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 ≥ $1.0x10^{-6}$) 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 ≤ $1.0x10^{-6}$), 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/1.73m^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		

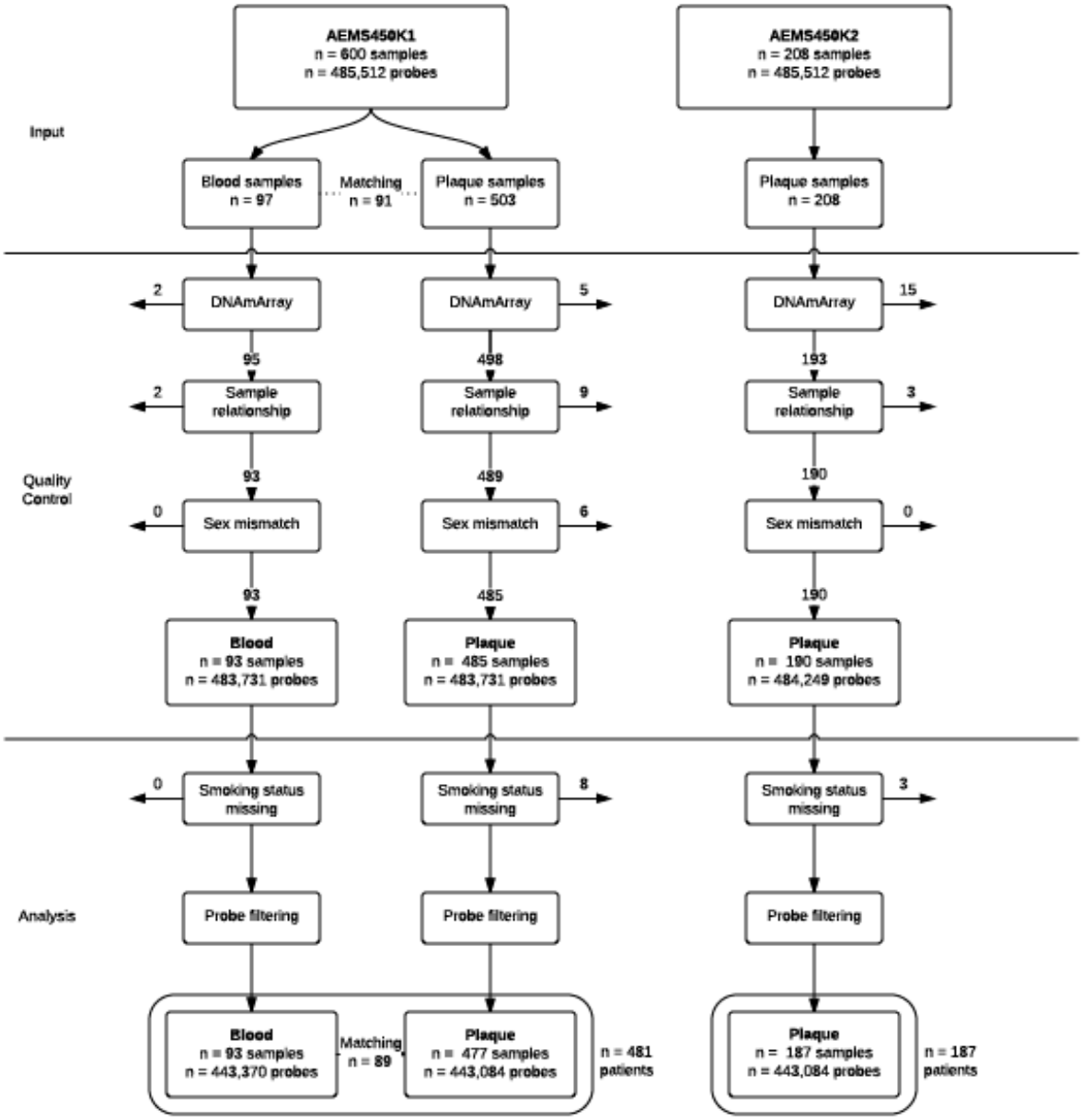
Supplemental Table 5: Association of current tobacco smoking with whole-tissue RNA expression in 30 plaques. *Mean counts*: average read count across all the samples. *Total counts*: total read counts across all samples. *Log_2FC*: log2-fold-change in gene expression associated to current tobacco smoking status. *SE*: standard error of log_2FC. P-value: associated p-value of association.

Gene	ENSEMBLID	Mean Counts	Total Counts	log_2FC	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

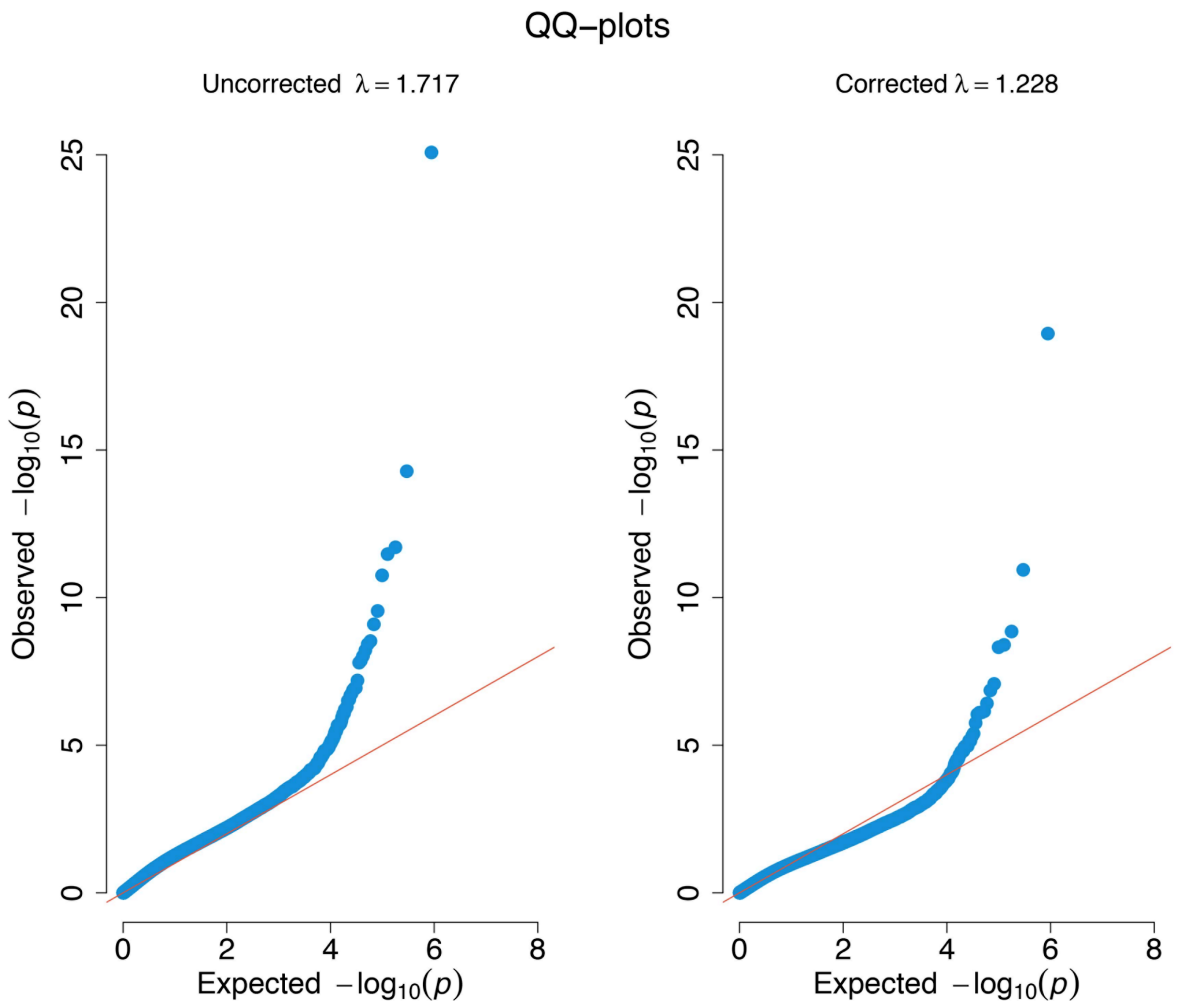
Supplemental Table 6: Gene-Gene expression associations in STAGE. Comparison of gene expressions between *AHRR* and *PLEKHG4B*, adjusted for age, sex and smoking status. *Beta*: effect size; *FDR*: false discovery rate of association; *AAW*, atherosclerotic arterial wall; *IMA*, internal mammary artery;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

Tissue	Beta	FDR
AAW	0.23	0.13
IMA	0.39	$3.30x10^{-4}$
Liver	0.17	0.16
SF	0.70	$7.00x10^{-3}$
SM	0.07	0.47
VF	0.06	0.45
WB	0.15	0.03

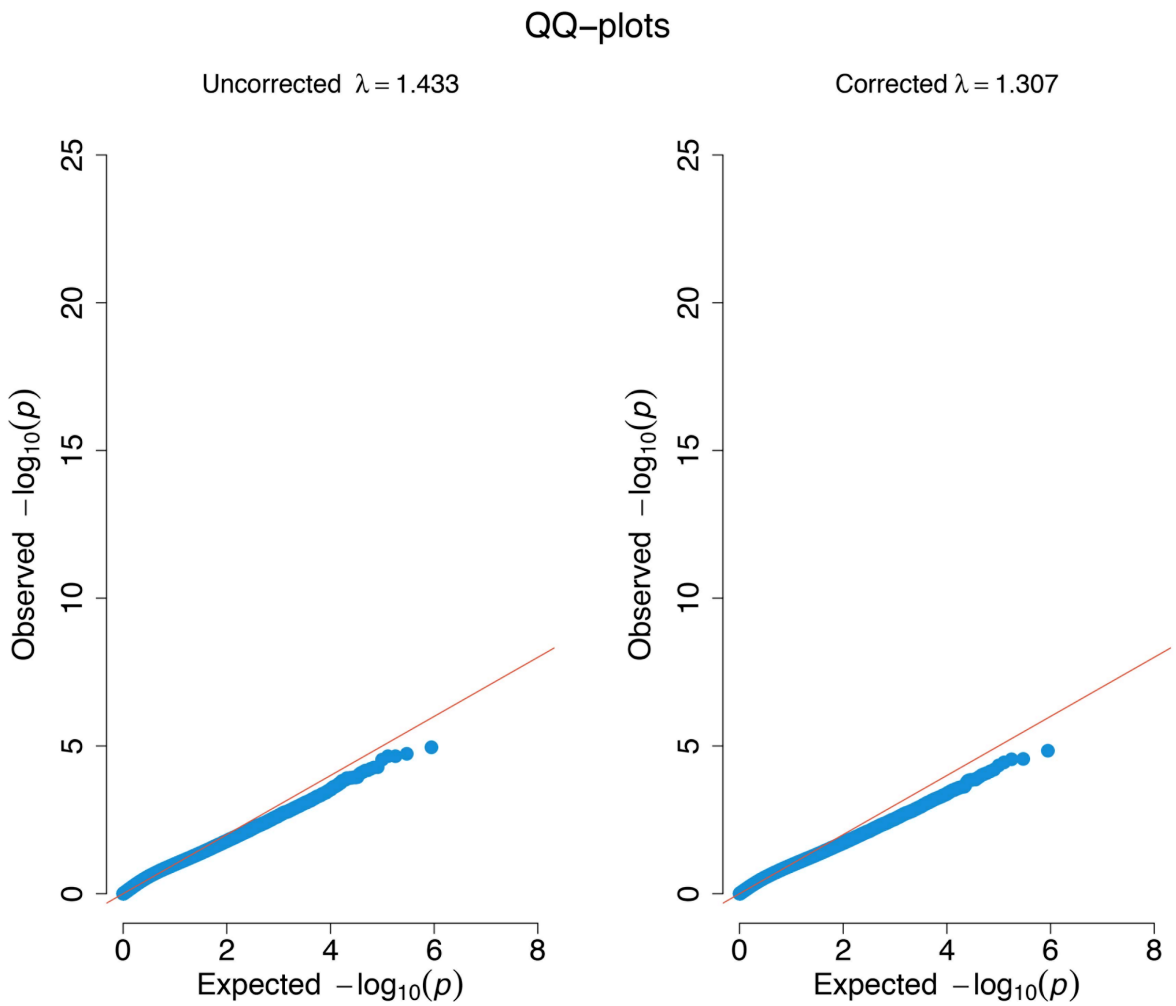
Supplemental Figures



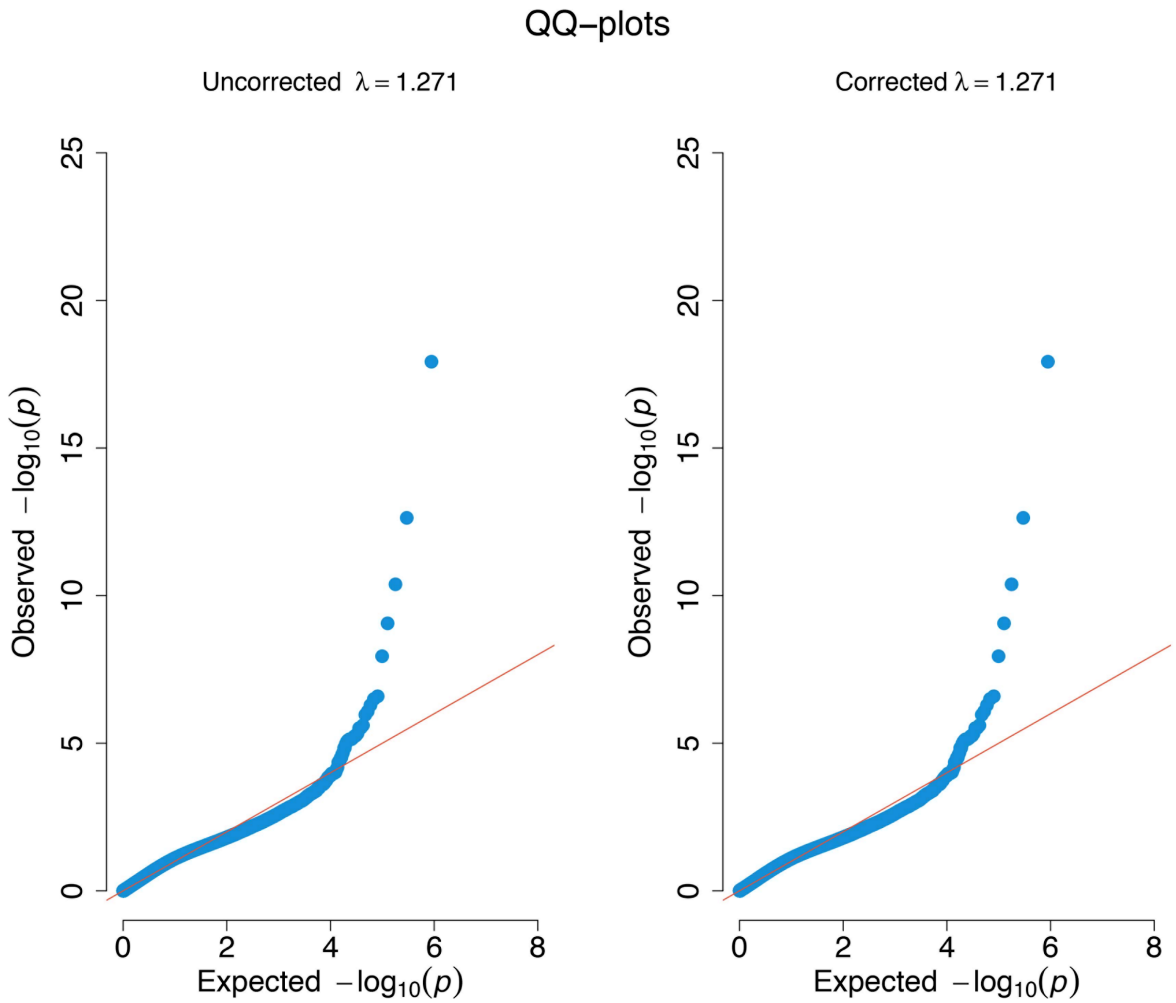
Supplemental Figure 1: Flowchart of samples used in the analysis after quality control. Flow-chart depicting the number of *input* samples, and *quality control* and *analysis* sample removal. Technical outliers were identified using DNAmArray[5] which includes MethylAid[25]. *Sample relationships* were 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).



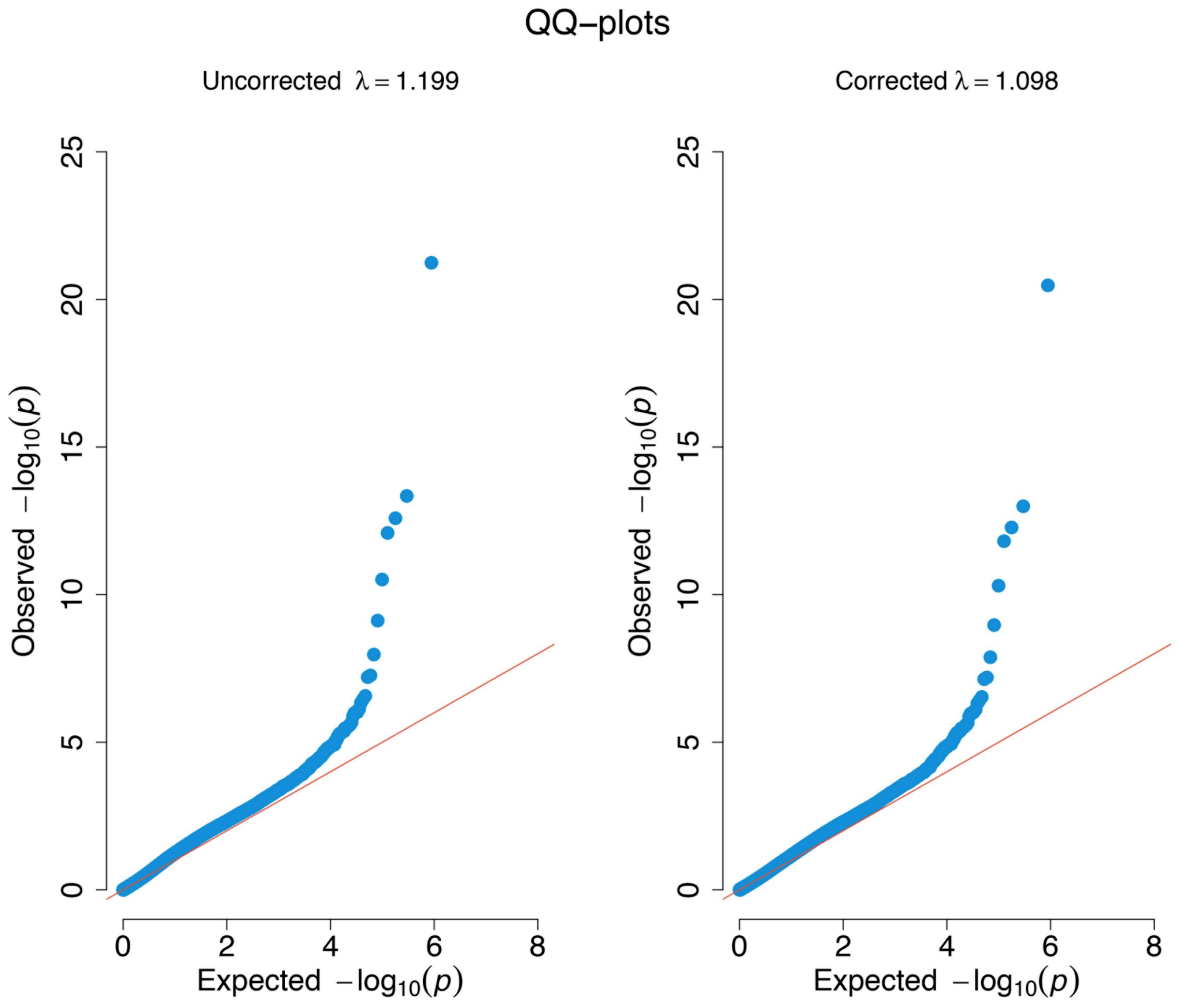
Supplemental Figure 2: Quantile-quantile plots of EWAS on current tobacco smoking with plaque-derived DNA methylation in discovery study (AEMS450K1). *Left*: QQ-plot of prior to bacon correction; *Right*: QQ-plot after bacon correction. Points show the relation between observed and expected $-\log_{10}$ (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).



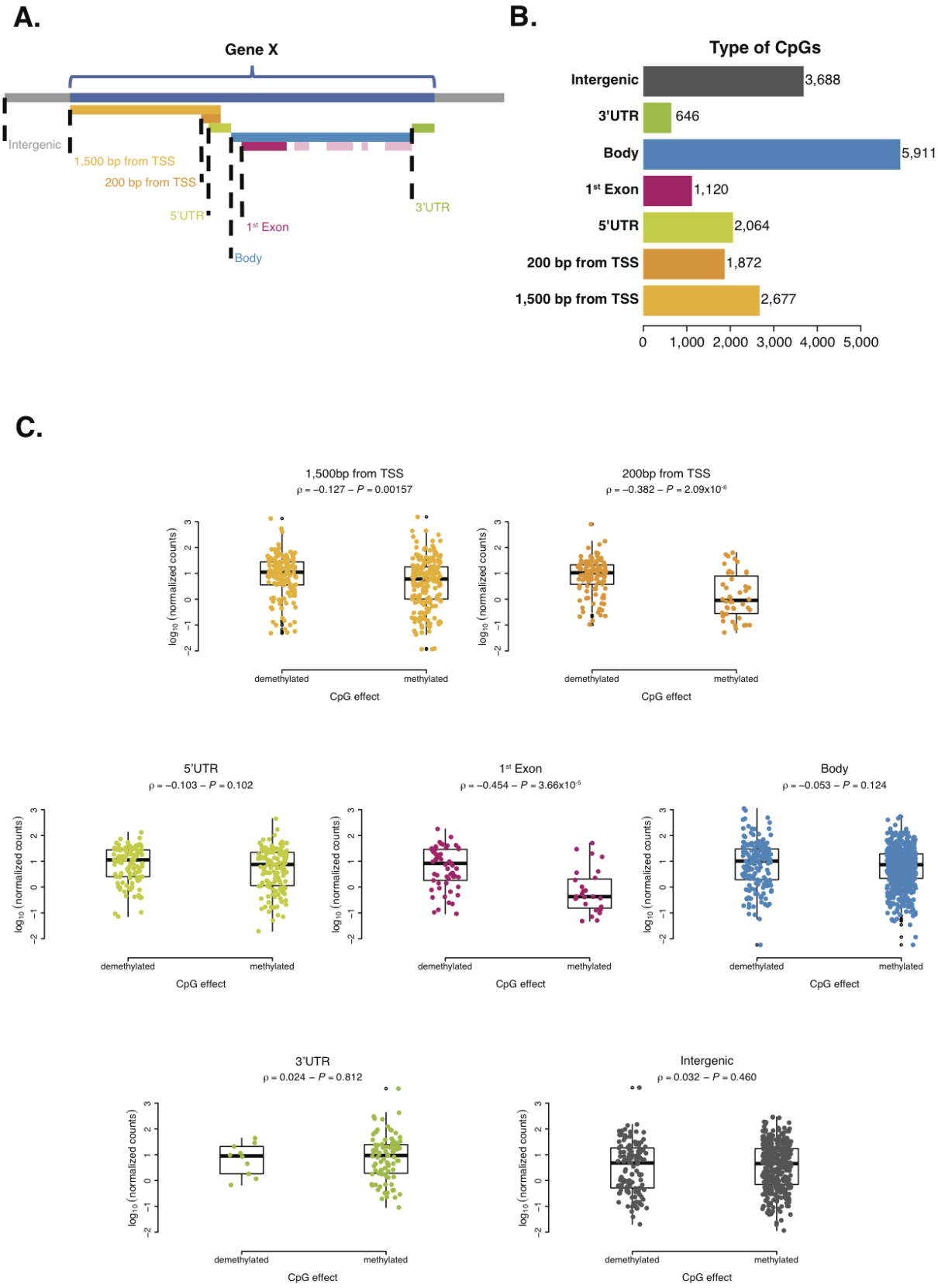
Supplemental Figure 3: Quantile-quantile plots of EWAS on current tobacco smoking with plaque-derived DNA methylation in replication study (AEMS450K2). *Left*: QQ-plot of prior to bacon correction; *Right*: QQ-plot after bacon correction. Points show the relation between observed and expected $-\log_{10}$ (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).



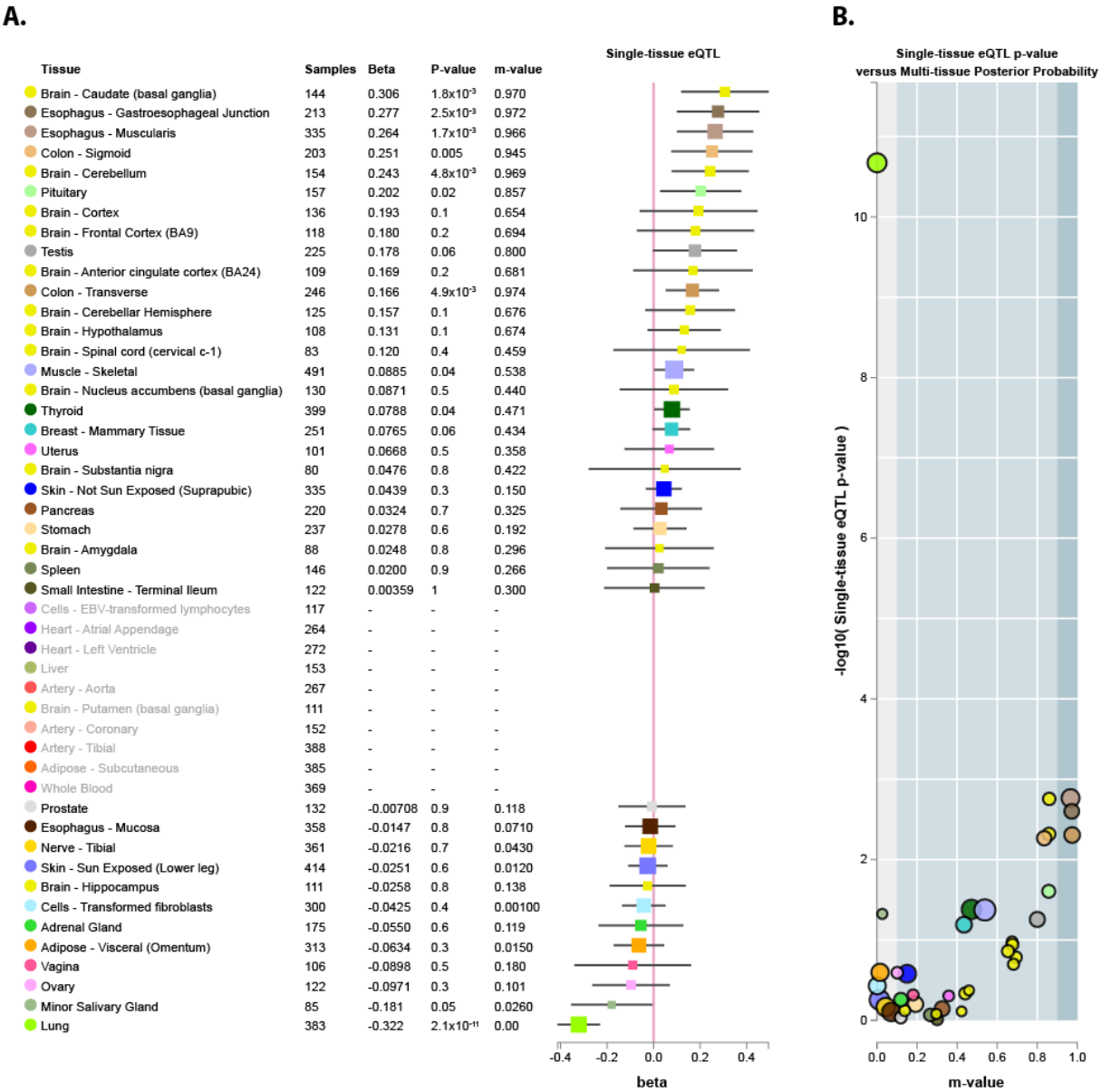
Supplemental Figure 4: Quantile-quantile plots of meta-analysis of the discovery and replication EWAS on current tobacco smoking with plaque-derived DNA methylation. *Left*: QQ-plot of prior to bacon correction; *Right*: QQ-plot after bacon correction. Points show the relation between observed and expected $-\log_{10}$ (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).



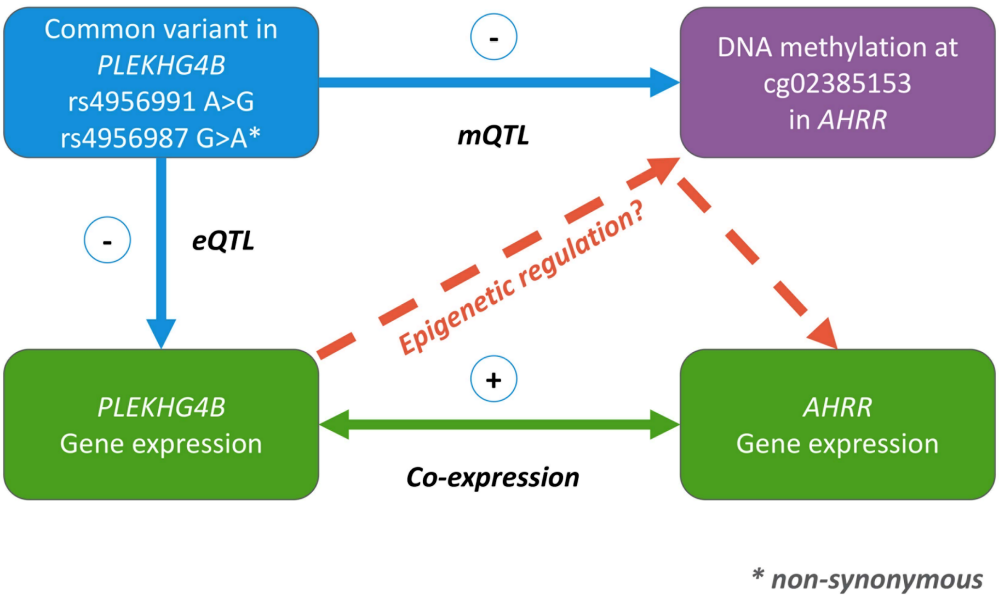
Supplemental Figure 5: Quantile-quantile plots of EWAS on current tobacco smoking with blood-derived DNA methylation. *Left*: QQ-plot of prior to bacon correction; *Right*: QQ-plot after bacon correction. Points show the relation between observed and expected $-\log_{10}$ (p-values) for each CpG. The 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)
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).



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 = 6.37×10^{-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).