ARIC Manuscript Proposal # 3131 (Amended)

PC Reviewed: 11/10/2020	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Association between CHIP genotype, Periodontal Disease and Atherosclerotic Cardiovascular disease

b. Abbreviated Title (Length 26 characters): **CHIP-Perio & CVD**

2. Writing Group:

Writing group members:
Kimon **Divaris**-UNC Adams School of Dentistry, Pediatric and Public Health;
Jim **Beck**-UNC Adams School of Dentistry, Periodontology;
Kevin **Moss**-UNC Adams School of Dentistry, Oral and Craniofacial Health Sciences;
Kari E. **North**-UNC Gillings School of Global Public Health, Epidemiology;
Di **Wu**-UNC Gillings School of Global Public Health, Biostatistics;
Wayne D **Rosamond**-UNC Gillings School of Global Public Health, Epidemiology;
Souvik **Sen**- Medical University of South Carolina School of Medicine, Neurology;
George **Hajishengallis**-U Penn Dental Medicine
Triantafyllos **Chavakis**-TU Dresden, Institute for Clinical Chemistry and Laboratory Medicine

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

First author: **Kimon Divaris, DDS, PhD**

Address: Adams School of Dentistry
Division of Pediatric and Public Health
Phone: 919-537-3556 Fax: 919-537-3950
E-mail: Kimon_Divaris@unc.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: **Jim Beck, PhD**

Address: UNC Adams School of Dentistry
Division of Comprehensive Oral Health, Periodontology
Phone: (919) 537-3241
E-mail: Jim_Beck@unc.edu

3. Timeline: Complete analyses by December 2022.

4. Rationale:

Recent evidence has identified a myeloid clonal expansion subset of cells which carry somatic mutations in 4 key genes. This genotype results in the presence of expanded somatic blood-cell clones without other hematological or leukemia abnormalities and is referred to as CHIP [Clonal Hematopoiesis of Indeterminate Potential]; and sometimes referred to as Clonal Hematopoesis of

Aging. CHIP is seen in approximately 10% of the population at age 60, with increasing prevalence with age. These 4 genes carry loss-of-function somatic mutations and include *DNMT3A, TET2, ASXL1* and *JAK2*. In the last few years exome sequence studies have identified and confirmed that CHIP doubles the risk for coronary disease and stroke. CHIP was initially described as a new risk factor for CVD that was independent of lipid pathways and inflammation, however *TET2* function has been recently identified as critical for chromatin remodeling enabling activation of inflammatory genes and inflammasome proteins. *TET2* knockout mice have an exaggerated inflammatory phenotype (excess IL-1, IL-6 and chemokines) and enhanced atherogenesis using BMDB *Tet2*$^{-/-}$ to reconstitute *Lldr* $^{-/-}$ mice [Cull et al. Exp.Hematol. 2017;55:56-70]. In humans, periodontal disease is associated with an enhanced inflammatory phenotype and is also associated with higher risk for CHD, MI and stroke.

5. Main Hypothesis/Study Questions:
This gives rise to our central hypothesis that severe periodontal disease is associated with the CHIP genotype and that there is significant interaction between the CHIP genotype and periodontal disease that "mediates" thick IMT, prevalent CVD (esp. stroke, CHF and MI) and incident events, independent of traditional risk factors.

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).

To be clear, we will be seeking to explore for the presence of somatic mutations in 4 genes that occur in nucleated blood samples collected in ARIC with the understanding that inflammatory infiltrates within periodontal tissues share a common origin as these blood cells- the bone marrow. We propose to use the whole exome (WES) and whole genome sequencing (WGS) datasets for those subjects in ARIC that have periodontal disease measures, focusing primarily on these 4 genes testing for loss-of-function-mutations, as described by Jaiswal et al. [N Engl J Med 2014;371:2488-2498 and N Engl J Med 2017;377:111-121] who also identified the CHIP genotype in blood with somatic mutations occurring in up to 20% of the cell populations. Using the criteria discussed in this report we will assign each subject as either CHIP negative or CHIP positive based upon the presence of loss-of-function mutations in these 4 genes including [deleterious frame shifts, indels, missense and splice variants, as flagged using CADD described in Kircher et al. [Nat Gen 2014;46(3):310-5]. We will examine the relationship between the CHIP positive vs CHIP negative subjects and the presence or absence of severe periodontal disease using Chi square and general logits models adjusting for relevant risk factors. We will confirm the association between CHIP and thick IMT and stroke (testing ischemic and hemorrhagic stroke separately) using visit 4 prevalence data. We anticipate null findings with hemorrhagic stroke and positive associations with ischemic stroke and will secondarily consider ischemic stroke subtypes (large artery, small vessel occlusive as well as cardio-embolic, as reported by Sen et al. [Stroke 2018;49(2):355-62]. We will examine for the potential for "causality" and directionality of association using mediation analyses. We will also explore the relationship of the CHIP genotype, periodontal status at visit 4 and incident CVD events using time-to event analyses.

Secondarily, since periodontal disease is associated with increased systemic dissemination of oral pathobionts that is enhanced by periodontal inflammation we will examine for the association

between CHIP and the presence or level of specific organisms using published methods (Divaris et al. J Dent Res 2012;91(7 Suppl):21S-28S.).

Detailed Methods: Few studies have developed and demonstrated the utility of using WGS data that contain low frequency and rare genetic variation for complex traits. Morrison and colleagues have used the WGS data from ARIC and demonstrated an analytical pipeline that emphasizes an *a priori* canonical selection of gene selection, based upon known biomarkers or established pathogenic pathways. We will use these methods selecting for those initial CHIP subset of 4 genes, but survey the additional 68 as described in Table S2 of Jaiswal et al. [N Engl J Med 2017;377:111-121]. As a novel addition to this analysis- since the dominant gene that differentiates the CHIP genotype involves somatic mutations of *TET2* which has recently been demonstrated to control chromatin structure to enhance accessibility and expression of inflammasomes, cytokines and chemokines, we will create a custom gene set and include a query that explores for association with loss-of-function mutations in a *TET2* effector series of canonical genes (using IPA) to include inflammasomal proteins and cytokines. This is particularly promising as we have identified genome-wide significant associations of *IFI16* and selected pyrin domain containing proteins (near significant) with biologically-informed periodontal disease traits [Offenbacher et al. Hum Mol Genet. 2016;25(10):2113-29]. For these analyses we will use the criteria for loss-of function described in Jaiswal et al. [N Engl J Med 2017;377:111-121]

Power calculation: Assuming prevalence of CHIP of 7% (corresponding to a mean age of 65 in the cohort), we expect 469 cases for which we have WGS data, cardiovascular data and periodontal data. We will identify 469 controls (1:1 ratio) that are frequency-matched for age (within 2 years) race and gender. If we assume a hazard ratio of 2.0 for coronary heart disease, we estimate approximately ~90% power to detect a difference at the 0.05 two-sided significance level. We will score a subject with one or more mutations in any one of the four genes as CHIP positive, CHIP negative will have no mutations in any of the four genes. We will use CADD [Kircher et al. Nat Gen 2014;46(3):310-5] for scoring of potentially deleterious genotypes. For other complex genotypes we will use Fisher's exact or logistic regression correcting for multiple testing based upon the numbers of genes in the *a priori* defined gene sets. Frameshift, nonsense, and splice-site mutations will be excluded if they occurred in the first or last 10% of the gene open reading frame, unless mutations in those regions had been previously reported, (e.g., *DNMT3A*), as described elsewhere [Jaiswal et al. N Engl J Med 2014;371:2488-2498 and N Engl J Med 2017;377:111-121]. We have a preselected gene set for levels of *P.gingivalis, T. denticola, Tannerella forsythia* and *Actinobacillus actinomycetemcomitans* (4 key pathobionts in periodontal disease) plus *Fusobacterium nucleatum--F.n,* This organism is critical for establishing dysbiosis and is a highly invasive pathobiont most commonly associated with systemic dissemination.

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

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)?
#1471 Divaris, Offenbacher, Beck, North
Uddin, M. Replication of CHIP associated EWAS loci in the ARIC Study
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 (1996.01) ___
_____B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

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

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.