ARIC Manuscript Proposal #3699

PC Reviewed: 9/8/20	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: DNA methylation markers for periodontal disease and tooth loss

b. Abbreviated Title (Length 26 characters): Epigenetics markers of periodontitis

2. Writing Group:

Dominique S. Michaud, Jiayun Lu, James D. Beck,. Eric Boerwinkle, Jan Bressler, Elizabeth A. Platz

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. DSM **[please confirm with your initials electronically or in writing]**

First author: Dominique S. Michaud
Address: Department of Public Health & Community Medicine
Tufts University School of Medicine
136 Harrison Avenue, M&V 253
Boston, MA 02111
(617) 636-0482
E-mail: Dominique.Michaud@tufts.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Name: Elizabeth A. Platz
Address: Department of Epidemiology, Rm E6132
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe St.
Baltimore, MD 21230

Phone: 410-614-9674 Fax: 410-614-2632
E-mail: eplatz1@jhu.edu

3. Timeline: Manuscript drafted by September 2021

4. Rationale: Most cohort studies with large numbers of cancers have no or limited data on periodontal disease status, making it difficult to assess the relationship between this condition and cancer risk. In ARIC, we previously observed that severe periodontal disease (HR=1.24, 95% CI 1.07-1.44, p-trend=0.004) and self-reported edentulism (no teeth; HR=1.28, 95% CI 1.09-1.50) were associated with increased cancer risk (1). These associations were stronger for cancer mortality (severe periodontal disease: HR=1.52, 95% CI 1.17-1.97, p-trend=0.002; edentulism HR=1.64, 95% CI 1.25-2.16). By cancer site, associations were strongest for lung cancer risk (HR=2.33, 95% CI 1.51-3.60, p-trend<0.0001), including possibly among never smokers. For colorectal cancer, associations were present for both white and black participants, especially when restricting to never smokers. The great strengths of this work compared other work in the small existing literature were: more accurate determination of periodontal disease by the use of a standardized dental examination rather than by self-report of the diagnosis; reduction in confounding by accounting for known cancer risk factors that are also the major causes of periodontal disease in the US, including smoking and diabetes, by adjustment and restriction; documentation that competing risks of death did not explain the findings; and addressing this association in both white and black participants. Our findings add to the number of studies supporting the hypothesis that periodontal disease increases cancer risk, observed in numerous cohort studies (2). To further address causality, we recently performed a Mendelian randomization study using genetic variants of periodontal disease to test for causal associations with lung, pancreatic and colorectal cancers and identified positive associations for colon cancer, but not for pancreatic and lung cancers (unpublished data). At this time, we are very interested in expanding our research on periodontal disease and cancer risk and mortality to other datasets, but need a better measure for the biological impact of different stages of periodontal disease.

In recent years, there has been tremendous promise in blood DNA methylation markers – providing new insight into risk factors, biological pathways, and disease processes. In cardiovascular disease, recent studies using blood DNA methylation levels from high dimensional arrays has resulted in identification of numerous new pathways (3, 4). Large studies have identified regions in the human genome that are differentially methylated in circulating leukocytes of subjects with elevated subclinical inflammation (5, 6), obesity (7, 8), type II diabetes (9), and smoking (10, 11). In one study, DNA methylation markers associated with smoking improved risk prediction of lung cancer in former smokers, demonstrating that biomarkers can add information on internal dose that cannot be reliably obtained with history of smoking data. For obesity, it was estimated that DNA methylation levels more often changed as a consequence of the phenotype (i.e., BMI), rather than being determined by genetic susceptibility (7). When not strongly driven by genetic susceptibility (e.g., imprinting), changes in DNA methylation levels in peripheral blood can serve as biological markers of exposure or disease occurrence (7-9). Therefore, **DNA methylation levels provide more integrated measures of exposure, genetic susceptibility and immune changes that occur in individuals across a life span and show promise as markers of integrated risk.** Identifying changes in methylation that occur in leukocytes as a result of periodontal disease progression could provide insight into biological mechanisms. These findings would be of relevance for periodontal disease, but also for other chronic conditions including diabetes, stroke and cancer.

5. Main Hypothesis/Study Questions:

In a cross-sectional study:

1. Determine whether DNA methylation levels at certain loci are strongly and consistently associated with periodontal disease and whether these methylation changes occur in a dose-response manner with increasing disease severity.

2. Determine whether DNA methylation levels at certain loci are strongly and consistently associated with tooth loss and whether these methylation changes occur in a dose-response manner with increasing tooth loss.

3. Examine overlap in CpGs between 1) and 2)

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Study design: Cross-sectional design

Analytic population: Men and women who self-reported being edentulous at Visit 4 or who attended the clinical dental examination at Visit 4, with DNA methylation data from Visit 3, and who consented to genetic studies and studies on chronic diseases.

Exposure: Available DNA methylation data in ARIC study from Visit 3. By the end of 2020, ARIC participants who consented to genetic studies and studies on cancer and other chronic diseases will have available DNA methylation data at Visit 3. The majority of participants will have data from the Illumina MethylationEPIC arrays (~11,500 are currently being arrayed by TOPMed).

CDC-AAP		Original ARIC	
Periodontal disease status	Measurement	Periodontal disease status	Measurement
No	No evidence of mild, moderate, or severe periodontitis		
Mild	≥2 interproximal sites with AL≥3mm, and ≥2 interproximal sites with PD≥4mm (not on same tooth) or one site with PD≥5mm	No/mild	10% of examined sites having AL≥3 mm
Moderate	≥2 interproximal sites with AL≥4mm (not on same tooth), or ≥2 interproximal sites with PD≥5mm (not on same tooth)	Moderate	≥10% to <30% of examined sites having AL≥3 mm
Severe	≥2 interproximal sites with AL≥6mm (not on same tooth) and ≥1 interproximal site with PD≥5mm	Severe	≥30% of examined sites with AL≥3 mm

Outcomes:
Aim 1. We will classify participants using two definitions of periodontal disease and data from the Visit 4 dental examination (see **table** above): 1) US Centers for Disease Control and Prevention - American Academy of Periodontology (CDC-AAP) definition developed for population-based surveillance of periodontitis, which uses both clinical attachment level and

pocket depth measurements (12); 2) the definition based only on clinical attachment level measurements used by Beck et al. (13) in ARIC previously. For definitions 1 and 2, we will also use self-reported edentulism at Visit 4. The reference for both definitions will be consist of participants with no PD and who were not edentulous.

Aim 2. Teeth number will be obtained from the Visit 4 oral examination and self-reported edentulism status.

Other variables: Age, race, derived BMI at Visit 3, current smoking status and packyears smoked by Visit; alcohol drinking at Visit 3 (never, former, or current drinker), diabetes status at Visit 3 (diagnosed: MD diagnosis, medications; undiagnosed: fasting glucose ≥126 mg/dL at any visit and/or glycated hemoglobin ≥6.5% at Visit 2; at risk for diabetes: fasting glucose of 100 to <126 mg/dL at visit 4; if not fasting, prior visit concentration will carried forward); ever use of hormone replacement therapy (women only; Visits 1 and 3); lifecourse SES calculated using data from ancillary study at Visit 4 as done previously in ARIC (14); US Census tract data on neighborhood income for the year 2000 (15).

Estimation of immune cell composition. Leukocyte subtypes proportions (i.e., CD4T, CD8T, natural killer cells (NK), B cells, monocytes (Mono) and neutrophils) will be estimated using the "*estimateCellCounts2*" function in the FlowSorted.Blood.EPIC Bioconductor package (16), which is based on previously published reference-based cell mixture deconvolution algorithm with reference library selection conducted using the IDOL methodology developed by our team (17). Immune cell ratios (e.g., Neutrophil-to-Lymphocyte ratio (NLR), CD4/CD8, etc.) can be ascertained using the cell proportion estimates (obtained via the "*estimateCellCounts2*" function).

Data analysis:
We will initially conduct linear regression models to identify CpGs most strongly associated with periodontal disease or tooth loss – for these models, CpGs will be the dependent variables, and surrogate variables for batch effects and cell proportions will be independent variables along with periodontal disease or tooth loss. Given the large number of tests being performed, multiplicity adjustment will be performed by computing the Benjamini-Hochberg false discovery rate (FDR) and CpGs with an FDR ≤ 0.05 will be considered statistically significant. Using the CpGs identified in those models, we will perform multivariate analyses by fitting polynomial regression models for periodontal disease and linear regression models for teeth number [continuous variable]) as the dependent variables and each CpG methylation level as independent variables (one at a time); each regression model will be adjusted for education (<high school, high school, >high school), field center*race (black from suburban Minneapolis, Forsyth County or Washington County; white from Forsyth County or Washington County; black from Jackson) and risk factors for periodontal disease, i.e., smoking (current, former, never; packyears smoked [continuous]), BMI (continuous), diabetes status (diagnosed, undiagnosed, at risk for diabetes, none), alcohol drinking (never, former, or current drinker), lifecourse SES, US Census tract data on neighborhood income, and estimated cell proportions. In addition, technical variables will be included to account for potential confounding by genetic ancestry, and surrogate variables (to control for unmeasured batch effects). We will repeat these analyses stratified by race and will test for statistical interaction between the significant CpGs and race using the likelihood ratio test. We will also consider a second analytical approach to identify genomic regions using the

DMRcate Bioconductor R package (18). DMRcate will be applied using the following settings: region length 2000bp; min of 2 significant CpGs; FDR pvalue<0.05 (adjusting for the same covariates as in the single CpG analyses).

Methodologic challenges:
We realize the DNA methylation levels are measured in blood from samples collected a few years prior to the periodontal disease assessments. Thus, we are making the assumption that DNA methylation levels are relatively stable during the progression of periodontal disease and that we are measuring disease status at slightly earlier stage than observed. We will conduct sub-analyses removing participants who have diabetes to determine whether it is a major explanatory factor in the changes in methylation levels associated with periodontal disease and/or teeth number. Similarly, we will conduct a separate analysis to examine associations among never smokers.

7.a. Will the data be used for non-CVD analysis in this manuscript? _X__ Yes ____ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = "CVD Research" for non-DNA analysis, and for DNA analysis RES_DNA = "CVD Research" would be used? _X__ Yes ____ No
(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? _X__ Yes ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"? _X__ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status.
ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/ARIC/search.php

__X___ Yes _____ No

No overlap

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

Many manuscript proposals mention periodontal disease and/or other dental-related measures, including MS1892, MS2191, MS2449, MS2453, MS942, MS1079, MS658, MS566, MS995, MS858, MS913, MS1593, MS852, and MS1937. A key investigator in these proposals is James D. Beck, who is an investigator on this current manuscript proposal.

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? __X__ Yes ____ No

11.b. If yes, is the proposal

_X__ A. primarily the result of an ancillary study (list number* 1996.0)

- Dental examination data generated as part of 1996.0 (Dental Study) – Dr. Beck

_X__ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* 1998.02, 2004.05)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

- Lifecourse SES data generated as part of 1998.02 (Life courses SES, social context, and CVD) – Dr. Heiss
- Census tract income data generated as part of 2004.05 (Socioeconomic characteristics of place of residence: impact on rates and trends in nonfatal and fatal CHD in the ARIC Surveillance Communities) – Drs. Heiss and Rose

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes _X__ No.

References

1. Michaud DS, Lu J, Peacock-Villada AY, Barber JR, Joshu CE, Prizment AE, et al. Periodontal Disease Assessed Using Clinical Dental Measurements and Cancer Risk in the ARIC Study. Journal of the National Cancer Institute. 2018.
2. Michaud DS, Fu Z, Shi J, Chung M. Periodontal Disease, Tooth Loss, and Cancer Risk. Epidemiol Rev. 2017;39(1):49-58.

3. Huan T, Joehanes R, Song C, Peng F, Guo Y, Mendelson M, et al. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. Nature communications. 2019;10(1):4267.
4. Agha G, Mendelson MM, Ward-Caviness CK, Joehanes R, Huan T, Gondalia R, et al. Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial Infarction and Coronary Heart Disease. Circulation. 2019;140(8):645-57.
5. Ligthart S, Marzi C, Aslibekyan S, Mendelson MM, Conneely KN, Tanaka T, et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. Genome biology. 2016;17(1):255.
6. Ahsan M, Ek WE, Rask-Andersen M, Karlsson T, Lind-Thomsen A, Enroth S, et al. The relative contribution of DNA methylation and genetic variants on protein biomarkers for human diseases. PLoS genetics. 2017;13(9):e1007005.
7. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2016;541:81.
8. Xu K, Zhang X, Wang Z, Hu Y, Sinha R. Epigenome-wide association analysis revealed that SOCS3 methylation influences the effect of cumulative stress on obesity. Biol Psychol. 2018;131:63-71.
9. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. The lancet Diabetes & endocrinology. 2015;3(7):526-34.
10. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Human molecular genetics. 2013;22(5):843-51.
11. Baglietto L, Ponzi E, Haycock P, Hodge A, Bianca Assumma M, Jung CH, et al. DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. International journal of cancer Journal international du cancer. 2017;140(1):50-61.
12. Eke PI, Page RC, Wei L, Thornton-Evans G, Genco RJ. Update of the case definitions for population-based surveillance of periodontitis. Journal of periodontology. 2012;83(12):1449-54.
13. Beck JD, Elter JR, Heiss G, Couper D, Mauriello SM, Offenbacher S. Relationship of periodontal disease to carotid artery intima-media wall thickness: the atherosclerosis risk in communities (ARIC) study. Arteriosclerosis, thrombosis, and vascular biology. 2001;21(11):1816-22.
14. Shoham DA, Vupputuri S, Diez Roux AV, Kaufman JS, Coresh J, Kshirsagar AV, et al. Kidney disease in life-course socioeconomic context: the Atherosclerosis Risk in Communities (ARIC) Study. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2007;49(2):217-26.
15. Foraker RE, Patel MD, Whitsel EA, Suchindran CM, Heiss G, Rose KM. Neighborhood socioeconomic disparities and 1-year case fatality after incident myocardial infarction: the Atherosclerosis Risk in Communities (ARIC) Community Surveillance (1992-2002). Am Heart J. 2013;165(1):102-7.

16. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome biology. 2018;19(64).
17. Koestler DC, Jones MJ, Usset J, Christensen BC, Butler RA, Kobor MS, et al. Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). BMC bioinformatics. 2016;17:120.
18. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, R VL, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin. 2015;8:6.
19. Wong ND, Zhao Y, Patel R, Patao C, Malik S, Bertoni AG, et al. Cardiovascular Risk Factor Targets and Cardiovascular Disease Event Risk in Diabetes: A Pooling Project of the Atherosclerosis Risk in Communities Study, Multi-Ethnic Study of Atherosclerosis, and Jackson Heart Study. Diabetes care. 2016;39(5):668-76.