Published in final edited form as:

Hum Genet. 2013 September ; 132(9): 1027–1037. doi:10.1007/s00439-013-1311-6.

Epigenomic Association Analysis Identifies Smoking-related DNA Methylation Sites in African Americans

Yan V. Sun^{1,2}, Alicia K. Smith³, Karen N. Conneely⁴, Qiuzhi Chang¹, Weiyan Li¹, Alicia Lazarus⁵, Jennier A. Smith⁵, Lynn M. Almli³, Elisabeth B. Binder^{3,6}, Torsten Klengel⁶, Dorthie Cross³, Stephen T. Turner⁷, Kerry J. Ressler³, and Sharon L.R. Kardia⁵

¹Department of Epidemiology, Rollins School of Public Health, Emory University

²Department of Biomedical Informatics, School of Medicine, Emory University

³Departments of Psychiatry and Behavioral Sciences, School of Medicine, Emory University

⁴Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA, USA

⁵Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA

⁶Max-Planck Institute of Psychiatry, Munich, Germany

⁷Division of Nephrology and Hypertension, Mayo Clinic, Rochester, Minnesota, USA

Abstract

Cigarette smoking is an environmental risk factor for many chronic diseases, and disease risk can often be managed by smoking control. Smoking can induce cellular and molecular changes, including epigenetic modification, but the short-term and long-term epigenetic modifications caused by cigarette smoking at the gene level have not been well understood. Recent studies have identified smoking-related DNA methylation (DNAm) sites in Caucasians. To determine whether the same DNAm sites associate with smoking in African Americans, and to identify novel smoking-related DNAm sites, we conducted a methylome-wide association study of cigarette smoking using a discovery sample of 972 African Americans, and a replication sample of 239 African Americans with two array-based methods. Among fifteen DNAm sites significantly associated with smoking after correction for multiple testing in our discovery sample, five DNAm sites are replicated in an independent cohort, and fourteen sites in the replication sample have effects in the same direction as in the discovery sample. The top two smoking-related DNAm sites in *F2RL3* (factor II receptor-like 3) and *GPR15* (G-protein-coupled receptor 15) observed in African Americans are consistent with previous findings in Caucasians. The associations between the replicated DNAm sites and smoking remain significant after adjusting for genetic background. Despite the distinct genetic background between African Americans and Caucasians, the DNAm from the two ethnic groups shares common associations with cigarette smoking, which suggests a common molecular mechanism of epigenetic modification influenced by environmental exposure.

Corresponding author: Yan V. Sun, Ph.D., Department of Epidemiology, Emory University, 1518 Clifton Road NE #3049, Atlanta, GA 30322, Phone: (404) 727-9090, Fax: (404) 727-8737, yvsun@emory.edu.

Conflict of Interest

The authors declare that they have no conflict of interest.

Supplementary Table 1: Significant smoking-related DNAm sites with FDR q-value less than 0.05.

Supplementary Table 2: Replication of smoking-related DNAm on the level of genes using human methylation 450k Array.

Keywords

Methylome; epigenetic epidemiology; leukocyte; replication; cigarette smoking

Introduction

DNA methylation (DNAm) at CpG sites provides molecular adaptability (Manolio et al. 2009; Suzuki and Bird 2008; van der Maarel 2008) and complexity (Lister et al. 2009) in the human genome by allowing gene expression to respond to environmental changes. This flexibility may contribute to disease risk by dysregulating gene expression in an exposure-specific manner (Feinberg 2007, 2008; Feinberg and Tycko 2004; Jones and Baylin 2002). Persistent effects of tobacco smoking on pulmonary (Bouloukaki et al. 2011; Willemse et al. 2005), cardiovascular (Conen et al. 2011; Kawachi et al. 1993) and malignant diseases (Mani et al. 2012; Wilhelm-Benartzi et al. 2011) have been widely observed and is indicative of epigenetic reprogramming. In particular, alterations in DNA methylation at both a global (Flom et al. 2011; Furniss et al. 2008; Smith et al. 2007; Terry et al. 2008) and locus-specific (Breitling et al. 2011; Wan et al. 2012) level have been associated with various diseases and are one potential mechanism mediating the adverse effects caused by smoking.

Tobacco smoking is a leading public health concern affecting more than 1 billion people worldwide and accounts for an estimated 3 million deaths per year (Peto et al. 1996). The molecular alterations of DNA induced by tobacco smoking have been extensively studied. The prolonged impact of tobacco smoke on various malignant (Luo et al. 2011; Moore et al. 2008; Oka et al. 2009; Smith et al. 2007; Tekpli et al. 2011) and non-malignant (Braber et al. 2010; Rea et al. 2002; Willemse et al. 2005) chronic diseases has suggested a role for epigenetic reprogramming. Previous studies have identified that global DNA methylation was associated with tobacco smoking in cancer-related tissues (Furniss et al. 2008; Smith et al. 2007), however the association between global DNA methylation (e.g. LINE-1 and Alu) and tobacco smoking is not evident in normal biosamples (Figueiredo et al. 2009; Kim et al. 2010; Rusiecki et al. 2008; Zhu et al. 2012). DNAm of several genes has been investigated to explore the epigenetic consequences of smoking. For example, the DNAm sites in the promoter region of *COMT* associate with smoking status in two independent samples (Xu et al. 2010). Different methylation levels between smokers and non-smokers were also reported on the untranslated region (UTR) of *MAOA* (Philibert et al. 2010). Breitling et al. screened over 27,000 DNAm sites covering about 14,000 human genes for epigenetic association with cigarette smoking using peripheral blood. (Breitling et al. 2011) One DNAm site located in *F2RL3* (factor II receptor-like 3) was robustly associated with smoking status, and this finding was replicated in two independent samples of European ancestry using two experimental assays. Using the same array-based measurement of DNAm, a study of two Caucasian cohorts replicated the smoking-related DNAm site of *F2RL3*, and identified another DNAm site in *GPR15* (G-protein-coupled receptor 15) associated with cigarette smoking (Wan et al. 2012). Most recently, Shenker et al. used a higher density array (over 480,000 DNAm sites) to conduct an epigenome-wide association study of cigarette smoking in 374 Europeans. (Shenker et al. 2012) This study replicated the smoking-related DNAm site of *F2RL3*, and identified three additional smoking-related loci including the aryl hydrocarbon receptor repressor gene (*AHRR*). These studies demonstrated that the methylome-wide discovery approach can identify novel genes related to cigarette smoking through epigenetic mechanisms.

Although the reported smoking-related DNAm sites are statistically significant and have been independently replicated, current evidence is limited to Caucasian populations. To

determine whether the same epigenetic associations with cigarette smoking are relevant in other ethnic groups, and to discover novel smoking-related DNAm sites, we conducted a methylome-wide association study of smoking exposure in peripheral blood samples of 972 African Americans from Genetic Epidemiology Network of Arteriopathy (GENOA) study, and pursued replication in an independent cohort of 239 African Americans.

Methods

Discovery Cohort

In the GENOA study, the status of cigarette smoking was collected from a self-reported questionnaire as “smoked within the past year”, “not smoked within the past year”, “never smoked”. Two binary variables, “current smoker” (*i.e.* 1 for individuals smoked within the past year) and “ever smoker” (*i.e.* 1 for individuals smoked more than 100 cigarettes) were derived (Giovino 2002). The number of pack-years was also derived based on the number of years of smoking history and the average amount of cigarette smoking (*i.e.* multiplying the number of packs of cigarettes smoked per day by the number of years the person has smoked). Age, sex and other phenotypic data were collected from the physical examination and laboratory assessment at the time of the Phase II study visit. The GENOA study was approved by the Institutional Review Boards of all participating institutions. Each participant gave written informed consent.

DNA Methylation Data

Genomic DNA was extracted from stored peripheral leukocytes (PLC) of 1,008 GENOA (Daniels et al. 2004) Phase II African American (AA) participants, bisulfite converted and then epityped for methylation profiling of 27,578 CpG loci using the Illumina Infinium HumanMethylation27 (27K) BeadChip (Illumina, San Diego, CA) as previously described (Sun et al. 2010). On each Illumina 27K BeadChip, there are 56 control probes on each chip representing a) *sample independent* measures of staining, hybridization, target removal, and DNA extension and b) *sample dependent* measures of bisulfite conversion, G/T mismatch, non-polymorphic and negative controls. The sample independent controls allow for the evaluation of the quality of the chip processing steps, while the sample dependent controls allow for the evaluation of the performance across samples. Seven samples were removed from the analysis due to poor bisulfite conversion control efficiency, measured by bisulfite conversion control intensity of less than 4,000. An additional twenty-nine samples were removed from the association analysis due to extreme control probe values, assessed as having at least one control probe with a value of greater than 4 standard deviations from its mean value. The cleaned data set included DNAm profiles of 972 AA individuals from 493 sibships.

We removed sites with control probe values greater than 4 standard deviations from their mean values. To reduce batch effects across samples, the correlation structure among fifty-six control probes was calculated within each color channel to identify the most parsimonious subset of probes that explained the maximum amount of batch and chip variation. We conducted the normalization by linearly regressing the thirteen selected probes (5 probes in the red channel and 8 probes in the green channel) onto the intensity signals from the methylated and unmethylated bead types separately across each DNAm site. We also excluded sites that are located on the X and Y chromosomes. Sites that are deemed multimodal based on the Dip Test proposed by Hartigan (J. A. Hartigan and Hartigan 1985), using a cut-off of $p < 0.001$ on either the methylated or non-methylated signal intensities were set aside for more specialized methods of normalization that take into account their modality. Finally, we flagged the 2,984 sites identified by Chen et al. (Chen et al. 2011) as having non-specific binding probes (over 10% of the probes map to highly homologous

genomic sequences at 40 or more of their base pairs), as well as the 875 sites that have probes overlapping with SNPs reported in dbSNP, which may influence the methylation levels reported by the microarray. We separated out these nonspecific and polymorphic sites after all analyses to aid in interpretation of results. The data set used in the final analysis included 22,927 autosomal DNAm sites.

Replication Cohort

The replication sample consisted of 262 African American subjects recruited as part of the Grady Trauma Project (GTP), a large study investigating the influence of genetic and environmental factors on response to stressful life events in an urban population of low socioeconomic status (Gillespie et al. 2009). Briefly, research participants were approached in the waiting rooms of the primary care clinic or obstetrical-gynecological clinic of a large, public hospital while either waiting for their medical appointments or while waiting with others who were scheduled for medical appointments. The Kreek–McHugh–Schluger–Kellogg scale (KMSK) was used to quantify self-exposure to tobacco use (Kellogg et al. 2003). This validated psychometric instrument was used to assess frequency, duration, and amount of tobacco consumed in last 30 days as a single quantitative score. All procedures in this study were approved by the Institutional Review Boards of Emory University School of Medicine and Grady Memorial Hospital.

Genomic DNA samples were extracted from PLC, bisulfite converted and interrogated using the Infinium HumanMethylation450 (450K) BeadChip (Illumina, San Diego, CA) according to the manufacturer-specified protocol. One sample of male DNA was included on each BeadChip as a technical control throughout the experiment, and reproducibility was assessed with Pearson correlation coefficients. Consistent with the strategy detailed above, a series of sample-dependent and sample-independent control probes were used to assess the quality of each sample as well as chip processing. The QC strategy detailed below was implemented using CpGassoc (Barfield et al. 2012). Samples with probe detection call rates <90% were excluded as were those with an average intensity value of either <50% of the experiment-wide sample mean or <2,000 arbitrary units. Unsupervised hierarchical clustering revealed no outliers. In total, four samples were removed based on these criteria; the cleaned data set included DNAm profiles of 239 unrelated AA individuals. Data points with detection p-values >0.001 were set to missing, but we limited additional processing (*i.e.* elimination of probes on the X or Y chromosomes) because this dataset was to be used only for replication of smoking-associated CpG sites identified in the discovery cohort.

The principal components of genome-wide SNP genotype data from the Illumina HumanOmni Express and Omni1-Quad BeadChips (Illumina Inc.) were included in the analysis to adjust for potential population stratification. PLINK (Purcell et al. 2007) was used to generate summary statistics for basic quality control determination. The inclusion criteria were as follows: samples with call rates > 95%, and SNPs with a Hardy-Weinberg equilibrium (HWE) p-value 10^{-5} and a minor allele frequency 0.05. We also examined the genotype data for each sample using PLINK to identify any potential duplicates among the cases and controls, defined as individuals whose genotypes are identical for close to 100% of SNPs at a significance level of $p < 0.0001$. All duplicated samples were excluded prior to analysis. We also used PLINK to evaluate the identity-by-state pattern indicative of undisclosed close genetic relation to another sample, defined as a π -hat >0.12. Of the 239 subjects with DNA methylation data, none were duplicated or related to another subject. To control for population stratification, we computed principal components on 56,309 independent SNPs pruned from the 589,375 autosomal SNPs passing quality control.

Statistical Methods

For each individual sample and DNAm site, the signals from methylated (M) and unmethylated (U) bead types were used to calculate a beta value as $\beta = M/(U + M)$. In the discovery phase, we modeled beta values as the dependent variable, with the variables “current smoker” and pack-years defined above as the primary independent variables in the association analysis. Linear mixed models were implemented in a multiple regression framework to adjust for the relatedness of the studied individuals, including age and sex as covariates. A Bonferroni-corrected p-value of 0.05 (nominal p-value of 2.18×10^{-6}) and False Discovery Rate (FDR) q-value of 0.05 (Giovino 2002) were applied to adjust for multiple testing of 22,927 autosomal DNAm sites. For the DNAm sites that were found to be significantly associated with current smoking status, the beta values were plotted against smoking status, stratified by sex. We mapped smoking-related DNAm sites to known human genes based on chromosomal location (NCBI 36.1). The top ten principal components of 668,293 genome-wide SNP markers were computed among 850 GENOA AAs with both GWAS and DNAm data as previously described (Arnett et al. 2011). The first two principal components were included as covariates in the regression model to adjust for potential effects of genetic admixture on smoking-related DNAm sites in both discovery and replication cohorts.

DNAm sites from genes with one or more smoking-associated DNAm sites in the GENOA cohort were selected for replication in the GTP sample. For each DNAm site, MethLAB (Kilaru et al. 2012) was used to examine the association between the methylation beta values and KMSK tobacco use score via a linear regression of beta values (dependent variable) on KMSK score (independent variable), adjusting for age, sex, chip/positional effects and population stratification using the first two principal components computed from genome-wide SNP data (see previous sub-section).

We estimated the proportions of PLC subtypes for each sample using an algorithm developed for 27K chip data (Houseman et al.). The proportions of six different cell types, including granulocyte, monocyte, natural killer cells (NK), B cell, CD4⁺ and CD8⁺ T cells, were projected based on top 100 cell-type specific DNAm sites. We included the proportions of PLC subtypes as covariates to assess the association between DNAm and cigarette smoking in a multiple regression model.

All statistical analyses were performed in the R statistical environment version 2.14.1 (<http://www.r-project.org/>). The authors had full access to the data and take responsibility for its integrity.

Results

Among 972 AA participants from the GENOA study, 58.3% had never smoked (“never smoker”), 29.1% had smoked but not within one year before the visit (“former smoker”), and 12.6% had smoked within the past one year (“current smoker”). Males (29.3%) have higher rates of “former smoker” and “current smoker” than females (Table 1). Within the GENOA sample, age is not significantly different between males and females. Females had significantly higher BMI than males, and 80.1% of males and 83.5% of females were diagnosed as having hypertension at the time of visit.

Using stringent Bonferroni correction for multiple testing in the analyses of linear mixed models to identify smoking-related DNAm sites, we identified 15 autosomal DNAm sites significantly associated with current smoking using a Bonferroni corrected p-value of 0.05 (Table 2). There are eighty-nine DNAm sites associated with current smoking with FDR q-value less than 0.05 (Supplementary Table 1). Among 15 smoking-related DNAm sites,

86.7% (13 out of 15 sites) exhibited hypomethylation among “current smokers”. Five of these fifteen (and 23 of the 89 using FDR correction) DNAm sites also associated with current smoking in an independent AA sample with 239 participants ($p < 0.05$; Table 1 and Table 2). Although a limited number of CpG sites were tested in the replication sample, the two sites in *F2RL3* and *GRP15* demonstrated association at or near conventional levels of genome-wide significance in both the replication sample as well as the discovery sample. Three DNAm sites, cg13668129 in *HNRPU1*, cg01500140 in *LIM2* and cg11314684 in *AKT3*, had p-value less than 0.05 in the GTP cohort (Table 2). However, using a Bonferroni-corrected p-value threshold of 0.0033 (15 tests in the replication stage), these three sites were not significantly associated with current smoking status in the GTP cohort. Notably, thirteen out of fifteen DNAm sites shared the same direction of effects between the discovery and replication samples (sign test $p < 0.001$). With additional adjustment for PLC subtypes, all top associations had p-values less than 4×10^{-6} (Table 2). Eleven of the fifteen DNAm sites remained genome-wide significant (nominal p-value less than 2.18×10^{-6}) after Bonferroni correction for multiple testing.

Eighteen unique DNAm sites were previously reported in Caucasians using 27K BeadChip. These DNAm sites were reported based on the top five associations with smoking (Breitling et al.) or the significant associations with FDR-adjusted p-value less than 0.05 (Wan et al.). Eight out of these eighteen sites were also significant (Bonferroni-corrected p-value < 0.05 , nominal p-value threshold of 2.78×10^{-3}) in our AA discovery sample after correcting for multiple testing (18 tests), and had the same direction of effects (Table 3). In addition to DNAm sites located in *F2RL3* and *GRP15*, one site located on *LIM2* also passed the significance threshold in the discovery sample (*i.e.* Bonferroni-corrected p-value < 0.05), and two sites in *NCAPD3* and *ARHGAP25* had $p < 1 \times 10^{-5}$ (Table 3). In addition to the DNAm site in *F2RL3*, eight DNAm sites in three loci were significantly associated with cigarette smoking in an European population using Illumina 450K chip. (Shenker et al. 2012) In our replication cohort, all DNAm sites were significantly associated (*i.e.* Bonferroni-corrected p-value < 0.05) with current smoking status using the same 450K platform (Table 3). The most significant smoking-related DNAm site (cg05575921) located on *AHRR* with p-value of 9.51×10^{-19} .

Among the two replicated smoking-related DNAm sites, cg03636183 (*F2RL3*) and cg19859270 (*GPR15*) were both hypomethylated among current smokers (*i.e.* lower beta value) compared to non-current smokers. The associations remained significant after adjusting for population substructure using the top two principal components of GWAS data with reduced sample size ($p = 8.07 \times 10^{-35}$ for cg03636183 and $p = 5.88 \times 10^{-10}$ for cg19859270). In addition, these two DNAm sites were also associated ($p = 1.41 \times 10^{-22}$ and 5.79×10^{-11} respectively) in a dichotomous comparison of ever smokers versus never smokers. After adjusting for multiple testing, no other DNAm site was significantly associated with ever smoker using either Bonferroni or FDR correction. The same trend of methylation (never smoker $>$ former smoker $>$ current smoker) was consistent in both males and females (Figure 1). Among 405 smokers, the two DNAm sites, cg03636183 (*F2RL3*) and cg19859270 (*GPR15*) were both associated with smoking intensity measured by pack-years ($p = 6.04 \times 10^{-5}$ and 2.16×10^{-4} respectively).

Methylation sites cg11314684 (*AKT3*) and cg13668129 (*HNRPU1*) exemplify a different type of pattern where former smokers and non-smokers have comparable levels of methylation. However, current smokers have significantly decreased levels of methylation. These DNAm sites suggest fast recovery of methylation status after smoking cessation (*i.e.* similar levels of methylation between “never smokers” and “former smokers”).

Discussion

In this study, the associations of DNAm sites in *F2RL3* and *GPR15* with current smoking have been replicated in two African American populations using two array-based platforms. With different definitions of current smoking status between the discovery and the replication cohorts, the replicated associations strongly support the true connection between cigarette smoking and these gene-specific DNA methylation sites. *F2RL3* plays a role in platelet activation and cell signaling. A recent study showed that the DNAm in *F2RL3* was strongly associated with mortality among patients with stable coronary heart disease adjusted for confounders. (Breitling et al. 2012) *GPR15* is homologous to the angiotensin II AT1, AT2 receptors, the interleukin 8 receptor beta, the orphan receptors GPR1 and angiotensin receptor-like 1, (Heiber et al. 1996) and functions as a co-receptor for the human immunodeficiency virus. (Okamoto and Shikano 2011) The hypomethylation of DNAm sites cg03636183 (*F2RL3*) (Breitling et al. 2011; Shenker et al. 2012; Wan et al. 2012) and cg19859270 (*GPR15*) (Breitling et al. 2011; Wan et al. 2012) have been associated with cigarette smoking in epigenetic studies of Caucasian populations. The two sites ranked as the top two smoking-related DNAm sites and shared similar effects in two studies (Breitling et al. 2011; Wan et al. 2012), however, DNAm site in *GPR15* is not significantly associated with smoking in the third study. (Shenker et al. 2012) In our AA samples, we also replicated DNAm sites in *LIM2*, *NCAPD3*, *ARHGAP25*, *APBA2*, *SLAMF1* and *LRRN3* using Illumina 27K chip, and in *AHRR* and inter-genic loci of 2q37.1 and 6p21.33 using Illumina 450K chip, which were previously reported in studies of Caucasian populations. Comparing to the recently study of epigenetic association with smoking exposure during pregnancy (Joubert et al. 2012), the top associations are mostly different from the smoking-related DNAm sites identified and replicated in this study of adults. However, two DNAm sites close to *AHRR* gene on chromosome 5, cg05575921 and cg21161138 (26 kb apart), were associated with cigarette smoking in both newborn and adult samples, and were replicated in our GTP cohort. These sites were measured on the 450K chip, but not on the 27K chip which was used in our discovery cohort of GENOA.

The replication cohort was genotyped by Infinium 450K chip which includes more DNAm sites proximate to known genes. For the fifteen genes with smoking-related DNAm sites in the discovery sample, we also looked for additional smoking-related DNAm sites within the same genes (Supplementary Table 2). In the replication sample, besides the two replicated DNAm sites in *F2RL3* and *GPR15*, a different DNAm site (cg05843841) in leucine rich repeat containing 32 (*LRR32*) is strongly associated with current smoking ($p=5.96 \times 10^{-5}$) while the DNAm site (cg13633560) significant in the discovery sample is only marginally significant ($p=0.057$). *LRR32* is located in 11q13 region and encodes a type I membrane protein which is a receptor for latent TGF- β , and regulates immune functions. Interestingly, a locus close to *LRR32* was associated with asthma, and was also implicated to be a genetic risk factor for allergic asthma in a large-scale genome-wide association study of asthma (Ferreira et al. 2011).

In addition to modification by environmental factors, the DNAm profile is also partially shaped by genetic variants (Bell et al. 2011; Numata et al. 2012). Due to the divergent genetic background and heterogeneous gene-environment interactions across ethnic groups, genetic associations often do not replicate across ethnic groups, particularly between Caucasians and African Americans (Barbalic et al. 2011; Ehret et al. 2011). We have shown that the association between smoking and DNAm sites in *F2RL3* and *GPR15* replicates not only across independent populations within the same ethnic group, but also across different ethnic groups. Considering the low transferability of genetic association between Caucasians and African Americans, these replicated smoking-related epigenetic changes are more likely to slowly reflect effects of the common environmental factor (*i.e.* cigarette smoking).

It has been reported that race is associated with global DNAm (Hsiung et al. ; Terry et al. ; Zhang et al.) and individual DNAm sites (Kwabi-Addo et al. ; Lam et al.), but the differences of DNAm between racial groups are not consistent across studies (Axume et al. 2007; Hsiung et al. 2007; Zhang et al. 2011). Race is not only confounded with genetic factors, but also may be confounded with environmental factors due to differences in culture and lifestyle. The association between race and DNAm may be driven by either genetic, or environmental factors, or both. In this study, we used the top two principal components derived from GWAS data as a proxy for genetic admixture. After additional adjustment for admixture, the DNAm associations with smoking status remain significant, which suggests that the genetic admixture in AAs is unlikely to drive the association between smoking and DNAm change. To further address the possibility that genetic variants could influence smoking-related DNAm, we searched for mQTL (*i.e.* SNPs associated with DNAm) within 10 kb flanking regions of *F2RL3* and *GPR15*. We did not identify any mQTL with significant association ($p < 0.001$) with the two replicated smoking-related DNAm sites in either the discovery samples or the replication samples. This result further supports that the association between DNAm and cigarette smoking is unlikely to be driven by genetic effects correlated with DNAm.

Early-life and adulthood psychosocial environment may affect epigenetics through stress-induced inflammatory process. Lam *et al.* recently conducted an epigenome-wide association study in a community cohort, and reported that psychosocial factors (e.g. perceived stress, and cortisol output) as well as early-life socioeconomic status (SES) were associated with DNA methylation (Lam et al. 2012). Limited by the availability of similar measurements in the GENOA study, we were not able to assess the associations between the psychosocial and early-life SES factors and DNAm in this study in AAs. Because AAs suffer disadvantageous SES and psychosocial environment, further epigenetic study of these factors may reveal novel biological mechanisms of health disparity in AAs.

DNAm profiles are tissue and cell type-specific (Pai et al. 2011; Sun et al. 2010). DNA methylation profiles have been commonly studied in PLC due to the easy access to the biosample. The choice of PLC is meaningful to study certain environmental exposures such as smoking, and chronic conditions involving the circulation and immune system. However, since PLC comprise a mixture of multiple cell types, it is possible that the results reported here and elsewhere reflect smoking-related DNAm changes that influence a single cell type component of PLC. If this is the case, it could potentially explain why the associations between smoking and sites in *F2RL3* and *GPR15* have a relatively low effect size (Figure 1; ΔBeta in Table 2–Table 3) but are highly significant and replicable. Cigarette smoking is associated with shifts in peripheral blood cell populations and among smokers differs by current smoking and time since quitting (Kocaman et al. 2011; Schwartz and Weiss 1994; Smith et al. 2003). Since different DNAm profiles have been observed in distinct leukocyte subtypes using dozens of samples (Houseman et al. 2012; Koestler et al. 2012), the association between DNA methylation and cigarette smoking can be confounded by differences in the proportion of leukocyte subtypes between samples. We applied the method developed by Houseman et al., (Houseman et al. 2012) and adopted by Liu et al. (Liu et al. 2013) to project the subtype proportions of each GENOA samples using the 27K DNAm data. We re-examined the association between DNAm sites and cigarette smoking adjusted for these subtypes of leukocytes including granulocyte, monocyte, natural killer cells (NK), B cell, CD4⁺ and CD8⁺ T cells. The top smoking-related DNA sites remain significant after the additional adjustment for the proportions of leukocyte subtypes. For example, the p-values of the three replicated DNAm sites, cg03636183 (*F2RL3*), cg19859270 (*GPR15*), and cg13668129 (*HNRPU1*), were 6.90×10^{-42} , 6.04×10^{-16} , 9.13×10^{-8} respectively, all significant after Bonferroni correction of multiple testing. The additional adjustment did not change the significant associations between DNAm sites and

smoking dramatically in the GTP cohort either (Table 2 and Table 3). Although we cannot completely rule out the impact of shift of cell populations on DNAm, these replicated results of smoking-related DNAm are not likely to be driven by the shift of leukocyte populations. Future studies of smoking-associated DNAm in a single targeted cell population would be valuable for identifying cell-type specific associations between DNAm and cigarette smoking.

Although we have observed and replicated statistically significant associations between DNAm and cigarette smoking across ethnic groups, we have limited knowledge about the functional consequences of these epigenetic changes, and how they relate to cigarette smoking. Since DNA methylation located in the promoter region often correlates with gene expression, systematic research of both epigenomic and transcriptomic data in the same biological samples may fill this knowledge gap, and improve our understanding of the complex molecular system underlying the pathophysiology of diseases. Along with continuous efforts to promote smoking cessation, elucidating the molecular mechanisms linking tobacco smoke exposure to disease remains important and may provide potential targets for development of adjunctive clinical interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully acknowledge the participants of the study. Genetic Epidemiology Network of Arteriopathy (GENOA) study is supported by the National Institutes of Health, grant numbers HL100185 and HL100245 from National Heart, Lung, Blood Institute. Grady Trauma Project (GTP) is supported, in part, by NIH/NIDA T32 DA15040, and support for AKS was provided by MH085806. The authors thank Richard Barfield and Varun Kilaru for technical assistance.

References

- Arnett DK, Meyers KJ, Devereux RB, Tiwari HK, Gu CC, Vaughan LK, Perry RT, Patki A, Claas SA, Sun YV, Broeckel U, Kardia SL. Genetic variation in NCAM1 contributes to left ventricular wall thickness in hypertensive families. *Circ Res*. 2011; 108:279–283. [PubMed: 21212386]
- Axume J, Smith SS, Pogribny IP, Moriarty DJ, Caudill MA. Global leukocyte DNA methylation is similar in African American and Caucasian women under conditions of controlled folate intake. *Epigenetics*. 2007; 2:66–68. [PubMed: 17965592]
- Barbalic M, Reiner AP, Wu C, Hixson JE, Franceschini N, Eaton CB, Heiss G, Couper D, Mosley T, Boerwinkle E. Genome-wide association analysis of incident coronary heart disease (CHD) in African Americans: a short report. *PLoS Genet*. 2011; 7 e1002199.
- Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics*. 2012; 28:1280–1281. [PubMed: 22451269]
- Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol*. 2011; 12:R10. [PubMed: 21251332]
- Bouloukaki I, Tsiligianni IG, Tsoumakidou M, Mitrouska I, Prokopakis EP, Mavroudi I, Siafakas NM, Tzanakis N. Sputum and nasal lavage lung-specific biomarkers before and after smoking cessation. *BMC Pulm Med*. 2011; 11:35. [PubMed: 21635782]
- Braber S, Henricks PA, Nijkamp FP, Kraneveld AD, Folkerts G. Inflammatory changes in the airways of mice caused by cigarette smoke exposure are only partially reversed after smoking cessation. *Respir Res*. 2010; 11:99. [PubMed: 20649997]
- Breitling LP, Salzmann K, Rothenbacher D, Burwinkel B, Brenner H. Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. *Eur Heart J*. 2012; 33:2841–2848. [PubMed: 22511653]

- Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet.* 2011; 88:450–457. [PubMed: 21457905]
- Chen YA, Choufani S, Ferreira JC, Grafodatskaya D, Butcher DT, Weksberg R. Sequence overlap between autosomal and sex-linked probes on the Illumina HumanMethylation27 microarray. *Genomics.* 2011; 97:214–222. [PubMed: 21211562]
- Conen D, Everett BM, Kurth T, Creager MA, Buring JE, Ridker PM, Pradhan AD. Smoking, smoking cessation, [corrected] and risk for symptomatic peripheral artery disease in women: a cohort study. *Ann Intern Med.* 2011; 154:719–726. [PubMed: 21646555]
- Daniels PR, Kardia SL, Hanis CL, Brown CA, Hutchinson R, Boerwinkle E, Turner ST. Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *Am J Med.* 2004; 116:676–681. [PubMed: 15121494]
- Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang SJ, Pihur V, Vollenweider P, O'Reilly PF, Amin N, Bragg-Gresham JL, Teumer A, Glazer NL, Launer L, Zhao JH, Aulchenko Y, Heath S, Sober S, Parsa A, Luan J, Arora P, Dehghan A, Zhang F, Lucas G, Hicks AA, Jackson AU, Peden JF, Tanaka T, Wild SH, Rudan I, Igl W, Milaneschi Y, Parker AN, Fava C, Chambers JC, Fox ER, Kumari M, Go MJ, van der Harst P, Kao WH, Sjogren M, Vinay DG, Alexander M, Tabara Y, Shaw-Hawkins S, Whincup PH, Liu Y, Shi G, Kuusisto J, Tayo B, Seielstad M, Sim X, Nguyen KD, Lehtimäki T, Matullo G, Wu Y, Gaunt TR, Onland-Moret NC, Cooper MN, Platou CG, Org E, Hardy R, Dahgam S, Palmen J, Vitart V, Braund PS, Kuznetsova T, Uitterwaal CS, Adeyemo A, Palmas W, Campbell H, Ludwig B, Tomaszewski M, Tzoulaki I, Palmer ND, Aspelund T, Garcia M, Chang YP, O'Connell JR, Steinle NI, Grobbee DE, Arking DE, Kardia SL, Morrison AC, Hernandez D, Najjar S, McArdle WL, Hadley D, Brown MJ, Connell JM, Hingorani AD, Day IN, Lawlor DA, Beilby JP, Lawrence RW, Clarke R, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011; 478:103–109. [PubMed: 21909115]
- Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature.* 2007; 447:433–440. [PubMed: 17522677]
- Feinberg AP. Epigenetics at the epicenter of modern medicine. *JAMA.* 2008; 299:1345–1350. [PubMed: 18349095]
- Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer.* 2004; 4:143–153. [PubMed: 14732866]
- Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souef P, Danoy P, Baltic S, Nyholt DR, Jenkins M, Hayden C, Willemsen G, Ang W, Kuokkanen M, Beilby J, Cheah F, de Geus EJ, Ramasamy A, Vedantam S, Salomaa V, Madden PA, Heath AC, Hopper JL, Visscher PM, Musk B, Leeder SR, Jarvelin MR, Pennell C, Boomsma DI, Hirschhorn JN, Walters H, Martin NG, James A, Jones G, Abramson MJ, Robertson CF, Dharmage SC, Brown MA, Montgomery GW, Thompson PJ. Australian Asthma Genetics C. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet.* 2011; 378:1006–1014. [PubMed: 21907864]
- Figueiredo JC, Grau MV, Wallace K, Levine AJ, Shen L, Hamdan R, Chen X, Bresalier RS, McKeown-Eyssen G, Haile RW, Baron JA, Issa JP. Global DNA hypomethylation (LINE-1) in the normal colon and lifestyle characteristics and dietary and genetic factors. *Cancer Epidemiol Biomarkers Prev.* 2009; 18:1041–1049. [PubMed: 19336559]
- Flom JD, Ferris JS, Liao Y, Tehranifar P, Richards CB, Cho YH, Gonzalez K, Santella RM, Terry MB. Prenatal smoke exposure and genomic DNA methylation in a multiethnic birth cohort. *Cancer Epidemiol Biomarkers Prev.* 2011; 20:2518–2523. [PubMed: 21994404]
- Furniss CS, Marsit CJ, Houseman EA, Eddy K, Kelsey KT. Line region hypomethylation is associated with lifestyle and differs by human papillomavirus status in head and neck squamous cell carcinomas. *Cancer Epidemiol Biomarkers Prev.* 2008; 17:966–971. [PubMed: 18398038]
- Gillespie CF, Bradley B, Mercer K, Smith AK, Conneely K, Gapen M, Weiss T, Schwartz AC, Cubells JF, Ressler KJ. Trauma exposure and stress-related disorders in inner city primary care patients. *Gen Hosp Psychiatry.* 2009; 31:505–514. [PubMed: 19892208]
- Giovino GA. Epidemiology of tobacco use in the United States. *Oncogene.* 2002; 21:7326–7340. [PubMed: 12379876]

- Heiber M, Marchese A, Nguyen T, Heng HH, George SR, O'Dowd BF. A novel human gene encoding a G-protein-coupled receptor (GPR15) is located on chromosome 3. *Genomics*. 1996; 32:462–465. [PubMed: 8838812]
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics*. 2012; 13:86. [PubMed: 22568884]
- Hsiung DT, Marsit CJ, Houseman EA, Eddy K, Furniss CS, McClean MD, Kelsey KT. Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev*. 2007; 16:108–114. [PubMed: 17220338]
- Hartigan JA, Hartigan PM. The dip test of unimodality. *Annals of Statistics*. 1985; 13:70–84.
- Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet*. 2002; 3:415–428. [PubMed: 12042769]
- Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, Huang Z, Hoyo C, Midttun O, Cupul-Uicab LA, Ueland PM, Wu MC, Nystad W, Bell DA, Peddada SD, London SJ. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect*. 2012; 120:1425–1431. [PubMed: 22851337]
- Kawachi I, Colditz GA, Stampfer MJ, Willett WC, Manson JE, Rosner B, Speizer FE, Hennekens CH. Smoking cessation and decreased risk of stroke in women. *JAMA*. 1993; 269:232–236. [PubMed: 8417241]
- Kellogg SH, McHugh PF, Bell K, Schluger JH, Schluger RP, LaForge KS, Ho A, Kreek MJ. The Kreek-McHugh-Schluger-Kellogg scale: a new, rapid method for quantifying substance abuse and its possible applications. *Drug Alcohol Depend*. 2003; 69:137–150. [PubMed: 12609695]
- Kilaru V, Barfield RT, Schroeder JW, Smith AK, Conneely KN. MethLAB: a graphical user interface package for the analysis of array-based DNA methylation data. *Epigenetics*. 2012; 7:225–229. [PubMed: 22430798]
- Kim KY, Kim DS, Lee SK, Lee IK, Kang JH, Chang YS, Jacobs DR, Steffes M, Lee DH. Association of low-dose exposure to persistent organic pollutants with global DNA hypomethylation in healthy Koreans. *Environ Health Perspect*. 2010; 118:370–374. [PubMed: 20064773]
- Kocaman SA, Sahinarslan A, Kunak T, Balcioglu S, Cetin M, Cemri M, Timurkaynak T, Boyaci B, Cengel A. The particular interactions of the traditional cardiovascular risk factors with different circulating specific leukocyte subtype counts in blood: an observational study. *Anadolu Kardiyol Derg*. 2011; 11:573–581. [PubMed: 21911321]
- Koestler DC, Marsit CJ, Christensen BC, Accomando W, Langevin SM, Houseman EA, Nelson HH, Karagas MR, Wiencke JK, Kelsey KT. Peripheral blood immune cell methylation profiles are associated with nonhematopoietic cancers. *Cancer Epidemiol Biomarkers Prev*. 2012; 21:1293–1302. [PubMed: 22714737]
- Kwabi-Addo B, Wang S, Chung W, Jelinek J, Patierno SR, Wang BD, Andrawis R, Lee NH, Apprey V, Issa JP, Ittmann M. Identification of differentially methylated genes in normal prostate tissues from African American and Caucasian men. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2010; 16:3539–3547. [PubMed: 20606036]
- Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, Kobor MS. Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci U S A*. 2012; 109(Suppl 2):17253–17260. [PubMed: 23045638]
- Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009; 462:315–322. [PubMed: 19829295]
- Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, Alfredsson L, Klareskog L, Ekstrom TJ, Feinberg AP. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol*. 2013; 31:142–147. [PubMed: 23334450]
- Luo J, Margolis KL, Wactawski-Wende J, Horn K, Messina C, Stefanick ML, Tindle HA, Tong E, Rohan TE. Association of active and passive smoking with risk of breast cancer among postmenopausal women: a prospective cohort study. *BMJ*. 2011; 342:d1016. [PubMed: 21363864]

- Mani S, Szymanska K, Cuenin C, Zaridze D, Balassiano K, Lima SC, Matos E, Daudt A, Koifman S, Filho VW, Menezes AM, Curado MP, Ferro G, Vaissiere T, Sylla BS, Tommasino M, Pinto LF, Boffetta P, Hainaut P, Brennan P, Hecceg Z. DNA methylation changes associated with risk factors in tumors of the upper aerodigestive tract. *Epigenetics*. 2012; 7:270–277. [PubMed: 22430803]
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature*. 2009; 461:747–753. [PubMed: 19812666]
- Moore LE, Pfeiffer RM, Poscablo C, Real FX, Kogevinas M, Silverman D, Garcia-Closas R, Chanock S, Tardon A, Serra C, Carrato A, Dosemeci M, Garcia-Closas M, Esteller M, Fraga M, Rothman N, Malats N. Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study. *Lancet Oncol*. 2008; 9:359–366. [PubMed: 18339581]
- Numata S, Ye T, Hyde TM, Guitart-Navarro X, Tao R, Wininger M, Colantuoni C, Weinberger DR, Kleinman JE, Lipska BK. DNA methylation signatures in development and aging of the human prefrontal cortex. *Am J Hum Genet*. 2012; 90:260–272. [PubMed: 22305529]
- Oka D, Yamashita S, Tomioka T, Nakanishi Y, Kato H, Kaminishi M, Ushijima T. The presence of aberrant DNA methylation in noncancerous esophageal mucosae in association with smoking history: a target for risk diagnosis and prevention of esophageal cancers. *Cancer*. 2009; 115:3412–3426. [PubMed: 19472401]
- Okamoto Y, Shikano S. Phosphorylation-dependent C-terminal binding of 14-3-3 proteins promotes cell surface expression of HIV co-receptor GPR15. *J Biol Chem*. 2011; 286:7171–7181. [PubMed: 21189250]
- Pai AA, Bell JT, Marioni JC, Pritchard JK, Gilad Y. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet*. 2011; 7:e1001316.
- Peto R, Lopez AD, Boreham J, Thun M, Heath C Jr, Doll R. Mortality from smoking worldwide. *Br Med Bull*. 1996; 52:12–21. [PubMed: 8746293]
- Philibert RA, Beach SR, Gunter TD, Brody GH, Madan A, Gerrard M. The effect of smoking on MAOA promoter methylation in DNA prepared from lymphoblasts and whole blood. *Am J Med Genet B Neuropsychiatr Genet*. 2010; 153B:619–628. [PubMed: 19777560]
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–575. [PubMed: 17701901]
- Rea TD, Heckbert SR, Kaplan RC, Smith NL, Lemaitre RN, Psaty BM. Smoking status and risk for recurrent coronary events after myocardial infarction. *Ann Intern Med*. 2002; 137:494–500. [PubMed: 12230350]
- Rusiecki JA, Baccarelli A, Bollati V, Tarantini L, Moore LE, Bonefeld-Jorgensen EC. Global DNA hypomethylation is associated with high serum-persistent organic pollutants in Greenlandic Inuit. *Environ Health Perspect*. 2008; 116:1547–1552. [PubMed: 19057709]
- Schwartz J, Weiss ST. Cigarette smoking and peripheral blood leukocyte differentials. *Ann Epidemiol*. 1994; 4:236–242. [PubMed: 8055125]
- Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, Belvisi MG, Brown R, Vineis P, Flanagan JM. 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*. 2012
- Smith IM, Mydlarz WK, Mithani SK, Califano JA. DNA global hypomethylation in squamous cell head and neck cancer associated with smoking, alcohol consumption and stage. *Int J Cancer*. 2007; 121:1724–1728. [PubMed: 17582607]
- Smith MR, Kinmonth AL, Luben RN, Bingham S, Day NE, Wareham NJ, Welch A, Khaw KT. Smoking status and differential white cell count in men and women in the EPIC-Norfolk population. *Atherosclerosis*. 2003; 169:331–337. [PubMed: 12921986]

- Sun YV, Turner ST, Smith JA, Hammond PI, Lazarus A, Van De Rostyne JL, Cunningham JM, Kardia SL. Comparison of the DNA methylation profiles of human peripheral blood cells and transformed B-lymphocytes. *Hum Genet.* 2010; 127:651–658. [PubMed: 20238126]
- Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews.Genetics.* 2008; 9:465–476.
- Tekpli X, Zienolddiny S, Skaug V, Stangeland L, Haugen A, Mollerup S. DNA methylation of the CYP1A1 enhancer is associated with smoking-induced genetic alterations in human lung. *Int J Cancer.* 2011
- Terry MB, Ferris JS, Pilsner R, Flom JD, Tehranifar P, Santella RM, Gamble MV, Susser E. Genomic DNA methylation among women in a multiethnic New York City birth cohort. *Cancer Epidemiol Biomarkers Prev.* 2008; 17:2306–2310. [PubMed: 18768498]
- van der Maarel SM. Epigenetic mechanisms in health and disease. *Ann Rheum Dis.* 2008; 67(Suppl 3):iii97–iii100. [PubMed: 19022824]
- Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, Agusti A, Anderson W, Lomas DA, Demeo DL. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum Mol Genet.* 2012
- Wilhelm-Benartzi CS, Christensen BC, Koestler DC, Andres Houseman E, Schned AR, Karagas MR, Kelsey KT, Marsit CJ. Association of secondhand smoke exposures with DNA methylation in bladder carcinomas. *Cancer Causes Control.* 2011; 22:1205–1213. [PubMed: 21660454]
- Willemse BW, ten Hacken NH, Rutgers B, Lesman-Leegte IG, Postma DS, Timens W. Effect of 1-year smoking cessation on airway inflammation in COPD and asymptomatic smokers. *Eur Respir J.* 2005; 26:835–845. [PubMed: 16264044]
- Xu Q, Ma JZ, Payne TJ, Li MD. Determination of Methylated CpG Sites in the Promoter Region of Catechol-O-Methyltransferase (COMT) and their Involvement in the Etiology of Tobacco Smoking. *Front Psychiatry.* 2010; 1:16. [PubMed: 21423427]
- Zhang FF, Cardarelli R, Carroll J, Fulda KG, Kaur M, Gonzalez K, Vishwanatha JK, Santella RM, Morabia A. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. *Epigenetics.* 2011; 6:623–629. [PubMed: 21739720]
- Zhu ZZ, Hou L, Bollati V, Tarantini L, Marinelli B, Cantone L, Yang AS, Vokonas P, Lissowska J, Fustinoni S, Pesatori AC, Bonzini M, Apostoli P, Costa G, Bertazzi PA, Chow WH, Schwartz J, Baccarelli A. Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. *Int J Epidemiol.* 2012; 41:126–139. [PubMed: 20846947]

Figure 1. Modification of DNAm between nonsmoker, former and current smokers. The four DNAm sites exhibit hypomethylation among current smokers, consistent in males and females.

Table 1

Characteristics of study population

Variable	Male (n=285) N (%)	Female (n=687) N (%)	Total (n=972) N (%)	Current KMSK Tobacco Score		
				Male (n=68) Mean (SD)	Female (n=171) Mean (SD)	Total (n=239) Mean (SD)
Smoke						
Never	93 (32.6%)	474 (69.0%)	567 (58.3%)			
Not in last year	135 (47.4%)	148 (21.5%)	283 (29.1%)	4.34 (3.79)	2.85 (3.84)	3.27 (3.88)
Smoke within 1 year	57 (20.0%)	65 (9.5%)	122 (12.6%)			
Age (years)	Mean (SD) 66.8 (7.7)	Mean (SD) 66.1 (7.6)	Mean (SD) 66.3 (7.6)	Mean (SD) 43 (12)	Mean (SD) 40 (13)	Mean (SD) 41 (13)
Diastolic Blood Pressure (mmHg)	80.2 (11.2)	77.5 (10.8)	78.3 (11.0)	78.0 (12.5)	73.7 (11.7)	75.1 (12.1)
Systolic Blood Pressure (mmHg)	137.8 (20.8)	140.9 (21.6)	140.0 (21.4)	130.8 (18.4)	123.5 (16.9)	125.8 (17.7)
Body Mass Index (kg/m ²)	29.0 (4.8)	32.1 (6.6)	31.2 (6.3)	29.0 (8.0)	35.0 (8.9)	33.2 (9.0)

Table 2
Association of Fifteen Smoking-related DNAm sites in the Discovery and Replication Cohorts

DNAm Site	Gene	GENOA (n= 972)					GTP (n= 239)				
		ΔBeta	SE	t-stat	p-value	Adjusted p-value *	ΔBeta	SE	t-stat	p-value	Adjusted p-value *
Cg03636183	<i>F2RL3</i>	-0.0958	0.0064	-15.00	6.21×10^{-42}	6.90×10^{-42}	-0.030	0.006	-4.693	4.91×10^{-6}	4.76×10^{-5}
Cg19859270	<i>GPR15</i>	-0.0212	0.0031	-6.84	2.45×10^{-11}	6.04×10^{-16}	-0.022	0.003	-6.547	4.64×10^{-10}	2.56×10^{-9}
Cg04983977	<i>GPR25</i>	0.0214	0.0039	5.45	8.05×10^{-8}	3.34×10^{-7}	0.004	0.003	1.237	0.217	0.460
Cg13668129	<i>HNRPU1</i>	-0.0132	0.0025	-5.40	1.06×10^{-7}	9.13×10^{-8}	-0.005	0.002	-2.881	4.38×10^{-3}	0.010
Cg13500388	<i>CBFB</i>	-0.0256	0.0048	-5.37	1.24×10^{-7}	1.98×10^{-7}	-0.005	0.004	-1.469	0.143	0.169
cg13633560	<i>LRRC32</i>	-0.0238	0.0046	-5.20	3.00×10^{-7}	5.16×10^{-7}	-0.005	0.003	-1.701	0.091	0.207
cg03340878	<i>OR2B6</i>	-0.0232	0.0045	-5.17	3.44×10^{-7}	8.36×10^{-7}	-0.006	0.004	-1.298	0.196	0.204
cg01500140	<i>LIM2</i>	0.0221	0.0044	5.00	8.10×10^{-7}	6.92×10^{-8}	0.011	0.005	2.318	0.021	0.027
cg00353953	<i>ZNF384</i>	-0.0198	0.0040	-4.99	8.40×10^{-7}	2.97×10^{-6}	-0.004	0.003	-1.213	0.227	0.551
cg11314684	<i>AKT3</i>	-0.0194	0.0039	-4.96	9.83×10^{-7}	2.29×10^{-6}	-0.006	0.003	-2.347	0.020	0.023
cg03330058	<i>ABTB1</i>	-0.0366	0.0074	-4.93	1.15×10^{-6}	2.68×10^{-6}	-0.004	0.005	-0.887	0.376	0.431
cg13745870	<i>SPATA12</i>	-0.0182	0.0037	-4.87	1.52×10^{-6}	3.28×10^{-7}	-0.002	0.003	-0.536	0.593	0.191
cg17791651	<i>POU3F1</i>	-0.0387	0.0080	-4.86	1.62×10^{-6}	8.75×10^{-7}	-0.001	0.003	-0.360	0.719	0.912
cg14223444	<i>NCBPI</i>	-0.0213	0.0044	-4.83	1.85×10^{-6}	3.89×10^{-6}	0.000	0.003	0.030	0.976	0.935
cg26259865	<i>LOC124220</i>	-0.0186	0.0039	-4.80	2.18×10^{-6}	2.18×10^{-6}	0.003	0.003	1.126	0.261	0.100

* additional adjustment for PLC subtype proportions.

Table 3

Replication Results of Smoking-related DNAm from Caucasian Populations

Platform	DNAm site	Gene/Locus	t-stat	p-value	Adjusted p-value [#]
HumanMethylation27 (Breitling et al. 2011) (Wan et al. 2012)	cg03636183**	<i>F2RL3</i>	-15.0	6.21×10^{-42}	6.90×10^{-42}
	cg19859270	<i>GPR15</i>	-6.84	2.45×10^{-11}	6.04×10^{-16}
	cg02564523	<i>ORAI2</i>	-1.74	0.082	0.078
	cg09155905	<i>FNDCC8</i>	-1.01	0.313	0.207
	cg09084200	<i>NCAPD3</i>	-4.56	6.39×10^{-6}	1.86×10^{-5}
	cg09837977	<i>LRRN3</i>	-3.04	2.53×10^{-3}	2.40×10^{-7}
	cg01500140	<i>LIM2</i>	5.00	8.10×10^{-7}	6.92×10^{-8}
	cg13247990	<i>MYLK</i>	2.33	2.00×10^{-2}	0.025
	cg01988129	<i>ADHFE1</i>	0.86	0.389	0.103
	cg16254309	<i>CNTNAP2</i>	-2.89	4.07×10^{-3}	2.47×10^{-3}
	cg18881723	<i>SLAMF1</i>	-3.06	2.32×10^{-3}	0.036
	cg12044210	<i>APBA2</i>	NA*	NA*	NA*
	cg21917349	<i>APBA2</i>	3.07	2.25×10^{-3}	7.54×10^{-3}
	cg15691199	<i>CEBPE</i>	-2.87	4.29×10^{-3}	0.031
	cg24262469	<i>TIPARP</i>	-1.29	0.199	0.205
	cg05445326	<i>TM4SF19</i>	3.79	1.70×10^{-2}	3.00×10^{-3}
	cg15258980	<i>ARHGAP25</i>	-4.65	4.36×10^{-6}	1.84×10^{-4}
	cg10161121	<i>FASLG</i>	1.82	6.97×10^{-2}	0.920
HumanMethylation450 (Shenker et al. 2012)	cg06644428	<i>2q37.1</i>	-3.87	1.49×10^{-4}	5.75×10^{-4}
	cg05951221	<i>2q37.1</i>	-6.35	1.34×10^{-9}	2.30×10^{-8}
	cg21566642	<i>2q37.1</i>	-7.71	5.32×10^{-13}	4.84×10^{-11}
	cg01940273	<i>2q37.1</i>	-7.46	2.37×10^{-12}	4.02×10^{-11}
	cg23576855	<i>AHRR</i>	-8.34	1.06×10^{-14}	1.36×10^{-12}
	cg05575921	<i>AHRR</i>	-9.77	9.51×10^{-19}	4.84×10^{-18}
	cg21161138	<i>AHRR</i>	-5.84	2.04×10^{-8}	4.47×10^{-7}
	cg06126421	<i>6p21.33</i>	-6.81	1.04×10^{-10}	1.00×10^{-11}

* The DNAm site was excluded from analysis due to non-unique probe sequence.

** DNAm site (cg03636183) in F2RL3 has been previously reported using both 27K and 450K platforms.

additional adjustment for PLC subtype proportions.