quantitative assessment of coronary diseases and a good indicator of cardiovascular diseases [16-18]. B-mode ultra sound also gave a non-invasive way to measure the IMT [19]. According to the Atherosclerosis Risk in Communities (ARIC) Study, both active smoking and environmental tobacco smoke (passive smoking) induce the progression of atherosclerosis [20].

Smoking has been shown to affect human genetics such as DNA methylation pattern. DNA methylation is a mechanism which a methyl group is added to the cytosine

nucleotides and typically occurs in CpG island in adult, while non-CpG methylation in embryonic stem cells [21, 22]. In normal cells, DNA methylation is under control while this process can be influenced by endogenous factors and environmental stimuli [23, 24]. In cancer related tissues, particularly in human papillomavirus (HPV) 16 negative tumors, Furniss et al suggested that decreasing level of global DNA methylation was induced by cigarettes usage[25]. Ostrow et al. indicated that the tumor suppressor gene NISCH was hypermethylated among smokers[26]. Epigenome-wide association studies (EWAS) is an examination of epigenome-wide markers in many individuals to scan for epigenetic markers associated with a trait [27]. EWAS of cigarette smoking have repeatedly and consistently identified genomic loci harboring smoking-related DNAm sites across populations. From extensive literature review, 10 EWAS focused on DNA methylation association with cigarette smoking (Table 1). Zeilinger et al.[28], Shenker et al. [29], Monick et al. [30], Dogan et al.[31], Elliott et al. [32], and Besingi et al. [33] all reported cg05575921 located on Aryl-Hydrocarbon Receptor Repressor (AHRR) as the most significant site. Zeilinger et al.[28], Wan et al.[34], Sun et al.[35], Breitling et al.[36], Shenker et al.[29], and Elliott et al.[32] reported the factor II receptor-like 3 gene (F2RL3) methylation was significantly related to smoking. More researches have been done to explore the effect of smoking on maternal health. Using the whole blood, Harlid [37]and Markunas [38] found differences between smoking effect on adults and maternal smoking effect on newborns. All the major EWAS about smoking effects used two platforms. One is Infinium HumanMethylation27 (27K) BeadChip and another one is Infinium HumanMethylation450 (450K) BeadChip. The latter one interrogates more than

485,000 methylation sites per sample at single nucleotide resolution [39]. Sun et al.[35] and Dogan et al.[31] studied the smoking-related DNA methylation in African American populations, and found cg19859270 in G protein-coupled receptor 15 (GPR15), cg21566642 in alkaline phosphatase, placental-like 2 (ALPPL2) and cg03991871 in AHRR as significant CpG sites in addition to cg03636183 in F2RL3.

Studies also showed that epigenetic modification such as DNA methylation induced diseases, especially the atherosclerosis[40]. For example, genes with anti-proliferative function are destroyed by DNA methylation in the process of atherosclerosis. P53 as a tumor suppressor protein, inhibits cell proliferation to exert its anti-atherosclerotic effect [40, 41]. Additionally, the finding of methylation of gene estrogen receptor-α (ERα) linked the atherosclerosis with the CpG island methylation and further confirmed the ERα control important cell activity by estrogen activation[42].

METHODS

Data source

The DNA methylation data is part of the Emory Twin Study (ETS) of cardiovascular disease, and examined in pairs at the Emory University General Clinical Research Center between 2002 and 2010. The ETS contains 307 middle-aged male (born between 1946 and 1956) Caucasians monozygotic and dizygotic twin pairs in the Vietnam Era Twin (VET) Registry [43-45]. The VET Registry is one of the largest twin registries in the United States [45]. Genomic DNA samples were extracted from peripheral blood of 140 monozygotic twins, and were epityped using Illumina 450K methylation chip in the

study. Zygosity screening was completed by DNA analysis. ETS was approved by the Emory Institutional Review Board, and informed consents had been signed by all participants.

DNA methylation methods

The EZ DNA Methylation Kit (Zymo Research, Orange CA) is based on bisulfite-conversion, which chemically converts unmethylated cytosine to uracil. It converted 0.5 μg of genomic DNA per sample from peripheral blood leukocytes (PBLs) and the genomic DNA were amplified, fragmented by enzyme, and finally purified. Samples with hybridization in 12 batches to the BeadChip. We used BeadChip with locus specific DNA oligomers to hybridize the genomic samples. Each bead site was fluorescently stained, scanned, and assessed for fluorescence intensities. Then BeadStudio software was used to generate β-values, which represented the fluorescence intensity of the methylated and unmethylated sites.

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

In the numerator, maximum of methylated intensity and zero was chosen to avoid possible negative value caused by global subtraction. A constant value α was used as compensate of small absolute values of methylated and unmethylated intensities [46]. The β-values were continuous variables ranging from 0 to 1. We also generated quantile-normalized β-values to adjust for known technical shift between methylation signals in different probe categories [47-49]. According to the results, there was no sample with control probe values greater than 4 standard deviations from their mean

values. Methylation sites were excluded from analyses if they overlapped with SNPs, or were not uniquely mapped to the reference genome [50]. The final methylome-wide association analyses included 409,968 autosomal CpG sites from 134 participants with 67 monozygotic twin pairs.

Phenotypic Measurements

The twins were given the same diet the night before the test without overnight diet. The twins were restricted from smoking the night before the test. All the measurements were performed at the same time in the morning. In addition, after obtaining and examining medical records and physical exam data, all medications were stopped 24 hours before the test and biochemical assays were in the same analytical circle.

The physical exam data of height and weight were used to calculate the body mass index (BMI). The participants were recorded as current smoker (any number of cigarettes) versus never or past smoker. The carotid intima-media thickness (IMT) was measure by ultra sound to refer to the thickness of carotid wall. According to Houseman et al., DNA methylation pattern in cells may differ as a result of inconsistency of leukocyte subtypes. Therefore six cell types variables, including granulocyte, monocyte, natural killer cells (NK), B cell, CD4+ and CD8+ T cells, were calculated by the method in Houseman's essay [51] , and included as covariates in multiple regression models.

Statistical Analysis

We use the Carotid intima-media thickness (IMT) to measure the development of atherosclerosis. Age, BMI and leukocyte subtypes are included as covariates in the

model. Smoking status (current vs. non-current), as a dichotomous variable, will be involved in the analysis process together with other covariates.

First, , as reported in Pan Du et al. [52], the beta value has severe heteroscedasticity at extreme values and suggests compression at the tail of the distribution. Therefore, we use following formula to transform beta values to m values to generate m-values which is approximately homoscedastic.

$$M - value = log2\left(\frac{\beta}{1-\beta}\right) \quad (1)$$

Second, to replicate the effect of smoking on IMT, we conducted linear mixed effect model treating co-twin as a random effect.

$$IMT = \beta_0 + \beta_1 * SMK + \beta_2 * AGE + \beta_3 * BMI + E \quad (2)$$

In this model, SMK represents whether the sample was current smoker, AGE denotes the age in years, BMI is the body mass index, and E suggests the error term.

Third, based on the fact that smoking as the risk factor of DNA methylation, we primarily focused on the association between the smoking and DNA methylation. Linear mixed effect model (3) was implemented to adjust for random effect of Beadchips and twins.

$$M - value = \beta_0 + \beta_1 * SMK + \beta_2 * BMI + \beta_3 * AGE + \beta_4 * Gran + \beta_5 * NK + \beta_6 * B - cell + \beta_7 * CD4 + \beta_8 * CD8 + E \quad (3)$$

In model (3) and other models later introduced, cell traits were included. Gran is granulocyte, NK is NK killer cell, CD4 is CD4+ T-cell, CD8 is CD8+ T-cell.

In order to explore the effect of DNA methylation level on IMT, we conducted the unconditional linear mixed effect model with the same random effects as the model 2.

$$IMT = \beta_0 + \beta_1 * M - value + \beta_2 * BMI + \beta_3 * AGE + \beta_4 * Gran + \beta_5 * NK + \beta_6 * B - cell + \beta_7 * CD4 + \beta_8 * CD8 + E \quad (4)$$

Next, we aimed to assess the independent epigenetic effect on IMT by adjusting the smoking status in the multiple-regression framework. The following linear mixed effect model 5 were implemented and performed.

$$IMT = \beta_0 + \beta_1 * M - value + \beta_2 * SMK + \beta_3 * BMI + \beta_4 * AGE + \beta_5 * Gran + \beta_6 * NK + \beta_7 * B - cell + \beta_8 * CD4 + \beta_9 * CD8 + E \quad (5)$$

All statistical analyses were performed in the R statistical environment version 3.1.2 (http://www.r-project.org/). R package *nlme* was used to implement linear mixed effect model.

RESULTS

Among 134 subjects, 47 (34.3%) were current smoker and 87 (65.7%) were not smoking at the time of tests. The minimum intima-media thickness is 0.6675 mm, while the maximum is 1.22 mm. The mean thickness of intima media is 0.7671 mm. The linear mixed effect model (1) showed that current smoking status as potential risk factors of atherosclerosis with effect size of 0.035 and p-value of 0.095.

Based on the publications mentioned in Table 1, criteria had been established to select candidate loci. Since the study done by Zeilinger et al. [28] has the largest sample size

(2,272) and better platform that covered more loci (Infinium 450K BeadChip), we selected the most significant CpG sites in loci located in the same region. Then we used threshold of p-value smaller than 10^{-5} to select loci from a second publication and merge in table 2. A total of 17 cytosine-phosphate-guanine (CpG) sites were selected as potential mediation sites and their effect size, standard error, t statistics and p-value were listed in Table 3.

As shown in Table 3, all the CpG sites that previously reported were successfully replicated (alpha level of 0.1) in our twin sample except cg01500140 in LIM2. In the meantime, another two sites (cg03274391, cg09837977 in LRRN3) had p values marginally significant (p-value less than 0.1). Using model (4), 7 CpG sites were significantly associated with IMT (alpha level of 0.05). In the model (5) conditioned on smoking status, among these 8 sites that were significant in the first two models, only two of them (cg03991871 in AHRR and cg06126421 in 6p21.33) were significant in both model (3) and model (5).

DISCUSSION

In summary, we found that cigarette smoking was associated with significant change in DNA methylation, and the top smoking-related DNA methylation sites were replicated in male Caucasian twins in the Emory Twin Study (ETS).

The most significant loci associated with IMT after adjustment for smoking are CpG site cg03991871 located in Aryl-Hydrocarbon Receptor Repressor (AHRR) gene and CpG site cg06126421 located on 6p21.33. As a feedback inhibition modulator of the aryl

hydrocarbon receptor, AHRR binds with AHR nuclear translocator or with xenobiotic response elements in AHR regulated genes [53], and finally affects AHR in oncogenesis and altered immune function [54]. Atherosclerosis occurs while immune competent cells in lesions producing mainly pro-inflammatory cytokines [55]. As a result of abnormal immune function, dead cells and oxidized forms of low density lipoproteins (oxLDL) are collected in atherosclerotic plaques [56, 57]. Therefore, our findings may suggest a possible role of the altered DNA methylation level on AHRR in atherosclerosis. As shown in Figure 1 generated by NCBI epigenomics browser (http://www.ncbi.nlm.nih.gov/epigenomics), the second locus 6p21.33 (chr6: 30,720,080) is located on a non-annotated region on the short arm (i.e. p arm) of chromosome 6, and maps onto a DNase I hypersensitivity site and transcription factor binding site, associated with an H3K27 acetylated chromatin site[29, 58]. The closest gene is the immediate early response 3 (IER3) gene which protects cells from Fas- or tumor necrosis factor type alpha-induced apoptosis (provided by RefSeq, Jul 2008).

In Table 2, more than 30% of twins are current smokers. The relatively comparative size of current smoker and non-smoker enforced the validity of the population. However, each group only contained 67 subjects which may possibly affect the accuracy of the estimates and lead to larger standard deviation.

Due to relatively small sample size, the association between smoking status and IMT is only marginally significant (p value = 0.095). However, the coefficient of 0.035 suggested a positive direction of the effect of smoking on IMT. In other words, smoking increases carotid intima-media thickness.

In Table 3, the directions of some epigenetic associations were different between two models with and without adjustment of smoking status. The negative effect in model 3 and 5 suggested that higher methylation level causes lower IMT, while positive effect in model 4 showed higher methylation level causes higher IMT. Possible explanation would be: when conditioned on smoking status, the DNA methylation level had reverse effect on IMT. However, smoking affected IMT in a larger scale through mechanisms other than DNA methylation. Therefore, in Model 4, the effect size was positive.

In terms of effect sizes in Model 4, 1 unit change of DNA methylation M-value is associated with 0.003mm to 0.005mm change of IMT, which indicating weak relationship. However, after conditioned on smoking status, the effect sizes were larger. Among the two significant sites, 1 unit increase of M-value of DNA methylation on cg03991871 is associated with 0.05mm decrease of IMT; 1 unite increase of M-value of DNA methylation on cg06126421 is associated with 0.06mm decrease of IMT.

Strength

First, the candidate loci are from multiple studies using whole blood as the source tissue. Since different tissues have different DNA methylation profile, it is critical to investigate the same tissue for replication purpose. Second, we ranked the reported associations based on the sample size of each study. With the largest sample size, study conducted by Zeilinger et al. [28] contributed the majority of smoking-related DNA methylation loci with relatively accurate results. Third, we adjusted for cell traits (i.e. computed proportions of leukocyte subtypes) to control potential confounding effect

due to the shift of leukocyte populations, including granulocyte, monocyte, natural killer cells (NK), B cell, CD4+ and CD8+ T cells. Fourth, since the samples are monozygotic twins, the co-twins within one twin pair share genetic profile and other familial environment. Therefore, we implemented a linear mixed model (co-twin as a random effect) to correct for the correlation structure between co-twins.

Limitation

Due to the limited resources, we only have epigenome-wide data of 67 pairs of monozygotic twins. Insufficient sample size may result in reduced power to detect significant results. In addition, small sample size may not represent the whole population.

As reported by Houseman et al.[51], DNA methylation levels differ by tissues/cell types. In terms of understanding the effect of DNA methylation on atherosclerosis, peripheral blood may only capture the epigenetic mechanism involving leukocytes but not epigenetic modifications in heart or aterial tissue, which are essential to understand the complete pathophysiology of atherosclerosis.

In addition, the current study uses a cross-sectional design. Thus, we can only estimate the DNA methylation level at the time of examination. The change of DNA methylation and its effect on the relation between smoking and the progression of atherosclerosis cannot be investigated. Moreover, since intima-media thickness can be affected by many mechanisms including environmental factor, and genetic factors, we are not able to fully interpret the functional connection between the DNA methylation and

atherosclerosis in current study. However, the possible causation relationship between DNA methylation and IMT is the change of DNA methylation level up-regulates atherosclerosis-susceptible genes and down-regulates atherosclerosis-protective genes[42]. Also, the change of DNA methylation may be associated with chronic inflammation, higher cholesterol level, and hypertension which are all risk factors of atherosclerosis [14].

FUTURE DIRECTION

In the future, more functional studies are needed to explore the molecular mechanism of the mediation effect of DNA methylation linking smoking exposure and atherosclerosis. Based on the two CpG sites we found significantly related to IMT, we can study their coded protein to further understand their functional roles in specific causal pathways of atherosclerosis. Once the results could be replicated in other studies, we may consider including these two CpG sites to predict the risk of atherosclerosis, and identify potential patients clinically.

DNA methylation data from other tissue such as heart and artery would improve the mechanistic understanding of the effect of DNA methylation on atherosclerosis. Ideally, we would study sorted cell types to further explore cell type and tissue-specific effects. For a more concrete conclusion, we plan to genotype more samples in the cross sectional study to improve the reliability of the conclusion. Longitudinal data can also help to understand the epigenetic effect on atherosclerosis progression and how will the DNA methylation level change over time. More importantly, we would be able to

rule out potential reverse causation between DNA methylation and atherosclerosis using the longitudinal design. Given that multiple environmental factors may modify epigenetic profile, we will investigate potential epigenetic mediation of other CVD risk factors such as the change of diet, change of behavioral factors, aging and social economic status, using this study of smoking-related DNA methylation as an example.

REFERENCES

1. Peto, R., et al., *Mortality from smoking worldwide.* Br Med Bull, 1996. **52**(1): p. 12-21.
2. Gilman, S.L. and X. Zhou, *Smoke : a global history of smoking*. 2004, London: Reaktion Books. 408 p.
3. Jamal, A., et al., *Current cigarette smoking among adults--United States, 2005-2013.* MMWR Morb Mortal Wkly Rep, 2014. **63**(47): p. 1108-12.
4. Howard, G., et al., *Active and passive smoking are associated with increased carotid wall thickness. The Atherosclerosis Risk in Communities Study.* Arch Intern Med, 1994. **154**(11): p. 1277-82.
5. Glantz, S.A. and W.W. Parmley, *Passive smoking and heart disease. Epidemiology, physiology, and biochemistry.* Circulation, 1991. **83**(1): p. 1-12.
6. Duriez, Q., F. Crivello, and B. Mazoyer, *Sex-related and tissue-specific effects of tobacco smoking on brain atrophy: assessment in a large longitudinal cohort of healthy elderly.* Front Aging Neurosci, 2014. **6**: p. 299.
7. Doll, R. and R. Peto, *Mortality in relation to smoking: 20 years' observations on male British doctors.* Br Med J, 1976. **2**(6051): p. 1525-36.
8. Centers for Disease, C., *The Surgeon General's 1989 Report on Reducing the Health Consequences of Smoking: 25 Years of Progress.* MMWR Morb Mortal Wkly Rep, 1989. **38 Suppl 2**: p. 1-32.
9. Koop, C.E. and J. Luoto, *"The Health Consequences of Smoking: Cancer," overview of a report of the Surgeon General.* Public Health Rep, 1982. **97**(4): p. 318-24.
10. Hjermann, I., et al., *Effect of diet and smoking intervention on the incidence of coronary heart disease. Report from the Oslo Study Group of a randomised trial in healthy men.* Lancet, 1981. **2**(8259): p. 1303-10.
11. Celermajer, D.S., et al., *Non-invasive detection of endothelial dysfunction in children and adults at risk of atherosclerosis.* Lancet, 1992. **340**(8828): p. 1111-5.
12. Leone, A., *Relationship between cigarette smoking and other coronary risk factors in atherosclerosis: risk of cardiovascular disease and preventive measures.* Curr Pharm Des, 2003. **9**(29): p. 2417-23.
13. Natori, S., et al., *Cardiovascular function in multi-ethnic study of atherosclerosis: normal values by age, sex, and ethnicity.* AJR Am J Roentgenol, 2006. **186**(6 Suppl 2): p. S357-65.
14. Garcia, J.H. and H. Khang-Loon, *Carotid atherosclerosis. Definition, pathogenesis, and clinical significance.* Neuroimaging Clin N Am, 1996. **6**(4): p. 801-10.
15. Stout, R.W., *Diabetes and atherosclerosis.* Biomed Pharmacother, 1993. **47**(1): p. 1-2.
16. Simon, A., et al., *Intima-media thickness: a new tool for diagnosis and treatment of cardiovascular risk.* J Hypertens, 2002. **20**(2): p. 159-69.
17. Bots, M.L. and D.E. Grobbee, *Intima media thickness as a surrogate marker for generalised atherosclerosis.* Cardiovasc Drugs Ther, 2002. **16**(4): p. 341-51.
18. de Groot, E., et al., *Measurement of arterial wall thickness as a surrogate marker for atherosclerosis.* Circulation, 2004. **109**(23 Suppl 1): p. III33-8.
19. Simon, A. and G. Chironi, *The relationship between carotid intima-media thickness and coronary atherosclerosis revisited.* European Heart Journal, 2007. **28**(17): p. 2049-2050.
20. Howard, G., et al., *Cigarette smoking and progression of atherosclerosis: The Atherosclerosis Risk in Communities (ARIC) Study.* JAMA, 1998. **279**(2): p. 119-24.
21. Dodge, J.E., et al., *De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation.* Gene, 2002. **289**(1-2): p. 41-8.

22. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.* Nat Genet, 2003. **33 Suppl**: p. 245-54.
23. Apostolou, E. and K. Hochedlinger, *Chromatin dynamics during cellular reprogramming.* Nature, 2013. **502**(7472): p. 462-71.
24. Waterland, R.A. and R.L. Jirtle, *Transposable elements: targets for early nutritional effects on epigenetic gene regulation.* Mol Cell Biol, 2003. **23**(15): p. 5293-300.
25. Furniss, C.S., et al., *Line region hypomethylation is associated with lifestyle and differs by human papillomavirus status in head and neck squamous cell carcinomas.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(4): p. 966-71.
26. Ostrow, K.L., et al., *Cigarette smoke induces methylation of the tumor suppressor gene NISCH.* Epigenetics, 2013. **8**(4): p. 383-8.
27. Sun, Y.V., *The Influences of Genetic and Environmental Factors on Methylome-wide Association Studies for Human Diseases.* Curr Genet Med Rep, 2014. **2**(4): p. 261-270.
28. Zeilinger, S., et al., *Tobacco smoking leads to extensive genome-wide changes in DNA methylation.* PLoS One, 2013. **8**(5): p. e63812.
29. Shenker, N.S., et al., *Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking.* Hum Mol Genet, 2013. **22**(5): p. 843-51.
30. Monick, M.M., et al., *Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers.* Am J Med Genet B Neuropsychiatr Genet, 2012. **159B**(2): p. 141-51.
31. Dogan, M.V., et al., *The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women.* BMC Genomics, 2014. **15**: p. 151.
32. Elliott, H.R., et al., *Differences in smoking associated DNA methylation patterns in South Asians and Europeans.* Clin Epigenetics, 2014. **6**(1): p. 4.
33. Besingi, W. and A. Johansson, *Smoke-related DNA methylation changes in the etiology of human disease.* Hum Mol Genet, 2014. **23**(9): p. 2290-7.
34. Wan, E.S., et al., *Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome.* Hum Mol Genet, 2012. **21**(13): p. 3073-82.
35. Sun, Y.V., et al., *Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans.* Hum Genet, 2013. **132**(9): p. 1027-37.
36. Breitling, L.P., et al., *Tobacco-smoking-related differential DNA methylation: 27K discovery and replication.* Am J Hum Genet, 2011. **88**(4): p. 450-7.
37. Harlid, S., et al., *CpG sites associated with cigarette smoking: analysis of epigenome-wide data from the sister study.* Environ Health Perspect, 2014. **122**(7): p. 673-8.
38. Markunas, C.A., et al., *Identification of DNA methylation changes in newborns related to maternal smoking during pregnancy.* Environ Health Perspect, 2014. **122**(10): p. 1147-53.
39. Clark, C., et al., *A Comparison of the Whole Genome Approach of MeDIP-Seq to the Targeted Approach of the Infinium HumanMethylation450 BeadChip (R) for Methylome Profiling.* Plos One, 2012. **7**(11).
40. Guevara, N.V., et al., *The absence of p53 accelerates atherosclerosis by increasing cell proliferation in vivo.* Nat Med, 1999. **5**(3): p. 335-9.
41. Rodriguez-Campos, A., et al., *Mitogen-induced p53 downregulation precedes vascular smooth muscle cell migration from healthy tunica media and proliferation.* Arterioscler Thromb Vasc Biol, 2001. **21**(2): p. 214-9.
42. Dong, C., W. Yoon, and P.J. Goldschmidt-Clermont, *DNA methylation and atherosclerosis.* J Nutr, 2002. **132**(8 Suppl): p. 2406S-2409S.

43. Vaccarino, V., et al., *Association of major depressive disorder with serum myeloperoxidase and other markers of inflammation: a twin study.* Biol Psychiatry, 2008. **64**(6): p. 476-83.
44. Vaccarino, V., et al., *Depressive symptoms and heart rate variability: evidence for a shared genetic substrate in a study of twins.* Psychosom Med, 2008. **70**(6): p. 628-36.
45. Goldberg, J., et al., *The Vietnam Era Twin Registry.* Twin Res, 2002. **5**(5): p. 476-81.
46. Bibikova, M., et al., *High-throughput DNA methylation profiling using universal bead arrays.* Genome Res, 2006. **16**(3): p. 383-93.
47. Dedeurwaerder, S., et al., *Evaluation of the Infinium Methylation 450K technology.* Epigenomics, 2011. **3**(6): p. 771-84.
48. Wang, D., et al., *IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data.* Bioinformatics, 2012. **28**(5): p. 729-30.
49. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.* Bioinformatics, 2014. **30**(10): p. 1363-9.
50. Chen, Y.A., et al., *Sequence overlap between autosomal and sex-linked probes on the Illumina HumanMethylation27 microarray.* Genomics, 2011. **97**(4): p. 214-22.
51. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.
52. Du, P., et al., *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.* BMC Bioinformatics, 2010. **11**: p. 587.
53. Haarmann-Stemmann, T., et al., *Analysis of the transcriptional regulation and molecular function of the aryl hydrocarbon receptor repressor in human cell lines.* Drug Metab Dispos, 2007. **35**(12): p. 2262-9.
54. Opitz, C.A., et al., *An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor.* Nature, 2011. **478**(7368): p. 197-203.
55. Frostegard, J., et al., *Cytokine expression in advanced human atherosclerotic plaques: dominance of pro-inflammatory (Th1) and macrophage-stimulating cytokines.* Atherosclerosis, 1999. **145**(1): p. 33-43.
56. Tabas, I., K.J. Williams, and J. Boren, *Subendothelial lipoprotein retention as the initiating process in atherosclerosis: update and therapeutic implications.* Circulation, 2007. **116**(16): p. 1832-44.
57. Camejo, G., et al., *Characterization and properties of a lipoprotein-complexing proteoglycan from human aorta.* Atherosclerosis, 1980. **35**(3): p. 307-20.
58. Guida, F., et al., *Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation.* Hum Mol Genet, 2015. **24**(8): p. 2349-59.

TABLES

Table 1 Smoking Effect on DNA Methylation Studies Summary [27]

Race	Source of DNA	Plateform	Sample Size	Study
European Ancestry	Peripheral Blood	450K	1,793+479	Zeilinger [28]
			374+180	Shenker [29]
			432	Besingi [33]
		27K	1,085+369	Wan [34]
			177+316	Breitling [36]
	Whole Blood	450K	908+200	Harlid [37]
	LCL[a]+PAM[b]	450K	119+19	Monick [31]
African Ancestry	Peripheral Blood	450K	111	Dogan [31]
		27K+450K	972+239	Sun [35]
South and East Asian Ancestry	Peripheral Blood	450K	192	Elliott [32]

a. Lymphoblastoid cell line; b. Pulmonary alveolar macrophages

Table 2. Smoking Related CpG Sites Summary from Publications

CpG Site	Chromo-some	Gene	Zeilinger [28]	Dogan [31]	Harlid [37]	Shenker [29]	Elliott [32]	Sun [35]	Wan [34]	Besingi [33]	others
cg09935388	1	GFI1	3.27E-24								
cg15542713	1	HIVEP3	2.54E-08								
cg04885881	1	X[a]	1.35E-28								
cg21566642	2	ALPPL2[b]	6.90E-138	2.99E-10		3.73E-13	3.00E-27	4.84E-11		2.57E-35	
cg19859270	3	GPR15		1.19E-15				6.04E-16	1.92E-22		
cg02657160	3	CPOX		3.97E-12	1.10E-11						
cg03274391	3	X[a]	1.67E-25								
cg15417641	3	CACNA1D	1.25E-16								
cg05575921	5	AHRR	2.54E-182	6.17E-19		2.31E-15	6.10E-59	9.51E-19		7.19E-70	1.34E-06[a]
cg03991871	5	AHRR	4.84E-41								
cg06126421	6	6p21.33		2.91E-10		4.96E-11				5.51E-32	
cg14753356	6	X[a]	8.14E-24								
cg00931843	6	TIAM2	1.69E-10								
cg22132788	7	MYO1G	1.99E-34				3.10E-17				
cg09837977	7	LRRN3							3.88E-11		
cg03636183	19	F2RL3	2.42E-80		2.00E-07	8.38E-11	1.00E-17	6.90E-42	2.36E-46		2.68E-31[b]
cg01500140	19	LIM2							1.88E-10		

a. Monick et al.[30]
b. Breitling et al.[36]

Table 3. Smoking Related CpG Sites Summary

CpG Site	Chromo -some	Gene	Smoking Status				IMT				IMT (Cond. Smk)			
			CSmoke_E	CSmoke_SE	CSmoke_t	CSmoke_P	cg_E	cg_SE	cg_T	cg_P	cg_E	cg_SE	cg_t	cg_P
cg09935388	1	GFI1	-0.5110	0.0812	-6.2939	<0.001	0.0043	0.0020	2.1608	0.0350	-0.0229	0.0220	-1.0391	0.3033
cg15542713	1	HIVEP3	0.2804	0.0797	3.5173	<0.001	0.0036	0.0020	1.8176	0.0745	0.0017	0.0233	0.0707	0.9439
cg04885881	1	X[a]	-0.2492	0.0564	-4.4175	<0.001	0.0035	0.0020	1.7221	0.0906	0.0307	0.0325	0.9438	0.3494
cg21566642	2	ALPPL2[b]	-0.6299	0.0796	-7.9141	<0.001	0.0039	0.0020	1.9322	0.0584	0.0039	0.0234	0.1687	0.8666
cg19859270	3	GPR15	-0.1837	0.0470	-3.9095	<0.001	0.0038	0.0020	1.9036	0.0621	-0.0083	0.0390	-0.2121	0.8328
cg02657160	3	CPOX	-0.1448	0.0449	-3.2291	<0.001	0.0037	0.0020	1.8740	0.0661	-0.0085	0.0374	-0.2261	0.8219
cg03274391	3	X[a]	0.1593	0.0889	1.7927	0.0781	0.0037	0.0019	1.9120	0.0610	0.0143	0.0171	0.8356	0.4070
cg15417641	3	CACNA1D	0.3455	0.0693	4.9837	<0.001	0.0039	0.0020	2.0084	0.0494	0.0186	0.0219	0.8488	0.3997
cg05575921	5	AHRR	-1.2714	0.1137	-11.1857	<0.001	0.0044	0.0020	2.1905	0.0327	-0.0140	0.0162	-0.8685	0.3889
cg03991871	5	AHRR	-0.3020	0.0792	-3.8123	<0.001	0.0046	0.0019	2.3801	0.0207	-0.0503	0.0221	-2.2776	0.0267
cg06126421	6	6p21.33	-0.4512	0.0771	-5.8499	<0.001	0.0046	0.0019	2.4064	0.0194	-0.0604	0.0222	-2.7168	0.0088
cg14753356	6	X[a]	-0.1939	0.0389	-4.9866	<0.001	0.0038	0.0020	1.9547	0.0556	-0.0345	0.0477	-0.7229	0.4728
cg00931843	6	TIAM2	0.2278	0.0651	3.4992	<0.001	0.0042	0.0019	2.1694	0.0343	0.0487	0.0272	1.7891	0.0791
cg22132788	7	MYO1G	0.6419	0.1401	4.5830	<0.001	0.0047	0.0020	2.3913	0.0202	0.0252	0.0127	1.9848	0.0522
cg09837977	7	LRRN3	-0.0595	0.0355	-1.6784	0.0985	0.0037	0.0020	1.8445	0.0704	-0.0022	0.0491	-0.0441	0.9650
cg03636183	19	F2RL3	-0.5939	0.0820	-7.2446	<0.001	0.0049	0.0020	2.4238	0.0186	-0.0393	0.0221	-1.7806	0.0805
cg01500140	19	LIM2	0.0347	0.0308	1.1298	0.2631	0.0038	0.0020	1.9090	0.0614	0.0243	0.0539	0.4509	0.6538

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

b. According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites, but SNPs within the same region (shore of a CpG Island) have a predicted function on the ALPPL2 gene, which is located several kb apart from this CpG island.

FIGURES

Figure 1. Genomic Features of 6p21.33