Smoking is Associated to DNA Methylation in Atherosclerotic Carotid Lesions

Siemelink, M.A.; Laan, S.W. van der; Haitjema, S.; Koeverden, I.D. van; Schaap, J.; Wesseling, M.; ... ; Pasterkamp, G.

Citation

Siemelink, M. A., Laan, S. W. van der, Haitjema, S., Koeverden, I. D. van, Schaap, J., Wesseling, M., … Pasterkamp, G. (2018). Smoking is Associated to DNA Methylation in Atherosclerotic Carotid Lesions. *Circulation: Genomic And Precision Medicine*, *11*(9). doi:10.1161/CIRCGEN.117.002030

Version: Not Applicable (or Unknown)
License: Leiden University Non-exclusive license
Downloaded from: https://hdl.handle.net/1887/87148

Note: To cite this publication please use the final published version (if applicable).

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

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.

Keywords: Epigenetics, carotid endarterectomy, smoking, cardiovascular diseases, atherosclerosis

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 ≤ 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* ($\beta = -0.020 \pm 0.003$ standard error (s.e.), $p = 1.52\text{x}10^{-9}$ which equals FDR = $6.51\text{x}10^{-8}$, **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* ($\beta = -0.32$, $p = 2.1\text{x}10^{-11}$, **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, $\beta = -0.012 \pm 0.003$ standard error (s.e.), $p = 2.40\text{x}10^{-4}$ which equals $\text{FDR} = 3.64\text{x}10^{-3}$). 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.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to Aisha Gohar for her help with the manuscript.

We thank Utrecht Sequencing Facility for providing the sequencing service and data. Utrecht Sequencing Facility is subsidized by the University Medical Center Utrecht, Hubrecht Institute and Utrecht University.

FUNDING SOURCES

The authors received no specific funding for this work and none of the funding sources had any influence in any aspect of this study. GP is the corresponding author and had full access to the data. The authors affirm that this manuscript is an honest, accurate and transparent account of the study performed. MAS acknowledges funding by the European Union (BiomarCaRE, grant number: HEALTH-2011-278913) and technology foundation STW (Stichting voor de Technische Wetenschappen – Danone partnership program, Project 11679). SWvdL was funded through grants from the Netherlands CardioVascular Research Initiative (“GENIUS”, CVON2011-19) and the Interuniversity Cardiology Institute of the Netherlands (ICIN, 09.001). SWvdL and SH were supported by the FP7 EU project CVgenes@target (HEALTH-F2-2013-601456). FWA is supported by the University College London Hospitals’ Biomedical Research Centre, and by a Dekker Scholarship (Junior Staff Member 2014T001) from the Dutch Heart Foundation. The work for STAGE was supported by PROCARDIS in the 6th EU-framework program (LSHM-CT-2007-037273), the Swedish Heart-Lung Foundation (JLMB), the King Gustaf V and Queen Victoria’s Foundation of Freemasons (JLMB), the Swedish Society of Medicine (JLMB). STAGE was also supported by grant from University of Tartu (SP1GVARENG, JLMB), the Estonian Research Council (ETIS, JLMB), the Roslin Institute Strategic Grant Funding from the Biotechnology and Biological Sciences Research Council (TM) and by Clinical Gene Networks AB. EA and SY were supported by the Academy of Finland, University of Eastern Finland spearhead program, Kuopio University Hospital, and Sigrid Juselius Foundation.

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.

	Discovery (AEMS450K1, n = 477)				**Replication (AEMS450K2, n = 187)**			
Characteristic	**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	2.24x10^{-25}	5.37x10^{-20}
cg05575921	**chr5:373378**	***AHRR***	**-0.319**	**0.052**	**3.41x10^{-12}**	**7.33x10^{-10}**
cg03991871	**chr5:368399**	***AHRR***	**-0.346**	**0.059**	**3.70x10-^{11}**	**4.63x10^{-9}**
cg16650073	chr16:2089849	*NTHL1*	-0.519	0.077	2.28x10^{-14}	1.53x10^{-11}
cg12806681	**chr5:368346**	***AHRR***	**-0.222**	**0.043**	**6.63x10^{-9}**	**2.59x10^{-7}**
cg05284742	**chr14:93552080**	***ITPK1***	**-0.212**	**0.047**	**3.84x10^{-7}**	**6.10x10^{-6}**
cg02385153	chr5:404766	*AHRR*	0.228	0.048	2.99x10^{-8}	2.56x10^{-6}
cg05951221	chr2:233284402	*ALPI*	-0.258	0.055	1.33x10^{-7}	2.67x10^{-6}
cg22702618	chr19:18705064	*CRLF1*	0.649	0.109	2.44x10^{-12}	2.29x10^{-9}
cg19505196	chr3:128080273	*EEFSEC*	0.225	0.043	5.14x10^{-10}	1.24x10^{-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**	**2.97x10^{-5}**	**5.13x10^{-5}**
cg03991871	**chr5:368399**	***AHRR***	**-0.300**	**0.095**	**1.14x10^{-3}**	**1.50x10^{-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

			blood, n = 93				
CpG	**Chr:BP**	**Gene**	**Beta**	**SE**	**P**	**P$_{corr}$**	**FDR**
cg05575921	chr5:373378	***AHRR***	-1.485	0.302	**5.76x10^{-22}**	**3.38x10^{-21}**	**1.50x10^{-15}**
cg03991871	chr5:368399	***AHRR***	-0.93	0.377	**4.53x10^{-14}**	**1.02x10^{-13}**	**2.26x10^{-8}**
cg12806681	chr5:368346	***AHRR***	-0.736	0.462	**2.60x10^{-13}**	**5.36x10^{-13}**	**7.91x10^{-8}**
cg21161138	chr5:399312	*AHRR*	-0.529	0.630	8.14x10^{-13}	1.59x10^{-12}	1.76x10^{-7}
cg26703534	chr5:377358	*AHRR*	-0.479	0.644	3.07x10^{-11}	5.03x10^{-11}	4.46x10^{-6}
cg03636183	chr19:17000537	*F2RL3*	-0.639	0.448	7.58x10^{-10}	1.07x10^{-9}	7.92x10^{-5}
cg23079012	chr2:8343662	*LINC00299*	-0.901	0.295	1.07x10^{-8}	1.34x10^{-8}	8.49x10^{-4}
cg03450842	chr10:80834947	*ZMIZ1*	-0.370	0.683	5.45x10^{-8}	6.40x10^{-8}	3.55x10^{-3}
cg23916896	chr5:368756	*AHRR*	-0.905	0.278	6.25x10^{-8}	7.29x10^{-8}	3.59x10^{-3}
cg05951221	chr2:233284402	*ALPI*	-0.454	0.527	2.72x10^{-7}	2.99x10^{-7}	0.013
cg21566642	chr2:233284613	*ALPI*	-0.597	0.397	3.54x10^{-7}	3.85x10^{-7}	0.016
cg03358636	chr3:197473958	*RUBCN*	-0.513	0.457	4.61x10^{-7}	4.96x10^{-7}	0.018
cg17295878	chr17:77924665	*TBC1D16*	-0.982	0.234	7.42x10^{-7}	7.83x10^{-7}	0.027
cg05284742	chr14:93552080	***ITPK1***	-0.445	0.512	**9.42x10^{-7}**	**9.86x10^{-7}**	**0.030**
cg14817490	chr5:392920	*AHRR*	-0.731	0.312	9.57x10^{-7}	1.00x10^{-6}	0.030
cg11660018	chr11:86510915	*OR7E2P*	-0.303	0.749	1.04x10^{-6}	1.08x10^{-6}	0.030
cg03371962	chr12:1772275	*MIR3649*	-0.651	0.346	1.31x10^{-6}	1.36x10^{-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 $FDR \leq 0.05$ 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 $-\log_{10}$(-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].

Supplemental Material

Supplemental Material and Methods

Patient inclusion

The Athero-Express Biobank Study (AE) is an ongoing longitudinal biobank study including patients that undergo arterial endarterectomy in two Dutch tertiary referral centers since 2002. A detailed description of the cohort study design has previously been published[1]. For the present study, subsequent patients were included who underwent carotid endarterectomy (CEA) and of which genotyping data were available. Clinical data were extracted from patient medical files and standardized questionnaires. Current tobacco smoking (*i.e.* including [hand rolled] cigarettes, cigars, *etc.*) was defined as smoking within 1 year prior to admission for CEA and was assessed by questionnaire. We estimated the number of pack years smoking based on a categorical question regarding the number of cigarettes smoked and define the "estimated Pack Years Smoking" = (number of cigarettes smoked per day x number of years smoked)/20; where 1 pack is defined as 20 cigarettes.

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.

Sample collection

Blood samples were obtained prior to surgery and stored at -80℃. Carotid plaque specimens were removed during surgery and immediately processed in the laboratory. Specimens were cut transversely into segments of 5 mm. The culprit lesion (the region with most severe stenosis) was identified, fixed in 4% formaldehyde, embedded in paraffin, and processed for histological examination. Plaque histological features were routinely scored through chemical- and immunohistochemical techniques as described below. Remaining segments were stored at -80℃.

Atherosclerotic plaque histology

The carotid plaque segments containing the culprit lesions were processed according to a standardized protocol, as previously described[2]. In short, 10 micron cross-sections of the paraffin-embedded segments were cut using a microtome and examined under a microscope. Microscopy-slides were stained with hematoxylin and eosin for assessment of calcifications, atheroma, and plaque hemorrhage. Picro Sirius Red was used to stain for collagen. Immunohistochemical staining was performed for assessment of macrophages (CD68), smooth-muscle cells (alpha-actin), and microvessels (CD34). The presence of atheroma was classified as either more or less than 40% of the plaque area. The amount of collagen, calcifications, and plaque hemorrhage were classified as minor or major. Plaque microvessels were quantitatively assessed as average number of vessels over three microscopy field. Plaque smooth-muscle cells and macrophages were quantitatively assessed as percentage of the microscopy field area by computerized analysis using AnalySIS 3.2 software (Soft Imaging Systems GmbH, Münster, Germany). All histological observations were performed by the same dedicated technician and interobserver analyses have been reported previously[3]. Associations of current tobacco smoking with histology were determined by linear or logistic regression modeling where appropriate, adjusting for age, sex, BMI, eGFR (based on the MDRD formula), diabetes, hypertension, history of coronary artery disease, history of peripheral artery disease, lipid levels and medication use.

DNA extraction and methylation experiment

DNA was extracted from stored plaque segments and stored blood samples of patients using standardized in-house protocols as described before in Van der Laan et al[4]. DNA purity and concentration were assessed using the Nanodrop 1000 system (Thermo Scientific, Massachusetts, USA). DNA concentrations were equalized at 600 ng, randomized over 96-well plates and bisulfite converted using a cycling protocol, and the EZ-96 DNA methylation kit (Zymo Research, Orange County, USA). Subsequently, DNA methylation was measured on the Infinium HumanMethylation450 Beadchip Array (HM450k, Illumina, San Diego, USA), which was performed at the Erasmus Medical Center Human Genotyping Facility in Rotterdam, the Netherlands. Processing of the sample and array was performed according to the manufacturer's protocol. Following these protocols, we isolated DNA of 509 patients across 503 plaque samples and 97 blood samples in the discovery study, called Athero-Express Methylation Study 1 (AEMS450K1). The replication study, called Athero-Express Methylation Study 2 (AEMS450K2), included 208 plaque samples (**Supplemental Figure 1**).

Quality control of methylation data

Quality control (QC) of the HM450k array data was performed following the workflow from the `DNAmArray` R-package[5] (https://github.com/molepi/DNAmArray) using default settings, controlling for sample-dependent and probe-dependent parameters. Bisulfate conversion efficiency was determined using dedicated probes on the HM450k. We performed a principal component (PC) analysis for exploratory data analysis using the `irlba` R-package[6] (https://github.com/bwlewis/irlba) and to determine the number of PCs to use for normalization. 'Functional Normalization'[7] with 4 control-probe principal components was used for normalization and correction of batch effects. We computed sex based on sex-chromosome beta-value distribution and compared this to the known sex-status in order to determine possible sample mix-ups. We further assessed sample relations using beta-value extracted genotypes as calculated by the `omicsPrint` R-package (https://github.com/molepi/omicsPrint and https://bioconductor.org/packages/release/bioc/html/omicsPrint.html)[8]. Where available we also compared genotype data to the raw data of the 65 SNPs included on the HM450k array, to determine possible mix-up (as indicated by R ≤ 0.8 across these 65 SNPs). All samples for which sample mix-up could not be confidently ruled out were excluded from further analysis. A total of 42,428 probes were excluded based on above QC steps and the intersection of AEMS450K1 and AEMS450K2, with 443,084 probes (91.3 %) of good quality remaining. After QC, imputation of missing data (average 0.14% and 0.07% missing in AEMS450K1 and AEMS450K2, respectively) was performed using the `knn` algorithm in the `impute` R package (http://bioconductor.org/packages/release/bioc/html/impute.html). For analyses we also excluded probes containing SNPs or which mapped to multiple locations[9].

Samples with missing smoking status or covariates (*i.e.* age, sex, hospital of inclusion) were excluded. After quality control, 485 plaque samples and 93 blood samples obtained from 485 unique patients were remaining in AEMS450K1. The replication dataset AEMS450K2 consisted of 190 plaque samples from an equal number of patients, following quality control. A flow-chart summarizing quality control of samples is presented in **Supplemental figure 1**.

Epigenome-wide (meta-)analysis of current smoking

Epigenome-wide association analysis was done using logistic regression modeling with `limma`[10] following the workflow as included in the `DNAmArray` R-package[5]; we used normalized beta-values (M-values) to ensure maximal power of regression modeling. Regression modeling was performed with covariates age, sex, and hospital of inclusion. The bias and inflation of the resulting test-statistics were controlled using a Bayesian method based on the empirical null distribution as implemented in the R package `bacon` that we recently developed[11]. We also used `bacon` to perform the fixed-effects meta-analysis of the discovery (AEMS450K1) and replication (AEMS450K2) samples.

Given that the discovery and replication samples contain 443,084 overlapping CpGs, we conservatively set a p-value threshold at $p \leq 1.13 \times 10^{-7}$ (0.05/443,084) to claim epigenome-wide significance during discovery. Upon meta-analysis we controlled for multiple testing by correcting p-values using the Benjamini-Hochberg False-Discovery Rate (FDR), and considered FDR Q-values ≤ 0.05 statistically significant[12]. We used the Bioconductor packages TxDb.Hsapiens.UCSC.hg19.knownGene (version 3.2.2, http://bioconductor.org/packages/release/data/annotation/html/TxDb.Hsapiens.UCSC.hg19.knownGene.html) and FDb.InfiniumMethylation.hg19 (version 2.2.0, https://bioconductor.org/packages/release/data/annotation/html/FDb.InfiniumMethylation.hg19.html) to map and annotate CpGs and genes to the genome (GRCh37, Hg19). Statistical analyses were performed with R (v3.4.1) in R Studio (v1.0.143, http://www.rstudio.com/).

Genotyping

DNA was isolated from stored samples and genotyping was performed in two series using commercially available genotyping arrays[4]. The first series (Athero-Express Genomics Study 1, AEGS1) was genotyped using Affymetrix Genome-Wide Human SNP Array 5.0, the second (Athero-Express Genomics Study 2, AEGS2) was genotyped using the Affymetrix Axiom® GW CEU 1 Array. We adhered to community standard quality control and assurance procedures to clean the genotype data obtained in AEGS1 and AEGS2[13]. We used phased haplotypes from the 1000 Genomes Project (phase 3, version 5)[14] merged with haplotypes from the Genome of the Netherlands (GoNL5)[15] as the reference panel for genotype imputation using IMPUTE2[16,17].

RNA-sequencing and differential expression analysis

We isolated RNA from 30 atherosclerotic plaques of the AE using in-house standardized protocols. The RNA-sequencing was performed on the polyadenylated mRNA fraction, which covers all protein coding genes and major part of non-coding RNAs. Sequencing libraries (median length of 350bp) were prepared using the Rapid Directional RNA-Seq Kit (NEXTflex) and sequenced at the Utrecht Sequencing Facility on Illumina NextSeq500 and produced single-end 75 base long reads with up to 15 million reads per library. RNA-seq reads were aligned to the reference genome using STAR (GRCh37, version 74). Transcript abundances were quantified with HTSeq-count using the union mode. Subsequently, reads per kilobase of transcript per million reads sequenced were calculated following the instructions in the Bioconductor workflow “RNA-seq workflow at the gene level” (version r131992, https://www.bioconductor.org/help/workflows/rnaseqGene/), thus DESeq2 was used for downstream analysis[18]. We excluded 9 samples that had low percentage of mRNA mapping to the reference (<5%), and <90% correct strand reads.

Methylation quantitative trait locus (mQTL) analysis

We used fastQTLToolKit (https://github.com/swvanderlaan/fastQTLToolKit)[19] which is based on fastQTL[20] (http://fastqtl.sourceforge.net) to identify variants associated to methylation, *i.e.* methylation quantitative trait loci (mQTL). For the mQTL analysis we considered only high-quality imputed variants (minor allele frequency (MAF) ≥ 0.05; imputation quality ≥ 0.9; Hardy-Weinberg Equilibrium (HWE) p value ≥ 1.0×10^{-6}) in *cis*, *i.e.* within 500 kb of the CpG. For the mQTL analysis we only used overlapping imputed genotypes of 444 patients in the discovery study (AEMS450K1). A linear regression model as implemented in fastQTL[20] was used for the mQTL analysis and corrected by age, sex, SNP array type, genotyping principal components 1 through 10, and current tobacco smoking status.

The Stockholm Atherosclerosis Gene Expression (STAGE) Study

General background on the STAGE Study

In the STAGE Study, seven vascular and metabolic tissues of well-characterized coronary artery disease (CAD) patients were sampled during coronary artery bypass grafting (CABG)[21]. The samples from atherosclerotic arterial wall (AAW), internal mammary artery (IMA), liver, skeletal muscle (SM), subcutaneous fat (SF), visceral fat (VF), and fasting whole blood (WB) were obtained during CABG and used for DNA and RNA isolation. Patients were included if they were eligible for CABG and had no other severe systemic diseases (*e.g.* widespread cancer or active systemic inflammatory disease).

Expression quantitative trait locus (eQTL) analysis in the STAGE Study

In order to prepare inferred genotypes in STAGE for genotype imputation, SNPs were quality controlled for minor allele frequency (MAF ≤ 5%), Hardy-Weinberg equilibrium (HWE; p ≤ 1.0×10^{-6}), and call rate (100%). Thereafter, genotypes for the STAGE study were imputed using IMPUTE2 with 1000 Genomes EUR[22] as the reference[16,17]. Quality control measures for imputed genotypes used an additional filter of IMPUTE2 Info score (≤ 0.3). This yielded a total of 5,473,585 SNPs. Thereafter, methylation quantitative trait loci (mQTLs) passing quality control were selected for expression quantitative trait locus (eQTL) analysis. eQTL analysis was performed for the mQTLs using the Matrix eQTL R package[23], by adding age, gender, smoking status, as covariates. Analysis was performed for eQTL effects on gene-expression against all 17,952 gene-expression profiles available. The eQTL analysis as well as association of gene expression association among genes of interest, were done using MATLAB and R[24]. Significance of the associations were determined after correction for multiple testing based on the total number of associations over all tissues.

Supplemental Tables

Supplemental Table 1: Correlation of the top 4 associated CpGs with the estimated number of pack years smoking. The top 4 CpGs (associated to current tobacco smoking) were associated to estimated number of pack years smoking, using a linear regression model corrected for age, sex and hospital. *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. *FDR*: the false discovery rate adjusted Q-value of association.

			Meta-Analysis ePackYearsSmoking			
			discovery plus replication, n = 595			
CpG	**Chr:BP**	**Gene**	**Beta**	**SE**	P^{corr}	**FDR**
cg05575921	chr5:373378	*AHRR*	-0.0028	0.0008	8.88×10^{-4}	0.340
cg03991871	chr5:368399	*AHRR*	-0.0011	0.0009	0.237	0.974
cg12806681	chr5:368346	*AHRR*	-0.0004	0.0007	0.575	1.000
cg05284742	chr14:93552080	*ITPK1*	-0.0017	0.0008	0.039	0.758

Supplemental Table 2: Association of current smoking with carotid plaque histological features. Current tobacco smoking was associated to histological features of carotid plaques, using a linear- or logistic regression model where appropriate. Data are presented as model odds ratio (*OR*), 95% confidence interval (*CI*) and associated *p-value*. Smooth-muscle cells (SMCs), macrophages, and vessel density were scored quantitatively; calcification, collagen, atheroma, and intraplaque hemorrhage (IPH) were dichotomized.

Trait	OR	95% CI	P-value	N
Calcification	1.42	[1.13-1.81]	0.0034	1,840
Collagen	1.47	[1.09-1.97]	0.0112	1,839
Fat 40%	0.89	[0.68-1.16]	0.3860	1,843
Fat 10%	1.10	[0.85-1.43]	0.4525	1,843
IPH	1.11	[0.87-1.41]	0.3974	1,841

	Beta	95% CI	P-value	N
Macrophages	-0.07	[-0.11- -0.02]	0.0109	1,791
SMCs	-0.04	[-0.11- 0.01]	0.1234	1,786
Vessel density	0.06	[0.01- 0.19]	0.0244	1,655

Supplemental Table 3: Association of current tobacco smoking-associated CpGs with carotid plaque histological features. Methylation at current tobacco smoking-associated CpGs was associated to histological features of carotid plaques, using a linear- or logistic regression model where appropriate. Data are presented as model effect size (*Beta*), and standard error (*SE*) and associated *p-value*. Smooth-muscle cells (SMCs), macrophages, and vessel density were scored quantitatively; calcification, collagen, atheroma, and intraplaque hemorrhage (IPH) were dichotomized.

				Calcification			
CpG	**Chr:BP**	**Strand**	**Gene**	**Beta**	**SE**	**P_{corr}**	**FDR**
cg05575921	chr5:373378	+	*AHRR*	-0.060	0.031	0.051	0.618
cg03991871	chr5:368399	+	*AHRR*	-0.053	0.035	0.126	0.738
cg12806681	chr5:368346	+	*AHRR*	-0.065	0.024	0.007	0.306
cg05284742	chr14:93552080	-	*ITPK1*	-0.050	0.030	0.090	0.688

				Collagen			
CpG	**Chr:BP**	**Strand**	**Gene**	**Beta**	**SE**	**P_{corr}**	**FDR**
cg05575921	chr5:373378	+	*AHRR*	-0.028	0.060	0.633	0.893
cg03991871	chr5:368399	+	*AHRR*	-0.079	0.068	0.242	0.795
cg12806681	chr5:368346	+	*AHRR*	-0.058	0.046	0.208	0.780
cg05284742	chr14:93552080	-	*ITPK1*	-0.076	0.059	0.197	0.780

				Atheroma			
CpG	**Chr:BP**	**Strand**	**Gene**	**Beta**	**SE**	**P_{corr}**	**FDR**
cg05575921	chr5:373378	+	*AHRR*	-0.059	0.035	0.097	0.802
cg03991871	chr5:368399	+	*AHRR*	0.021	0.040	0.598	0.976
cg12806681	chr5:368346	+	*AHRR*	-0.042	0.028	0.133	0.865
cg05284742	chr14:93552080	-	*ITPK1*	0.002	0.033	0.958	0.996

				IPH			
CpG	**Chr:BP**	**Strand**	**Gene**	**Beta**	**SE**	**P_{corr}**	**FDR**
cg05575921	chr5:373378	+	*AHRR*	-0.011	0.032	0.729	0.996

cg03991871	chr5:368399	+	*AHRR*	0.025	0.037	0.494	0.996
cg12806681	chr5:368346	+	*AHRR*	0.044	0.025	0.078	0.844
cg05284742	chr14:93552080	-	*ITPK1*	0.005	0.031	0.880	0.996

				Macrophages			
CpG	**Chr:BP**	**Strand**	**Gene**	**Beta**	**SE**	**P_{corr}**	**FDR**
cg05575921	chr5:373378	+	*AHRR*	0.002	0.030	0.938	0.981
cg03991871	chr5:368399	+	*AHRR*	-0.018	0.034	0.600	0.914
cg12806681	chr5:368346	+	*AHRR*	-0.045	0.024	0.057	0.778
cg05284742	chr14:93552080	-	*ITPK1*	-0.029	0.029	0.315	0.883

				SMCs			
CpG	**Chr:BP**	**Strand**	**Gene**	**Beta**	**SE**	**Pcorr**	**FDR**
cg05575921	chr5:373378	+	*AHRR*	-0.053	0.063	0.398	0.940
cg03991871	chr5:368399	+	*AHRR*	-0.031	0.071	0.664	0.957
cg12806681	chr5:368346	+	*AHRR*	0.016	0.049	0.739	0.976
cg05284742	chr14:93552080	-	*ITPK1*	0.011	0.059	0.847	0.990

				Vessel density			
CpG	**Chr:BP**	**Strand**	**Gene**	**Beta**	**SE**	**P_{corr}**	**FDR**
cg05575921	chr5:373378	+	*AHRR*	-0.404	0.239	0.091	0.607
cg03991871	chr5:368399	+	*AHRR*	-0.638	0.261	0.015	0.607
cg12806681	chr5:368346	+	*AHRR*	-0.280	0.170	0.099	0.607
cg05284742	chr14:93552080	-	*ITPK1*	-0.264	0.257	0.305	0.712

Supplemental Table 4: Patient characteristics of 89 blood samples in the discovery dataset. Patient characteristics at time of inclusion in the dataset, 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. *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.

	Discovery			
	(AEMS450K1 – blood, n = 93)			
Characteristic	**Former smokers [n = 53]**	**Current smokers [n = 40]**	**Missing %**	**P**
Age (years [s.e.])	69.1 [8.9]	65.2 [7.1]	*0.0%*	*0.023*
Males (%)	66.0	67.5	*0.0%*	1.000
SBP (mmHg [s.e.])	155.7 [21.3]	155.9 [25.6]	*6.5%*	0.967
DBP (mmHg [s.e.])	84.2 [10.1]	79.7 [10.5]	*6.5%*	*0.049*
*eGFR (mL/min/1.73*m^2 *[s.e.])*	74.9 [19.8]	79.2 [23.6]	*1.1%*	0.340
BMI (kg/m^2 *[s.e.])*	27.2 [4.6]	25.3 [4.2]	*5.4%*	*0.046*
ePackyears (years [s.e.])	18.7 [16.8]	24.3 [15.6]	*5.4%*	0.117
Comorbidities (%)				
Diabetes	18.9	10.0	*0.0%*	0.373
Hypertension	94.3	77.5	*0.0%*	*0.037*
Medication use (%)				
Hypertensive drugs	84.9	70.0	*0.0%*	0.140
Anti-coagulants	13.2	12.5	*0.0%*	1.000
Anti-platelet drugs	90.6	92.5	*0.0%*	1.000
LLDs	71.7	72.5	*0.0%*	1.000
Symptoms† (%)			0.0%	0.566
Asymptomatic	22.6	12.5		
Ocular	9.4	15.0		
TIA	49.1	50.0		
Stroke	18.9	22.5		

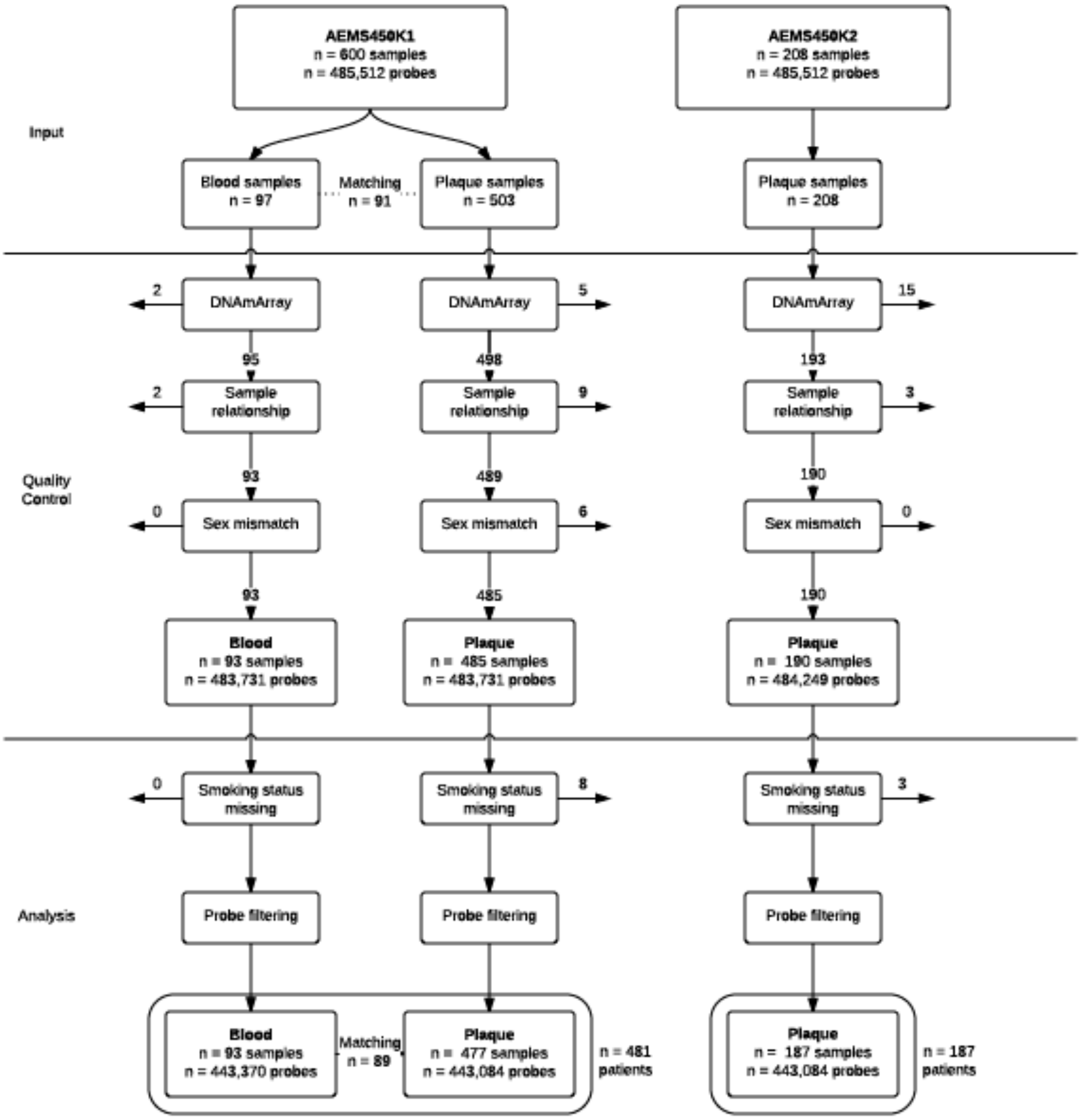
Supplemental Table 5: Association of current tobacco smoking with whole-tissue RNA expression in 30 plaques. *Mean counts*: average read count across all the samples. *Total counts*: total read counts across all samples. *Log$_2$FC*: log2-fold-change in gene expression associated to current tobacco smoking status. *SE*: standard error of log$_2$FC. P-value: associated p-value of association.

Gene	**ENSEMBLID**	**Mean Counts**	**Total Counts**	**log$_2$FC**	**SE**	**P-value**
AHRR	ENSG00000063438	1.38	41.52	1.26	1.20	0.293
ITPK1	ENSG00000100605	55.89	2432.15	-0.41	0.25	0.100

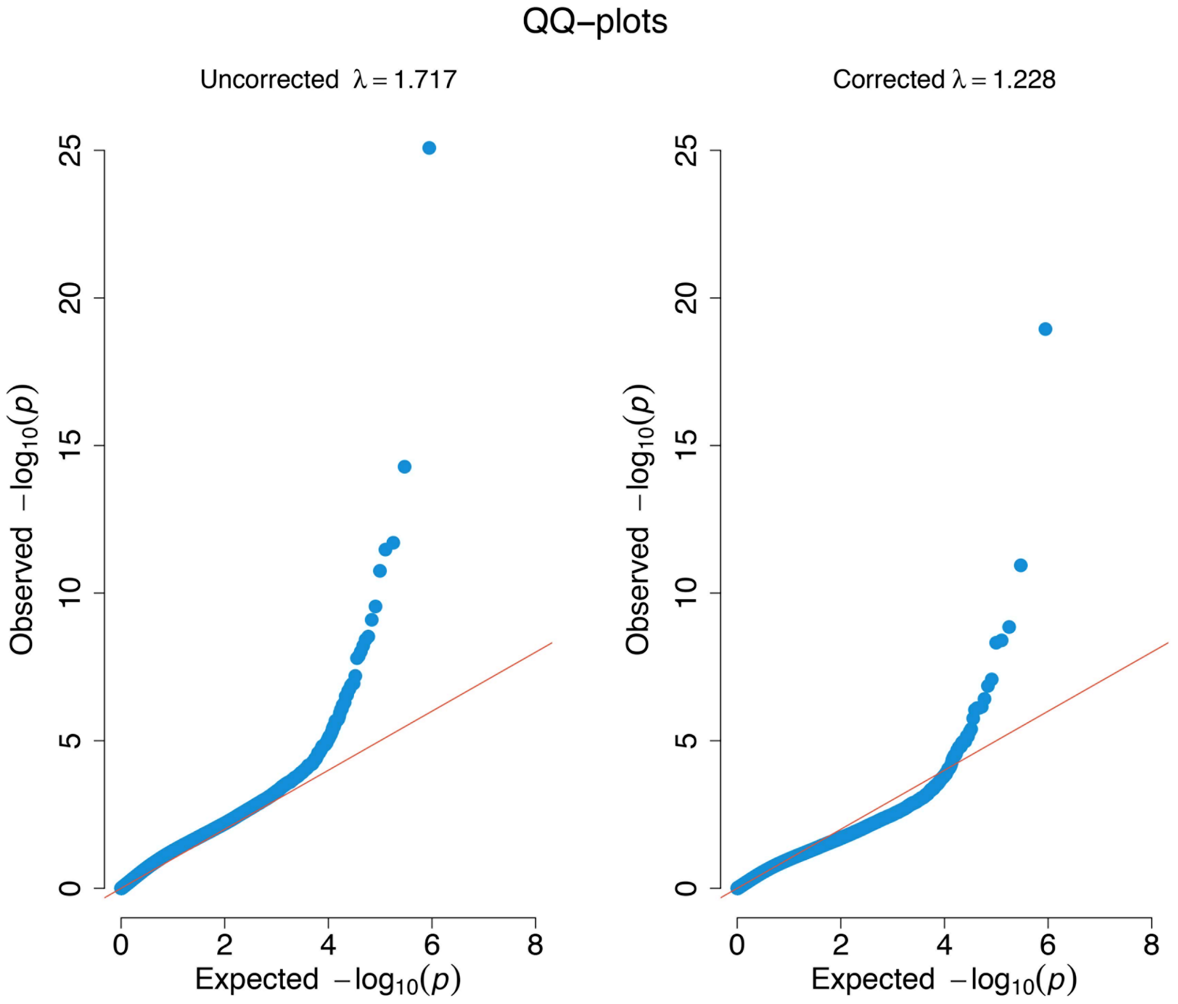
Supplemental Table 6: Gene-Gene expression associations in STAGE. Comparison of gene expressions between *AHRR* and *PLEKHG4B*, adjusted for age, sex and smoking status. *Beta*: effect size; *FDR*: false discovery rate of association; *AAW*, atherosclerotic arterial wall; *IMA*, internal mammary artery; *SM*, skeletal muscle; *SF*, subcutaneous fat; *VF*, visceral fat; *WB*, whole blood.

Tissue	**Beta**	**FDR**
AAW	0.23	0.13
IMA	0.39	3.30×10^{-4}
Liver	0.17	0.16
SF	0.70	7.00×10^{-3}
SM	0.07	0.47
VF	0.06	0.45
WB	0.15	0.03

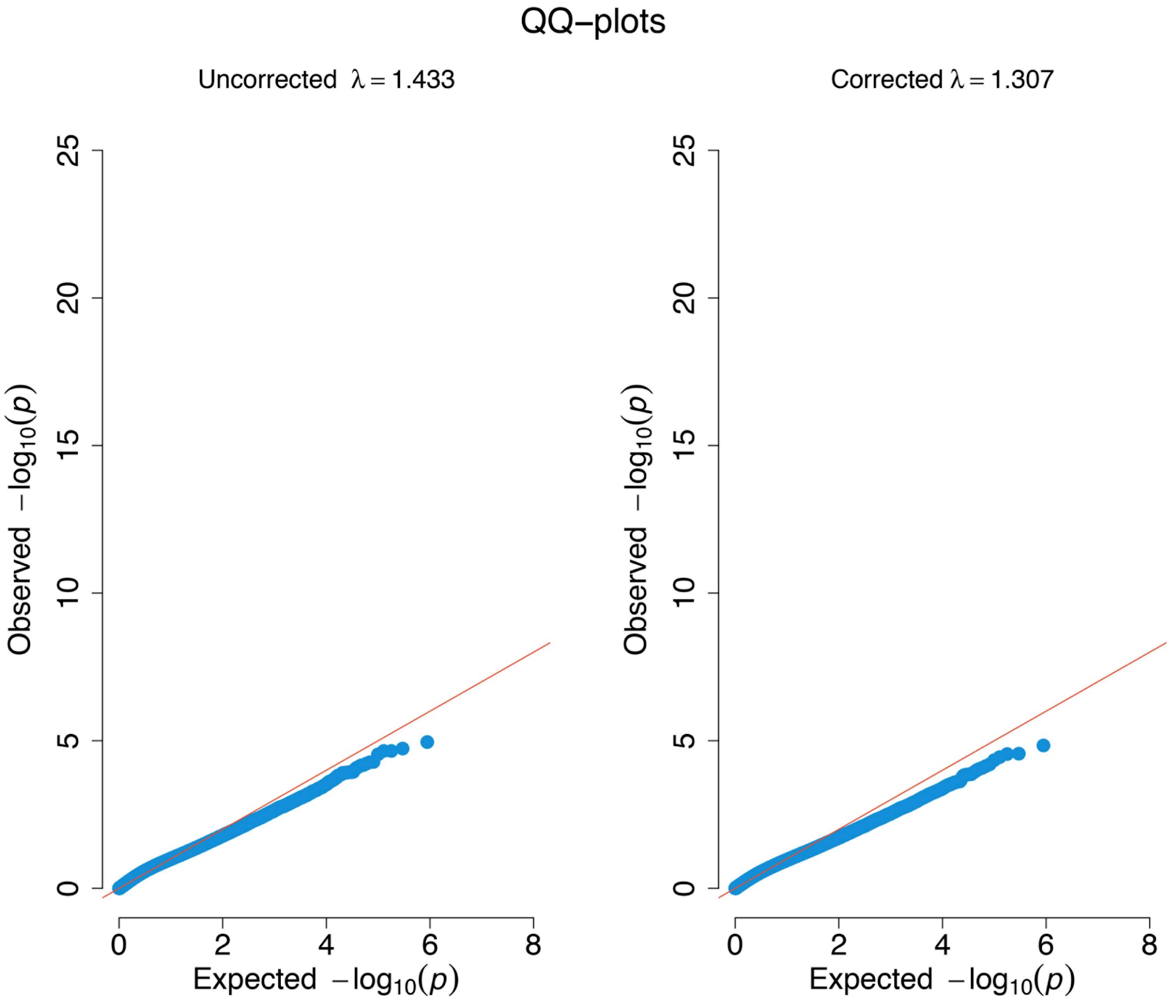
Supplemental Figures



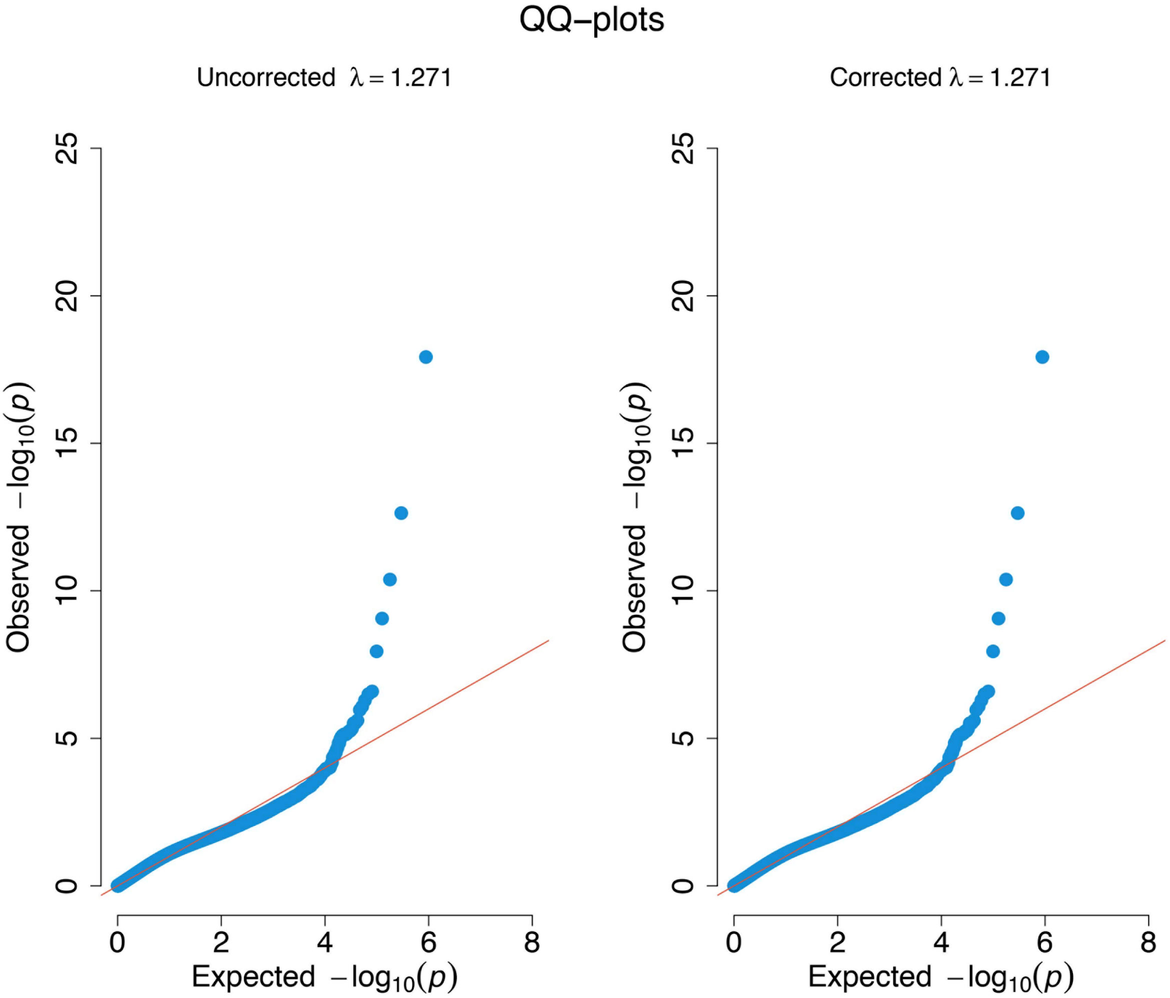
Supplemental Figure 1: Flowchart of samples used in the analysis after quality control. Flow-chart depicting the number of *input* samples, and *quality control* and *analysis* sample removal. Technical outliers were identified using DNAmArray[5] which includes MethylAid[25]. *Sample relationships* were identified through correlation of methylation data derived genotypes based on work by Chen *et al.*[26] and Zhou *et al.*[9]; where available we also compared the raw data of the 65 SNPs included on the HM450k array with those of SNP-chip derived data using the `--genome` function in PLINK[27], and samples with poor correlation (pi-hat ≤ 0.8, indicative of possible mix-up) across these 65 SNPs were excluded. In addition, sex mismatches were identified by comparing sex-chromosomes (X and Y) beta-value distribution with the sex status derived from the medical records. *Matching* shows number of patients with both plaque and blood data in AEMS450K1 (n = 89).



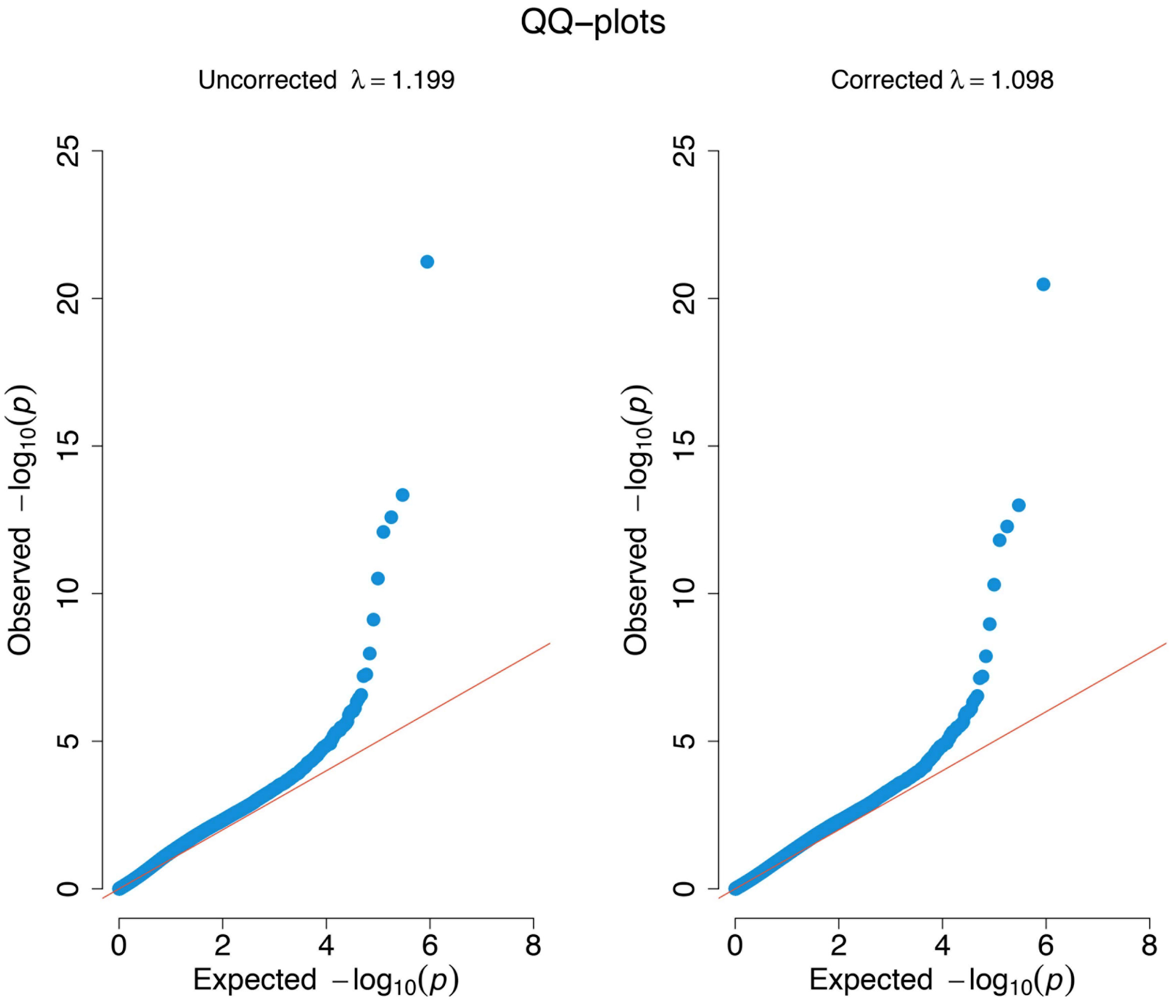
Supplemental Figure 2: Quantile-quantile plots of EWAS on current tobacco smoking with plaque-derived DNA methylation in discovery study (AEMS450K1). *Left*: QQ-plot of prior to bacon correction; *Right*: QQ-plot after bacon correction. Points show the relation between observed and expected –$\log_{10}$ (p-values) for each CpG. The solid red line shows expected p-values under the normal distribution. The blue dots show the analysis results in the discovery study (AEMS450K1). Inflation (λ) prior to correction = 1.433; after correction λ = 1.307).



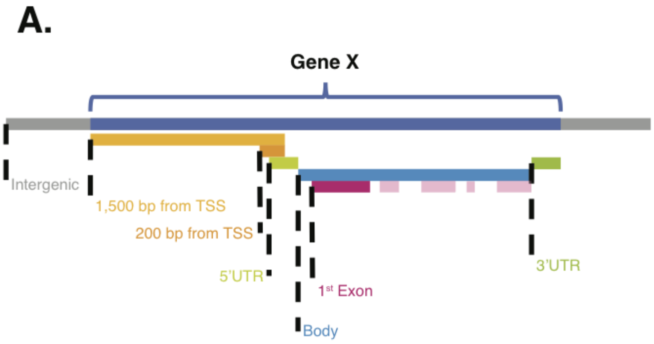
Supplemental Figure 3: Quantile-quantile plots of EWAS on current tobacco smoking with plaque-derived DNA methylation in replication study (AEMS450K2). *Left*: QQ-plot of prior to bacon correction; *Right*: QQ-plot after bacon correction. Points show the relation between observed and expected $-\log_{10}$ (p-values) for each CpG. The solid red line shows expected p-values under the normal distribution. The blue dots show the analysis results in the replication study (AEMS450K2). Inflation (λ) prior to correction = 1.433; after correction λ = 1.307).

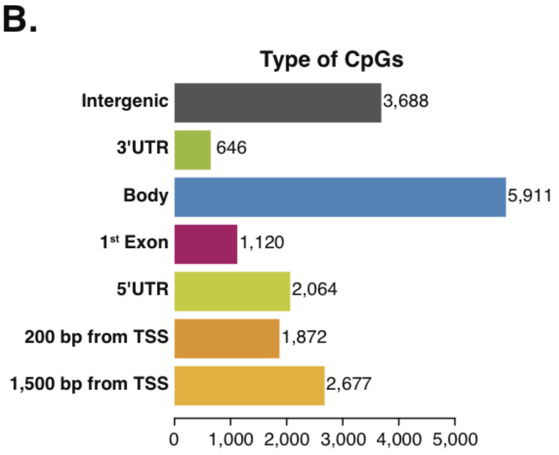


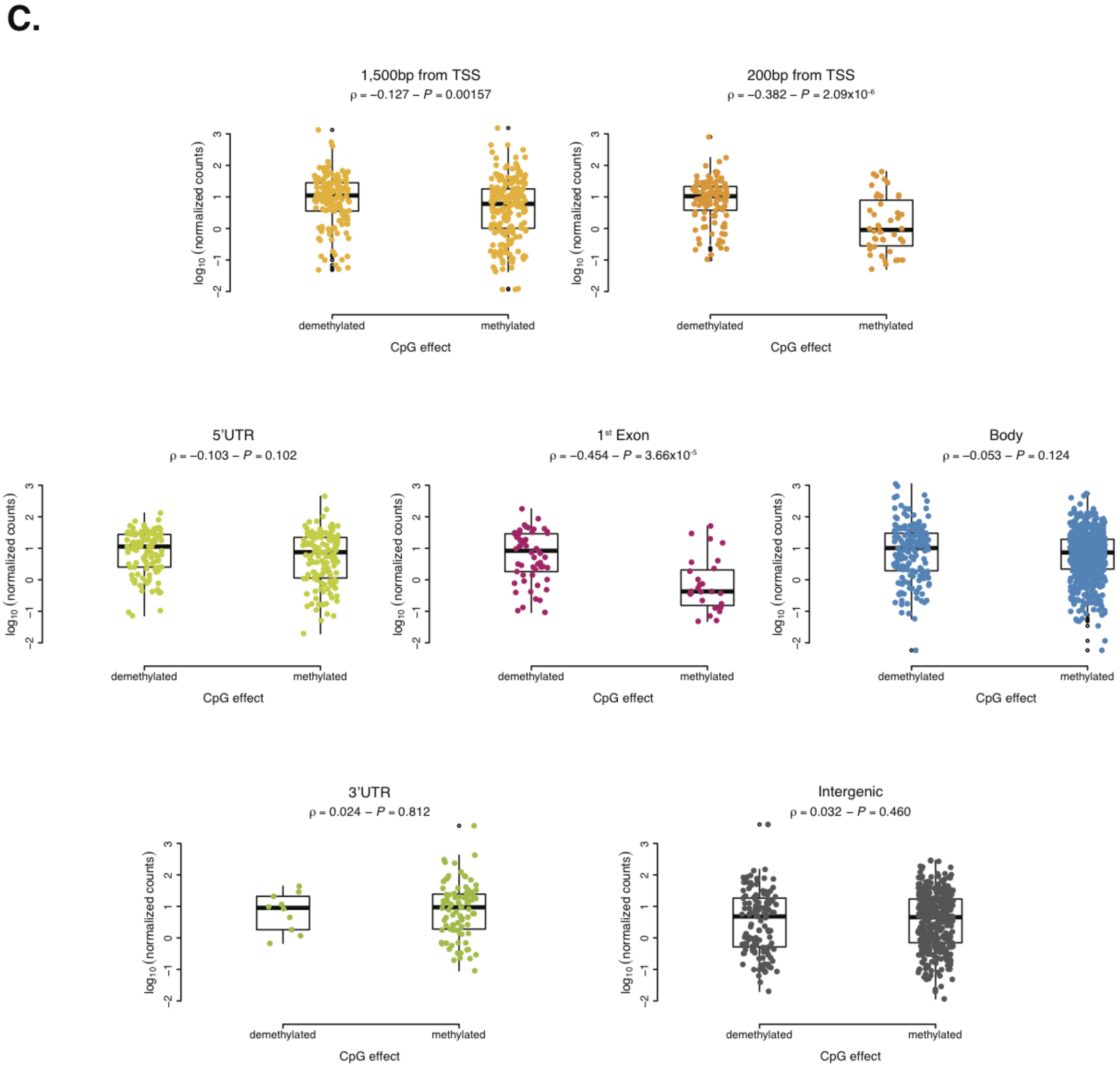
Supplemental Figure 4: Quantile-quantile plots of meta-analysis of the discovery and replication EWAS on current tobacco smoking with plaque-derived DNA methylation. *Left*: QQ-plot of prior to bacon correction; *Right*: QQ-plot after bacon correction. Points show the relation between observed and expected –$\log_{10}$ (p-values) for each CpG. The solid red line shows expected p-values under the normal distribution. The blue dots show the meta-analysis results. Inflation (λ) prior to correction = 1.717; after correction λ = 1.228).



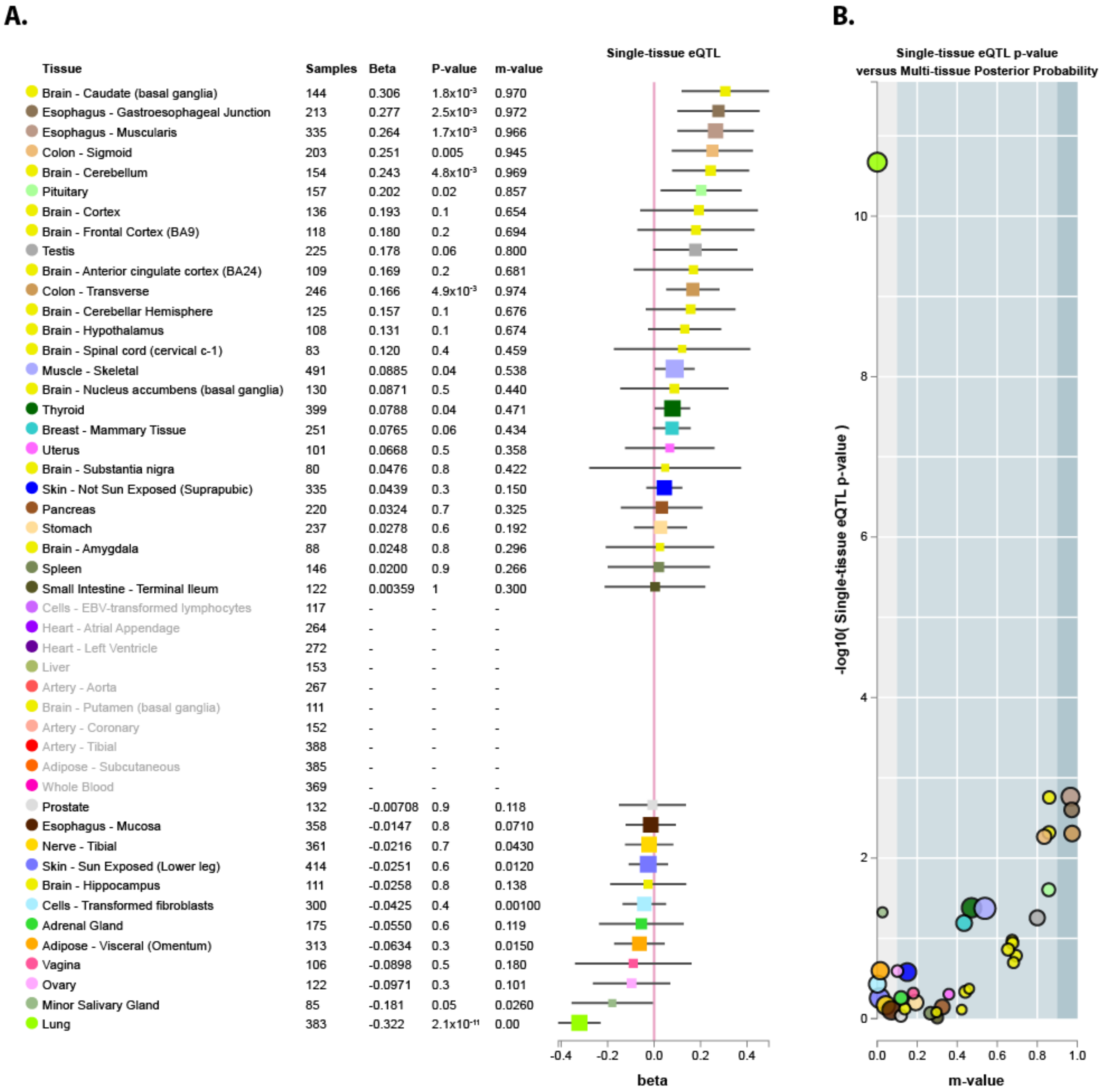
Supplemental Figure 5: Quantile-quantile plots of EWAS on current tobacco smoking with blood-derived DNA methylation. *Left*: QQ-plot of prior to bacon correction; *Right*: QQ-plot after bacon correction. Points show the relation between observed and expected –$\log_{10}$ (p-values) for each CpG. The solid red line shows expected p-values under the normal distribution. The blue dots show the results from AEMS450K1 (n = 93). Inflation (λ) prior to correction = 1.199; after correction λ = 1.097).


A.
Gene X
Intergenic
1,500 bp from TSS
200 bp from TSS
5'UTR
1st Exon
Body
3'UTR


B.
Type of CpGs
Intergenic 3,688
3'UTR 646
Body 5,911
1st Exon 1,120
5'UTR 2,064
200 bp from TSS 1,872
1,500 bp from TSS 2,677
0
1,000
2,000
3,000
4,000
5,000

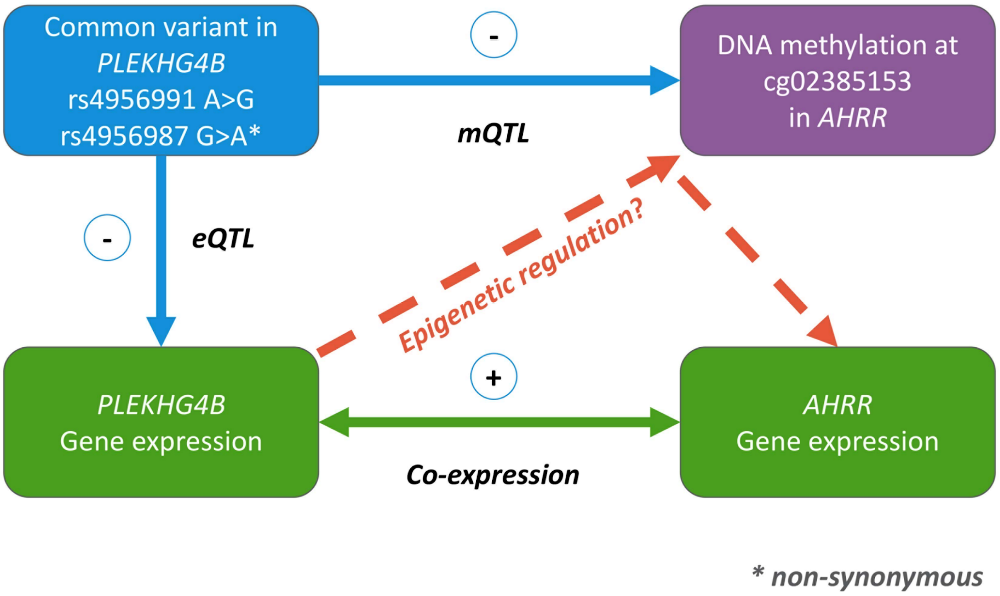

C.
1,500bp from TSS
ρ = −0.127 − P = 0.00157
200bp from TSS
ρ = −0.382 − P = 2.09x10⁻⁶
5'UTR
ρ = −0.103 − P = 0.102
1st Exon
ρ = −0.454 − P = 3.66x10⁻⁵
Body
ρ = −0.053 − P = 0.124
3'UTR
ρ = 0.024 − P = 0.812
Intergenic
ρ = 0.032 − P = 0.460
log10 (normalized counts)
−2
−1
0
1
2
3
demethylated
methylated
CpG effect

Supplemental Figure 6: The correlation in direction of effects between differential methylation and gene expression in 21 plaques. ***A.*** For an arbitrary gene, the 7 different regions are indicated to which CpGs are mapped in the Illumina Methylation 450K Annotation File (`IlluminaHumanMethylation450kanno.ilmn12.hg19`)[14]. ***B.*** For each gene region the number of mapped CpGs, nominally associated to smoking in carotid plaques after meta-analysis (p-value ≤ 0.05), are given. ***C.*** For each region, CpGs are mapped to genes and associated to the expression of the same genes. For this we calculated the median M-value per CpG of all nominal CpGs (associated to current tobacco smoking after the meta-analysis) across the 21 samples of which we also had RNAseq data. We also calculated the average read count for each gene across all 21 samples. We then mapped each CpG to a gene and subsequently grouped those CpGs per gene-region (5'UTR, body, etc.), for each of these groupings we calculated the median M-value. Thus, we obtained a per-gene-per-region CpG M-value and dichotomized these into demethylated and methylated. We performed a Wilcox-rank test to calculate the p-value (*p*) of association with the average gene read count, and calculated the correlation using Spearman's rho (ρ).



Tissue	Samples	Beta	P-value	m-value
Brain - Caudate (basal ganglia)	144	0.306	$1.8x10^{-3}$	0.970
Esophagus - Gastroesophageal Junction	213	0.277	$2.5x10^{-3}$	0.972
Esophagus - Muscularis	335	0.264	$1.7x10^{-3}$	0.966
Colon - Sigmoid	203	0.251	0.005	0.945
Brain - Cerebellum	154	0.243	$4.8x10^{-3}$	0.969
Pituitary	157	0.202	0.02	0.857
Brain - Cortex	136	0.193	0.1	0.654
Brain - Frontal Cortex (BA9)	118	0.180	0.2	0.694
Testis	225	0.178	0.06	0.800
Brain - Anterior cingulate cortex (BA24)	109	0.169	0.2	0.681
Colon - Transverse	246	0.166	$4.9x10^{-3}$	0.974
Brain - Cerebellar Hemisphere	125	0.157	0.1	0.676
Brain - Hypothalamus	108	0.131	0.1	0.674
Brain - Spinal cord (cervical c-1)	83	0.120	0.4	0.459
Muscle - Skeletal	491	0.0885	0.04	0.538
Brain - Nucleus accumbens (basal ganglia)	130	0.0871	0.5	0.440
Thyroid	399	0.0788	0.04	0.471
Breast - Mammary Tissue	251	0.0765	0.06	0.434
Uterus	101	0.0668	0.5	0.358
Brain - Substantia nigra	80	0.0476	0.8	0.422
Skin - Not Sun Exposed (Suprapubic)	335	0.0439	0.3	0.150
Pancreas	220	0.0324	0.7	0.325
Stomach	237	0.0278	0.6	0.192
Brain - Amygdala	88	0.0248	0.8	0.296
Spleen	146	0.0200	0.9	0.266
Small Intestine - Terminal Ileum	122	0.00359	1	0.300
Cells - EBV-transformed lymphocytes	117	-	-	-
Heart - Atrial Appendage	264	-	-	-
Heart - Left Ventricle	272	-	-	-
Liver	153	-	-	-
Artery - Aorta	267	-	-	-
Brain - Putamen (basal ganglia)	111	-	-	-
Artery - Coronary	152	-	-	-
Artery - Tibial	388	-	-	-
Adipose - Subcutaneous	385	-	-	-
Whole Blood	369	-	-	-
Prostate	132	-0.00708	0.9	0.118
Esophagus - Mucosa	358	-0.0147	0.8	0.0710
Nerve - Tibial	361	-0.0216	0.7	0.0430
Skin - Sun Exposed (Lower leg)	414	-0.0251	0.6	0.0120
Brain - Hippocampus	111	-0.0258	0.8	0.138
Cells - Transformed fibroblasts	300	-0.0425	0.4	0.00100
Adrenal Gland	175	-0.0550	0.6	0.119
Adipose - Visceral (Omentum)	313	-0.0634	0.3	0.0150
Vagina	106	-0.0898	0.5	0.180
Ovary	122	-0.0971	0.3	0.101
Minor Salivary Gland	85	-0.181	0.05	0.0260
Lung	383	-0.322	$2.1x10^{-11}$	0.00

Supplemental Figure 7: eQTL analysis of rs4956991 with PLEKHG4B expression multiple tissues from the GTEx Project. ***A.*** The forest plot shows the per-tissue correlation of rs4956991 with *PLEKHG4B* expression (ENSG00000153404.9) in various tissues (random-effects meta-analysis p-value = $6.37x10^{-16}$ across all tissues)[28]. ***B.*** Shows the METASOFT[28,29] based posterior probability that an eQTL exists in each tissue, *i.e.* a large m-value indicates that the variant is predicted to be an eQTL for *PLEKHG4B* in that tissue. Data obtained from GTEx Portal[30].



Supplemental Figure 8: Schematic view of smoking-associated CpGs with putative epigenetic gene regulation. Shows association of common variants in *PLEKHG4B*, one of which is non-synonymous encoding predicted to alter the PLEKHG4B protein, and the association of these variants with DNA methylation at *AHRR* (top). It also shows the association between *PLEKHG4B* expression and *AHRR* expression (bottom). In both situations, this may indicate gene-regulation through epigenetic mechanisms. Positive- and negative signs indicate positive- or negative direction of effect. *mQTL*, methylation quantitative trait locus; *eQTL*, expression quantitative trait locus.

Supplemental References

1. Verhoeven, B. *et al.* Athero-express: differential atherosclerotic plaque expression of mRNA and protein in relation to cardiovascular events and patient characteristics. Rationale and design. *European journal of* (2004).
2. van den Borne, P. *et al.* Leukotriene B4 levels in human atherosclerotic plaques and abdominal aortic aneurysms. *PLoS One* **9,** e86522 (2014).
3. Hellings, W. E. *et al.* Intraobserver and interobserver variability and spatial differences in histologic examination of carotid endarterectomy specimens. *J. Vasc. Surg.* **46,** 1147–1154 (2007).
4. Van Der Laan, S. W. *et al.* Variants in ALOX5, ALOX5AP and LTA4H are not associated with atherosclerotic plaque phenotypes: The Athero-Express Genomics Study. *Atherosclerosis* **239,** 528–538 (2015).
5. van Iterson, M. *Quality control, probe/sample filtering and normalization of Infinium HumanMethylation450 BeadChip data: 'The Leiden Approach'*. (2016). doi:10.5281/zenodo.158908
6. Baglama, J. & Reichel, L. Augmented Implicitly Restarted Lanczos Bidiagonalization Methods. *SIAM J. Sci. Comput.* **27,** 19–42 (2005).
7. Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15,** 503 (2014).
8. van Iterson, M., Cats, D., Hop, P., BIOS Consortium & Heijmans, B. T. omicsPrint: detection of data linkage errors in multiple omics studies. *Bioinformatics* **34,** 2142–2143 (2018).
9. Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* **45,** e22 (2017).
10. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47 (2015).
11. van Iterson, M., van Zwet, E. W., BIOS Consortium & Heijmans, B. T. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.* **18,** 19 (2017).
12. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57,** 289–300 (1995).
13. Laurie, C. C. *et al.* Quality control and quality assurance in genotypic data for genome-

wide association studies. *Genet. Epidemiol.* **34,** 591–602 (2010).

14. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
15. Boomsma, D. I. *et al.* The Genome of the Netherlands: design, and project goals. *Eur. J. Hum. Genet.* 1–7 (2013).
16. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39,** 906–913 (2007).
17. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** e1000529 (2009).
18. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
19. van der Laan, S. W. *swvanderlaan/fastQTLToolKit: Acutuncus antarcticus*. (2017). doi:10.5281/zenodo.996000
20. Ongen, H., Buil, A., Brown, A., Dermitzakis, E. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *bioRxiv* (2015).
21. Hägg, S. *et al.* Multi-organ expression profiling uncovers a gene module in coronary artery disease involving transendothelial migration of leukocytes and LIM domain binding 2: the Stockholm Atherosclerosis Gene Expression (STAGE) study. *PLoS Genet.* **5,** e1000754 (2009).
22. 1000 Genomes Project Consortium *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
23. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28,** 1353–1358 (2012).
24. R Development Core Team. R: A Language and Environment for Statistical Computing. (2012).
25. van Iterson, M. *et al.* MethylAid: visual and interactive quality control of large Illumina 450k datasets. *Bioinformatics* **30,** 3435–3437 (2014).
26. Chen, Y.-A. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8,** 203–209 (2013).
27. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
28. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88,** 586–598 (2011).

29. Han, B. & Eskin, E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet.* **8,** e1002555 (2012).
30. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–585 (2013).