SUPPLEMENTAL MATERIAL

Supplemental Methods

Purification of Monocytes

Centralized training of technicians, standardized protocols, and extensive quality control (QC) measures were implemented for collection, on-site processing, and shipment of MESA specimens, and routine calibration of equipment was performed. Blood was initially collected in sodium heparin-containing Vacutainer CPTTM cell separation tubes (Becton Dickinson, Rutherford, NJ) to separate peripheral blood mononuclear cells from other elements within 2 hours from blood draw. Subsequently, monocytes were isolated with anti-CD14 monoclonal antibody coated magnetic beads, using autoMACS automated magnetic separation unit (Miltenyi Biotec, Bergisch Gladbach, Germany). Initially flow cytometry analysis of 18 monocyte samples collected from all four MESA field centers was performed to assess the cell purification quality across the labs and technicians. The purity was > 90% for all samples.

DNA/RNA extraction

DNA and RNA were isolated from samples simultaneously using the AllPrep DNA/RNA Mini Kit (Qiagen, Inc., Hilden, Germany). DNA and RNA QC metrics included optical density (OD) measurements, using a NanoDrop spectrophotometer and evaluation of the integrity of 18s and 28s ribosomal RNA using the Agilent 2100 Bioanalyzer with RNA 6000 Nano chips (Agilent Technology, Inc., Santa Clara, CA) following manufacturer's instructions. RNA with RIN (RNA Integrity) scores > 9.0 was used for global expression microarrays. The median of RIN for our 1,264 samples was 9.9.

Epigenome-Wide Methylation Quantification

The Illumina HumanMethylation450 BeadChip and HiScan reader were used to perform the epigenome-wide methylation analysis. Illumina HumanMethylation450 BeadChip microarray was utilized (12 samples per chip). The EZ-96 DNA Methylation^{TM} Kit (Zymo Research, Orange, CA) was used for bisulfite conversation with 1μg of input DNA (at 45 μl). 4 μl of bisulfite-converted DNA were used for DNA methylation assays, following the Illumina Infinium HD Methylation protocol. This consisted of a whole genome amplification step followed by enzymatic end-point fragmentation, precipitation, and resuspension. The resuspended samples were hybridized on HumanMethylation 450 BeadChips at 48°C for 16 h. To avoid potential biases due to batch, chip, and position effects, a stratified random sampling technique was used to assign individual samples (including five common control samples for the first 480 samples) to specific BeadChips and chip position. This methylation data has been deposited in the NCBI Gene Expression Omnibus and is accessible through GEO Series accession number (GSE56046).

Quality Control and Pre-Processing of Microarray Data

Data pre-processing and quality control (QC) analyses were performed in *R* (http://www.r-project.org/) using *Bioconductor* (http://www.bioconductor.org/) packages. Bead-level methylation data were summarized in *GenomeStudio*. Because the Illumina HumanMethylation450 BeadChip technology employs a two-channel system and uses both Infinium I and II assays, normalization was performed in several steps using the *lumi* package [1]. We first adjusted for color bias using "smooth quantile normalization". Next, the data were background adjusted by subtracting the median intensity value of the negative control probes. Lastly, data were normalized across all samples by standard quantile normalization applied to the bead-type intensities and combined across Infinium I and II assays and both colors. QC measures

included checks for sex and race/ethnicity mismatches, and outlier identification by multidimensional scaling plots. The final methylation value for each methylation probe was computed as the M-value, essentially the log ratio of the methylated to the unmethylated intensity [2]. The M-value is well suited for high-level analyses and can be transformed into the beta-value, an estimate of the percent methylation of an individual CpG site that ranges from 0 to 1 (thus M is logit(beta-value)).

The Illumina HumanMethylation450 BeadChip included probes for 485K CpG sites. Statistical analyses were limited to CpG sites which passed the following filters: "detected" methylation levels <90% of MESA samples using a detection p-value cut-off of 0.05, existence of any SNPs within 10 base pairs of the targeted CpG site, or overlap with a repetitive element or region. Pre-processing with global normalization was used to remove large position and chip effects across all probes.

To estimate residual sample contamination for data analysis, we generated separate enrichment scores for neutrophils, B cells, T cells, monocytes, and natural killer cells. We implemented a Gene Set Enrichment Analysis [3] to calculate the enrichment scores using the gene signature of each blood cell type in the ranked list of expression values for each MESA sample. The cell type-specific signature genes were selected from previously defined lists [4] and passed the following additional filters: at least four-fold more highly expressed in the targeted cell type than in other cell populations and low expression levels in the targeted cells.

mRNA quantification using RNA-seq

Total RNA samples were enriched for mRNA, by depleting rRNA using the MICROBExpress kit from Ambion and following the manufacturer's instructions. Poly(A)

mRNA was enriched, and Illumina compatible, strand-specific libraries were constructed using Illumina's TruSeq Stranded mRNA HT Sample Prep Kit (Illumina, RS-122-2103). 1 ug of total RNA with RIN ≥ 8.0 was converted into a library of stranded template molecules suitable for subsequent cluster generation and sequencing by Illumina HiSeq. The libraries generated were validated using Agilent 2100 Bioanalyzer and quantitated using Quant-iT dsDNA HS Kit (Invitrogen) and qPCR. Six individually indexed cDNA libraries were pooled and sequenced on Illumina HiSeq, resulting in an average of close to 30 million reads per sample. Libraries were clustered onto flow cells using Illumina's TruSeq PE Cluster Kit v3 (PE-401-3001) and sequenced 2X100 cycles using TruSeq SBS Kit -HS (FC-401-3001) on an Illumina HiSeqTM 2500. A total of 64 lanes were run to generate approximately 30 million 2 x 101 Paired End reads per sample. The Illumina HiSeq Control Software (HCS v2.0.12) with Real Time Analysis (RTA v1.3.61) was used to provide the management and execution of the HiSeq 2500.

Illumina sequencing runs were processed to de-multiplex samples and generate FastQ files using the Illumina provided *configureBclToFastq.pl* script to automate running CASAVA 1.8.4 using default parameters for removal of sequencing reads failing the chastity filter (yes) and mismatches in the barcode read (0). Following generation of FastQ files, reads were trimmed to remove poor quality reads (or read tails) using *Btrim* (5 base sliding window average with $Q > 15$) [5] and then trimmed to remove any adaptor sequence present in the reads using custom perl scripts (trim sequences containing 11 base tag of adaptor, final length >40 bases). The *Ensembl* GRCh37 *Homo Sapiens* reference file, annotations and Bowtie2 indexes were downloaded from the *igenomes.com* website (10-Apr-2013) for mapping of the sequencing reads to the genome and read counting. *Bowtie2 (2.1.0)* and *TopHat2 (2.0.8)* were used to map the sequencing reads to the genome using a mate-inner-distance of 100 bp and '*firststrand*' options [6, 7] . Following

alignment, *bam* files were merged using the *samtools* (0.1.19) merge function [8], and read counts per gene were obtained using *HTSeq* (0.5.4p3) (http://www-huber.embl.de/users/anders/HTSeq /doc/ overview.html). The '*intersection-strict*' overlap resolution mode and '*stranded reverse*' options were used in *HTSeq*.

Data pre-processing and QC analyses were performed in *R* (http://www.r-project.org/) using *Bioconductor* (http://www.bioconductor.org/) packages. The transcript-based raw count data files for each sample from *TopHat2* were combined into a count matrix with 56,303 features (rows) and 374 MESA samples (columns). The median total count per sample was 28.8 million. Reads denoted by *TopHat2* as "no_feature","ambiguous", "too_low_aQual", "not_aligned", "alignment_not_unique" were removed. Counts were converted to Counts Per Million (CPM) using the *cpm* function of the *edgeR* package [9] , and all features with CPM ≤ 0.25 in $\geq$90% of the 374 MESA samples were removed. Features assigned to the mitochondrial genome were removed as well. Using the *biomaRt* package and querying the *Ensembl BioMart* database, *Entrez Gene ID*s, Gene Symbols, genome coordinates, gene length and percent GC content were obtained for 12,585 features which had a corresponding *Entrez ID or* Illumina HumanHT-12 v4 probe ID. To be able to continue to use the flexible and computationally efficient linear modeling functions in *R*, we transformed the raw count data to log counts per million (y = logCPM) as recommended by Law et al (2013) [10]:

$$y_{gs} = log_2\left(\frac{c_{gs} + 0.5}{T_s + 1} \cdot 10^6\right)$$

where c_{gs} is the raw count of gene transcript g in sample s, and T_s is the normalized total count of sample s, using the Trimmed Mean of M-values (TMM) normalization method [11] as implemented in the *calcNormFactors* function in the *edgeR Bioconductor* package [9]. We either

performed only this TMM normalization, or we applied quantile normalization (QN) to the logCPM values. Because the logCPM values' variance tends to decrease with increasing count for smaller counts, we used the voom function of the *limma* package [12] to estimate the mean-variance trend non-parametrically and to predict the residual variance of each individual observation for each gene. Then we incorporated the inverse residual variances into the linear modeling (*lm*) as weights in a standard manner. For the logCPM data, we imposed the same low variance filter that we had used for the microarray data, removing another 192 features with the lowest variance and retaining 12,380 features for analysis. We then performed weighted linear model analyses with the otherwise exact same models as for the microarray data.

Replication study:

Collection of samples for DNA methylation analysis

As previously described, liver samples were obtained from the PDAY study [13] from a subset of the total population, selected for a separate study, which consisted of subjects with the lowest 25th (controls) and highest 10th (cases) percentile of non-HDL cholesterol. Samples were from 72 European American and 72 African-American males, 15 to 34 years of age, who died of violent causes within 72 hours after injury and underwent autopsy in one of the cooperating medical examiners' laboratories. DNA was isolated from liver samples that had been stored at −80°C. Five hundred to 700 mg of thawed liver tissues were homogenized with a Dispomix Drive (Medic Tools AG, Switzerland) and genomic DNA extracted with a MagneSil Genomic, Large Volume system (Promega, USA) process that had been automated on a Freedom EVO liquid handler (Tecan, Switzerland). Extracted DNA was quantitated with PicoGreen reagent (Molecular Probes, USA) and verified as high molecular weight (>50 Kb) by agarose gel

electrophoresis. This study used DNA obtained from liver samples acquired at autopsy. Since all study subjects were deceased at the time of study, use of these specimens is not considered Human Subjects research.

DNA methylation analysis

As previously described [14] samples were evaluated using the Illumina HumanMethylation450 BeadChip, which assays 485,577 unique CpG sites. The average beta (essentially the ratio of the methylated to unmethylated signal) for each site was used to test for differences by genotype. Association analysis was performed in the European American and African-American samples separately using a generalized linear model (proc glm), as implemented in SAS (Cary, NC). Age, chip, and chip position were included in the model as covariates. Meta analysis was performed using METAL [15], weighting by sample size and accounting for direction of effect.

Measurement of risk factors

Risk factors were determined and have been previously described in detail [16, 17]. Briefly, determination of total serum cholesterol and high density lipoprotein cholesterol (HDL-C) was accomplished using blood collected at autopsy from the vena cava, heart, or aorta. Non-HDL-C was calculated by subtracting HDL-C from total cholesterol. Body mass index (BMI) was defined as weight (in kilograms) divided by height (in meters) squared. Glycohemoglobin was measured by affinity column chromatography (Helena Laboratories) after a sample of thawed cell hemolysate was mixed with hemolysate reagent to ensure complete lysis. The column was an insoluble cellulose resin bound to dihydroxyboryl groups with an affinity for cis diol groups present in glucose [16]. Smoking status was determined based on the levels of post-mortem serum

thiocyanate; a smoker was defined as having a serum thiocyanate level equal to or greater than 90 μmol/L [17].

Extent of fatty streaks

The methods for dissection and preservation of arteries have previously been detailed [13, 17-19]. Pathologists, blinded to demographic, clinical, and pathological observations, evaluated the right coronary arteries and left halves of the aortas. They visually estimated the extent of intimal surface involved with fatty streaks, fibrous plaques, complicated lesions, and calcified lesions by procedures developed in the International Atherosclerosis Project[20]. A fatty streak was a flat or slightly elevated intimal lesion stained by Sudan IV and without other underlying changes[17]. Consensus grading of lesions was the average of independent gradings by three pathologists.

Statistical analyses

Tests for associations between extent of fatty streaks and cg05575921 methylation measured in hepatic samples were performed using linear regression, including the following risk factors: age, race, lipids (HDL, non-HDL), BMI, and glucose levels. Analysis included 141 males aged 17 – 34, 49% African American, 51% Caucasian.

Reduced Representation Bisulfite Sequencing (RRBS):

To investigate the relationship between smoking and CpG methylation not captured by microarray, RRBS libraries of CD14+ monocytes from two smokers and two nonsmokers were generated as previously described [21] with minor modifications. Briefly, samples of 20ul genomic DNA (200ng) were spike-in with 1ul (50pg) phage lambda DNA for bisulfite conversion control. Samples were digested with Msp I (C^CGG, New England Biolabs, Cat. NO. R0106L) in a 30 ul

reaction containing 3 ul 10x NEB buffer 2, 1 ul of MspI (20U/ul) and 5 ul H_2O, and incubated at 37 °C overnight. Digested DNA samples were purified by 2x volume of AMPure XP beads (BECKMAN COULTER, Item NO. A63881) followed by end repairing, A-tailing, and indexed adapter ligation using TruSeq DNA Sample Preparation Kit v2 (illumina, Cat. NO. FC-121-2001). Ligated DNA samples were purified with 2x volume of AMPure XP beads followed by two consecutive bisulfite conversions using EpiTect Bisulfite kit (48) (Qiagen, Cat. NO. 59104). Bisulfite converted DNA samples were purified with 2.5x volume of AMPure XP beads before 9 cycles of PCR amplification for library generation. After the PCR was completed, PCR primers and adapter dimers were removed with 1.2x volume of AMPure XP beads clean-up. To further minimize adapter dimers, a second round of clean-up was conducted with 1.5x volume of AMPure XP beads. The final library DNA samples were eluted with 32.5 ul resuspension buffer from TruSeq DNA Sample Preparation Kit. Libraries were quantified by a Qubit fluorometer (Invitrogen, Cat. NO. Q32857). The distribution and concentration of DNA fragments of libraries were further examined by running High Sensitivity DNA Chip (Agilent Technologies, Kit-Reorder NO. 5067-4626) on Agilent 2100 Bioanalyzer (Cat. NO. G2938C). Four final RRBS libraries were pooled together for further QC by running on Illumina MiSeq machine before sending for HiSeq run to generate final sequencing data.

RRBS-seq Processing

We filtered the raw reads to only include those with a median Phred quality score of 20 or greater. We used Trim Galore! version 0.2.8 to trim any reads containing adapter sequence. In order to avoid counting unmethylated cytosines that were introduced during the end repair step of library preparation, we trimmed off the first two bases of every mate 2 read and trimmed off the last two bases of any mate 1 read that had been trimmed of adapter sequence. We aligned the

filtered and trimmed reads to the hg19 genome assembly using Bismark version 0.9.0. To avoid double counting bases that originated from the same fragment but were present in both paired-end reads, we trimmed all bases from any mate 2 read that overlapped with its paired mate 1 read. We extracted the methylation status of all CpGs within the reads, and we assembled a data set with the counts of methylated and unmethylated cytosines for each CpG and sample. To avoid spurious results due to SNP differences being quantified as methylation differences, we eliminated any CpGs that were located at known common SNPs, which were defined as being present in dbSNP build 138 at a population frequency of 1% or greater.

RRBS - DMR Identification

To identify genomic regions that are differentially methylated (DMRs) between smokers and nonsmokers, we used the method described in Ziller *et al.* [22]. Briefly, this method first uses the methylated and unmethylated counts to identify a set of dynamic CpGs, which are CpGs that are significantly differentially methylated between groups. To do this, it uses a model based on the beta difference distribution. Next, it merges any dynamic CpGs whose genomic locations are close to each other into a set of CpG clusters. Finally, all clusters that are significantly differentially methylated between groups are retained as the final set of DMRs. We defined the DMR location as the region between the outermost CpGs in the cluster. We excluded any DMRs that had fewer than 5 CpGs with nonzero methylation counts, and we also excluded any DMRs that had an overall methylation difference of less than 10%.

RT-qPCR:

Reverse Transcription Quantitative Polymerase Chain Reaction (RTqPCR) was used for replication of *AHRR* gene expression in CD14+ monocytes from nonsmokers and smokers

collected at the NIEHS Clinical Research Unit. CD14+ monocytes were isolated from whole blood using CD14+ antibody coated magnetic beads (Life Technologies). RNA and DNA was isolated using the ALLPrep DNA/RNA/miRNA Universal Kit (Qiagen, Catalog # 80224). cDNA was generated using the SuperScript® III First-Strand Synthesis SuperMix for qRT-PCR (Life Technologies, Catalog# 11752-050). For each individual RNA sample, target and reference genes (AHRR and ß-actin) were amplified in triplicate using FAM probes (Life Technologies FAM-MGB Catalog # 4331182, AHRR: Hs01005075_m1; ß-actin: Hs01060665_g1) designed to span exon junctions using PCR Master Mix (Life Technologies), and ABI 7900HT Realtime PCR machine.

Supplemental Tables

Supplemental Table 1: Association of cg05575921 methylation and other model covariates with carotid plaque score in 472 current or former smokers

	Beta	Lower CI	Upper CI	Std. Beta	$R^{2\,\dagger}$	t	p-value	sig
(Intercept)	-1.453	-2.257	-0.650			-3.56	4.16E-04	***
cg05575921 methylation	-0.127	-0.224	-0.030	-0.13	0.009	-2.58	0.01	*
Current smoker	-0.095	-0.346	0.156	-0.04	0.000	-0.74	0.46	
Pack-years	0.005	0.001	0.008	0.13	0.010	2.71	7.04E-03	**
Urine cotinine	8.75E-06	-2.21E-05	3.96E-05	0.03	0.000	0.56	0.58	
Age	0.031	0.023	0.038	0.37	0.105	8.03	8.32E-15	***
Gender (male)	0.097	-0.037	0.231	0.06	0.002	1.43	0.15	
Race (AFA)	-0.307	-0.493	-0.120	-0.16	0.015	-3.22	1.35E-03	**
Race (HIS)	-0.113	-0.270	0.043	-0.07	0.002	-1.42	0.16	
BMI	-0.004	-0.017	0.008	-0.03	0.000	-0.71	0.48	
LDL cholesterol	0.004	0.001	0.006	0.16	0.016	3.26	1.21E-03	**
Diabetes (impaired)	0.039	-0.119	0.197	0.02	0.000	0.49	0.63	
Diabetes (untreated)	-0.051	-0.424	0.322	-0.01	0.000	-0.27	0.79	
Diabetes (treated)	0.264	0.084	0.443	0.13	0.012	2.88	4.13E-03	**
Hypertension	0.161	0.021	0.301	0.10	0.006	2.26	0.02	*
Statin use	0.248	0.099	0.397	0.16	0.016	3.27	1.17E-03	**
Site 2	-0.116	-0.405	0.172	-0.04	0.000	-0.79	0.43	
Site 3	0.014	-0.171	0.199	0.01	0.000	0.15	0.88	
Site 4	0.081	-0.104	0.265	0.05	0.000	0.86	0.39	

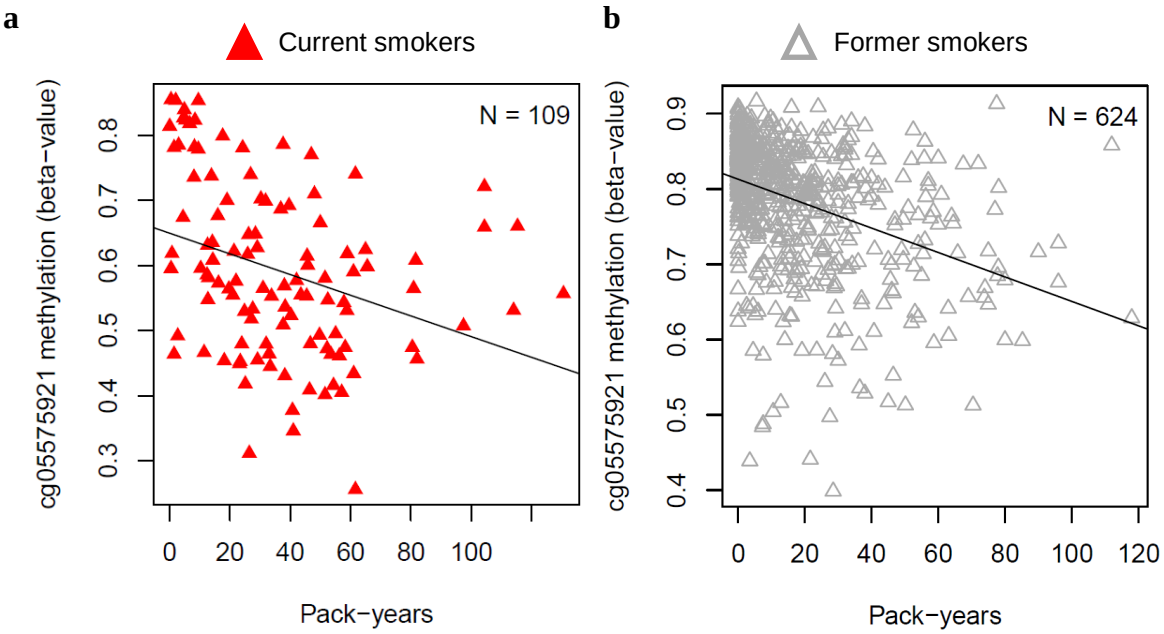
Overall model $R^2 = 0.293$; Adjusted $R^2 = 0.265$; Residual Std. Error (0.664, df = 453); F statistic 10.419 (df = 18; 453)

CI, confidence interval; Std. Beta, standardized beta; AFA, African American; HIS, Hispanic

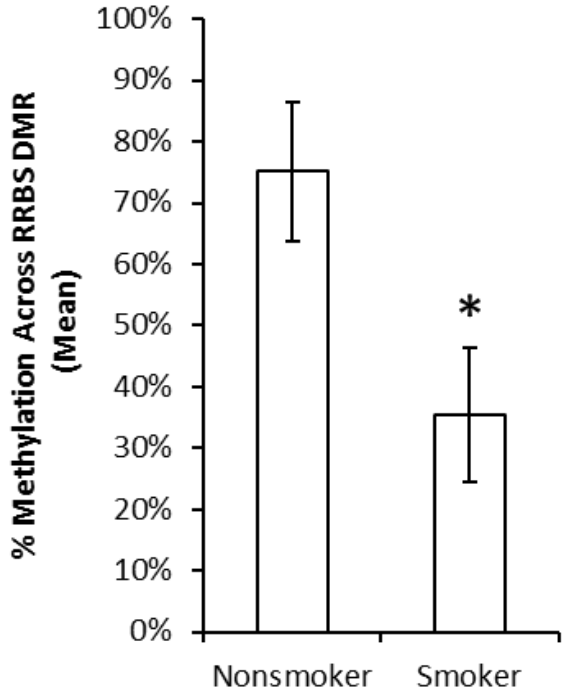
† Calculated as difference in Adjusted R^2 with and without predictor in the model

* $p < 0.1$; ** $p < 0.01$; *** $p < 0.001$

Supplemental Figures and Figure Legends

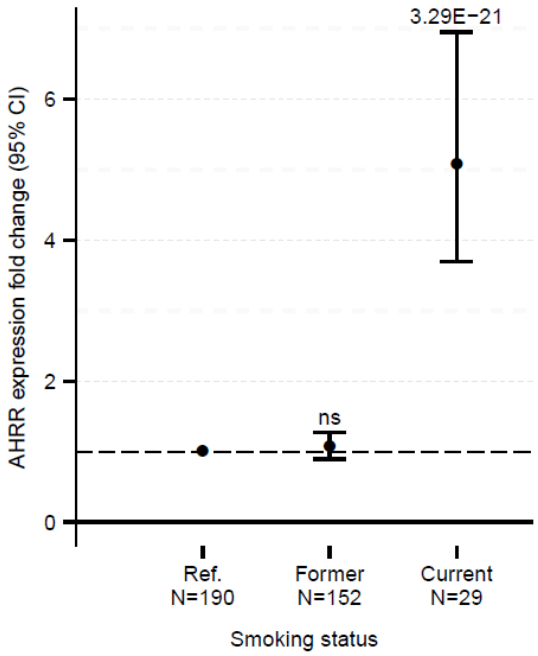


Supplemental Figure 1. DNA methylation of cg05575921 in monocytes compared to self-reported pack-years in current and former smokers. Methylation (y-axis) of the *AHRR* CpG dinucleotide cg05575921 (chr5:373,378; hg19) significantly declines with cumulative smoking exposure (x-axis: pack-years) in **(a)** current smokers (cor = -0.34, 95% CI: -0.50 – -0.17, p = 3.97×10^{-4}), and (**b**) former smokers (cor = -0.35, 95% CI: -0.42 – -0.28, p < 2×10^{-16}), after adjusting for age, sex, race, and study site.

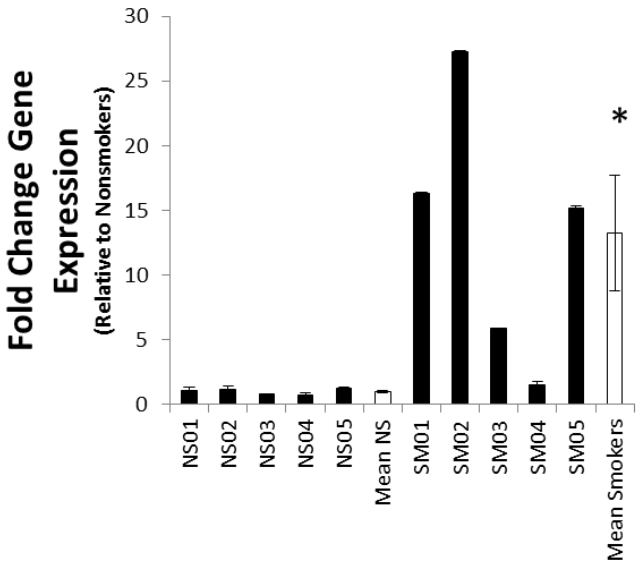


Supplemental Figure 2. Confirmation of smoking-associated methylation of *AHRR* in monocytes. Plotted is the average CpG methylation measured for the differentially methylated region (DMR) in *AHRR* (chr5:373378-373556 hg19) containing 8 CpG sites (including cg05575921), that was identified using Reduced Representation Bisulfite Sequencing (RRBS, >20X coverage for DMR CpGs in all samples) in non-MESA monocyte samples. Data shown are the mean values from two non-smokers and two smokers ($p= 2.39 \times 10^{-5}$).

a



b



Supplemental Figure 3. Up-regulation of *AHRR* expression in monocytes from smokers. **a)** In MESA samples, *AHRR* expression measured using RNA-sequencing was up-regulated in monocytes from current smokers compared with never smokers (Ref.), but not in former smokers compared to never smokers. **b)** In non-MESA samples, relative increases in *AHRR* expression between 2 and 26-fold (mean FC = 13.2; p=0.0025) were detected in monocytes from smokers (SM, n = 5) compared to non-smokers (NS, n = 5) using real-time polymerase chain reaction (RT-PCR).

Supplemental References:

1. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* 2008;24:1547-1548.

2. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC. Bioinformatics.* 2010;11:587.

3. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 2005;102:15545-15550.

4. Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F et al. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun.* 2005;6:319-331.

5. Kong Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics.* 2011;98:152-153.

6. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10:R25.

7. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25:1105-1111.

8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078-2079.

9. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139-140.

10. Law CW, Chen Y, Shi W, Smyth GK. Voom! Precision weights unlock linear model analysis tools for RNA-seq read counts. Preprint available at http://www.statsci.org/smyth/pubs/

11. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11:R25.

12. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*.New York: Springer; 2005. p. 397-420.

13. Wissler RW. USA Multicenter Study of the pathobiology of atherosclerosis in youth. *Ann. N. Y. Acad Sci.* 1991;623:26-39.

14. Bressler J, Shimmin LC, Boerwinkle E, Hixson JE. Global DNA methylation and risk of subclinical atherosclerosis in young adults: the Pathobiological Determinants of Atherosclerosis in Youth (PDAY) study. *Atherosclerosis.* 2011;219:958-962.

15. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26:2190-2191.

16. McGill HC, Jr., McMahan CA, Malcom GT, Oalmann MC, Strong JP. Relation of glycohemoglobin and adiposity to atherosclerosis in youth. Pathobiological Determinants of Atherosclerosis in Youth (PDAY) Research Group. *Arterioscler. Thromb. Vasc. Biol.* 1995;15:431-440.

17. McGill HC, Jr., McMahan CA, Malcom GT, Oalmann MC, Strong JP. Effects of serum lipoproteins and smoking on atherosclerosis in young men and women. The PDAY Research Group. Pathobiological Determinants of Atherosclerosis in Youth. *Arterioscler. Thromb. Vasc. Biol.* 1997;17:95-106.

18. Strong JP, Malcom GT, McMahan CA, Tracy RE, Newman WP, III, Herderick EE et al. Prevalence and extent of atherosclerosis in adolescents and young adults: implications for prevention from the Pathobiological Determinants of Atherosclerosis in Youth Study. *JAMA.* 1999;281:727-735.

19. Homma S, Troxclair DA, Zieske AW, Malcom GT, Strong JP. Histological changes and risk factor associations in type 2 atherosclerotic lesions (fatty streaks) in young adults. *Atherosclerosis.* 2011;219:184-190.

20. Guzman MA, McMahan CA, McGill HC, Jr., Strong JP, Tejada C, Restrepo C et al. Selected methodologic aspects of the International Atherosclerosis Project. *Lab Invest.* 1968;18:479-497.

21. Boyle P, Clement K, Gu H, Smith ZD, Ziller M, Fostel JL et al. Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol.* 2012;13:R92.

22. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature.* 2013;500:477-481.