

Original investigation

Relationships of Long-Term Smoking and Moist Snuff Consumption With a DNA Methylation Age Relevant Smoking Index: An Analysis in Buccal Cells

Jamaji C. Nwanaji-Enwerem PhD¹, Andres Cardenas PhD²,
Peter R. Chai MD, MMS³, Marc G. Weisskopf PhD⁴,
Andrea A. Baccarelli MD, PhD⁵, Edward W. Boyer MD, PhD³

¹Department of Environmental Health, Harvard T.H. Chan School of Public Health and MD-PhD Program, Harvard Medical School, Boston, MA; ²Division of Chronic Disease Research Across the Lifecourse, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim HealthCare Institute, Boston, MA; ³Division of Medical Toxicology, Department of Emergency Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; ⁴Department of Environmental Health and Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; ⁵Department of Environmental Health Sciences, Columbia Mailman School of Public Health, New York, NY

Corresponding Author: Jamaji C. Nwanaji-Enwerem, PhD, Department of Environmental Health, Harvard T.H. Chan School of Public Health, Biological Sciences in Public Health, Building 1, Room G7, 665 Huntington Avenue, Boston, MA 02115, USA. Telephone: (+1) 617-432-2050; E-mail: jamaji_nwanaji-enwerem@hms.harvard.edu

Abstract

Introduction: Currently, there is no widely accepted, non-self-report measure that simultaneously reflects smoking behaviors and is molecularly informative of general disease processes. Recently, researchers developed a smoking index (SI) using nucleated blood cells and a multi-tissue DNA methylation-based predictor of chronological age and disease (DNA methylation age [DNAm-age]). To better understand the utility of this novel SI in readily accessible cell types, we used buccal cell DNA methylation to examine SI relationships with long-term tobacco smoking and moist snuff consumption.

Methods: We used a publicly available dataset composed of buccal cell DNA methylation values from 120 middle-aged men (40 long-term smokers, 40 moist snuff consumers, and 40 nonsmokers). DNAm-age (353-CpGs) and SI (66-CpGs) were calculated using CpG sites measured using the Illumina HumanMethylation450 BeadChip. We estimated associations of tobacco consumption habits with both SI and DNAm-age using linear regression models adjusted for chronological age, race, and methylation technical covariates.

Results: In fully adjusted models with nonsmokers as the reference, smoking ($\beta = 1.08$, 95% CI = 0.82 to 1.33, $p < .0001$) but not snuff consumption ($\beta = .06$, 95% CI = -0.19 to 0.32 , $p = .63$) was significantly associated with SI. SI was an excellent predictor of smoking versus nonsmoking (area under the curve = 0.92, 95% CI = 0.85 to 0.98). Four DNAm-age CpGs were differentially methylated between smokers and nonsmokers including cg14992253 [*EIF3I*], which has been previously shown to be differentially methylated with exposure to long-term fine-particle air pollution ($PM_{2.5}$).

Conclusions: The 66-CpG SI appears to be a useful tool for measuring smoking-specific behaviors in buccal cells. Still, further research is needed to broadly confirm our findings and SI relationships with DNAm-age.

Implications: Our findings demonstrate that this 66-CpG blood-derived SI can reflect long-term tobacco smoking, but not long-term snuff consumption, in buccal cells. This evidence will be useful as the field works to identify an accurate non-self-report smoking biomarker that can be measured in an easily accessible tissue. Future research efforts should focus on (1) optimizing the relationship of the SI with DNAm-age so that the metric can maximize its utility as a tool for understanding general disease processes, and (2) determining normal values for the SI CpGs so that the measure is not as study sample specific.

Introduction

Exposure to cigarette smoke is a well appreciated modifiable risk factor for mortality and morbidity from diseases including asthma, chronic obstructive pulmonary disease, stroke, and lung cancer.¹ Interventions that promote smoking cessation rely on accurate assessments of smoking exposure. Unfortunately, most health care professionals rely on individual self-report to assess exposure to tobacco smoke. Self-report measures among smokers tend to underestimate and misrepresent the degree of exposure.² Evidence continues to suggest that self-reports are unacceptable particularly when questionnaires are limited (eg, when asking a nonsmoker about everyday unconscious secondhand smoking exposures) and for certain demographics for whom smoking is socially unacceptable (eg, adults with young children or pregnant women).^{3–6} In these situations where vulnerable populations may be at risk, in situations such as organ transplantation where smoking can impact resource allocation and therapeutic outcomes,⁷ and even in less insidious situations such as individuals simply being unable to accurately recall their smoke exposures, a great need exists for a more objective yet readily accessible measure of tobacco smoke exposure.

To address this critical gap, many researchers have focused their efforts on identifying biomarkers of tobacco smoke exposure. One of the leading biomarkers, to date, is cotinine. Cotinine, one of the major metabolites of nicotine, can be measured in numerous body tissues (eg, blood, saliva, hair, and urine), and has been widely used to validate tobacco smoke exposure self-reports.⁸ Nevertheless, the precise tissue cotinine concentrations that serve as cutoffs for smoking status remain controversial. Although cotinine is a good biomarker of tobacco smoke exposure, it is simply a metabolite of nicotine and is not particularly informative for understanding the biology of tobacco smoke-related disease processes.⁵ For instance, cotinine has been associated with adverse health outcomes (eg, early abortions, atrial fibrillation, and all-cause mortality), but the underlying biology of these relationships remains poorly understood.^{6,9,10}

DNA methylation, an epigenetic modification, is a promising alternative to cotinine. Not only do DNA methylation levels reflect tobacco smoke exposure and smoking-related diseases, but existing technologies already enable researchers to explore nucleotide resolution changes in DNA methylation across the genome.^{11,12} Ultimately, this allows for the identification of specific genes that are differentially methylated and specific pathways that may be implicated in diseases related to tobacco smoke exposure.^{12,13} Of the identified loci (CpG sites) whose methylation has been associated with smoking, one of the most promising (cg05575921) is located in the aryl hydrocarbon receptor repressor (*AHRR*) gene.^{14,15} The aryl hydrocarbon receptor is involved in cellular pathways such as

cell cycle regulation and is well known for mediating the toxicity of environmental contaminants.¹⁶ *AHRR* represses signal transduction from aryl hydrocarbon receptor.¹⁷

In 2015, Gao et al.¹³ performed a systematic review of the cg05575921 CpG site and other top CpGs whose methylation was related to tobacco smoking. The group later used blood cells to develop and validate a smoking index (SI) composed of smoking-related methylation CpGs. They began with 150 CpGs that were related to active smoking and identified at least two times in previous smoking epigenome-wide association studies. Next, they identified CpGs that were significantly associated with DNA methylation age (DNAm-age), a multi-tissue predictor of chronological age associated with all-cause mortality and many disease processes.^{18–20} Of the 150 CpGs, 66 had significant relationships with DNAm-age and were used to build the final validated SI. Among the 66 SI CpGs is the cg05575921 site as well as seven other sites in *AHRR*. As a consequence of the manner in which it was created, the 66 CpG SI is unique in that it should not only reflect tobacco smoke exposure but also potentially have relationships with multi-tissue disease processes.

Given the need for a smoking biomarker that accurately reflects exposure and is informative about disease processes, we designed this study to better understand the utility of this novel SI. We were particularly interested in the performance of the SI across cell types that may be more readily available than blood (eg, buccal cells). Buccal cell retrieval does not require a trained phlebotomist and can be done painlessly—making it less prone to patient and/or participant evasion. Moreover, evidence demonstrates that the quality of buccal genomic DNA is comparable to that of blood genomic DNA.²¹ In this study, we calculated the SI using DNA methylation measurements from buccal cells in a cohort of male middle-aged long-term smokers, moist snuff consumers, and nonsmokers. We then examined if either tobacco consumption habit was associated with SI and how well SI could differentiate smokers from nonsmokers. We also investigated associations of tobacco consumption habits with DNAm-age.

Methods

Study Population

This study was based on a publically available Illumina HumanMethylation450 BeadChip dataset (Series GSE94876) located on the National Center for Biotechnology Information Gene Expression Omnibus Web site (<https://www.ncbi.nlm.nih.gov/geo/>). The dataset is composed of 120 buccal cell DNA methylation samples from 120 middle-aged men (40 long-term smokers, 40 moist

snuff users, and 40 nonsmokers). The study sample has been previously described. Briefly, the study is composed of generally healthy men aged 35–60 years who were enrolled at a single site in High Point, North Carolina. The study subjects were enrolled in parallel into one of the three aforementioned groups. Given the low rate of smokeless tobacco use by women in the United States, the study personnel did not recruit any female subjects. All informed consent, institutional review board approval, and ethical conduct of study information regarding the collection of this data can be found in the initial publication.²² In addition to tobacco consumption/smoking status, methylation values, technical DNA methylation measurement variables, and basic demographic characteristics were available for all study subjects (eg, age and race).

Smoking Status/Tobacco Consumption Habits

In this dataset, (A) *smokers* were defined as individuals who (1) were exclusive cigarette smokers of any brand at least 6-mg “tar” measured by Cambridge Filter Pad method,²² (2) reported smoking at least 10 cigarettes/day for at least 3 years, and (3) had an expired carbon monoxide of 10–100 ppm. (B) *Moist snuff consumers* were individuals who (1) exclusively used moist snuff of any brand, (2) reported using at least two cans of moist snuff per week for at least 3 years, and (3) had an expired carbon monoxide of 0–5 ppm. (C) *Nonsmokers* were defined as individuals who reported abstinence from any nicotine-containing or tobacco products for at least 5 years and had an expired carbon monoxide of 0–5 ppm.²²

DNA Methylation, DNAm-age, Age Acceleration, and SI

Study subjects fasted from food and tobacco for 2 hours and then rinsed their mouths with Scope mouth wash followed by a water rinse. Buccal cells were collected in water via vigorous swishing. After being centrifuged, the cell pellet was washed in phosphate buffered saline and DNA extraction was later performed. DNA samples were bisulfite converted before being hybridized to the Illumina HumanMethylation450 BeadChips as per Infinium HD Methylation protocol. DNA global methylation profiling was performed by Expression Analysis on the Illumina HumanMethylation450 BeadChip arrays. Normalization via Illumina recommended normalization procedures including background estimation using negative control probes and normalization of intensities to housekeeping genes was performed. The resulting normalized β values represent the methylation intensity at a particular CpG site. DNA methylation β values range from 0 (completely unmethylated) to 1 (completely methylated). These normalized β values were publically downloadable from National Center for Biotechnology Information Gene Expression Omnibus and were used in the present analyses.

DNAm-age was calculated using a publically available online calculator (<https://dnamage.genetics.ucla.edu>).²³ In brief, a penalized regression elastic net model was used to regress a calibrated version of chronological age on 21 369 CpG probes shared by Illumina HumanMethylation27 and HumanMethylation450 BeadChip arrays. The elastic net selected 353 CpGs that correlate with age (193 positively and 160 negatively). The calculator predicts the age of each DNA sample (DNAm-age) using regression coefficients of the 353 CpGs resulting from the elastic net regression model trained from a number of training datasets. The calculator maintains predictive accuracy (age correlation 0.97, error = 3.6 years) across body tissues including blood. In the initial DNAm-age publication, buccal cells have a reported age correlation of $r = .67$ in the test data.²³

The age acceleration (AgeAccel) variable is simply the residuals that result from regressing DNAm age on chronological age.²⁰

SI was calculated using 66 CpGs and the formula described by Gao et al.¹⁹ Using their method, the mean β value (μ_c) and standard deviation (σ_c) across the never-smokers of the given dataset were first computed. We then defined the SI as where W_c is +1(–1) if the smoking-associated CpG, c , is hypermethylated (hypomethylated) in smokers and where β_c is the β value of this CpG in samples s .

$$SI(s) = \frac{1}{n} \sum_c W_c \frac{\beta_{cs} - \mu_c}{\sigma_c}$$

SI was calculated for all the subjects in the study. Importantly, no CpGs are shared between DNAm-age and SI.

Statistical Analyses

Main Analyses

We used linear models to evaluate the relationships of tobacco consumption habits with SI. We ran models that made comparisons across all three categories (smokers, moist snuff consumers, and nonsmokers). The same model frameworks were used to examine the relationship of tobacco consumption modalities with DNAm-age and AgeAccel. Each set of models was fully adjusted for age (years), race, and methylation technical covariates (ie, chip, row, and column).

Using the `Cpg.assoc` function and package in R,²⁴ we investigated if DNA methylation values at each of the 353 DNAm-age component CpG sites were differentially expressed when comparing (1) smokers to nonsmokers; (2) snuff consumers to nonsmokers; (3) smokers to snuff consumers; and (4) tobacco users to nonsmokers. False discovery rate correction using the Benjamini–Hochberg procedure was performed to account for multiple hypotheses testing for all CpG methylation analyses. Gene ontology analyses were performed on significant CpG results using the publically available Go TermFinder platform.²⁵

We also performed logistic regression and a receiver operating characteristic analysis to evaluate the performance of SI in classifying smokers from nonsmokers.²⁶ Here, SI was the predictor and status as a smoker or nonsmoker (dichotomous) was the outcome. From this analysis, we were able to calculate a sensitivity and specificity for SI in buccal cells.

Sensitivity Analyses

To examine how the results from this publically available methylation dataset compared to previously published methylation data on smoking status and DNA methylation, we identified the top five smoking-associated CpG sites in the existing literature. We then examined the association of smoking status with methylation at each of these sites in the present dataset. Finally, we compared the resulting effect sizes from these models with the effect sizes that have been published in the literature.

As an additional sensitivity analysis, we tested the associations of well-known cg05575921, annotated to the *AHRR* gene, with DNAm-age and AgeAccel using fully adjusted linear models.

We performed all statistical analyses using R Version 3.4.1 (R Core Team, Vienna, Austria) and considered a p value less than .05 to be statistically significant.

Results

Baseline Characteristics and Descriptive Statistics

The demographic and biomarker descriptive statistics for the study participants in each of the three cohorts are presented in Table 1. All

participants were male. Forty participants were sampled from each smoking habit stratum, resulting in a total sample size of 120 people. Nonsmokers had an average chronological age \pm SD of approximately 47.2 ± 8.28 years. Snuff consumers and smokers had average chronological ages \pm SD of 44.7 ± 6.98 years and 46.8 ± 7.73 years, respectively. The majority of participants ($\geq 70\%$) in each cohort were Caucasians. African Americans made up 30%, 10%, and 23% of the nonsmoker, moist snuff user, and smoker cohorts, respectively. The mean \pm SD DNAm-ages for nonsmokers, moist snuff users, and smokers were 48.6 ± 7.17 years, 46.5 ± 6.06 years, and 46.9 ± 7.29 years. The mean \pm SD SI for the same cohorts were $1.66e-16 \pm 0.49$, 0.07 ± 0.54 , and 1.02 ± 0.55 . [Supplementary Figure 1](#) presents the Pearson correlations between DNAm-age, AgeAccel, chronological age, and SI. In all study subjects, DNAm-age was highly correlated with chronological age ($r = .67$). SI did not have any significant relationships with DNAm-age or AgeAccel. [Supplementary Figure 2](#) depicts boxplots of SI by each tobacco consumption habit.

Long-Term Smoking and Moist Snuff Consumption as Predictors of SI, DNAm-age, and AgeAccel

[Table 2](#) summarizes the results of fully adjusted linear models where tobacco consumption habits were modeled as predictors of SI. Using nonsmokers as a reference, status as a smoker was significantly associated with having a higher SI value ($\beta = 1.08$, 95% CI = 0.82 to 1.33, $p < .0001$). Status as a snuff consumer was not significantly associated with SI ($\beta = 0.06$, 95% CI = -0.19 to 0.32 , $p = .63$). [Table 2](#) also summarizes the results of fully adjusted linear models where tobacco consumption habits were modeled as predictors of DNAm-age and AgeAccel. When compared to nonsmokers, status as a smoker ($\beta = -1.44$, 95% CI = -3.82 to 0.94 , $p = .23$) or snuff consumer ($\beta = -1.67$, 95% CI = -4.07 to 0.74 , $p = .17$) was not significantly associated with DNAm-age. The observed relationships of tobacco consumption habits with AgeAccel, mirrored those of tobacco consumption habits with DNAm-age.

Associations Between Smoking Status and Methylation Values at the 353 DNAm-Age Component CpG Sites

We investigated if average DNA methylation values at each of the 353 component DNAm-age CpG sites were significantly different when comparing (1) smokers to nonsmokers, (2) snuff consumers to nonsmokers, (3) smokers to snuff consumers, and (4) tobacco users to nonsmokers. Following false discovery rate correction, there were four significant differentially methylated CpGs between smokers and

nonsmokers ([Supplementary Figure 3](#)). One of the four CpGs was hypermethylated in smokers compared to nonsmokers (cg10266490 [ACOT11]). The remaining three CpGs (cg14992253 [EIF3I], cg08030082 [POMC], and cg15804973 [MAP3K5]) were hypomethylated in smokers compared to nonsmokers ([Table 3](#)). Gene ontology analyses of these four sites did not return any significant results. No significant differentially methylated CpGs were found in the comparisons of snuff consumers to nonsmokers, smokers to snuff consumers, and tobacco users to nonsmokers.

SI Performance (Receiver Operating Characteristic Analysis)

In our study sample of buccal cells, SI had a sensitivity (ie, the fraction of smokers that SI correctly identified as smokers) of 0.90, and a specificity (ie, the fraction of nonsmokers that SI correctly identified as nonsmokers) of 0.85 ([Supplementary Table 1](#)). [Figure 1](#) depicts the receiver operating characteristic curve from this analysis. The area under the receiver operating characteristic curve was 0.92 (95% CI = 0.85 to 0.98).

Sensitivity Analysis Comparing Top Smoking-Related CpG Methylation Values in the Study Sample to Published Data

[Supplementary Table 2](#) presents the results of a sensitivity analysis where we first identified the top five CpGs (cg06126421 [6p21.33], cg05575921 [AHRR], cg23576855 [AHRR], cg03636183 [F2RL3], and cg09935388 [GFI1]) consistently related to smoking in the literature and demonstrating the largest methylation differences between smokers and nonsmokers. We then evaluated the relationships of smoking status in our study sample for each of these CpG sites. Finally, we compared our effect sizes to those published in the literature. Two of these top five CpGs (cg06126421 [6p21.33] and cg05575921 [AHRR]) are included as part of the SI. All of the CpG sites were hypomethylated in smokers compared to nonsmokers but showed no significant relationships for snuff consumers. Furthermore, all CpGs had effect sizes in our study sample that were within ± 0.03 of the β value or range of β values reported in the existing literature.

Sensitivity Analysis Investigating the Association of cg05575921 [AHRR] with DNAm-Age and AgeAccel

In our study sample using fully adjusted liner models, methylation at cg05575921 [AHRR] was not significantly associated with DNAm-age ($\beta = 4.88$, 95% CI = -1.88 to 11.6 , $p = .16$). Methylation at

Table 1. Characteristics of Study Population (N = 120)

	Nonsmokers (N = 40)	Moist snuff users (N = 40)	Smokers (N = 40)
Demographic variables			
Chronological age (years), mean (SD)	47.2 (8.28)	44.7 (6.98)	46.8 (7.73)
Race			
African American, N (%)	12 (30)	4 (10)	9 (23)
Caucasian, N (%)	28 (70)	35 (88)	30 (75)
Other, N (%)	0 (0)	1 (2)	1 (2)
Main variables			
Age acceleration (residuals), mean (SD)	0.69 (5.08)	0.09 (4.05)	-0.78 (5.97)
DNA methylation age (years), mean (SD)	48.6 (7.17)	46.5 (6.06)	46.9 (7.29)
Smoking index, mean (SD)	1.66e-16 (0.49)	0.07 (0.54)	1.02 (0.55)

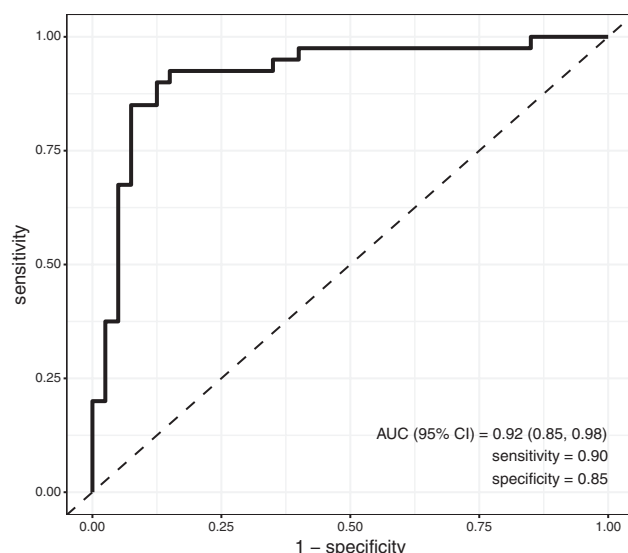


Figure 1. Receiver operating characteristic (ROC) curve of smoking index (SI) as a predictor of smoking versus nonsmoking in buccal cells. This figure depicts the ROC curve of SI as predictor of smoking versus nonsmoking status in buccal cell samples. SI had a sensitivity and specificity of 0.90 and 0.85, respectively. The overall area under the ROC curve was 0.92. AUC = area under the curve.

Table 2. Long-Term Smoking and Moist Snuff Consumption as Predictors of SI, DNA Methylation Age, and Age Acceleration in Buccal Cells ($N = 120$)

Predictor	Difference in outcome (95% CI)	<i>p</i>
Smoking index model		
Nonsmokers	Ref	Ref
Snuff consumers	0.06 (−0.19 to 0.32)	.63
Smokers	1.08 (0.82 to 1.33)	<.0001
DNA methylation age model		
Nonsmokers	Ref	Ref
Snuff consumers	−1.67 (−4.07 to 0.74)	.17
Smokers	−1.44 (−3.82 to 0.94)	.23
Age acceleration model^a		
Nonsmokers	Ref	Ref
Snuff consumers	−1.99 (−4.42 to 0.43)	.11
Smokers	−1.49 (−3.91 to 0.93)	.23

All models adjusted for chronological age, race, and methylation technical covariates (ie, chip, row, and column).

^aAge acceleration model includes all covariates except for chronological age.

cg05575921 [*AHRR*] was also not significantly associated with AgeAccel ($\beta = 4.31$, 95% CI: −2.44 to 11.1, $p = .21$).

Discussion

To the best of our knowledge, this is the first successful application of this novel blood cell-derived 66-CpG SI in buccal cell DNA. We demonstrated a specific positive association between tobacco smoking, but not all tobacco exposures, and SI. We are also the first to report that SI is an excellent predictor of smoking versus nonsmoking status in buccal cell samples. SI had a sensitivity (ie, the fraction of smokers that SI correctly identified as smokers) of 0.90 and a specificity (ie, the fraction of nonsmokers that SI correctly identified as nonsmokers) of 0.85. Furthermore, we replicated null associations that have previously been reported of tobacco smoking with DNAm-age and AgeAccel. Even with the overall null association of smoking with DNAm-age, we found that four DNAm-age CpGs are differentially methylated when comparing smokers and nonsmokers.

To construct the 66-CpG SI, Gao et al.¹⁹ began with 150 CpGs that were associated with active smoking in blood cells. They then identified and validated 66 CpGs that were associated with DNAm-age to create an SI.¹⁹ DNAm-age is a 353-CpG predictor of chronological age that has also been shown to reflect general disease risk and disease processes.²³ To date, DNAm-age has been associated with a host of diseases including lung cancer, Parkinson's disease, and HIV infection.^{18,27–29} Moreover, DNAm-age has been linked to molecular process involving metabolism, the immune system, cellular aging, and cancer.^{30–35} We hypothesized that this molecularly informative 66-CpG SI could also be applied to buccal cells. Buccal cells are of particular interest for tobacco smoke exposure, because they are among the few cell types directly exposed to tobacco smoke. Moreover, in real-world settings, buccal cells are easier to obtain than blood cells.³⁶ Again, this is the first study to successfully apply this novel 66-CpG SI across cell types and demonstrate that this SI is specifically related to combustion tobacco exposures. This latter point is critical because it suggests specific relationships between combustion-related tobacco exposures and methylation-related measures. This point has also been made regarding the relationships of smoking and smokeless tobacco with general DNA methylation, and should be carefully considered when this SI is applied in the future.³⁷

Still, our study is not the first study comparing smoking-related DNA methylation relationships between blood and buccal cells. An earlier research group performed a smoking epigenome-wide association study in matched buccal and blood cell samples from 152 women from the United Kingdom Medical Research Council National Survey of Health and Development birth cohort.³⁸ This

Table 3. DNA methylation age Component CpG Probes Differentially Methylated between Long-term Smokers and Nonsmokers

CpG	Gene	Process/function	Sample size methylation β value mean (SD)	Difference in methylation β value for smokers (95% CI)	False discovery rate adjusted <i>p</i>
cg10266490	<i>ACOT11</i>	Lipid binding	0.15 (0.06)	0.06 (0.03 to 0.09)	.01
cg14992253 ^a	<i>EIF3I</i>	Translation initiation	0.12 (0.02)	−0.02 (−0.03 to −0.01)	.01
cg08030082	<i>POMC</i>	Regulation of hormone activity	0.42 (0.06)	−0.05 (−0.08 to −0.03)	.03
cg15804973	<i>MAP3K5</i>	Protein kinase activity	0.27 (0.07)	−0.05 (−0.08 to −0.02)	.04

CpG associations are from a model that is fully adjusted for chronological age, race, and methylation technical covariates (ie, chip, row, and column).

^aCpG associated with long-term fine-particle ($PM_{2.5}$) levels in a prior publication.

study found that the top-ranked CpGs were similarly hypomethylated across both tissue types although many more significant associations were identified in buccal cells. Ultimately, this suggested that although both tissue types could be used for assessing smoking relationships, buccal cells may be more sensitive. Our results further support the assertion of buccal cell sensitivity. These same authors then created a 1501-CpG SI to discriminate normal tissues from cancer tissues.³⁸ This 1501-CpG SI has weaker relationships with DNAm-age measures and thus, is less comprehensively associated with general disease biology than the 66-CpG SI.¹⁹

In contrast to the findings of Gao et al.,¹³ we observe no strong relationships between DNAm-age and SI even though DNAm-age prediction performed well in our study sample. To help understand why this may have occurred, we examined relationships of the top five smoking-related CpGs from the existing literature. These five sites represent the CpGs that have been identified in multiple epigenome-wide association studies and that have the largest reported differential methylation between smokers and nonsmokers.¹³ Two of these CpGs are part of the 66-CpG SI. From this analysis, we found that all of the CpGs were hypomethylated in our study sample, just as they were in the published literature, and that the effect sizes that we estimated were comparable to those previously reported. We next returned to the original DNA-age article to compare the performance of the DNAm-age in peripheral blood cells versus buccal cells. We found that DNAm-age performs slightly better in blood cells: $r = .96$ versus $r = .83$.²³ Still, we do not fully attribute the lack of a DNAm-age and SI association to this performance difference. We also have a much smaller sample size than the original 66-CpG SI cohort (120 versus 1509) and we do not use the exact same model framework to examine the association due to a lack of information regarding alcohol consumption, physical activity, and body mass index in our publicly available data. The lack of an observed relationship between DNAm-age and SI in buccal cells may also be biologically informative. The relationship with disease risk that was gained by choosing DNAm-age-related CpGs for the SI is one characteristic that makes the SI unique, particularly for potential clinical applications. An inability to conserve this feature across cell types may imply that even though the SI demonstrates multi-tissue utility for detecting tobacco smoke exposure, the 66 CpGs may specifically be sensitive for blood cell pathological processes. Ultimately, a larger cohort will be needed to validate the findings from our pilot study and truly determine the relationship of the 66-CpG SI and DNAm-age in buccal cells.

In line with the consensus established by the literature,^{19,23,39} we did not observe significant relationships of DNAm-age or AgeAccel with smoking status in buccal cells. We also did not observe any significant association of cg05575921 [AHRR] methylation with DNAm-age or AgeAccel. Still, we looked to see if any of the 353 DNAm-age component CpG sites had significant associations with smoking status. In our comparison of smokers to nonsmokers we identified four CpGs related to four genes: cg10266490 [ACOT11], cg14992253 [EIF3I], cg08030082 [POMC], and cg15804973 [MAP3K5]. Some of these genes have known relationships with smoking.⁴⁰ For instance, transcription of *POMC* produces the protein pro-opiomelanocortin (POMC), which is eventually cleaved into a number of peptides including adrenocorticotrophic hormone. Nicotine is thought to decrease appetite through activation of *POMC* in neurons.⁴¹ Importantly, cg14992253 [EIF3I] was previously identified in a study examining the association of long-term fine-particle exposures with DNAm-age.⁴² The EIF3I protein is involved in the initiation of eukaryotic translation, and its identification in this

study suggests that some degree of overlap exists between DNAm-age-related biology involving air pollution and combustible tobacco smoke. This provides additional evidence for the characterization of smoking as a form of personal air pollution and the utility of DNAm-age-related measures in assessing such exposures.^{43,44}

Our study presents a number of novel findings despite some limitations. First, this is a relatively small cross-sectional analysis that was focused on determining if a novel SI could even be utilized in buccal cells. Our findings suggest that the SI is indeed applicable across buccal cells and blood. Still, a larger and more extensive population-based study is necessary to confirm the SI sensitivity and specificity we report. This will be critical for utilizing the SI in future settings including but not limited to scientific research. We also note that our findings are based on a middle-age cohort of predominantly Caucasian males who were recruited in the southeastern United States. Additional studies involving other demographic groups and in different environments will be necessary to confirm our findings more broadly. Finally, we used a publicly available dataset with limited information on lifestyle factors, other environmental exposures, and other important covariates. Even though we adjust for the important covariates that we do have (eg, chronological age, race, and technical covariates), we cannot rule out the possibility of unknown or residual confounding in our analyses.

In conclusion, our study provides novel evidence that a blood-derived 66-CpG SI can be utilized in buccal cells to differentiate long-term tobacco smokers from nonsmokers. Importantly, we also provide evidence that the SI is not related to other tobacco practices such as long-term moist snuff consumption. Our findings represent early steps in the process of identifying an accurate, easily accessible biomarker that reflects tobacco smoke exposure and provides some insight to disease risks. Still, a number of issues must be addressed before the clinical utility of the SI can be actualized. Future research in larger cohorts will need to determine normal values for the SI CpGs so that the measure is not as study sample specific (ie, not based on nonsmokers in each respective study sample). Moreover, it will be interesting to assess the utility of the SI in differentiating long-term smokers from former smokers, individuals who recently began smoking, and individuals who smoke irregularly. Future efforts should also focus on optimizing the relationship of the SI with DNAm-age so that the SI metric maximizes its utility as a tool for understanding general disease processes.

Supplementary Material

Supplementary data is available at *Nicotine & Tobacco Research* online

Funding

JCN is supported by an NIH/NIA Ruth L. Kirschstein National Research Service Award (1 F31AG056124-01A1).

Declaration of Interests

None declared.

Acknowledgments

The authors also acknowledge Jessen WJ, Borgerding MF, and Prasad GL from RAI Services Company (401 N Main Street, Winston-Salem, NC) for making the dataset publicly available.

References

- Do E, Maes H. Narrative review of genes, environment, and cigarettes. *Ann Med*. 2016;48(5):337–351.
- Connor Gorber S, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. *Nicotine Tob Res*. 2009;11(1):12–24.
- Florescu A, Ferrence R, Einarson T, Selby P, Soldin O, Koren G. Methods for quantification of exposure to cigarette smoking and environmental tobacco smoke: focus on developmental toxicology. *Ther Drug Monit*. 2009;31(1):14–30.
- Russell T, Crawford M, Woodby L. Measurements for active cigarette smoke exposure in prevalence and cessation studies: why simply asking pregnant women isn't enough. *Nicotine Tob Res*. 2004;6(suppl 2):S141–S151.
- Kim S. Overview of cotinine cutoff values for smoking status classification. *Int J Environ Res Public Health*. 2016;13(12):1236.
- Flores RM, Liu B, Taioli E. Association of serum cotinine levels and lung cancer mortality in non-smokers. *Carcinogenesis*. 2016;37(11):1062–1069. doi:10.1093/carcin/bgw094.
- Corbett C, Armstrong MJ, Neuberger J. Tobacco smoking and solid organ transplantation. *Transplantation*. 2012;94(10):979–987.
- Schick SF, Blount BC, Jacob P Rd, et al. Biomarkers of exposure to new and emerging tobacco delivery products. *Am J Physiol Lung Cell Mol Physiol*. 2017;313(3):L425–L452.
- Zuo H, Nygård O, Vollset SE, et al. Smoking, plasma cotinine and risk of atrial fibrillation: The Hordaland Health Study. *J Intern Med*. 2018;283(1):73–82.
- Zhao R, Wu Y, Zhao F, et al. The risk of missed abortion associated with the levels of tobacco, heavy metals and phthalate in hair of pregnant woman: a case control study in Chinese women. *Medicine (Baltimore)*. 2017;96(51):e9388.
- Talikka M, Siervo N, Ivanov NV, et al. Genomic impact of cigarette smoke, with application to three smoking-related diseases. *Crit Rev Toxicol*. 2012;42(10):877–889.
- Chatterjee A, Rodger EJ, Morison IM, Eccles MR, Stockwell PA. Tools and strategies for analysis of genome-wide and gene-specific DNA methylation. *Methods Mol Biol*. 2017;1537:249–277.
- Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin Epigenetics*. 2015;7(1):113.
- Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG. AHRH (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax*. 2017;72(7):646–653.
- Guida F, Sandanger TM, Castagné R, et al. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum Mol Genet*. 2015;24(8):2349–2359.
- Cella M, Colonna M. Aryl hydrocarbon receptor: linking environment to immunity. *Semin Immunol*. 2015;27(5):310–314.
- Hahn ME, Allan LL, Sherr DH. Regulation of constitutive and inducible AHR signaling: complex interactions involving the AHR repressor. *Biochem Pharmacol*. 2009;77(4):485–497.
- Declerck K, Vanden Bergh W. Back to the future: epigenetic clock plasticity towards healthy aging. *Mech Ageing Dev*. 2018.
- Gao X, Zhang Y, Breitling LP, Brenner H. Relationship of tobacco smoking and smoking-related DNA methylation with epigenetic age acceleration. *Oncotarget*. 2016;7(30):46878–46889.
- Chen BH, Marioni RE, Colicino E, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)*. 2016;8(9):1844–1865.
- Woo JG, Sun G, Haverbusch M, et al. Quality assessment of buccal versus blood genomic DNA using the Affymetrix 500 K GeneChip. *BMC Genet*. 2007;8(1):79.
- Prasad GL, Jones BA, Chen P, Gregg EO. A cross-sectional study of biomarkers of exposure and effect in smokers and moist snuff consumers. *Clin Chem Lab Med*. 2016;54(4):633–642.
- Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115.
- Barfield RT, Kilari V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics*. 2012;28(9):1280–1281.
- Boyle EI, Weng S, Gollub J, et al. GO:TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*. 2004;20(18):3710–3715.
- Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. *Bioinformatics*. 2005;21(20):3940–3941.
- Horvath S, Ritz BR. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY)*. 2015;7(12):1130–1142.
- Levine ME, Hosgood HD, Chen B, Absher D, Assimes T, Horvath S. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY)*. 2015;7(9):690–700.
- Horvath S, Levine AJ. HIV-1 Infection accelerates age according to the epigenetic clock. *J Infect Dis*. 2015;212(10):1563–1573.
- Lin Q, Wagner W. Epigenetic aging signatures are coherently modified in cancer. *PLoS Genet*. 2015;11(6):e1005334.
- Quach A, Levine ME, Tanaka T, et al. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (Albany NY)*. 2017;9(2):419–446.
- Nwanaji-Enwerem JC, Bind MA, Dai L, et al. Editor's highlight. Modifying role of endothelial function gene variants on the association of long-term PM2.5 exposure with blood DNA methylation age: The VA Normative Aging Study. *Toxicol Sci*. 2017;158(1):116–126.
- Nwanaji-Enwerem JC, Colicino E, Dai L, et al. Impacts of the mitochondrial genome on the relationship of long-term ambient fine particle exposure with blood DNA methylation age. *Environ Sci Technol*. 2017;51(14):8185–8195.
- Lowe D, Horvath S, Raj K. Epigenetic clock analyses of cellular senescence and ageing. *Oncotarget*. 2016;7(8):8524–8531.
- Nwanaji-Enwerem JC, Weisskopf MG, Baccarelli AA. Multi-tissue DNA methylation age: molecular relationships and perspectives for advancing biomarker utility. *Ageing Res Rev*. 2018;45:15–23.
- Proia NK, Paszkiewicz GM, Nasca MA, Franke GE, Pauly JL. Smoking and smokeless tobacco-associated human buccal cell mutations and their association with oral cancer—a review. *Cancer Epidemiol Biomarkers Prev*. 2006;15(6):1061–1077.
- Besing W, Johansson A. Smoke-related DNA methylation changes in the etiology of human disease. *Hum Mol Genet*. 2014;23(9):2290–2297.
- Teschendorff AE, Yang Z, Wong A, et al. Correlation of smoking-associated DNA methylation changes in buccal cells with DNA methylation changes in epithelial cancer. *JAMA Oncol*. 2015;1(4):476–485.
- Marioni RE, Shah S, McRae AF, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol*. 2015;44(4):1388–1396.
- Lee MK, Hong Y, Kim SY, London SJ, Kim WJ. DNA methylation and smoking in Korean adults: epigenome-wide association study. *Clin Epigenetics*. 2016;8(1):103.
- Mineur YS, Abizaid A, Rao Y, et al. Nicotine decreases food intake through activation of POMC neurons. *Science*. 2011;332(6035):1330–1332.
- Nwanaji-Enwerem JC, Dai L, Colicino E, et al. Associations between long-term exposure to PM2.5 component species and blood DNA methylation age in the elderly: The VA normative aging study. *Environ Int*. 2017;102:57–65.
- Nwanaji-Enwerem JC, Colicino E, Trevisi L, et al. Long-term ambient particle exposures and blood DNA methylation age: findings from the VA normative aging study. *Environ Epigenet*. 2016;2(2):1–10.
- Ward-Caviness CK, Nwanaji-Enwerem JC, Wolf K, et al. Long-term exposure to air pollution is associated with biological aging. *Oncotarget*. 2016;7(46):74510–74525.