TITLE

Smoking is associated to DNA methylation in atherosclerotic carotid lesions.

RUNNING TITLE: DNA methylation in carotid lesions of smokers.

AUTHORS

Marten A. Siemelink, MD PhD[1*], Sander W. van der Laan, PhD[1*], Saskia Haitjema, MD PhD[2], Ian D. van Koeverden, MD[1], Jacco Schaap BSc[1], Marian Wesseling, MSc[1], Saskia C.A. de Jager, PhD[1,3], Michal Mokry, PhD[4], Maarten van Iterson, PhD[5], Koen F. Dekkers, MSc[5], René. Luijk, MSc[5,6], Hassan Foroughi Asl, MSc[7], Tom Michoel, PhD[8], Johan L.M. Björkegren, PhD[7], Einar Aavik, PhD[9], Seppo Ylä-Herttuala, PhD[9,10], Gert Jan de Borst, MD PhD[11], Folkert W. Asselbergs, MD PhD[12,13,14], Hassan el Azzouzi, PhD[12], Hester M. den Ruijter, PhD[1], Bas T. Heijmans, PhD[5], Gerard Pasterkamp, MD PhD[2].

* these authors contributed equally

AFFILIATIONS

[1]Laboratory of Experimental Cardiology, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands;

[2]Laboratory of Clinical Chemistry and Hematology, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands;

[3]Laboratory of Translational Immunology, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands;

[4]Division Pediatrics, Wilhelmina Children's Hospital, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands;

[5]Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands;

[6]Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, the Netherlands;

[7]Cardiovascular Genomics Group, Division of Vascular Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden;

[8]Division of Genetics and Genomics, the Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom;

[9]Department of Biotechnology and Molecular Medicine, A.I. Virtanen Institute, University of Eastern Finland, Kuopio, Finland;

[10]Science Service Center and Gene Therapy Unit, Kuopio University Hospital, Kuopio, Finland;

[11]Department of Vascular Surgery, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands;

[12]Department of Cardiology, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands;

[13]Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, the Netherlands;

[14]Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom.

CORRESPONDING AUTHORS

Prof. dr. Gerard Pasterkamp

Laboratory of Clinical Chemistry, University Medical Center Utrecht, Utrecht University

Heidelberglaan 100, 3508 GA, Utrecht, the Netherlands

Telephone: +31 (0) 88 75 571 55

Fax: +31 (0)30 252 26 93

E-mail: g.pasterkamp@umcutrecht.nl

dr. Sander W. van der Laan

Laboratory of Experimental Cardiology, University Medical Center Utrecht, Utrecht University

Heidelberglaan 100, 3508 GA, Utrecht, the Netherlands

1

Telephone: +31 (0) 88 75 676 96

Fax: +31 (0)30 252 26 93

E-mail: s.w.vanderlaan-2@umcutrecht.nl

MANUSCRIPT CONTENTS

Manuscript of 4,044 words, 5 tables, 3 figures and Supplemental Material.

JOURNAL SUBJECT TERMS

Epigenetics, Genetics, Atherosclerosis, Cardiovascular disease

ABSTRACT

Background: Tobacco smoking is a major risk factor for atherosclerotic disease and has been associated with DNA methylation (DNAm) changes in blood cells. However, whether smoking influences DNAm in the diseased vascular wall is unknown but may prove crucial in understanding the pathophysiology of atherosclerosis. In the this study we associated current tobacco smoking to epigenome-wide DNAm in atherosclerotic plaques from patients undergoing carotid endarterectomy (CEA).

Methods: DNAm at commonly methylated sites (CpGs) was assessed in atherosclerotic plaque samples and peripheral blood samples from 485 CEA patients. We tested the association of current tobacco smoking with DNAm corrected for age, and sex. To control for bias and inflation due to cellular heterogeneity we applied a Bayesian method to estimate an empirical null distribution as implemented by the R package `bacon`. Replication of the smoking associated methylated CpGs in atherosclerotic plaques was executed in a second sample of 190 CEA patients, and results were meta-analyzed using a fixed-effects model.

Results: Tobacco smoking was significantly associated to differential DNAm in atherosclerotic lesions of 4 CpGs (FDR < 0.05) mapped to 2 different genes (*AHRR*, *ITPK1*), and 17 CpGs mapped to 8 genes and RNAs in blood. The strongest associations were found for CpGs mapped to the gene *AHRR*, a repressor of the aryl hydrocarbon receptor transcription factor involved in xenobiotic detoxification. One of these methylated CpGs were found to be regulated by local genetic variation.

Conclusions: The risk factor tobacco smoking associates with DNA methylation at multiple loci in carotid atherosclerotic lesions. These observations support further investigation of the relationship between risk factors and epigenetic regulation in atherosclerotic disease.

INTRODUCTION

Tobacco smoking is a major risk factor for the development of atherosclerosis and subsequent cardiovascular disease (CVD), such as myocardial infarction and stroke. Tobacco smoke contains over 5,000 toxic chemicals which may jointly contribute to CVD risk[1]. Smoking activates the immune system, facilitates pro-atherogenic lipid profiles, and induces a prothrombotic state[2,3]. Moreover, smoking affects the vascular wall, leading to endothelial dysfunction and atherosclerosis[4]. Histological examination of plaques of smokers have shown increased atheroma, decreased fibrous volume[5], more plaque hemorrhage[6], and increased inflammation and tissue destruction[7]. All these changes contribute to a plaque composition that is more vulnerable to rupture and more likely to cause cardiovascular events.

Yet, a detailed understanding of the pathophysiological mechanisms underlying these changes remains elusive. Such an understanding may help to identify patients at increased risk due to smoking and may contribute to cessation and preventative treatment strategies. Of equal importance, it may show common pathophysiological pathways of atherosclerosis, shared by multiple risk factors, which may be important for identification of new drug targets.

Large-scale genetic association studies (GWAS) have proven instrumental in the investigation of many cardiovascular risk factors and susceptibility to CVD[8]. Smoking has been shown to directly impact CVD risk[2,3], and indirectly by modulating the effect of genetic variants on cardiovascular risk factors[9–13]. Genome-wide genetic studies of smoking have mainly focused on behavioral traits of smoking[14]. Identification of the pathophysiology caused by environmental exposures, such as smoking-induced cardiovascular risk, may require other approaches.

Epigenetics refers to the study of gene expression modifications not caused by changes in the DNA sequence but rather external factors[15]. Epigenetic alterations can be influenced by age, environment, and lifestyle, and aberrant modifications can lead to diseases like cancer and neurodevelopmental disorders. DNA methylation (DNAm) is a key mechanism of epigenetic regulation, whereby a methyl group is added to the cytosine (C) or adenine (A)

nucleotides in the DNA molecule; in humans, the most common DNA methylation is at cytosine in CpG dinucleotides.

DNAm in blood cells has been associated to cardiovascular risk factors such as body mass index (BMI)[16] and blood lipid levels[17]. Chemicals in tobacco smoke may change gene expression through DNAm, either adaptive or pathologic. Such epigenetic changes have predominantly been shown in circulating cells, in which CpGs were associated to smoking as identified through epigenome-wide association studies (EWAS)[18–25]. Conceivably, the most important insights in vascular pathology may be obtained by scrutinizing the effect of tobacco smoking on DNAm in the vascular lesion itself. To our knowledge, this has not been studied to this date.

In the current study, we performed a two-stage EWAS of tobacco smoking in carotid atherosclerotic plaques of patients undergoing carotid endarterectomy (CEA), reporting 4 loci near *AHRR* and *ITPK1* that are differentially methylated in plaques. Together our findings point to vascular epigenetic mechanisms of smoking-induced cardiovascular disease.

MATERIAL AND METHODS

This study complies with the Declaration of Helsinki and all participants provided informed consent. The medical ethical committees of the respective hospitals approved these studies. Detailed *Material and Methods* are available in the **Supplemental Material**.

The data, analytic methods, and study materials will be made available to other researchers for purposes of reproducing the results or replicating the procedure. The raw omics data are available through the European Genome-Phenome Archive (EGA). The main scripts used for the quality control and the (meta-)analysis of the data are available through GitHub (https://github.com/swvanderlaan/publications under doi:10.5281/zenodo.1069531).

RESULTS

We performed a two-stage epigenome-wide association study of plaque-derived DNA methylation with current tobacco smoking in carotid endarterectomy patients from the Athero-Express Biobank Study (AEMS450K1 discovery study and AEMS450K2 replication study, **Table 1**, **Supplemental Figure 1)**. In the discovery study 10 CpGs across 6 genes (**Table 2**, **Figure 1A**, **Supplemental Figure 2**) were associated to tobacco smoking (at $p \leq 1.13 \times 10^{-6}$ (FDR $\leq$ 0.05)). To assess the validity of these associations we performed a second methylation experiment (**Figure 1B**, **Supplemental Figure 3**), and replicated 4 CpGs (at p = 0.05 /10 = 0.005) (**Table 3**). We then performed a fixed-effects meta-analysis of these datasets and found 4 CpGs that were associated to current tobacco smoking in plaques at FDR < 0.05 mapping to 6 different genes (**Table 4**, **Figure 1C**, **Supplemental Figure 4**). All of these 4 CpGs showed reduced DNA methylation in current smokers as compared to former or never smokers (**Figure 2**). A sensitivity analysis on the number of estimated pack-years of smoking showed two of these 4 nominally associated (cg05575921 near *AHRR* and cg05284742 near *ITPK1*, **Supplemental Table 1**).

To study the possible effect of smoking-induced methylation changes on the carotid atherosclerotic plaque in more detail, we investigated histological features of the plaques. Considering data from the whole Athero-Express Biobank (n = 2,319), current tobacco smoking behaviour was associated with more calcification (OR = 1.42 [1.13-1.81], p = 0.0034), and collagen deposition (OR = 1.47 [1.09-1.97], p = 0.0112) in atherosclerotic plaques (**Supplemental Table 2**). However, none of the 4 CpGs associated to smoking was associated to specific plaques characteristics (**Supplemental Table 3**).

DNA methylation in Blood

In addition to the analysis in plaque specimens, we performed an EWAS between blood-derived DNA methylation and current tobacco smoking in 89 blood samples (**Figure 3**,

Supplemental Figure 5, **Supplemental Table 4**). We identified 17 significant (FDR ≤ 0.05)

CpGs in blood, mapping to 8 genes, one long-non-coding RNA, and one miRNA (**Table 5**), all

of which showed lower DNA methylation in current smokers compared to former or never

smokers. Of these 17 CpGs, 8 have previously been associated with smoking in blood and

other tissues (**Table 5**)[18–21,23,26–29], confirming the relevance of previously reported loci in

patients with severe atherosclerotic disease.

Correlations to RNA

To investigate possible effects of the tissue specific CpGs on local gene expression we

performed a pilot RNA-sequencing experiment using plaque-derived whole-tissue RNA (n =

21). None of the genes mapped to the 4 plaque-derived CpGs were significantly associated

to current smoking status (**Supplemental Table 5**). However, when comparing the direction

of effects of all nominal significant CpGs with the gene expression, the correlation was

significant for CpGs mapped to 1,500 or 200 bp from the transcription start site, and for

CpGs mapped to the first exon (**Supplemental Figure 6**).

Genetic variation

The susceptibility of CpGs to undergo epigenetic modifications due to environmental factors

may be modified by genetic variation. Therefore, we associated DNA methylation at the

smoking-associated CpGs in plaque with nearby common DNA sequence variation. We

identified a common variant, rs4956991 (c.*1078A>G, effect allele frequency = 0.65) in the 3'

UTR of *PLEKHGB4*, that associated to methylation at the cg02385153 in *AHRR* (β = -0.020 ±

0.003 standard error (s.e.), p = 1.52×10^{-9} which equals FDR = 6.51×10^{-8}, **Figure 4**). This

suggests that DNA methylation at the smoking-related CpG cg02385153 may also affected

by genetic variation 221,251 bp upstream of the *AHRR* gene.

To investigate if this methylation quantitative trait locus (mQTL) also indicate co-regulatory

gene-gene interaction, we determined the relationships between common genetic variation,

CpGs, and the expression of the involved genes (*PLEKHG4B* and *AHRR*). Exploring data from GTEx Portal (https://www.gtexportal.org)[31] we found rs4956911 also to be an expression quantitative trait loci (eQTL) of *PLEKHG4B* in multiple tissues but not of *AHRR* (β = -0.32, p = 2.1x10^{-11}, **Supplemental Figure 7**). In addition, while exploring data from gnomAD[32] we found one non-synonymous variant, rs4956987, that may alter the function of the PLEKHG4B protein (p.Arg1076Gln, β = -0.012 ± 0.003 standard error (s.e.), p = 2.40x10^{-4} which equals FDR = 3.64x10^{-3}). Finally, we show positive associations between expression of the genes *AHRR* and *PLEKHG4B*, in multiple CVD related tissues in the STAGE-cohort (**Supplemental Table 6**). In light of these results, we speculate that *PLEKHG4B* may be a co-regulatory gene of *AHRR* expression (**Supplemental Figure 8**).

DISCUSSION

We performed a two-stage epigenome-wide association study on smoking in 664 carotid atherosclerotic plaque samples. This study shows that smoking is strongly associated with differential DNA methylation in carotid atherosclerotic plaques. As far as we know, this is the first study reporting 4 CpGs differentially methylated in DNA derived from plaques due to tobacco smoking (**Table 4**). In addition, we could replicate 8 CpG loci known to associate with smoking in circulating cells (**Table 5**)[18–21,23,26–29,33].

This study provides supporting evidence for an effect of smoking on epigenetic regulation in atherosclerotic vascular tissue. This is strengthened by the partial similarity observed in DNA methylation patterns between blood and plaque. For example, multiple associations with smoking were observed at CpG loci near *AHRR*, a regulator of the *aryl hydrocarbon receptor* (*AhR*) transcription factor and its pathway. Differential DNA methylation at this locus has been associated with smoking on numerous occasions and various tissues, including pulmonary macrophages and neonatal cord-blood[18–21,23–29,33]. Furthermore, this relationship has also been shown in a mouse model in which lower DNA methylation at the *AHRR* gene was associated with higher *AHRR* expression[21]. The *AhR* transcription factor is a xenobiotic receptor, sensitive to some endogenous ligands as well as many exogenous toxins. These toxins include polycyclic aromatic hydrocarbons and dioxins both of which are important constituents of tobacco smoke[34] and lead to upregulation of enzymes involved in the detoxifying metabolism of these substances.

The other smoking associated CpG locus is located near *ITPK1* (a gene encoding for inositol-tetrakisphosphate 1-kinase) and earlier studies had associated the same locus to differential methylation in circulating blood cells[18,35,36]. The ITPK1 enzyme functions as a key regulator of the rate limiting step in the inositol metabolic pathway pivotal in the formation of phosphorylated forms of inositol[37]. Inositol has been implicated in neural tube defects[38] and has a role in transcriptional regulation[39]. Although differential methylation at *ITPK1* has been implicated with smoking before, the exact biological implications and the role of *ITPK1* or inositol in the response to smoking remains unknown.

Furthermore, smoking was associated with several CpGs in our discovery dataset, that were identified in literature before. For instance, we identified cg16650073 near *NTHL1* encoding for endonuclease III-like protein 1, an enzyme that is involved in nucleotide base-excision repair of DNA. Interestingly, *NTHL1* expression was shown to be reduced in lung tumor biopsies in humans[40], and smoke exposure was shown to reduce NTHL1 protein expression in mice lung fibroblasts[41]. Our results suggest that smoke exposure may inhibit DNA-repair in vascular tissue through down-regulation of *NTHL1* expression. This notion is further supported by a study showing that reactive oxygen species (ROS) can induce DNA oxidation, leading to aberrant regulation of *NTHL1*[42]. Indeed, it is known that both tobacco smoking and ROS cause vascular endothelial dysfunction leading to endothelial activation and vascular smooth muscle cell proliferation, and ultimately atherosclerosis. Our data adds to this by supporting a role for epigenetic regulation in atherosclerotic lesions through demethylation of *NTHL1*, *AHRR* and other cellular maintenance genes. These data imply that epigenetic changes may adversely affect vascular tissue and thereby affect atherosclerotic lesion development and progression.

In addition, it is remarkable that our results in blood-derived DNAm also indicate a significant association at cg03636183 in the *F2RL3*-gene (coagulation factor II receptor-like 3)[20,43]. Indeed, hypomethylation at this locus in blood cells has been reported to associate strongly with current and long-time tobacco smoke exposure[44]. This protease-activated receptor is involved in cardiovascular pathophysiological processes including thrombin-induced platelet-aggregation[45] as well as inflammation[46]. Also, methylation at *F2RL3* in blood cells is shown to be a predictor for lung cancer[47] and mortality[48].

Genetic variation may affect methylation status of specific genes. Using mQTL analysis, we found strong associations between lesion CpGs and nearby SNPs, showing that some of the smoking-associated CpG methylation may be affected by genetic variation. Therefore, these SNPs are of particular interest since they may reveal hereditary susceptibility to toxicity in the

vascular wall. Not much is known about the biological functions of the *PLEKHG4B* genes, and further research should focus on their relationship with smoking.

Our observations in pilot data suggest that smoking affects the atherosclerotic vascular lesion at the epigenetic level, which may affect local gene expression levels (**Supplemental Figure 6**). Although the concept of transcriptional regulation by DNA methylation has been abundantly shown[49], the effect of a particular CpG on local gene-expression is complex. Elucidation of the effects of CpGs on gene expression within the atherosclerotic vascular wall tissue in larger samples may offer important insights into the biological mechanisms by which tobacco smoking confers an increased cardiovascular risk.

Most epigenetic smoking studies to date have focused on blood-derived DNA-methylation. To gain better insight in the tissue specificity of the methylation results obtained in atherosclerotic plaques and to verify consistency with pre-existing studies, we also performed an EWAS in blood samples from the same patients. Furthermore, we carefully scrutinized literature investigating blood or other tissues. The combined results of the literature search and our experimental data, suggest vasculature-specific methylation differences induced by tobacco smoking. This emphasizes the importance of investigating DNA methylation in the vascular lesion itself, as well as the need for further validation in external studies.

Limitations of the current study. Our analyses are based on patients' current smoking behavior, which will not reflect time-dependent effects of smoking on plaque methylation[50], as patients may be light or heavy smokers in the past. Thus, our results may apply specifically to active or recent (< 1 year) smokers. Although we show strong associations and correct for inflation and bias using Bayesian modeling, it is impossible to exclude residual confounding, or misclassification bias as a consequence of self-reported smoking behaviour. This is complicated by the differences in DNA methylation between cell-types in the sample, indeed the limited replication (4 out 10 CpGs are significant) are indicative of cellular heterogeneity.

Future studies focused on single-cell methylation and tissue-specific spatial methylation can aid in determining the relevant cell types in tobacco smoking-induced epigenetic regulation in vascular lesions.

Gene regulation and expression are thought to act on cellular and tissue function, and thus ultimately on intermediate phenotypes. Yet, we did not find an association between the smoking associated DNAm and plaque characteristics. This may be a reflection of low power due to the heterogeneity of the tissue in which we measured methylation. In addition, it is uncertain what the correlation is between methylation and protein levels that ultimately affect cellular function and intermediate phenotypes.

Furthermore, it should be noted that the Athero-Express Biobank is a cohort of patients with advanced atherosclerotic disease. Therefore, it merits careful consideration to draw inferences on earlier stages of atherosclerotic disease. This selected group of atherosclerotic patients with advanced stages of disease may also explain the lack of association with plaque characteristics.

Finally, our replication dataset was of limited sample size, reducing power in the meta-analysis. Future studies should aim to include more samples for discovery and replication.

In summary, we performed a two-stage epigenome-wide association study of current smoking in 664 atherosclerotic plaque samples and 89 peripheral blood samples derived from 668 carotid endarterectomy patients. We show that tobacco smoking is associated with DNA methylation at 4 loci in atherosclerotic lesions of carotid endarterectomy patients. Future studies should verify these findings, and focus on the underlying mechanisms of *AHRR* and *ITPK1* methylation in the vasculature as a response to smoking.

DISCLOSURES

JLMB is founder, main shareholder and chairman of the board for Clinical Gene Networks AB (CGN; Stockholm, Sweden) and TM is shareholder. CGN has an invested interest in microarray data generated from the STAGE cohort. CGN had no part in this study, neither in the conception, design and execution of this study, nor in the preparation and contents of this manuscript.

REFERENCES

1. Talhout R, *et al*. Hazardous compounds in tobacco smoke. *Int J Environ Res Public Health*. 2011;8:613–628.

2. Messner B, *et al*. Smoking and Cardiovascular Disease: Mechanisms of Endothelial Dysfunction and Early Atherogenesis. *Arterioscler Thromb Vasc Biol*. 2014;34:509–515.

3. Csordas A, *et al*. The biology behind the atherothrombotic effects of cigarette smoke. *Nat Rev Cardiol*. 2013;10:219–230.

4. Herder M, *et al*. Risk factors for progression of carotid intima-media thickness and total plaque area: a 13-year follow-up study: the Tromsø Study. *Stroke*. 2012;43:1818–1823.

5. Kumagai S, *et al*. Impact of cigarette smoking on coronary plaque composition. *Coron Artery Dis*. 2015;26:60–65.

6. van den Bouwhuijsen QJA, *et al*. Determinants of magnetic resonance imaging detected carotid plaque components: the Rotterdam Study. *Eur Heart J*. 2012;33:221–229.

7. Kangavari S, *et al*. Smoking increases inflammation and metalloproteinase expression in human carotid atherosclerotic plaques. *J Cardiovasc Pharmacol Ther*. 2004;9:291–298.

8. Nikpay M, *et al*. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* [Internet]. 2015;47. Available from: http://dx.doi.org/10.1038/ng.3396

9. Zeller T, *et al*. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One*. 2010;5:e10693.

10. Polfus LM, *et al*. Genome-wide association study of gene by smoking interactions in coronary artery calcification. *PLoS One*. 2013;8:e74642.

11. van Setten J, *et al*. Genome-wide association study of coronary and aortic calcification

implicates risk loci for coronary artery disease and myocardial infarction. *Atherosclerosis*. 2013;228:400–405.

12. Hamrefors V, *et al*. Smoking modifies the associated increased risk of future cardiovascular disease by genetic variation on chromosome 9p21. *PLoS One*. 2014;9:e85893.

13. Justice AE, *et al*. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun*. 2017;8:14977.

14. The Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet*. 2010;42:441–447.

15. Venolia L, *et al*. Comparison of transformation efficiency of human active and inactive X-chromosomal DNA. *Nature*. 1983;302:82–83.

16. Dick KJ, *et al*. DNA methylation and body-mass index: a genome-wide analysis. *Lancet*. 2014;383:1990–1998.

17. Pfeiffer L, *et al*. DNA Methylation of Lipid-Related Genes Affects Blood Lipid Levels. *Circ Cardiovasc Genet*. 2015;8:334–342.

18. Zeilinger S, *et al*. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One*. 2013;8:e63812.

19. Besingi W, *et al*. Smoke-related DNA methylation changes in the etiology of human disease. *Hum Mol Genet*. 2014;23:2290–2297.

20. Breitling LP, *et al*. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet*. 2011;88:450–457.

21. Shenker NS, *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;22:843–851.

22. Allione A, *et al*. Novel epigenetic changes unveiled by monozygotic twins discordant for smoking habits. *PLoS One*. 2015;10:e0128265.

23. Dogan MV, *et al*. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. *BMC Genomics*. 2014;15:151.

24. Richmond RC, *et al*. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Hum Mol Genet*. 2015;24:2201–2217.

25. Joubert BR, *et al*. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect*. 2012;120:1425–1431.

26. Elliott HR, *et al*. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics*. 2014;6:4.

27. Monick MM, *et al*. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *Am J Med Genet B Neuropsychiatr Genet*. 2012;159B:141–151.

28. Sun YV, *et al*. Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans. *Hum Genet*. 2013;132:1027–1037.

29. Wan ES, *et al*. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum Mol Genet*. 2012;21:3073–3082.

30. Pruim RJ, *et al*. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336–2337.

31. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580–585.

32. Lek M, *et al*. Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv*. 2015;

33. Reynolds LM, *et al*. DNA Methylation of the Aryl Hydrocarbon Receptor Repressor Associations With Cigarette Smoking and Subclinical Atherosclerosis. *Circ Cardiovasc Genet*. 2015;8:707–716.

34. Klingbeil EC, *et al*. Polycyclic aromatic hydrocarbons, tobacco smoke, and epigenetic remodeling in asthma. *Immunol Res*. 2014;58:369–373.

35. Besingi W, *et al*. Smoke-related DNA methylation changes in the etiology of human disease. *Hum Mol Genet*. 2014;23:2290–2297.

36. Zhang Y, *et al*. Self-reported smoking, serum cotinine, and blood DNA methylation. *Environ Res*. 2016;146:395–403.

37. Zhang C, *et al*. Regulation of inositol 1,3,4-trisphosphate 5/6-kinase (ITPK1) by reversible lysine acetylation. *Proc Natl Acad Sci U S A*. 2012;109:2290–2295.

38. Wilson MP, *et al*. Neural tube defects in mice with reduced levels of inositol 1,3,4-trisphosphate 5/6-kinase. *Proc Natl Acad Sci U S A*. 2009;106:9831–9835.

39. Odom AR, *et al*. A role for nuclear inositol 1,4,5-trisphosphate kinase in transcriptional control. *Science*. 2000;287:2026–2029.

40. Radak Z, *et al*. Lung cancer in smoking patients inversely alters the activity of hOGG1 and hNTH1. *Cancer Lett*. 2005;219:191–195.

41. Deslee G, *et al*. Cigarette smoke induces nucleic-acid oxidation in lung fibroblasts. *Am J Respir Cell Mol Biol*. 2010;43:576–584.

42. Fleming AM, *et al*. Oxidative DNA damage is epigenetic by regulating gene transcription

via base excision repair. *Proc Natl Acad Sci U S A*. 2017;114:2604–2609.

43. Breitling LP. Current genetics and epigenetics of smoking/tobacco-related cardiovascular disease. *Arterioscler Thromb Vasc Biol*. 2013;33:1468–1472.

44. Zhang Y, *et al*. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. *Environ Health Perspect*. 2014;122:131–137.

45. Wu C-C, *et al*. The roles and mechanisms of PAR4 and P2Y12/phosphatidylinositol 3-kinase pathway in maintaining thrombin-induced platelet aggregation. *Br J Pharmacol*. 2010;161:643–658.

46. McDougall JJ, *et al*. Triggering of proteinase-activated receptor 4 leads to joint pain and inflammation in mice. *Arthritis & Rheumatology*. 2009;60:728–737.

47. Zhang Y, *et al*. F2RL3 methylation, lung cancer incidence and mortality. *Int J Cancer*. 2015;137:1739–1748.

48. Zhang Y, *et al*. F2RL3 methylation in blood DNA is a strong predictor of mortality. *Int J Epidemiol*. 2014;43:1215–1225.

49. Suzuki MM, *et al*. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*. 2008;9:465–476.

50. Guida F, *et al*. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum Mol Genet*. 2015;24:2349–2359.

Table 1: Patient characteristics of the discovery and replication datasets. Patient characteristics at time of inclusion in both datasets, stratified by smoking status. Patients without data on current smoking were excluded. *Symptoms at presentation, before carotid endarterectomy. Significance shown as p-values (*P*) without FDR adjustment. Abbreviations: *SBP*, systolic blood pressure; *DBP*, diastolic blood pressure; *eGFR*, estimated glomerular filtration rate by MDRD-formula; *BMI*, body-mass index; *LLDs*, use of lipid-lowering drugs; *Ocular*, retinal infarction and amaurosis fugax.

Characteristic	Discovery (AEMS450K1, n = 477)				Replication (AEMS450K2, n = 187)			
	Former or never smokers [n = 283]	Current smokers [n = 194]	Missing %	P	Former or never smokers [n = 131]	Current smokers [n = 56]	Missing %	P
Age (years [s.e.])	70.0 [9.0]	64.9 [8.3]	0.0%	<0.001	70.2 [8.7]	65.2 [9.0]	0.0%	0.001
Males (%)	71.7	67.5	0.0%	0.377	84.7	80.4	0.0%	0.601
SBP (mmHg [s.e.])	155.7 [24.7]	155.4 [27.9]	11.3%	0.917	153.9 [20.9]	149.7 [22.3]	17.1%	0.274
DBP (mmHg [s.e.])	82.2 [12.9]	83.0 [13.7]	11.3%	0.558	81.9 [12.4]	80.6 [11.7]	17.1%	0.537
eGFR (mL/min/1.73m2[s.e.])	69.5 [19.4]	76.6 [22.0]	2.5%	<0.001	73.6 [19.2]	75.4 [23.5]	4.8%	0.587
BMI (kg/m2[s.e.])	26.8 [3.7]	26.2 [4.2]	3.4%	0.085	27.0 [3.8]	25.9 [4.3]	3.7%	0.095
ePackyears (years [s.e.])	22.1 [22.2]	26.5 [19.3]	57.2%	0.033	20.5 [22.4]	26.4 [19.6]	10.7%	0.101
Comorbidities (%)								
Diabetes	21.9	22.2	0.0%	1	23.7	16.1	0.0%	0.335

Hypertension	91.9	80.9	0.0%	0.001	89.3	75	0.0%	0.022
Medication use (%)								
Hypertensive drugs	83.7	69.1	0.0%	<0.001	82.4	66.1	0.0%	0.023
Anti-coagulants	13.1	10.3	0.0%	0.441	18.3	10.7	0.0%	0.280
Anti-platelet drugs	90.1	90.7	0.0%	0.947	87.8	85.7	0.0%	0.881
LLDs	75.6	77.3	0.0%	0.749	77.9	78.6	0.0%	1
Symptoms† (%)			0.2%	0.444			0.0%	0.470
Asymptomatic	17.4	14.9			13	5.4		
Ocular	11.7	16.5			17.6	26.8		
TIA	44	44.3			43.5	37.5		
Stroke	27	24.2			26	30.4		

Table 2: CpGs associated with current tobacco smoking in carotid plaque after discovery. *Chr:BP*: chromosome base-pair position of the methylation probes (CpG). *Strand*: strand position of the methylation site. *Gene* the gene mapped to the CpG. *Beta*: effect size. *SE*: standard error. P: p-value of association prior to `bacon` correction. P_{corr}: p-value of association after `bacon` correction.

| | | | **Discovery** | | | |
| | | | **(AEMS450K1, n = 477)** | | | |
CpG	**Chr:BP**	**Gene**	**Beta**	**SE**	**P**	**Pcorr**
cg25648203	chr5:395396	*AHRR*	-0.294	0.032	2.24×10^{-25}	5.37×10^{-20}
cg05575921	**chr5:373378**	***AHRR***	**-0.319**	**0.052**	**3.41×10^{-12}**	**7.33×10^{-10}**
cg03991871	**chr5:368399**	***AHRR***	**-0.346**	**0.059**	**3.70×10^{-11}**	**4.63×10^{-9}**
cg16650073	chr16:2089849	*NTHL1*	-0.519	0.077	2.28×10^{-14}	1.53×10^{-11}
cg12806681	**chr5:368346**	***AHRR***	**-0.222**	**0.043**	**6.63×10^{-9}**	**2.59×10^{-7}**
cg05284742	**chr14:93552080**	***ITPK1***	**-0.212**	**0.047**	**3.84×10^{-7}**	**6.10×10^{-6}**
cg02385153	chr5:404766	*AHRR*	0.228	0.048	2.99×10^{-8}	2.56×10^{-6}
cg05951221	chr2:233284402	*ALPI*	-0.258	0.055	1.33×10^{-7}	2.67×10^{-6}
cg22702618	chr19:18705064	*CRLF1*	0.649	0.109	2.44×10^{-12}	2.29×10^{-9}
cg19505196	chr3:128080273	*EEFSEC*	0.225	0.043	5.14×10^{-10}	1.24×10^{-7}

Table 3: CpGs associated with current tobacco smoking in carotid plaque after replication. *Chr:BP*: chromosome base-pair position of the methylation probes (CpG). *Gene* the gene mapped to the CpG. *Beta*: effect size. *SE*: standard error. *P*: p-value of association prior to `bacon` correction. P_{corr}: p-value of association after `bacon` correction.

| | | | **Replication** | | | |
| | | | **(AEMS450K2, n = 187)** | | | |
CpG	**Chr:BP**	**Gene**	**Beta**	**SE**	**P**	**Pcorr**
cg25648203	chr5:395396	*AHRR*	-0.082	0.081	0.33	0.313
cg05575921	**chr5:373378**	***AHRR***	**-0.333**	**0.082**	**2.97×10^{-5}**	**5.13×10^{-5}**
cg03991871	**chr5:368399**	***AHRR***	**-0.300**	**0.095**	**1.14×10^{-3}**	**1.50×10^{-3}**

cg16650073	chr16:2089849	NTHL1	-0.062	0.144	0.713	0.666
cg12806681	**chr5:368346**	**AHRR**	**-0.173**	**0.062**	**4.47×10^{-3}**	**5.38×10^{-3}**
cg05284742	**chr14:93552080**	**ITPK1**	**-0.313**	**0.117**	**6.30×10^{-3}**	**7.41×10^{-3}**
cg02385153	chr5:404766	AHRR	0.200	0.102	0.031	0.050
cg05951221	chr2:233284402	ALPI	-0.223	0.118	0.058	0.059
cg22702618	chr19:18705064	CRLF1	0.018	0.166	0.838	0.912
cg19505196	chr3:128080273	EEFSEC	0.071	0.061	0.189	0.244

Table 4: Methylation of CpGs in carotid plaques associated to current tobacco smoking status after meta-analysis of discovery and replication cohorts. *Chr:BP*: chromosome base-pair position of the methylation probes (CpG). *Gene* the gene mapped to the CpG. *Beta*: effect size. *SE*: standard error. P: p-value of association prior to `bacon` correction. P_{corr}: p-value of association after `bacon` correction. *FDR*: the false discovery rate adjusted Q-value of association.

					Meta-Analysis			
					discovery plus replication, n = 664			
CpG	Chr:BP	Gene	CpG Island	Relation to Island	Beta	SE	P_{corr}	FDR
cg05575921	chr5:373378	AHRR	chr5:373842-374426	N_Shore	-0.323	0.044	1.71×10^{-13}	3.80×10^{-8}
cg03991871	chr5:368399	AHRR	chr5:370185-370422	N_Shore	-0.333	0.05	2.90×10^{-11}	4.28×10^{-6}
cg12806681	chr5:368346	AHRR	chr5:370185-370422	N_Shore	-0.206	0.035	5.95×10^{-9}	5.27×10^{-4}
cg05284742	chr14:93552080	ITPK1		OpenSea	-0.226	0.044	2.05×10^{-7}	0.015

Table 5: Methylation of CpGs in blood associated to current tobacco smoking status in AEMS450K1. In bold the CpGs that were also significant in the final meta-analysis of plaque-derived DNAm. *Chr:BP*: chromosome base-pair position of the methylation probes (CpG). *Gene* the gene mapped to the CpG. *Beta*: effect size. *SE*: standard error. *P*: p-value of association prior to `bacon` correction. P_{corr}: p-value of association after `bacon` correction. *FDR*: the false discovery rate adjusted Q-value of association.

AEMS450K1

| CpG | Chr:BP | Gene | blood, n = 93 | | | | |
			Beta	SE	P	P_{corr}	FDR
cg05575921	chr5:373378	***AHRR***	-1.485	0.302	**5.76×10^{-22}**	**3.38×10^{-21}**	**1.50×10^{-15}**
cg03991871	chr5:368399	***AHRR***	-0.93	0.377	**4.53×10^{-14}**	**1.02×10^{-13}**	**2.26×10^{-8}**
cg12806681	chr5:368346	***AHRR***	-0.736	0.462	**2.60×10^{-13}**	**5.36×10^{-13}**	**7.91×10^{-8}**
cg21161138	chr5:399312	*AHRR*	-0.529	0.630	8.14×10^{-13}	1.59×10^{-12}	1.76×10^{-7}
cg26703534	chr5:377358	*AHRR*	-0.479	0.644	3.07×10^{-11}	5.03×10^{-11}	4.46×10^{-6}
cg03636183	chr19:17000537	*F2RL3*	-0.639	0.448	7.58×10^{-10}	1.07×10^{-9}	7.92×10^{-5}
cg23079012	chr2:8343662	*LINC00299*	-0.901	0.295	1.07×10^{-8}	1.34×10^{-8}	8.49×10^{-4}
cg03450842	chr10:80834947	*ZMIZ1*	-0.370	0.683	5.45×10^{-8}	6.40×10^{-8}	3.55×10^{-3}
cg23916896	chr5:368756	*AHRR*	-0.905	0.278	6.25×10^{-8}	7.29×10^{-8}	3.59×10^{-3}
cg05951221	chr2:233284402	*ALPI*	-0.454	0.527	2.72×10^{-7}	2.99×10^{-7}	0.013
cg21566642	chr2:233284613	*ALPI*	-0.597	0.397	3.54×10^{-7}	3.85×10^{-7}	0.016
cg03358636	chr3:197473958	*RUBCN*	-0.513	0.457	4.61×10^{-7}	4.96×10^{-7}	0.018
cg17295878	chr17:77924665	*TBC1D16*	-0.982	0.234	7.42×10^{-7}	7.83×10^{-7}	0.027
cg05284742	chr14:93552080	***ITPK1***	-0.445	0.512	**9.42×10^{-7}**	**9.86×10^{-7}**	**0.030**
cg14817490	chr5:392920	*AHRR*	-0.731	0.312	9.57×10^{-7}	1.00×10^{-6}	0.030
cg11660018	chr11:86510915	*OR7E2P*	-0.303	0.749	1.04×10^{-6}	1.08×10^{-6}	0.030
cg03371962	chr12:1772275	*MIR3649*	-0.651	0.346	1.31×10^{-6}	1.36×10^{-6}	0.035

FIGURE LEGENDS

Figure 1. Manhattan plots of the association of DNA methylation in carotid atherosclerotic plaques with current tobacco smoking in A) the discovery (AEMS450K1), B) the replication (AEMS450K2) cohorts, and C) the meta-analysis (n = 664). Each point represents an individual CpG, with the x-axis shows the genomic location of each CpG and the y-axis shows the observed $-\log_{10}(p$-value) of the association with current tobacco smoking after meta-analysis. Loci with CpGs that were epigenome-wide significant after replication at are shown in grey.

Figure 2. Top 4 replicated associations stratified by current tobacco smoking status in the discovery (AEMS450K1). Each boxplot shows the association of current tobacco smoking status (x-axis) with the methylation of a CpG (y-axis).

Figure 3. Manhattan plot of the association of DNA methylation in whole-blood blood with current tobacco smoking in AEMS450K1. Each point represents an individual CpG, with the x-axis shows the genomic location of each CpG and the y-axes shows the observed $-\log_{10}(p\text{-value})$ of the association with current tobacco smoking. CpGs that were epigenome-wide significant after false-discovery rate correction at FDR ≤ 0.05 are shown in grey.

Figure 4: The association of genetic variants near *AHRR* with methylation of cg02385153. The strongest association was for rs4956991 (G-allele, $p = 5.2 \times 10^{-9}$, see main text, purple). The x-axis shows the chromosomal position relative to 1000G (March 2012, Hg19). The lower panel shows the refSeq canonical genes from UCSC (the black arrow indicates the direction of transcription). The left y-axis shows the $-\log_{10}(\text{-value})$ of the association with the methylation of cg02385153 (in the body of *AHRR*). The right y-axis shows the recombination rate (grey line in the middle panel). The middle panel shows each associated variants colored by the linkage disequilibrium r^2 relative to rs4956991; the legend in the upper right corner shows the r^2 color scale. Made using LocusZoom version 1.3[30].