1.a. Full Title: GWAS of Dental Traits at Surface-Level Resolution

b. Abbreviated Title (Length 26 characters): RESOLUTION-GWAS

2. Writing Group:
   Writing group members: Jim Beck and Kevin Moss- Periodontology, Kari North, Kimon Divaris, Cary Agler- Genetic Epidemiology, Dmitry Shungin - Broad Institute of MIT and Harvard

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. KD

First author: Kimon Divaris, DDS, PhD
Address: Adams School of Dentistry
Department of Pediatric and Public Health
Brauer 228, CB #7450, Chapel Hill, NC 27599
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
Address: Adams School of Dentistry
Department of Periodontology
Phone: Fax:
E-mail: Jim_Beck@unc.edu


4. Rationale:
The genetic contribution to oral health outcomes and the heritability of dental caries and periodontitis have been reported to be as high as 50%[1-3]. However, the nature of this contribution remains poorly characterized despite the promise that increased understanding of genetic factors can bring with respect to etiology dissection and theoretically, the clinical management of these diseases.
Genome-wide association studies (GWAS) for dental caries have investigated measures including overall caries experience [4,5], specific presentations of disease [6] and the presence or absence of disease in pediatric populations [7,8]. To date, only few reliable genetic association signals have been published and this is likely due to different measurement approaches used, the complex genetic architecture [9] of dental caries or limited statistical power to detect associations. GWAS for periodontitis have investigated measures including the presence or absence of disease [10], quantitative measures of periodontal status [11], severe presentations of disease [12,13], molecular and microbial intermediates of disease [14] and composite phenotypes such as principal components which aim to capture multiple facets of periodontal health [15]. These studies have not yielded consistent evidence of specific genetic contributions to periodontitis [8].

We have performed a large-scale meta-analysis of clinical and self-reported caries and periodontitis data in more than ½ million individuals (Shungin et al, Nature Communications, in review). This study discovered 47 novel loci implicated in caries at the robust 5x10^{-8} P-values significance threshold when combining data from studies with clinically-assessed summary indices of caries experience (DMFS and DFS) and self-reported data. We have observed heterogeneity of genetic effects for DMFS and DFS that might be explained by variability in genetic underpinnings intra-oral caries patterns (e.g., occlusal disease versus smooth surface disease) implying tooth and surface-specificity of genetics effects. This is also supported by differences in heritability of clustered caries surfaces. Such surface-specific genetic effects for caries have not been studied in the context of a large-scale GWAS meta-analysis previously. For periodontitis defined using the CDC/AAP classification criteria and in self-reported data we identified only one locus in more than 510,000 participants in similar, combined meta-analyses. Such low number of loci highlighted in such a large sample size is unusual for a polygenic trait with SNP-based heritability estimates ranging between 10 and 30% and is very likely explained by the inherent imprecision of the periodontitis definition when data from multiple sites are combined into one definition.

5. Main Hypothesis/Study Questions:
Here we propose to utilize the full spectrum of available dental data using tooth- and surface-level (or site-level) information for caries and periodontitis within the GWAS context.

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).
We will perform GWAS analysis of tooth surface- and periodontal site-level data.

Detailed Methods:
*Genetic association analyses*
Single variant association analyses will be performed using linear or logistic regression models, depending on the distribution of the outcome variable, with appropriate adjustment for population structure using either genetic principal components as covariates (up to 20 first PCs) or using genetic correlation matrix within mixed models implemented in GCTA. Lead genomic loci for a given region will be selected using approximate conditional analysis with GCTA. The level of genome-wide significance will be set at 5x10^{-8}, and will be further corrected for the number of
sub-analyses for caries and periodontitis. Gene-based analyses will be performed using the MAGMA generalized gene-set analysis tool [16].

We will interrogate the results for both directionality and magnitude of association with other traits that have publicly available GWAS or are in the UK biobank data at the level of genome-wide significance using a set of in-house and online tools.

Results of association analyses will be functionally annotated using a variety of bioinformatics methods including pathway analyses (DEPICT, MAGENTA, etc.), imputation of gene transcript levels using elastic net models fitted on GTEx tissues implemented in S-PrediXcan17 tool, chromatin and eQTL mapping, etc.

We will also attempt to estimate the shared genetic contribution across a range of traits with available genome-wide association results using LD-score regression, as well as to calculate partitioned heritability by functional annotation.

**Phenotypes**

**Periodontal health:** We will use probing depth and clinical attachment loss data for periodontal site, both as continuous trait and using 4, 5 and 6mm cut-off groups, as separate phenotypes, as well as utilize several approaches of clustering of sites to obtain aggregate phenotypes.

**Caries:** We will use caries experience data for each tooth/surface as a separate phenotype (decayed, filled or missing) and utilize several approaches of clustering of surfaces to obtain aggregate phenotypes.

**Power calculation:** Results from ARIC will be meta-analyzed with other studies, effectively comprising the majority, if not all, studies with GWAS and dental data available up to date.

**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

**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)?

ms1471

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* 1996.01_________)
     ___     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.

References