EXPLORING THE GENETIC BASIS OF CHRONIC PERIODONTITIS: A GENOME-WIDE APPROACH

Kimon Divaris

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department Epidemiology of the Gillings School of Global Public Health.

Chapel Hill
2011

Approved by:
Andrew F. Olshan
Ethan M. Lange
Keri L. Monda
Kari E. North
Steven Offenbacher
ABSTRACT

KIMON DIVARIS: Exploring the genetic basis of chronic periodontitis: a genome-wide approach
(Under the direction of Andrew F. Olshan)

Chronic periodontitis (CP) is a common-complex oral disease that affects the majority of the adult population and is a major cause of tooth loss. The disease is characterized by an oral biofilm pathological shift that contributes to cascade of events leading to periodontal destruction. Factors modulating the establishment of a dysbiotic oral microbiome or affecting the host immunity and inflammatory response are promising preventive and therapeutic targets. Although a substantial genetic component of CP is theorized and numerous candidate-gene studies have been completed, to-date no whole-genome association (GWA) studies have been performed.

We performed a GWA analysis of CP in well-defined cohort of 4500 white subjects who were participants of the Atherosclerosis Risk In Communities study. Traits of interest were the three-level disease CDC/AAP classification (healthy/mild, moderate, severe CP) and a continuous extent of disease [proportion of sites exhibiting ≥3 mm attachment loss (EAL)] measure. Additionally, we examined three traits of high bacterial colonization defined as the highest quintile of the distribution of “red” and “orange” complex bacteria, and Aggregatibacter actinomycetemcomitans that were quantified using DNA-DNA
checkerboard hybridization in a subset of 1020 white study subjects. Genotyping was performed using the Affymetrix 6.0 platform. Imputation to 2.5 million markers was based on HapMap II-CEU and a multiple-test corrected significance threshold was applied (P<5x10^{-8}).

We detected no genome-wide significant signals. However, we found suggestive evidence of association (P<5x10^{-6}) for CP with ten loci including \textit{NPY}, \textit{NIN}, \textit{WNT5A} for severe, \textit{NCR2}, \textit{EMR1} for moderate CP, and \textit{TBX18}, \textit{ETS1}, \textit{DYN2H1}, \textit{TTC26} and \textit{ZC3HAV1} for EAL. Additionally, thirteen loci including \textit{KCNK1}, \textit{FBXO38}, \textit{IL33}, \textit{RUNX2}, \textit{CAMTA1} and \textit{VAMP3} provided suggestive signals of association (P<5x10^{-6}) with the examined “high” bacterial colonization traits. The \textit{NPY} (7p15) locus was replicated in an independent cohort of whites of European descent. These genome-wide scan results from a large well-defined cohort provide information on multiple candidate regions for interrogation in genetic studies of CP. Future investigations providing further replication of these findings may lead to an improved understanding of the complex nature of host-biofilm and -bacteria interactions that characterizes states of health and disease.
C. Analytical strategy......................................................................................................... 42
D. Replication plan for GWA findings.............................................................................. 52
E. Tables ............................................................................................................................ 54
F. References...................................................................................................................... 55

IV. MANUSCRIPT #1: EXPLORING THE GENETIC
BASIS OF CHRONIC PERIODONTITIS: A GENOME-WIDE APPROACH.................... 59
A. Title and authors............................................................................................................ 59
B. Manuscript #1 summary............................................................................................. 60
C. Introduction ................................................................................................................... 61
D. Materials and methods ............................................................................................... 62
E. Results ........................................................................................................................... 65
F. Discussion ...................................................................................................................... 66
G. Tables ............................................................................................................................ 71
H. Figures ........................................................................................................................... 73
I. References ...................................................................................................................... 82
J. Manuscript #1 Supplemental material......................................................................... 85
K. Manuscript #1 Supplemental Tables.......................................................................... 88
L. Manuscript #1 Supplemental Figures......................................................................... 94
M. Manuscript #1 Supplemental References.................................................................. 104

V. MANUSCRIPT #2: RISK LOCI FOR COLONIZATION WITH
PERIODONTAL PATHOGENS: A GENOME-WIDE APPROACH................................. 105
A. Title page..................................................................................................................... 105
B. Manuscript #2 summary............................................................................................. 106
C. Introduction ................................................................................................................... 107
LIST OF TABLES

Table

**Table 1.1** Clinical case definitions of Periodontitis proposed by the CDC working group for use in population-based surveillance of Periodontitis ........................................... 18

**Table 1.2**. Overview of published genetic effects from candidate-gene case control studies of the risk of chronic periodontitis among Caucasians ................................. 19

**Table 3.1**. Software applications used for the genome-wide association analysis, annotation and visualization of results............................................................................................. 54

**Table 4.1** Distribution in the total sample and bivariate associations of periodontal diagnosis (CDC/AAP disease classification) and periodontal attachment loss (extent score: percent of sites with $\geq$3mm attachment loss) with sex, smoking and diabetic status among the white participants of the Dental ARIC study ........................................................................................................... 71

**Table 4.2**. Genome-wide association analysis results of the CDC/AAP chronic periodontitis (CP) classification traits (severe CP vs. healthy and moderate CP vs. healthy) and extent of attachment loss trait (EAL-proportion of sites exhibiting attachment loss $\geq$3mm), among the white participants of the Dental ARIC study (n=4610). Single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF-HapMap II CEU) of $\geq$5% and associated $P<5\times10^{-6}$. The SNP with the lowest $P$-value per locus is presented; additional prioritized SNPs in each locus are presented in the footnote, along with corresponding $R^2$ (based on 1000 genomes pilot 1 release) with the top SNPs........................................................................... 72

**Table 4.3** Clinical case definitions of Periodontitis proposed by the CDC working group for use in population-based surveillance of Periodontitis ........................................... 88

**Table 4.4** Changes in estimate of top SNPs for the three chronic periodontitis traits after adjustment for smoking (never, former, current) and diabetic status (binary) among the white participants of the Dental ARIC study ........................................ 89

**Table 4.5** Distribution of genotype (based on number of minor allele copies of SNPs with $P<5\times10^{-6}$) by phenotypic classification (three disease traits and mean “extent” of disease) among the Dental ARIC participants ................................................ 90

**Table 4.6** Reported genes’ symbols [HUGO Gene Nomenclature Committee (HGNC) approved symbols], names and chromosomal locations ............................................... 91

**Table 4.7** Reported SNPs address according to the 1000 genomes (1000G) convention and proxy SNPs ($R^2>0.8$ where available) based on 1000G pilot 1 release. All SNPs were imputed based on HapMap II CEU with the exception of
rs2521634 with was directly genotyped on the Affymetrix 6.0 platform

Table 4.8 Reported SNPs as expression quantitative loci (eQTLs) and their association with gene expression (lymphoblastoid cell line-based expression) in populations of European descent with $P<10^{-4}$ in the Scan database (http://www.scandb.org) ........................................................................................................ 93

Table 5.1 Distribution of subjects’ demographic characteristics and bacterial “high colonization” (defined as the top quintile of the distribution) profile (with “Red”, “Orange” complex, and *Aggregatibacter actinomycetemcomitans*) overall, and stratified by periodontal diagnosis (CDC/AAP classification) among the Dental ARIC study participants ........................................................................................................... 117

Table 5.2 Genome-wide association analysis results of the high colonization traits (highest quintile of the distribution versus the other four; quantified with DNA-DNA “checkerboard” hybridization) for “Red”, “Orange” complex and *Aggregatibacter actinomycetemcomitans*, among the white participants of the Dental ARIC study (n=1020). Single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF-HapMap II CEU) of $\geq 5\%$ and associated $P<5x10^{-6}$. The SNP with the lowest $P$-value per locus is presented; additional prioritized SNPs in each locus are presented in the footnote, along with corresponding $R^2$ (based on 1000 genomes pilot 1 release) with the top SNPs................................................................................................................................ 118

Table 5.3 *Porphyromonas gingivalis* “high” colonization (quantified with DNA “checkerboard” hybridization) GWA results among the white participants of the Dental ARIC study (n=1020). Single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF-HapMap II CEU) of $\geq 5\%$ and $P<5x10^{-6}$. The SNP with the lowest $P$-value per locus is presented; additional prioritized SNPs in each locus are presented in the footnote, along with corresponding $R^2$ (based on 1000 genomes pilot 1 release) with the top SNPs................................................................................................................................ 136

Table 5.4 Reported SNPs as expression quantitative loci (eQTLs) and their association with gene expression (lymphoblastoid cell line-based expression) in populations of European descent with $P<10^{-4}$ in the Scan database (http://www.scandb.org) ........................................................................................................... 137

Table A1. Distribution in the total sample and bivariate associations of periodontal diagnosis (CDC classification) and periodontal attachment loss (extent score: percent of sites with $\geq 3$mm attachment loss) with brushing and flossing frequency, dental attendance reason and frequency, and education level among the Caucasian participants of the Dental ARIC study, for whom clinical and genetic data were matched after exclusions ................................................................................. 184

Table A2. Distribution in the total sample and bivariate associations of periodontal
diagnosis (CDC classification) and periodontal attachment loss (extent score: percent of sites with ≥3mm attachment loss) with quintiles of age and BMI among the Caucasian participants of the Dental ARIC study, for whom clinical and genetic data were matched after exclusions ................................................................. 185

**Table A3.** Distribution and bivariate associations of attachment loss (quintile-categorical and three-level classification of extent score) with periodontal diagnosis (CDC classification) among the Caucasian participants of the Dental ARIC study, for whom clinical and genetic data were matched after exclusions .............. 186

**Table A4.** Median (interquartile range, IQR) of sub-gingival microbial counts of eight periodontal pathogens obtained via DNA checkerboard immunoblotting among a subset of the Dental ARIC participants ............................................. 187

**Table A5.** Median (interquartile range, IQR) of sub-gingival microbial counts of eight periodontal pathogens obtained via DNA checkerboard immunoblotting among a subset of the Dental ARIC participants ................................................................. 188

**Table A6.** Median (interquartile range, IQR) values of sub-gingival microbial counts of the “red” and “orange” complex pathogens, and total microbial load obtained via DNA checkerboard immunoblotting, in the total sample, stratified by sex, smoking and diabetic status and their association with age and BMI among the Caucasian participants of the Dental ARIC study, for whom clinical and genetic data were matched after exclusions ................................................................. 189

**Table A7.** Median (interquartile range, IQR) values of sub-gingival microbial counts of the “red” and “orange” complex pathogens, and total microbial load obtained via DNA checkerboard immunoblotting, in the total sample, stratified by brushing and flossing frequency, dental attendance frequency and reason, and education level among the Caucasian participants of the Dental ARIC study, for whom clinical and genetic data were matched after exclusions ................................................................. 190

**Table A8.** Sample power calculation for SNP main effect among the ARIC cohort study whites. Assumptions are: MAF 0.1-0.5, estimated SNP effects of 1.1-1.6, and outcome prevalence 0.2 (severe periodontitis) .............................................................................................................................. 191

**Table A9.** Sample power calculation for SNP main effect among the ARIC cohort study whites. Assumptions are: MAF 0.1-0.5, estimated SNP effects of 1.2-1.6, and outcome prevalence 0.2 (highest quintile of bacterial colonization phenotypes) .............................................................................................................................. 192

**Table A10.** Sample power calculation for SNP*environment interaction among Caucasians. Assumptions are: MAF 0.1-0.5, estimated SNP effect of 1.4, environmental factor prevalence 0.1 (current heavy smoking), interaction effect 1.2-2.0 and outcome prevalence 0.2 (severe periodontitis) .............................................................................................................................. 193
Table A11. Sample power calculation for SNP*environment interaction among Caucasians. Assumptions are: MAF 0.1-0.5, estimated SNP effect of 1.4, environmental factor prevalence 0.5 (ever smoker), interaction effect 1.2-2.0 and outcome prevalence 0.2 (severe periodontitis) ............................................................... 194

Table A12. Sample power calculation for SNP main effect among African Americans. Assumptions are: MAF 0.1-0.5, estimated SNP effects of 1.3-1.7 and outcome prevalence 0.2 (severe periodontitis) ............................................................... 195

Table A13. Results of SNP interactions (interaction term P<10^{-5}) with sex, diabetes (binary) and smoking status (never, ever, current) in the GWA analysis of chronic periodontitis (“extent” of disease trait: proportion of sites exhibiting attachment loss ≥3 mm) among the white Dental ARIC participants ........................................ 196

Table A14. Genome-wide association analysis results of the CDC/AAP chronic periodontitis (CP) classification traits (severe CP vs. healthy and moderate CP vs. healthy) and extent of attachment loss trait (EAL-proportion of sites exhibiting attachment loss ≥3mm), among the white participants of the Dental ARIC study (n=4610). Supplemental list of single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF-HapMap II CEU) of ≥5% and that emerged below the P<5x10^{-5} in the “sensitivity” analysis that included logistic regression models adjusted for diabetes (binary variable) and smoking (never, former, current) ......................... 197

Table A15. Coded and non-coded alleles of prioritized SNPs for severe and moderate CP on the Affymetrix 6.0 (ARIC GWAS) and the Illumina Human 1M Duo (ver.3) (Health ABC GWAS) platforms ........................................................................................... 198

Table A16. Genome-wide association replication results [SNPs prioritized based on a P<5x10^{-6} threshold in the dental ARIC (n=4610)] of severe and moderate CP in the Health ABC cohort. Reported estimates correspond to the ARIC cohort ‘risk’ allele effect on severe and moderate CP and EAL in the Health ABC cohort ................................................................. 199

Table A17. Replication of GWAS results (P-values and effect estimates) of prioritized SNPs of the ARIC genome-wide association analysis (crude and exploratory smoking and diabetes-adjusted analyses) in the Health ABC cohort. Reported effect estimates correspond to the ‘risk’ allele in the ARIC cohort ...................... 200

Table A18. Meta-analysis of GWAS results (meta-analysis P-values and effect estimates based on inverse variance weighting) of prioritized SNPs for severe and moderate CP in the ARIC and the Health ABC studies. Reported effect estimates correspond to the ‘risk’ allele in the ARIC cohort. Pooled estimates based on inverse-variance meta-analysis are presented for results that did not demonstrate substantial heterogeneity (homogeneity X^2 P<0.2 criterion) ......................... 201
LIST OF FIGURES

Figure

**Figure 1.1** Classification of microbial periodontal pathogens according to Socransky (1998) ..................................................................................................................... 20

**Figure 4.1** Manhattan plot of the genome-wide association analysis results for severe chronic periodontitis among the 4610 white dental ARIC participants ...................... 73

**Figure 4.2** Manhattan plot of the genome-wide association analysis results for moderate chronic periodontitis among the 4610 white dental ARIC participants................... 74

**Figure 4.3** Manhattan plot of the genome-wide association analysis results for extent of attachment loss (proportion of sites exhibiting ≥3mm attachment loss) among the 4610 white dental ARIC participants .................................................................... 75

**Figure 4.4** Visualization of the 14q21 locus that was marked by rs12883458 (P=3.5x10^{-7}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) ............................................................................................................. 76

**Figure 4.5** Visualization of the 19p13.3 locus that was marked by rs3826782 (P=4.0x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) .............................................................................................................. 77

**Figure 4.6** Visualization of the 6p21.1 locus that was marked by rs7762544 (P=1.1x10^{-7}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) .............................................................................................................. 78

**Figure 4.7** Visualization of the 7p15 locus that was marked by rs2521634 (P=1.6x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium
values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) .............................................................................................................. 79

**Figure 4.8** Visualization of the 6q15 locus that was marked by rs17792917 (P=1.8x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) .............................................................................................................. 80

**Figure 4.9** Visualization of the 7q34 locus that was marked by rs10500130 (P=4.6x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) .............................................................................................................. 81

**Figure 4.10** Quantile-quantile plot of genome-wide association analysis results of severe chronic periodontitis among the 4610 white Dental ARIC participants .......... 94

**Figure 4.11** Quantile-quantile plot of genome-wide association analysis results of moderate chronic periodontitis among the 4610 white Dental ARIC participants.......... 95

**Figure 4.12** Quantile-quantile plot of genome-wide association analysis results of the extent of attachment loss trait (proportion of sites exhibiting ≥3 mm attachment loss) among the 4610 white Dental ARIC participants ........................................ 96

**Figure 4.13** Visualization of the 3p21 locus that was marked by rs11925054 (P=6.5x10^{-7}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) .............................................................................................................. 97

**Figure 4.14** Visualization of the 4q21.3 locus that was marked by rs17006135 (P=4.4x10^{-6} after adjustment for smoking and diabetic status) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) ......................... 98

**Figure 4.15** Visualization of the 1p21.1 locus that was marked by rs10493998
(P=2.2x10^{-6} after adjustment for smoking and diabetic status) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36). Figure 4.16 Visualization of the 10p15 locus that was marked by rs12260727 (P=6.0x10^{-7}) for moderate chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36). Figure 4.17 Visualization of the 12q15 locus that was marked by rs11615037 (P=4.4x10^{-6} after adjustment for smoking and diabetic status) for moderate chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36). Figure 4.18 Visualization of the 11q22 locus that was marked by rs7120142 (P=3.9x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36). Figure 4.19 Venn diagram representing genes adjacent to identified SNPs for the three chronic periodontitis traits. Genes with asterisk are associated with SNPs that emerged below the P<5x10^{-6} threshold in genetic models adjusted for smoking and diabetic status. Figure 5.1 Manhattan plot of the GWAS results (-log10 P-values of the ~2.5 million examined SNPs arranged by chromosome) for “high red” complex bacterial colonization among the 1020 white participants of the Dental ARIC Study cohort. Figure 5.2 Manhattan plot of the GWAS results (-log10 P-values of the ~2.5 million examined SNPs arranged by chromosome) for “high orange” complex bacterial colonization among the 1020 white participants of the Dental ARIC Study cohort. Figure 5.3 Manhattan plot of the GWAS results (-log10 P-values of the ~2.5 million examined SNPs arranged by chromosome) for “high”
Aggregatibacter actinomycetemcomitans bacterial colonization among the 1020 white participants of the Dental ARIC Study cohort

**Figure 5.4** Visualization of the 1p22 locus that was marked by rs12032672 (P=9.6x10^{-7}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36)

**Figure 5.5** Visualization of the 5q33 locus that was marked by rs10043775 (P=2.4x10^{-6}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36)

**Figure 5.6** Visualization of the 9p24 locus that was marked by rs16924631 (P=3.2x10^{-6}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36)

**Figure 5.7** Visualization of the 4p14 locus that was marked by rs10010758 (P=3.7x10^{-6}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36)

**Figure 5.8** Visualization of the 1p36.2 locus that was marked by rs1616122 (P=4.9x10^{-6}) for “high orange” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36)

**Figure 5.9** Visualization of the 1q42 locus that was marked by rs11800854 (P=4.0x10^{-6}) for “high” Aggregatibacter actinomycetemcomitans colonization among the subset of 1020 white participants of the Dental Atherosclerosis in
Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).

**Figure 5.10** Quantile-quantile plot of genome-wide association analysis results of “high red” complex bacterial colonization, among the 1020 Dental ARIC participants.

**Figure 5.11** Quantile-quantile plot of genome-wide association analysis results of “high orange” complex bacterial colonization, among the 1020 Dental ARIC participants.

**Figure 5.12** Quantile-quantile plot of genome-wide association analysis results of “high” *Aggregatibacter actinomycetemcomitans* colonization, among the 1020 Dental ARIC participants.

**Figure 5.13** Visualization of the 2p23.3 locus that was marked by rs1011108 ($P=2.0\times10^{-6}$) for “high” *Porphyromonas gingivalis* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).

**Figure 5.14** Visualization of the 9q33.2 locus that was marked by rs10760187 ($P=4.5\times10^{-7}$) for “high” *Porphyromonas gingivalis* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).

**Figure 5.15** Visualization of the 10p11.21 locus that was marked by rs1360573 ($P=1.5\times10^{-6}$) for “high” *Porphyromonas gingivalis* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).

**Figure A1.** Power calculation for severe CP among whites for MAF: 0.1-0.5.

**Figure A2.** Power calculation for the “high” bacterial colonization trait for MAF: 0.1-0.5 among the dental ARIC whites.

**Figure A3.** Power calculation for GxE interaction effects for severe CP, assuming
‘exposure’ prevalence 0.5 and MAF: 0.1-0.5 among the dental ARIC whites.......................... 204

**Figure A4.** Power calculation for GxE interaction effects for severe CP, assuming ‘exposure’ prevalence 0.5 and MAF: 0.1-0.5 among the dental ARIC whites.......................... 205

**Figure A5.** Density plot of average DST among the ARIC cohort study whites zoomed in the regions of DST cut-offs for related individuals (first degree DST ≥0.8, second degree 0.745 ≤ DST < 0.8). Red dots represent those 380 pairs of first degree and 207 pairs of second degree relatives ................................................................. 206

**Figure A6.** Density plot of average DST among the dental ARIC study whites. Red dots represent 16 genetic outliers based on average DST values............................. 207

**Figure A7** Scatter plot matrix of the ten principal components generated with the EIGENSTRAT program for the white ARIC cohort study genotyped participants...................................................................................................................... 208

**Figure A8** Catalogue of published GWA studies (n=951) through June 2011................. 209

**Figure A9** Visualization of the 14q21 locus that was marked by rs12883458 (P=3.5x10^{-7}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) ................................................................. 210

**Figure A10** Visualization of the 7p15 locus that was marked by rs2521634 (P=1.6x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) .............................................................................. 211

**Figure A11** Visualization of the 3p21 locus that was marked by rs11925054 (P=6.5x10^{-7}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) .............................................................................. 212

**Figure A12** Visualization of the 6p21.1 locus that was marked by rs7762544 (P=1.1x10^{-7}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) .............................................................................. 213

**Figure A13** Visualization of the 19p13.3 locus that was marked by rs3826782 (P=4.0x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a
linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) ................................................................. 214

**Figure A.14** Visualization of the 10p15 locus that was marked by rs12260727 (P=6.0x10^{-7}) for moderate chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) .............................................................................................. 215

**Figure A.15** Visualization of the 6q15 locus that was marked by rs17792917 (P=1.8x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) .............................................................................................. 216

**Figure A.16** Visualization of the 4q21.3 locus that was marked by rs17006135 (P=4.4x10^{-6} after adjustment for smoking and diabetic status) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) .............................................................................................. 217

**Figure A.17** Visualization of the 11q22 locus that was marked by rs7120142 (P=3.9x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) .............................................................................................. 218

**Figure A.18** Visualization of the 7q34 locus that was marked by rs10500130 (P=4.6x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) .............................................................................................. 219

**Figure A.19** Visualization of the 1q42 locus that was marked by rs11800854 (P=4.0x10^{-6}) for “high” *Aggregatibacter actinomycetemcomitans* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) .............................................................................................. 220

**Figure A.20** Visualization of the 1p22 locus that was marked by rs12032672 (P=9.6x10^{-7}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) .............................................................................................. 221

**Figure A.21** Visualization of the 5q33 locus that was marked by rs10043775
(P=2.4x10^{-6}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) ........................................................................................................ 222

**Figure A.22** Visualization of the 9p24 locus that was marked by rs16924631 (P=3.2x10^{-6}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) ........................................................................................................ 223

**Figure A.23** Visualization of the 4p14 locus that was marked by rs10010758 (P=3.7x10^{-6}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) ........................................................................................................ 224

**Figure A.24** Visualization of the 1p36.2 locus that was marked by rs1616122 (P=4.9x10^{-6}) for “high orange” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) ........................................................................................................ 225

**Figure A.25** Visualization of the 1q42 locus that was marked by rs11800854 (P=4.0x10^{-6}) for “high” *Aggregatibacter actinomycetemcomitans* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) ........................................................................................................ 226

**Figure A.26** Visualization of the 9q33.2 locus that was marked by rs10760187 (P=4.5x10^{-7}) for “high” *Porphyromonas gingivalis* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) ........................................................................................................ 227

**Figure A.27** Visualization of the 10p11.21 locus that was marked by rs1360573 (P=1.5x10^{-6}) for “high” *Porphyromonas gingivalis* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data) ........................................................................................................ 228

**Figure A.28** Predicted network of associations [physical interactions (grey line) and co-expression (blue line)], generated with the GeneMANIA module) among the genes adjacent to prioritized loci for the CP and the “high” bacterial traits in the Dental ARIC cohort ........................................................................................................ 229
**Figure A.29** Predicted network of associations [physical interactions (grey line), co-expression (dark blue line), pathway (light blue line), shared protein domain (green line)], generated with the GeneMANIA module) among the genes adjacent to prioritized loci for the CP and the “high” bacterial traits in the Dental ARIC cohort, along with genes that have been previously reported to be associated with periodontitis .......................... 230

**Figure A.30** Visualization of GWA results for severe CP in the ARIC cohort for loci that have been previously associated with periodontitis (IL1A, IL1RN, TLR4, MMP2, GLT6D1, GCGR3A). The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) ............................................................................... 231

**Figure A.31** Visualization of GWA results for moderate CP in the ARIC cohort for loci that have been previously associated with periodontitis (IL1A, IL1RN, TLR4, MMP2, GLT6D1, GCGR3A). The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) ............................................................................... 232

**Figure A.32** Visualization of GWA results for the extent of attachment loss trait (proportion of sites exhibiting ≥3mm attachment loss) CP in the ARIC cohort for loci that have been previously associated with periodontitis (IL1A, IL1RN, TLR4, MMP2, GLT6D1, FCGR3A). The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) ............................................................................... 233

**Figure A.33** Visualization of GWAS results in the IL6 (Interleukin 6) and VDR (vitamin D receptor) loci in the ARIC cohort for severe (left column) and moderate CP (right column). The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36) ............................................................................... 234

**Figure A.34** Visualization of the 6p21.1 locus (chr6:41,279,315-41,479,315; flanking 100Kb of rs7762544, which was prioritized for moderate CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters ................................................................. 235

**Figure A.35** Visualization of the 19p13.3 locus (chr19:6,787,736-6,987,736; flanking 100Kb of rs3826782, which was prioritized for moderate CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters ......................................................................... 236

**Figure A.36** Visualization of the 10p15 locus (chr10:10,238,329-10,438,329; flanking
100Kb of rs12260727, which was prioritized for moderate CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters........................................................................................... 237

**Figure A.37** Visualization of the 14q21 locus (chr14:51,179,379-51,379,379; flanking 100Kb of rs12883458, which was prioritized for severe CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters........................................................................................... 238

**Figure A.38** Visualization of the 7p15 locus (chr7:24,278,040-24,478,040; flanking 100Kb of rs2521634, which was prioritized for severe CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters........................................................................................... 239

**Figure A.39** Visualization of the 3p21 locus (chr3:55,290,886-55,490,886; flanking 100Kb of rs11925054, which was prioritized for severe CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters........................................................................................... 240

**Figure A.40** Visualization of the 6q15 locus (chr6:85,166,965-85,366,965; flanking 100Kb of rs17792917, which was prioritized for the “extent” of attachment loss trait) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters........................................................................................... 241

**Figure A.41** Visualization of the 11q24 locus (chr11:127,557,071-127,757,071; flanking 100Kb of rs10790919, which was prioritized for the “extent” of attachment loss trait) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters........................................................................................... 242

**Figure A.42** Visualization of the 11q22 locus (chr11:103,317,781-103,517,781; flanking 100Kb of rs7120142, which was prioritized for the “extent” of attachment loss trait) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters........................................................................................... 243

**Figure A.43** Visualization of the 7q34 locus (chr7:138,715,913-138,915,913; flanking 100Kb of rs10500130, which was prioritized for the “extent” of attachment loss trait) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters........................................................................................... 244
# LIST OF ABBREVIATIONS AND SYMBOLS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>Percent</td>
</tr>
<tr>
<td>≥</td>
<td>Greater than or equal to</td>
</tr>
<tr>
<td>≤</td>
<td>Less than or equal to</td>
</tr>
<tr>
<td>$k$</td>
<td>Kappa coefficient</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Lambda coefficient</td>
</tr>
<tr>
<td>$\mu$L</td>
<td>Micro liters</td>
</tr>
<tr>
<td>Aa</td>
<td>Aggregatibacter actinomycetemcomitans</td>
</tr>
<tr>
<td>AA</td>
<td>African-American</td>
</tr>
<tr>
<td>AAP</td>
<td>American Academy of Periodontology</td>
</tr>
<tr>
<td>AgP</td>
<td>Aggressive periodontitis</td>
</tr>
<tr>
<td>ARIC</td>
<td>Atherosclerosis Risk In Communities study</td>
</tr>
<tr>
<td>b</td>
<td>Beta coefficient</td>
</tr>
<tr>
<td>BGI</td>
<td>Biofilm-gingival interface</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>BOP</td>
<td>Bleeding upon probing</td>
</tr>
<tr>
<td>CAL</td>
<td>Clinical attachment loss</td>
</tr>
<tr>
<td>CARe</td>
<td>Candidate-gene Association Resource Consortium</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers for Disease Control and Prevention</td>
</tr>
<tr>
<td>CEU</td>
<td>Central European</td>
</tr>
<tr>
<td>CHD</td>
<td>Coronary heart disease</td>
</tr>
<tr>
<td>Chr</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>cM</td>
<td>Centimorgan (1/100&lt;sup&gt;th&lt;/sup&gt; of a Morgan)</td>
</tr>
<tr>
<td>CP</td>
<td>Chronic periodontitis</td>
</tr>
<tr>
<td>Cr</td>
<td><em>Campylobacter rectus</em></td>
</tr>
<tr>
<td>CVD</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>DM</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DST</td>
<td>Allelic sharing distance</td>
</tr>
<tr>
<td>DZ</td>
<td>Dizygotic</td>
</tr>
<tr>
<td>EAL</td>
<td>Extent of attachment loss</td>
</tr>
<tr>
<td>EMM</td>
<td>Effect measure modification</td>
</tr>
<tr>
<td>ESI</td>
<td>Extent and severity index</td>
</tr>
<tr>
<td>eQTL</td>
<td>Expression quantitative locus</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>Fn</td>
<td><em>Fusobacterium nucleatum</em></td>
</tr>
<tr>
<td>gAgP</td>
<td>Generalized aggressive periodontitis</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association study</td>
</tr>
<tr>
<td>HPFS</td>
<td>Health Professionals Follow-up Study</td>
</tr>
<tr>
<td>HWE</td>
<td>Hardy-Weinberg equilibrium</td>
</tr>
<tr>
<td>IBD</td>
<td>Identical by descent</td>
</tr>
<tr>
<td>IBS</td>
<td>Identical by state</td>
</tr>
<tr>
<td>IMT</td>
<td>intima-media thickness</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclass correlation coefficient</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase (1 million base pairs)</td>
</tr>
<tr>
<td>MMP</td>
<td>Matrix metalloproteinase</td>
</tr>
<tr>
<td>Kb</td>
<td>Kilobase (1 thousand base pairs)</td>
</tr>
<tr>
<td>LASSO</td>
<td>Least absolute shrinkage and selection operator</td>
</tr>
<tr>
<td>LD</td>
<td>Linkage disequilibrium</td>
</tr>
<tr>
<td>MAF</td>
<td>Minor allele frequency</td>
</tr>
<tr>
<td>mm</td>
<td>Millimeters</td>
</tr>
<tr>
<td>MZ</td>
<td>Monozygotic</td>
</tr>
<tr>
<td>N</td>
<td>Number</td>
</tr>
<tr>
<td>NHLBI</td>
<td>National Heart, Lung, and Blood Institute</td>
</tr>
<tr>
<td>NHANES</td>
<td>National Health and Nutrition Examination Survey</td>
</tr>
<tr>
<td>NIDCR</td>
<td>National Institute of Dental and Craniofacial Research</td>
</tr>
<tr>
<td>NIH</td>
<td>National Institutes of Health</td>
</tr>
<tr>
<td>OHRQoL</td>
<td>oral health-related quality of life</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>PC</td>
<td>Principal Component</td>
</tr>
<tr>
<td>PD</td>
<td>Probing depth</td>
</tr>
<tr>
<td>Pg</td>
<td><em>Porphyromonas gingivalis</em></td>
</tr>
<tr>
<td>Pi</td>
<td><em>Prevotella intermedia</em></td>
</tr>
<tr>
<td>Pn</td>
<td><em>Prevotella nigriscens</em></td>
</tr>
<tr>
<td>PR</td>
<td>Prevalence ratio</td>
</tr>
<tr>
<td>Q-Q</td>
<td>Quantile-quantile</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative trait locus</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>rho</td>
<td>Spearman’s rank correlation coefficient</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SE</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>SES</td>
<td>Socio-economic status</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>Td</td>
<td><em>Treponema denticola</em></td>
</tr>
<tr>
<td>Tf</td>
<td><em>Tannerella forsythia</em></td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor necrosis factor</td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
</tr>
<tr>
<td>VDR</td>
<td>Vitamin D receptor</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
CHAPTER 1

BACKGROUND

A. Introduction

Periodontitis, along with dental caries, is one of the two most common diseases of the oral cavity. It constitutes an inflammatory response to predominantly commensal oral bacteria. This relatively common chronic inflammatory disease is found in about 20% of the adult US population, is characterized by gingival pocket formation and clinical attachment loss, results in gradual destruction of periodontal tissues and tooth-supporting alveolar bone, and is considered the main cause of tooth loss among adults (1,2).

B. Definition and Pathophysiology of chronic periodontitis

1. Dental ecology - the host

The composition and complexity of the oral ecology has recently re-emerged as a focus of research, with the investigation of the oral microbiome’s composition and significance gaining increased attention (3). It is commonly assumed that an oral ecosystem in harmonious symbiosis with its host will likely be associated with health (4). Departures from this equilibrium due to pathogenic ecological shifts, up-regulated host inflammatory/immune responses, or both, are characteristics of disease (5). In periodontitis, highly organized oral microbial biofilms (dental plaque) interact with a plethora of host-specific factors.
(periodontal anatomy, oral hygiene, inflammatory response, and more) to result in tissue damage. Because tissue destruction appears to be a result of host responses to bacterial challenges rather than of the lytic effects of periodontal pathogens (6), the nature and extent of the microbiome-induced host response is considered an important determinant of the periodontitis phenotype. In fact, recent evidence indicates that autoimmune reactions may play role in the pathogenesis of the aggressive form of periodontitis (7).

2. Dental ecology- the oral microbiome

The “specific plaque” theory was articulated more than three decades ago, and emphasized the role of oral ecological shifts, essentially the bacterial insult in periodontitis (8, 9). In fact, Socransky summarized that in order for a pathogen to cause a disease the following conditions must be satisfied: 1) it must be a virulent type; 2) it must possess the chromosomal and extra-chromosomal genetic factors to initiate disease; 3) the host must be susceptible to the pathogen; 4) the pathogen must be in numbers sufficient to exceed the threshold for the host; 5) it must be located at the right place; 6) other bacterial species must foster, or at least not inhibit, the process; and 7) the local environment must be one which is conducive to the expression of the species' virulence properties (10).

The current state of knowledge with regard to the microbiology in periodontitis has not shifted from these core principles, as they were described in the 1970s and 1980s (11). The importance of the proliferating oral Gram-negative bacteria of the “orange” and the “red complex” is firmly established. Socransky in 1998 (12) defined these complexes as follows: “orange complex”: Prevotella intermedia (Pi), Campylobacter rectus (Cr), Fusobacterium
*nucleatum* (*Fn*) and *Prevotella nigriscens* (*Pn*). The “red complex” includes *Porphyromonas gingivalis* (*Pg*), *Tannerella forsythia* (*Tf*) and *Treponema denticola* (*Td*) (Figure 1).

It is commonly agreed that periodontitis manifests as a response to predominantly commensal microbes, but some authors have argued that *Pg* and *Aggregatibacter actinomycetemcomitans* (*Aa*) which are important periodontal pathogens (13) should be considered exogenous (14, 15) although this is not well established in the literature. Although it is known that the aforementioned bacteria are organized in colonies or biofilms on the tooth surfaces, the dental plaque, the question of whether innate factors may facilitate or prevent such an infection and colonization is intriguing.

Recent findings have added to our understanding of how periodontal pathogens harbor the oral cavity from early age, symbiose and get organized, and interact among themselves and with the host (16, 17). For example, it has been shown that *Fn* possesses the capacity to adhere to and invade oral epithelial cells (18). *Tf*, a “red complex” pathogen, has been shown to possess a glycosylated S-layer, which is responsible for evading the bacterium’s recognition by the host immune system (19).

Several lines of research have used the presence and colonization level of periodontal pathogens as a “refined” exposure of periodontitis (20, 21). Many investigations have used serum antibody levels (IgG) to periodontal pathogens as a measure of exposure and/or systemic response (17, 22-25). Fewer studies have used direct quantification methods to assess the counts of specific pathogens in the subgingival biofilms (microbial plaque) (21, 26, 27). Interestingly, in a study of the association of periodontal conditions with carotid intima-media thickness (IMT), Desvarieux et al (21) reported that total periodontal microbial
counts and specifically periodontal pathogens were associated with IMT, whereas periodontitis (as classified by CDC) was not. Beck et al (28) suggested that particularly “red complex” bacteria may pose a potential risk for the development of systemic disorders. For example, a recent report suggested that Pg and Aa may be associated with the risk for pro-thrombotic state (29), while Pg possesses properties known to be related to the pathogenesis of atherosclerosis (30). These investigations support the consideration of these specific microbial factors as a distinct exposure in investigations of periodontal, oral and systemic health, consistent with the paradigm of “periodontal medicine” (31, 32).

Interestingly, novel studies have provided insights on the induction by periodontopathogenic bacteria of epigenetic changes such as DNA hypermethylation (33). Epigenetics is one of the most rapidly expanding fields in biology, and the study of epigenetic mechanisms as well as the characterization of the human methylome alterations in health and disease, are priorities in biomedical research (34). Further research on the role of epigenetics in periodontal disease is needed.

3. Modifying factors

Despite important advances in our understanding of periodontal diseases the determinants of the inflammatory host response are not completely understood. While periodontal pathogens are necessary but not sufficient for disease activity to occur (10), several lines of research examine factors affecting or modulating the type and extent of inflammation evoked by the periodontal pathogens (35).

The fact that Diabetes Mellitus (DM) and severe periodontitis co-exist has been documented, and although a two-way relationship has been suggested (36, 37) causality has
not been demonstrated (38-40). While diabetic status modifies the host response to bacterial challenges and constitutes a risk factor for periodontitis, adequate glycemic control may be facilitated by controlling of an existing periodontal infection and inflammation (41). This has been demonstrated by a recent randomized controlled trial (42). It has been shown that although the periodontitis-DM relationship holds for both insulin-dependent and non-insulin dependent DM, it is more pronounced among patients with poorly controlled DM (43). Similar modifying effects have been shown for obesity (44).

Another example of host response-modification is thought to be the case of oral viral infections, which have been gaining attention with regard to CP (45-48). Noteworthy, Contreras et al. (49) showed that herpesviruses infections in combination with subgingival bacterial presence may exert periodontopathogenic potential by modifying the local host response.

Vitamin D has been reported as an important biological parameter with regard to the risk for periodontitis. Recently, Boggess et al. showed that the presence of moderate or severe periodontitis among pregnant women was associated with vitamin D insufficiency (50). Other lines of research indicate that innate antimicrobial factors such as lactoferrin, a salivary defense protein which is also secreted in gingival crevicular fluid during inflammation (51, 52), are important in the modulation of periodontitis (53).

Further elucidation of these biological interactions and pathway dissections can create new preventive and therapeutic opportunities for periodontitis. Moreover, mechanistic insights into the model of periodontitis may be key in understanding pathogenetic pathways relevant to other biologic conditions or chronic diseases, such as diabetes mellitus (DM) (37, 54), cardiovascular disease (CVD) (55, 56) and cancer (57-60).
C. Classification of periodontal diseases

The broad term of “periodontal disease” incorporates conditions other than periodontitis, such as gingivitis (61). The latter constitutes of an entirely reversible gingival inflammation (62) without periodontal destruction, and is typically diagnosed by bleeding upon periodontal probing (BOP). Although gingivitis had long been considered the entry-level stage of periodontitis and integral part of the disease continuum, this is not the case according to the current state of knowledge.

With regard to periodontitis, classification to two main types of chronic (CP) and aggressive periodontitis (AgP) was proposed in 1999 (61). The diagnosis between “aggressive” and “chronic” disease is only possible by knowledge of the rate of periodontal tissue destruction, or indirectly, by the assessment of tissue destruction level disproportional to a patient’s age. This type of sub-type classification may be problematic both in clinical practice (63) and surveillance or population-based studies. Although these two sub-types share several features and may appear close in pathogenetic features (64) distinct classification between AgP and CP is possible and feasible (65). For example, the consideration of an “aggressive” disease type, which likely corresponds to a more genetically penetrant or phenotypically expressed characteristic, may be worthwhile in genetic studies. On the other hand, it has been recognized that AgP cases may often represent manifestations of systemic conditions that interfere with resistance to bacterial infections (61). More recent evidence indicated the presence of auto-antibodies against some collagen types in patients with AgP but not CP (7). This phenomenon could pose a threat to validity due to disease misclassification in studies of periodontitis. At any rate, the classification of periodontitis by
sub-type has been a difficult exercise for clinicians and scientists over the last decades, and has been subject to numerous changes (65). Nonetheless, to aid in the design and interpretation of high-quality population-based studies of periodontitis, an accurate disease classification based on precise case definitions is warranted.

Chronic periodontitis (CP), previously known as “adult periodontitis” (61) is associated with irreversible periodontal tissue damage manifested by gradual periodontal attachment loss (Figure B). In order to establish the presence and quantify the extent and severity of periodontitis a number of clinical signs or measurements have been used: increased periodontal pocket depth (measured as probing depth - PD), loss of clinical attachment (CAL), alveolar bone loss (ABL) evidenced in radiographic examination, and bleeding upon probing (BOP) (65-68). Of those measurements, combined use of both CAL and PD has been recommended as the preferred approach in ascertaining the prevalence or progression of periodontitis in epidemiologic studies (66). Other attempts to quantify the extent and severity of the disease using the maximum amount of clinical information possible resulted in the introduction of the extent and severity scores, included in the Extent and Severity Index (ESI) by Carlos et al. in 1986 (69). Essentially, the ESI is represents the proportion of sites that exhibit disease expressed as a percent of pre-defined probing sites. Because extent and severity scores quantify the periodontal destruction as evidenced by the intraoral distribution of clinical measurements they can be regarded as an improvement over a categorical nosological model. However, these quantitative measures are also particularly vulnerable to bias due to tooth loss, even more so compared to more “crude” categorical disease models. For this reason, analytical approaches that consider extent scores typically include adjustments for tooth loss (70).
Because various researchers had not been using a universally accepted case definition of periodontitis, collaborative efforts in Europe and in the US were initiated to work towards agreement and convergence in the periodontitis case definition and disease progression standards. A European consensus was reached in the 5th European Workshop in Periodontology. This consensus statement was published in 2005 and recommended the use of a three-level (including health) disease case definition (71). Subsequently, in the US a joint workgroup that was initially commissioned in 2003 by the Centers for Disease Control and Prevention (CDC) and the American Academy of Periodontology (AAP) reviewed existing definitions and suggested another standard (66). Based on that work, an analogous three-level classification of periodontitis was proposed for use in population-based studies (Table 1.2). At the same time, Offenbacher (72) used a combination of clinical, microbial, inflammatory and host-response parameters to propose a five-level disease classification labeled BGI (Biofilm-Gingival Interface) that is defined by a combination of PD and BOP measurements. The BGI classification is considered a major improvement, because it reflects recent advances in our knowledge of the disease characteristics on both biological and clinical level taking into consideration the “biological systems model” (35).

For the purposes of the present investigation we will use the CDC definition and classification of CP. Based on the age structure of the studied population it is assumed that all periodontitis cases are representing chronic types and not AP. Continuous extent scores (e.g. percent of sites exhibiting ≥4mm CAL) will be considered for exploratory analyses.

The “CDC definition” of periodontitis uses the following taxonomy (Table 1.2):

1. **Severe periodontitis**: two or more interproximal sites not on the same tooth with ≥6 mm CAL and one or more interproximal sites with ≥5 mm PD;
2. **Initial** (moderate) *periodontitis*: two or more sites with 4 or 5mm CAL not on the same tooth;

3. **Health/Gingivitis**: individuals not meeting the above criteria.

D. **Epidemiology of periodontal diseases**

1. **Prevalence and methodological issues**

Most estimates of periodontitis indicate that advanced adult periodontitis does not exceed a 10-15% prevalence in most populations (1, 73). Estimates from the US are typically derived from the National Health and Nutrition Surveys. In the latest NHANES survey it was reported that 12.8% of the adult population had periodontitis, when the disease was defined as one or more sites with \( \text{CAL} \geq 5 \) mm. The prevalence was at a substantially higher 43.8% if the disease was defined as one or more sites with \( \text{CAL} \geq 3 \) mm, and this reflects the dramatic impact that classification criteria can have on the disease prevalence estimates (66). Moreover, the NHANES employed a partial-mouth examination protocol, essentially selecting a subset of teeth, and then sites within teeth to examine. The partial-mouth examination protocol, while providing gains in efficiency (maximum examination sites were 42 sites per individual in the 2001-2004 NHANES versus a maximum of 168 theoretically at risk sites per person) was found to substantially underestimate the disease prevalence by as much as 60% relative to the true prevalence, as determined by a full mouth examination (74). The impact of tooth loss on periodontitis estimates is also an important methodological consideration, and was outlined in a previous section. A more detailed investigation and quantification of the impact of tooth loss in population-based studies of periodontitis is warranted.
2. Disease burden

Although the initially reported estimate of approximately 20% prevalence of CP in
the adult population in the US is considerable, this figure is likely higher for the
methodological reasons outlined above. Moreover, most countries in the world do not have
adequate surveillance systems in place to monitor CP (1, 2). Another consideration is that
while care-seeking individuals will be diagnosed and subsequently treated for the disease,
others with sporadic pattern of or no dental attendance will not receive treatment and may
progress to substantial tooth loss at early ages. This phenomenon will increase estimates of
edentulism. Moreover, the increasing potential for dentition retention in older ages, combined
with an ageing population is expected to contribute to a continuously increasing prevalence
of CP in the next decades unless large scale awareness campaigns or interventions are
effectively implemented in communities. This population approach is warranted because
significant disparities in the prevalence of periodontitis exist between age and racial groups,
with individuals of African American and Hispanic ethnicity being disproportionately affected
(75).

A rapidly increasing number of reports associate CP with systemic conditions
(cardiovascular disease, pregnancy outcomes, diabetes control, oxidative stress, etc), with the
microbial load and the resultant inflammation being the likely mediators of these
relationships (76-78). Although definitive evidence that links the treatment of CP with
improved systemic health outcomes is limited, there are several plausible hypotheses and
numerous research reports pointing towards this direction (79).
3. Impact on the quality of life

The impact of periodontitis on quality of life has been well-documented (80-83). Periodontitis is commonly associated with increased impacts on daily activities, even after adjustment for socio-demographic and other clinical conditions, such as caries, tooth loss and prostheses (81). Noteworthy, patients under treatment maintenance were found to report less daily impacts on their oral health-related quality of life (OHRQoL) compared to newly diagnosed patients with CP (83). Of course tooth loss and edentulism, representing the end-stage of CP are disabling conditions that are associated with severe and multi-level impacts (84, 85). Specifically in the US, reports based on NHANES data indicate that oral disease disproportionately impacts disadvantaged groups (86). Another recent report based on NHANES data concluded that periodontitis, and particularly the level of serum antibodies against Pg, was associated with cognitive impairment among older US adults (22).

E. Risk factors for chronic periodontitis

Although a plethora of risk factors has been reported to be associated with periodontitis, the importance of each factor remains unknown. In fact person-level factors, such as age and race have been shown to be important determinants of periodontal inflammation (35, 87). Oral hygiene has been the longest considered “traditional” risk factor for periodontitis (1, 88). In fact, while gingivitis parallels the level of oral hygiene on a population-level, this is not the case with CP (75).

Smoking has been found to be the major risk factor for both periodontitis and tooth loss (89-91). This is consistent with the well-documented biological effects of smoking on periodontal health (92-94). Noteworthy, a substantial proportion of over 40% of periodontitis
cases in the US are attributable to smoking based on NHANES III estimates (89); a finding with great public health relevance. Prospective studies have also confirmed the causative role of smoking in the progression of periodontitis, as evidenced by CAL (95). Evidence also exists that smoking modifies the sub-gingival microbial colonization pattern (96) with a mechanism that is consistent with the observation that among smokers there is “higher infectious burden with periodontal pathogens but less inflammation” (97).

With regard to DM, numerous studies have shown that diabetics have a greater prevalence and extent of periodontal pockets (37, 41, 98, 99). Interestingly, it has been reported that inflammatory markers that are evident in periodontitis, were associated with incident type 2 DM (100). These findings emphasize the links between oral and systemic conditions and the need for a comprehensive consideration of such co-morbidities in the diagnosis and treatment of both conditions.

To date, there is no consensus in the literature regarding the role of alcohol consumption as a risk factor for CP (101). A report from the Health Professionals Follow-up Study (HPSF) suggested a weak association (in the range of RR=1.2-1.3) but relied upon self-reports of periodontitis (102). Another investigation suggested that increased periodontitis risk may be conferred by alcohol consumption among individuals with certain polymorphisms of the ALDH2 gene (103). Despite the documented multi-level detrimental effects of alcohol, it is well established that alcohol also exhibits antimicrobial effects that can confer protection from the microbial-plaque diseases. Evidence of such a protective effect has been found in both in vitro (104) and clinical (105) investigations.

Although some evidence suggests that dietary factors such as whole-grain and fiber intake may be associated with decreased risk of periodontitis (106) and nutrients such as
calcium (107) with decreased prevalence of the disease, most evidence indicates that impaired dental condition may be the cause of sub-optimal diet diversity and quality (108, 109).

Stress and other psychosocial factors such as coping skills and personality type have also been reported as risk factors for development or progression of periodontitis (110, 111). For example, in a study by Moss et al. (112), stress and depression traits were important risk factors for being a periodontitis case and manifesting disease progression, respectively.

F. The genetic basis of chronic periodontitis

1. Early reports of familial segregation

The genetic component of periodontitis was supported by early reports in the dental literature. These reports observed a familial aggregation of severe forms of the disease (113-117). Other reports, such as the one by Chung et al. (118) who carried out a path analysis using data from 241 nuclear families, did not detect any important evidence of heritability. These authors instead suggested that common family environment may be a major determinant of periodontal health. Michalowicz (119) later suggested that, although other periodontal diseases such as gingivitis have a substantial behavioral etiologic component, most of the observed heritability in periodontitis is attributable to genetics. The realization of the fact that “genes do not function in a vacuum” and that it is unlikely that a “master gene” for periodontitis will be discovered (120) underlines the necessity to investigate periodontitis risk loci within the full spectrum of common genetic variation.
2. **Twin studies**

An early report of a periodontitis study among twins was published by Ciancio (121) in 1969; two more reports were published by Michalowicz in the early 1990s (122, 123). In 1993 Corey et al. (124), published their investigation of self-reported periodontitis among twins. For their study they used a mailed survey among the members of a Virginia-based twin registry, and they reported concordance rates of 0.38 and 0.16 between monozygotic (MZ) and dizygotic (DZ) twins, respectively. The authors’ conclusion was that, based on their findings, genetic factors make an important contribution to the risk of adult-onset periodontitis. Subsequently, Michalowicz et al. in 2000 (119) improved the current state of knowledge by employing clinical examinations to diagnose periodontitis among 117 MZ and DZ pairs of adult twins. In interpreting their findings the authors estimated that periodontitis has approximately 50% heritability and thus concluded “approximately half of the variance in disease in the population is attributed to genetic variance”. A more recent but small investigation that was carried out among 10 pairs of MZ and 8 pairs of DZ twins (125) found a substantial amount of discordance in disease severity among the twins, with that discordance being greater among DZ pairs. The authors suggested that previous estimates of heritability may be exaggerated (125, 126).

3. **Candidate-gene studies**

A large number of publications have reported on the association of several gene polymorphisms with periodontitis. In these approaches the studied risk variants are thought to be associated with altered immune response to the bacterial insult and include, among other genes, the interleukin-1 (IL-1) (127), tumor necrosis factors (TNF) (128), matrix
metalloproteinases (129), estrogen (130) and vitamin D receptor (50). A summary of these findings is illustrated in Table 1.2. There have been very few candidate-gene investigations of CP among African Americans (131).

Previous candidate gene studies have found evidence for an association between periodontitis risk and interleukin (IL) gene polymorphisms (particularly IL-1), Fcγ, TNF, and matrix metalloproteinsases (128, 131, 132). A recent study reported an association between periodontitis and variants in FAM5C (133). Other candidate-gene studies did not detect any strong associations between AgP or CP and selected polymorphisms, and suggested the use of haplotype or genome-wide analyses as potentially more fruitful strategies (134). Moreover, the largest and best-powered candidate-gene study that was carried out in a Caucasian population (135) did not detect any important association between the IL-1 cluster and AgP risk. In spite of the high risk of bias from very small studies in the field, most reviews suggest that the balance of published evidence favors a causative role of IL gene cluster polymorphisms with periodontitis (127, 136). Two such meta-analyses summarized the evidence of increased risk for CP conferred by IL-1 (137) and Fcγ receptor polymorphisms (138). Nikolopoulos et al. summarized 53 published studies and concluded that there was a statistically significant association of IL-1A and IL-1B polymorphisms with CP risk. Dimou et al. suggested that there is accumulating evidence and supporting biological plausibility linking the FcγRIIIB NA1/NA2 polymorphism with increased risk for both AgP and CP (138). Finally, one recent investigation reported an association of IL-6 polymorphism with colonization with Aa but not Pg, among Caucasian CP patients (139).
4. Genome-wide association evidence

The opportunity to use whole-genome scans for the exploration of the association with various phenotypes was early recognized, in the era of family-based studies (140). Because current knowledge, technologies, and methods employed in GWAS allow a satisfactory coverage of the common genetic variation (at least among white populations), this approach has gained great popularity (141, 142) for the study of common complex diseases. Although the initial impact of GWAS has been less dramatic than initially postulated (141, 143), they have provided an unprecedented amount of new information regarding the susceptibility for and the pathogenesis of many diseases.

Schaefer et al. (144) were the first to report a genome-wide association hit with periodontal disease. Specifically, these investigators also reported the existence of a shared risk susceptibility locus on 9p21.3 (rs1333048), for both AgP and coronary heart disease (145). This region was mapped to the sequence of the large antisense noncoding RNA ANRIL, which partly overlaps regulatory and coding sequences of CDKN2A/CDKN2B. This association was recently replicated in an independent case-control sample with 130 cases of AgP and 339 controls (146), where the authors suggested that the most plausible genetic model underlying the association between the identified SNPs and AgP is the multiplicative one. Further, the same group of investigators in another publication reported that an intronic SNP rs1537415 located in the glycosyltransferase gene GLT6D1 was associated with AgP. In that study, the rare G allele showed 10% enrichment in cases (total ~280 cases). The authors suggested that the rare allele was associated with reduction of the binding affinity of the zinc-finder transcription factor GATA-3, which could be important in the pathogenesis of AgP.
While these GWA investigations provide the first insights into the genome-wide evidence of risk loci for periodontitis, their small sample sizes limited their statistical power. Moreover, both studies examined cases of AgP which represents the most aggressive form of periodontitis, and is found in about 3% or less among various populations (147, 148).

A recent review of published articles up to April 2009 on genetic polymorphisms associated with CP was published by Laine in 2010 (131). The authors noted that for their search strategy they used the keywords: Periodontitis, Periodontal disease in combination with the words: Genes, Mutation, or Polymorphism to identify articles written in the English language, employing a case-control design to study CP or adult periodontitis and reporting genotype distribution. The authors however did not provide an exact search string that could be replicated, information about which databases were searched and with what limits, the exact number of abstract and full-text articles identified from their searches, additional information for seeking additional articles, exclusions including removing duplicates, etc.
### G. Tables

**Table 1.1** Clinical case definitions of Periodontitis proposed by the CDC working group for use in population-based surveillance of Periodontitis\(^1\) (2)

<table>
<thead>
<tr>
<th>Disease Category</th>
<th>Clinical Attachment Level (CAL)</th>
<th>Probing Depth (PD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Severe periodontitis</strong></td>
<td>( \geq 2 ) interproximal sites(^2) with CAL ( \geq 6\text{mm} )</td>
<td>and ( \geq 1 ) interproximal site with PD ( \geq 5\text{mm} )</td>
</tr>
<tr>
<td><strong>Moderate periodontitis</strong></td>
<td>( \geq 2 ) interproximal sites(^2) with CAL ( \geq 4\text{mm} )</td>
<td>or ( \geq 2 ) interproximal sites(^2) with PD ( \geq 5\text{mm} )</td>
</tr>
<tr>
<td><strong>No or mild periodontitis</strong></td>
<td>Neither “moderate” nor “severe” periodontitis</td>
<td></td>
</tr>
</tbody>
</table>

1: third molars are excluded

2: not on the same tooth
Table 1.2. Overview of published genetic effects from candidate-gene case control studies of the risk of chronic periodontitis among Caucasians.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polymorphism</th>
<th>Race</th>
<th>Association with CP</th>
<th>No association with CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL1A</td>
<td>IL1A -889 (+4845) C&gt;T</td>
<td>Caucasian</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>IL1B +3954 (+3953) C&gt;T</td>
<td>Caucasian</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>IL1RN</td>
<td>IL1RN VNTR (+2018) C&gt;T</td>
<td>Caucasian</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>TFNA (-367) G&gt;A</td>
<td>Caucasian</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>IL6</td>
<td>IL6 (-572) C&gt;G</td>
<td>Caucasian</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>IL6 (-1363) G&gt;T</td>
<td>Caucasian</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IL10</td>
<td>IL10 -819 (-824) C&gt;T</td>
<td>Mixed (~80% Caucasians)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>IL10 -592 (-597) C&gt;A</td>
<td>Mixed &amp; Caucasians</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>FcγRIIα</td>
<td>IIα 131 H&gt;R</td>
<td>Caucasian</td>
<td>1 (among smokers only)</td>
<td>4</td>
</tr>
<tr>
<td>FcγRIIIb</td>
<td>IIIb NA1&gt;NA2</td>
<td>Japanese</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>VDR</td>
<td>VDR Taq1 T&gt;C</td>
<td>Caucasian</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CD14</td>
<td>CD14 -260 C&gt;T</td>
<td>Caucasian</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>TLR4</td>
<td>TLR4 Thr399Ile</td>
<td>Japanese</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>
Figures

**Figure 1.** Classification of microbial periodontal pathogens according to Socransky 1998 (12)

- **“Orange complex”**
  - *Prevotella intermedia*
  - *Campylobacter rectus*
  - *Fusobacterium nucleatum*
  - *Prevotella nigrescens*

- **“Red complex”**
  - *Porphyromonas gingivalis*
  - *Tannerella forsythia*
  - *Treponema denticola*

- *Aggregatibacter actinomycetemcomitans*
H. References


100. Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. JAMA 2001;286:327-34.


A. Rationale - the case for a GWAS of chronic periodontitis

Exploring and unraveling the genetic basis of complex diseases has promise in aiding therapy, prevention and care for such diseases, and can ultimately lead to improved health outcomes (1, 2). The opportunity to use whole-genome scans for the exploration of the association with various phenotypes was early recognized, in the era of family-based studies (3). Because current knowledge, technologies, and methods employed in GWAS allow a satisfactory coverage of the common genetic variation among whites (4), this approach has gained great popularity for the study of common complex diseases (5, 6). Although the impact of GWAS has been less dramatic than initially postulated (5, 7), they have generated an unprecedented amount of new information regarding the susceptibility for and the pathogenesis of many common complex diseases. Notably, a large number of nucleotide polymorphisms (SNPs) were found to be associated with a total of 165 traits and were reported in 951 published GWA studies through June 2011 (8).

In the field of dentistry, thus far there have been limited applications of GWAS. Three small GWA investigations have reported genome-wide associations of loci with aggressive periodontitis (9-11) and two have reported on suggestive risk loci for dental caries (12, 13). These investigations were hampered by small sample size and other methodological issues such as clinical examination procedures, disease definition, and more. We propose to
add on to the existing knowledge basis and improve on these previous investigations by carrying out a GWA analysis of chronic periodontitis in the context of a well-defined cohort, using a moderately-sized sample of approximately 4600 white individuals for whom detailed socio-demographic, anthropometric, periodontal, medical, and behavior data have been collected (Specific Aim 1). As a supplement to this aim, we will conduct exploratory analyses of genome-wide [gene] x [environment] interactions, by considering SNP interactions with sex, smoking, and diabetes mellitus.

With regard to microbial factors, the pathogenetic role of specific bacterial strains in periodontitis is well-established and has been characterized in detail. While there is a great diversity in the periodontal microbiome, few microorganisms have been directly implicated in the pathogenesis of periodontitis (14). It is well established that colonization with high amounts of “red complex” bacteria (Pi, Pg, Tf), as well as Aa, is more common in severe or aggressive forms of periodontitis. Pi causes periodontal tissue destruction by triggering the host immune response, inducing Prostaglandin E2 and increasing the expression of matrix metalloproteinases (15). Pg and Tf have been shown to be more common among cases with “refractory” periodontitis versus “treatment-responsive” or periodontally-healthy individuals (16). Tf, in particular, was recently shown to possess an external S-layer that likely is responsible for the attenuated host response to this pathogen (17). Moreover, periodontal pathogens have been associated with systemic morbidities such as CHD, and they have been used as a “refined” exposure in studies of oral-systemic health links over clinical classifications of periodontitis (18, 19). Thus, preventing or controlling oral ecological shifts towards pathogenic biofilms (proliferation of the “orange” or “red” complex) is critical. Moreover, it is intriguing to determine why some individuals harbor more or more
pathogenic periodontal bacteria, and to this end detect whether innate host factors are important. In this domain, one report of a positive association between an IL-6 polymorphism and harboring of Aa was based on a small sample of forty Caucasian patients (20). We will address this question by exploring whether there are genome-wide loci that are associated with increased susceptibility for colonization with CP-risk associated bacteria of the “red” and the “orange” complex, as well as Aa (Specific Aim 2). As an additional exploratory aim, due to the major role of Pg in the pathogenesis of CP (21), we will investigate risk loci for Pg colonization.

B. SPECIFIC AIMS

By performing the proposed genome-wide association analysis among the white ARIC study participants we aim:

Specific Aim 1: To identify susceptibility loci for chronic periodontitis among a moderate-sized sample of community-dwelling white adults using the CDC disease classification and a continuous measure of disease severity.

As an addition to Specific Aim 1, we will explore for effect measure modification or gene x environment interactions of SNPs with sex, smoking and diabetes mellitus.

Specific Aim 2: To identify susceptibility loci for colonization with periodontal microorganisms of the “orange” and “red” complex, and Aggregatibacter actinomycetemcominatis.
As an additional exploratory step for Specific Aim 2 we will examine for risk loci for another individual pathogen, *Porphyromonas gingivalis* which is member of the “red” complex.
REFERENCES


CHAPTER 3

RESEARCH METHODS

A. The Atherosclerosis Risk in Communities (ARIC) study

1. The ARIC study

The ARIC is a National Heart, Lung, and Blood Institute (NHLB)-funded prospective epidemiologic study conducted in four U.S. communities (Jackson, Mississippi; Washington County, Maryland; suburban Minneapolis, Minneapolis; Forsyth County, North Carolina). The study was IRB-approved for every participating site, and all participants had given informed consent. ARIC was designed to investigate the etiology and natural history of atherosclerosis, the etiology of clinical atherosclerotic diseases, and variation in cardiovascular risk factors, medical care and disease by race, gender, location, and date (1). Within the context of ARIC, an NIDCR-funded ancillary dental study was carried out, the Dental ARIC. The Dental ARIC aims were to determine the prevalence, extent, and severity of periodontal conditions in the dentate ARIC population, and to describe the associations between those conditions and prevalent coronary heart disease, coronary artery IMT, presence of carotid artery lesions, and atherosclerosis risk factors (2).
2. Study population

The ARIC Cohort Component began in 1987 and recruited a total of 15,792 community-dwelling participants aged 45-64 sampled from a defined population in their communities (1, 3). These participants were reexamined every three years with the first screen (baseline) occurring in 1987-89, the second in 1990-92, the third in 1993-95, and the fourth and last exam was in 1996-98. The Dental ARIC ancillary study took place during the fourth visit, between 1996 and 1998 among dentate ARIC subjects.

From the initial recruitment of 15,792 participants 11,656 were seen in visit 4. Of those, 6797 underwent the periodontal examination. After exclusions for various reasons a final sample of 6017 subjects formed the Dental ARIC sample (2). There participants had mean age of 62 years (range 52-75), 21% had severe periodontitis when using the study protocol criteria ($\geq 30\%$ of sites with AL $\geq 3\text{mm}$), 13% were diabetics and 44% were males.

3. Dental data

The Dental ARIC included a clinical oral examination, collection of gingival crevicular fluid, oral microbial plaque, and serum, as well as in-person interviews. Clinical measurements of PD and CAL were made on six sites on all teeth, and the number of missing teeth was recorded. For these measurements, very good accuracy and reproducibility is to be expected with trained and calibrated examiners (4-6), as was the case in ARIC. Studies have estimates that agreement between examiners for PD and CAL are within 1mm in 90% of the measurements taken (4, 7). In the Dental ARIC weighted kappa statistics ranged between 0.76-0.86 indicating excellent agreement with a standard examiner and intraclass correlation
coefficients (ICC) ranged between 0.76-0.90 indicating excellent to outstanding agreement (2).

Using the PD and CAL measurements and based the CDC/AAP criteria (Table 1.1), subjects were classified as having no disease, moderate or severe CP. Additionally, a continuous measure of disease severity or “extent” based on Carlos et al. (8) was defined as the proportion of sites exhibiting CAL ≥3mm. The distribution characteristics of these variables in the analytical sample overall and across strata of covariates are presented in Tables 3-7. Additional dental (time of and reason for last dental visit, tooth brushing frequency, oral hygiene practices) and anthropometric data (body mass index) were used for exploratory descriptive and bivariate analyses, and are presented in the Appendix. Details about the analytical strategy followed to analyze the CDC and the “extent” traits are presented as part of the analytical approach for Specific Aim 1.

4. Microbiological data

Microbiological data are available for a subset of approximately 1200 white participants (9). Samples of subgingival microbial plaque were obtained during the clinical examination. Subsequently, the levels of eight periodontal pathogens were determined by DNA-DNA checkerboard hybridization analysis, a technique initially described by Gunaratnam et al. (10) and Socransky et al. (11). The selection of these eight microbes was based on their reported and plausible implication in the pathogenesis of periodontitis (10-12). The method has a lowest detection threshold of 10³-10⁴ microbial counts, and has been shown to have good detection properties (10). Therefore, subgingival microbial counts were obtained for organisms of the “red” complex: *Pg*, *Tf*, and *Td*, the “orange” complex: *Pi*, *Pn*, *Cr*, *Fn*, as
well as Aa. Additional information about the analytical strategy with regard to the microbiological data is presented as part of the approach for Specific Aim 2.

5. Covariates

The ARIC study investigators have collected a comprehensive list of socio-demographic, behavior, anthropometric and biological measurements (clinical, laboratory, and other specialized tests). An extensive list of variables collected can be found on the ARIC website (http://www.cscc.unc.edu/aric/) and in previous publications (1). For the purposes of the GWA analyses we used the following ARIC study variables: examination center (modeled with two indicator variables), age (measured in years and modeled as a continuous variable) and sex (male or female). For exploratory “sensitivity” analyses we also used smoking (modeled as a three-level ordinal categorical variable were 0: never, 1: former, and 2: current smoker), and diabetic status (modeled as a dichotomous variable where 0: healthy and 1: fasting glucose levels of $\geq 126$ mg/dL, non-fasting of $\geq 200$ mg/dL, or pharmacological treatment for diabetes). To correct for population stratification (admixture) we included in all analyses ten principal components derived by the EIGENSTRAT method (see section 3.B.2).

B. Genotyping, quality control and exclusions

1. Genotyping and imputation

In the study population, DNA was extracted from blood samples drawn from an antecubital vein into tubes containing serum separator gel. Blood samples were analyzed at a central ARIC laboratory in Houston, TX. Genotyping was performed with the Affymetrix Genome-Wide Human SNP Array 6.0 chip. The platform offers 906,600 markers for SNPs.
The rigorous quality control procedures included initial blind duplicate genotyping and identification/flagging of SNPs with kappa<0.95 and reconciliation of unintentional duplicate samples (17 duplicates and one triplicate). Imputation to 2.5million markers was performed using 669,450 SNPs and MACH v1.0.16 (http://www.sph.umich.edu/csg/abecasis/MaCH/index.html), based on HapMap Phase II CEU build 36. The SNPs used for imputation were selected from 839,048 autosomal SNPs restricted to those with minor allele frequency (MAF)>0.01 (129,543 excluded), Hardy-Weinberg equilibrium (HWE) P>10^{-5} (12,432 excluded) and call rate >95% (1,693 excluded). We used the following SNP exclusion criteria for further analyses: quality score <0.8 and missing data rate >10% after imputation, and MAF of <5%.

2. Population stratification

To obtain estimates of relatedness and population stratification a subset of 85,947 “high quality” linkage-disequilibrium (LD)-pruned SNPs was selected. These SNPs met the following criteria among self-reported whites: MAF ≥0.1, call rate >99.5%, HWE P ≥10^{-3}, autosomal, with annotation in the platform annotation file, not labeled “AFFX” or “chromosome 0”, and not monomorphic. Using these SNPs identity-by-state (IBS) allele sharing distance (DST values) were computed using PLINK, as such: DST = IBS distance (IBS_2 + 0.5*IBS_1) / (n SNP pairs). First degree relative status was assigned to pairs of individuals with DST≥0.8 and second degree relatives were considered those with 0.763 ≤DST< 0.8. Among the white ARIC participants there were 380 pairs of first degree and 207 pairs of second degree relatives identified. To minimize exclusions, related pairs were broken by iterative selection of individuals with most relatives using a custom program.
Population stratification was further evaluated with principal component (PC) analysis using the EIGENSTRAT approach (13) and the EIGENSOFT program (14). The above chosen set of LD-pruned SNPs was used for the computation of ten principal components. Genetic outliers were considered those that were further than 8 standard deviations (SD) away from any of the ten PCs over ten runs of PC computation. Based on DST and PC criteria there were 716 subjects flagged from removal from the analysis (206 as genetic outliers based on PCs and 16 based on average DST values (“too little IBS sharing” with the rest of the sample), 351 first degree relatives and 143 second degree relatives. All but ten second degree relatives (whose relatives were excluded as genetic outliers) were re-entered in the dataset and were assigned PCs. After exclusion of 364 individuals (4%) there were 9349 whites who were included in the GWA analysis and of those, 4610 had periodontal phenotype data available as Dental ARIC participants.

C. Analytical strategy

Goldstein (15), Manolio (16), McCarthy (17) and Hirschorn (18) have summarized the key challenges in the analysis and interpretation of genome-wide data. We acknowledge the inherent limitations of GWAS, such as the low power to detect very small effects, and the consideration of common genetic variance for the detection of “single-polymorphism effects”. The problem of the “little variance explained” by the discovered SNPs for most diseases has led investigators to look for the “dark matter” of genetic effects in the so-called rare (<5% or <1%) variants (19); the study of the latter however, may be methodologically challenging unless whole-genome typing becomes more accessible, or when multi-stage (20) or novel, more powerful statistical methods (21) are implemented. However, GWAS are a
powerful discovery tool that has the potential to unveil previously unknown genetic risk loci and provide insights on novel mechanisms and pathways. Acknowledging the strengths and limitations of the GWA approach which simultaneously interrogates millions of SNPs across the genome, we embarked on this investigation with a “risk locus discovery” rather than an “effect estimation” approach.

1. **Analytical sample**

Primary analyses were conducted among the Dental ARIC white subjects with imputed data (‘freeze 3’ version of genomic dataset). After exclusions specified above (364 whites) and merging of the genetic and clinical datasets, our final analytical sample included 4610 white individuals for whom clinical and genomic data were successfully matched. The analytical sample description for Specific Aim 1 is presented in Tables 3-7, and for Aim 2 in Tables 4-11. The distribution of the phenotype classification (CP status: CDC definition and extent of attachment loss: percent of sites with attachment loss of 3mm or greater) by the study covariates is also presented in Table 3. Participants’ mean age is 62 years (range 52-75), 13% are diabetics and 44% are males. When Beck et al. used the CDC classification of periodontitis in a subset of ~5000 Dental ARIC participants, 42% were periodontally healthy, 41% had initial periodontitis and 17% had ‘severe’ periodontitis (22). In our analytical sample, these proportions are virtually identical, 43 and 17% respectively.

2. **Analytical strategy for Specific aim 1**

The primary phenotype of interest to address our Specific aim 1 was chronic periodontitis as defined by CDC, a three-level categorical classification. The secondary phenotype was the
“extent” of attachment loss, a continuous measure that is expressed as the proportion of site with attachment loss of 3mm or greater. The following traits were defined and considered for the GWA analyses:

1. Binary outcome: severe periodontitis (coded as 1) versus initial periodontitis or health (coded as 0). Statistical analysis will be based on a logistic regression model assuming multiplicative (log-additive) allelic effects.

2. Binary outcome: moderate periodontitis (coded as 1) versus initial periodontitis or health (coded as 0). As above, statistical analysis will be based on a logistic regression model assuming multiplicative (log-additive) allelic effects.

3. Continuous outcome: “extent” or “severity” of periodontitis (proportion of sites with CAL≥3mm). An appropriate transformation of this measure (z-score, normal curve equivalent transformation of the “crude” proportion of sites) was performed prior to conducting the GWAS. Statistical analysis was based on a linear regression model assuming multiplicative (log-additive) allelic effects.

“Minimally” adjusted models

As noted in the general methods section, all models were adjusted for age, sex and population substructure/stratification. Adjustment for population substructure in the ARIC study has been based on principal component analysis. Although authors have argued that population stratification may not be a serious threat to the validity of GWAS results (23), adjustment for population ancestry has become standard practice (13). Therefore, the “minimal” genetic models include 10 principal components obtained by the statistical package EIGENSOFT (13, 14), as well as for sex and age. Additional terms adjusting for
examination center (two indicator variables) were entered in all models. For implementation in the context of the GWAS we used the probABEL package, which is part of the geneABEL suite (24).

Interpretation of analysis results relied on the evidence that the data offer against the null hypothesis of no SNP (allelic) association with the traits of interest, and therefore was based on the associated P-values. When multiple SNPs emerged below the P-value threshold for prioritization (see following section on hypothesis testing), we presented the “top SNP” per locus, which was the one with the lowest P-value. The additional prioritized SNPs in the same locus were also presented, along with linkage disequilibrium metrics ($R^2$ with the “top SNP” in the locus) obtained with the SNAP application (25). Moreover, effect estimates (odds ratios for the disease classification and betas for the “extent” trait), as well as ‘model-predicted’ phenotypic estimates by genotype were also obtained and presented for the “top SNPs” in each locus.

**Hypothesis testing**

To determine genome-wide “significance” of the tested SNPs’ association with the examined phenotypes we considered a multiple-test correction. The Bonferroni has been most frequently used in published GWAS, it is commonly agreed however, that this approach is overly conservative. This issue is further discussed in the limitations part of the discussion section. After applying the multiple-test correction, assuming approximately a million independent tests, a genome-wide significance threshold of $P<5 \times 10^{-8}$ was set. We set another, less stringent threshold of $P<5 \times 10^{-6}$ for prioritizing SNPs for further investigation and locus exploration.
The FDR, as described by Storey and Tibshirani (26), an alternative method proposed to evaluate results from GWAS. In the FDR approach, a $q$ value (instead of a $p$) is calculated as a measure of significance in terms of false discovery rate, versus false positive (in the $p$ value setting). The interpretation of a $q$ value can be thought as the proportion of significant findings (genome-wide “hits”) that turn out to be false positives, after replication/verification (26). Along these lines, Shi et al recently reported a simulation study based on a two-stage FDR approach and use of a least absolute shrinkage and selection operator (LASSO) regression to reduce false positives (27).

**Sensitivity analysis - exploratory “fully” adjusted models**

Using the accepted definition of a confounder (28), no confounding of the association SNP→phenotype is to be expected by “environmental” or “behavioral” variables. This is supported by the fact that, under standard assumptions, these covariates do not affect the distribution of SNPs in the source population. However, because factors such as smoking and diabetes are risk factors for CP, one may argue that GWAS-identified signals for CP could be in fact, signals marking risk loci for these “intermediate” characteristics. In that scenario, adjusting for smoking and diabetes would “sanitize” the GWAS results from these results. On the other hand, in cases of pleiotropy (implying a true common genetic cause or risk locus of two traits) this result would reflect a valid association. Investigators have proposed methods of adjustment for correlated phenotypes (29), but “non-adjusted” analyses remain the current standard of practice to-date. Another view in favor of “adjusted” genetic modeling supports that the variance reduction that results from such adjustments is favorable
in the GWA “discovery” attempt, because there remains “less variance to be explained” by
the genetic effects.

Motivated by the above, we undertook a “sensitivity” analysis as an exploratory step
that is presented in the Appendix of Chapter 4. In these supplemental analyses we employed
smoking and diabetes-adjusted genetic models for the three traits examined in the first
manuscript (severe and moderate CP, and “extent” of attachment loss). We compared ‘crude’
and ‘adjusted’ effect estimates for the prioritized SNPs, using an empirical 10% change-in-
estimate criterion for ‘notable’ effect estimate changes, similar to settings of confounding
evaluation (30). Moreover, we inspected whether additional SNPs emerged below the
threshold of genome-wide significance (P<5x10^{-8}) or prioritization (P<5x10^{-6}), due to the
“available variance reduction” effect that was explained above.

Exploratory assessment of effect measure modification- “gene x environment”

Genome-wide explorations of gene x environment interactions are not common. As
noted in the general methods section, our study of approximately 4000 subjects was
underpowered to detect interaction effects. However, we conducted exploratory evaluations
of EMM by three variables: sex, smoking and diabetes. The rationale for the selection of
these three factors is supported by the fact that they represent important and well-
documented risk factors for CP: the disease is substantially more common among males,
smokers and diabetics. This is evident both in the literature, as well as our Dental ARIC
study population. We used the “extent” of attachment loss trait for these exploratory
analyses, because it represents a cumulative and non-reversible measure of periodontal destruction among the teeth present in the oral cavity.

EMM was evaluated on the multiplicative (log-additive) scale in the context of statistical interaction between SNPs and sex (binary variable), smoking (three-level variable) and diabetes (binary variable). Each interaction term was entered in one of three linear regression models that apart from age, sex, examination center and 10 PCs, included the risk factor main effect (i.e. smoking) and its interaction with the SNP (SNP*smoking). We conducted these analyses with the probABEL package, and considered evidence of EMM interaction term P-values of less than $10^{-5}$. This relatively high P-value threshold was chosen because interaction evaluations have traditionally low power. An alternative approach of EMM evaluation could have been stratified analyses (by sex, smoking and diabetes) followed by a between-strata homogeneity evaluation, which would also be subject to reduced power. We chose the statistical interaction approach over stratified analyses as a more efficient strategy.

**Annotation and visualization of GWA results**

Genome-wide significant and ‘prioritized’ SNPs (based on the P-value criteria that were set in the ‘hypothesis testing’ section) were annotated using the WGAViewer (31) and Snipper (http://csg.sph.umich.edu/boehnke/snipper/) programs. We explored and reported SNPs locations and their role (i.e. intronic or intergenic, representing a base pair change or a synonymous change, and more), as well as their physical distance (in Kb) from the two or three closest known genes. Additionally, in the Appendix we present associations of the prioritized SNPs with known expression quantitative loci (eQTL) (32), as well as their
interacting genes. This expression-association information was obtained from the Scan (33) database (http://www.scandb.org/newinterface/about.html) via the Snipper annotator. We report eQTL-gene associations found in lymphoblastoid cell lines (LCL) among central European (CEU) population samples and \( P \leq 10^{-4} \).

Quantile-quantile (Q-Q) plots were generated to compare the calculated versus expected empirical distributions of the test statistics (-log10 P-values) and detect any substantial deviations that could indicate residual population stratification. For this purpose we evaluated the Lambda (\( \lambda \)) inflation coefficients’ proximity to 1. Manhattan plots were also used to display the summary of the genome-wide analysis results (-log10 P-values) by chromosomal location. The Matlab\(^\text{®} \) program was used to display Q-Q plots, \( \lambda \) coefficients and Manhattan plots. To inspect genomic regions that appear associated with the examined phenotypes, we used LocusZoom\(^\text{®} \) version 1.1 (34) and Haploview v.4.2 (35). With these applications we were able to plot selected SNPs on their physical chromosome locations, along with their corresponding –log10 P-values, nearby gene locations, linkage disequilibrium (LD) and recombination rate statistics based on HapMap II-CEU. For missense changes, prediction of the possible impact of amino acid substitutions on protein structure and function was assessed using the PolyPhen-2 application (36). We used the USCF/Encode genome browser (http://genome-preview.ucsc.edu/cgi-bin/hgTracks?db=hg19) to visualize these prioritized loci for genomic areas that may include functionally relevant features, such as open chromatin and DNase I hypersensitivity annotation. Additionally, crude visualizations of functional pathways or interaction networks that identified genes may be implicated were generated with the GeneMANIA application (37). The complete set of
software that we used for annotation, visualization and additional exploration of identified SNPs and loci is presented in Table 3.1.

**Power calculations**

Several programs are available for power calculations in GWA studies (38). Because of its flexibility in enabling power calculations in the presence of GxE (gene by exposure) interactions, we used the QUANTO v. 1.2.4 program (http://hydra.usc.edu/GxE/).

Typical effect sizes of new SNPs detected by GWAS have been reported to range between 1.1 and 1.3 (15). Using our sample size of ~4,000 individuals and an outcome prevalence of 17 or 60% (depending on the definition) for Specific Aim 1, we had 80% power to detect effects of odds ratio size $\geq 1.4$ for minor allele frequencies (MAF) greater than 10%. As expected, we were less powered to detect interaction effects in the exploratory analyses, unless one assumes relative interaction effects of $\geq 1.8$ magnitude. Illustrations of various iterations of effect sizes, outcome prevalence values, MAF and power are presented in the Appendix.

**3. Analytical strategy for Specific aim 2**

Counts of the eight periodontal pathogens were derived with a semi-quantitative method, “checkerboard” DNA-DNA hybridization (10, 11). This method was applied on plaque samples to measure the extent of sub-gingival colonization with: *Prevotella intermedia* [American Type Culture Collection (ATCC) 25611], *Campylobacter rectus* (ATCC 33238), *Fusobacterium nucleatum* (ATCC 10953), and *Prevotella nigricens* (ATCC 33563) (belonging to the “orange” complex); *Porphyromonas gingivalis* (ATCC 33277), *Tannerella*
forsythia (ATCC 43037) and Treponema denticola (ATCC 35404) (“red” complex); and Aggregatibacter actinomycetemcomitans (ATCC 43718). In this method bacterial levels are expressed as counts relative to established microbial standards.

Three dichotomous traits of “high” colonization with “red” and “orange” complex, and Aggregatibacter actinomycetemcomitans were considered for analytical purposes. The two composite phenotypes were derived by the summation of bacterial count for each species belonging to the “red” (n=3) and “orange” (n=4) group, as described above. Because Porphyromonas gingivalis is considered the major periodontal pathogen implicated in periodontitis in adult populations, we explored for additional risk loci using its “high” colonization phenotype defined as above, as a separate trait, in analyses that we present in the Appendix.

Various approaches in defining the bacterial colonization profiles have been previously used, including summations of the absolute microbial counts (9), tertile-categorization (39), and five-level categorization of log10-transformed counts (40). To approach Specific Aim 2, we defined a “high” colonization trait as the top quintile (20%) of each trait’s distribution. The rationale for the selection of this phenotype is based on the fact that the “checkerboard” semi-quantitative method has a lower detection threshold of $\sim 10^3$-$10^4$ and reduced precision in the lower end of the distribution. For this reason, individuals with “high” bacterial colonization profile may be those with reduced or impaired host and at high risk for periodontal tissue destruction, thus the “high colonization cases” are a trait of interest. Additionally, selection of a smaller subset of individuals (i.e. top 5 or 10%) would reduce our already low power for the GWA analysis among the ~1000 subjects.
Genetic models, annotation and visualization of results, power calculations

Three logistic regression models assuming multiplicative (log-additive) genetic effects were employed to address Specific aim 2, one for each trait of interest: “high red” complex, “high orange” complex, and “high” Aa colonization. We used identical model specifications as in Aim 1, and included age, sex, examination center and 10 PCs as covariates. We followed identical post-analysis annotation and visualization procedures. The power analyses are presented in the Appendix. As noted in the Specific Aims section, we developed an additional fourth model to examine for risk loci for Pg as an exploratory step. Results of this analysis are presented in the Supplemental material of Chapter 5.

C. Replication plan for GWA findings

Replication of identified SNPs from any GWA study to other samples is a warranted validation step (17, 41). Some investigators have suggested that because replication efforts may fail to confirm even true associations when power considerations exist there is no reason for investigators to place unreasonably high expectations on replication studies (42). In spite of this, the field appears settled into routinely requesting replication of GWA findings.

Candidate replication datasets for the present GWA analysis will ideally include detailed ascertainment of chronic periodontitis based on the CDC taxonomy (or alternatively extent of attachment loss) using full-mouth six-sites per tooth periodontal examinations of an adequately sized sample of white subjects. The investigators’ team is actively seeking opportunities for collaboration and eventually replication of our findings. Replication of the microbiological findings (Specific Aim 2) appears less likely, but opportunities may arise in the future, as more investigators explore the host-oral microbiome interactions. Candidate
studies where a replication of our main findings could be performed if genotype data become available include the National Health And Nutrition Examination Survey (NHANES), the Health ABC study, and the Health Professionals Follow-up Study (HPFS). Meta-analysis of replicated results will be performed using the METAL program (43).
### Table 3.1. Software applications used for the genome-wide association analysis, annotation and visualization of results.

<table>
<thead>
<tr>
<th>Program name (version)</th>
<th>Functions</th>
<th>Website</th>
<th>Developer, year and citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haploview (ver.4.2)</td>
<td>Linkage disequilibrium (LD) and haplotype visualization and analysis</td>
<td><a href="http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview">http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview</a></td>
<td>Barrett, 2005</td>
</tr>
<tr>
<td>LocusZoom (ver.1.1)</td>
<td>Regional association results plotting</td>
<td><a href="http://csg.sph.umich.edu/locuszoom/">http://csg.sph.umich.edu/locuszoom/</a></td>
<td>Pruim, 2010</td>
</tr>
<tr>
<td>METAL (ver.4.2)</td>
<td>Meta-analysis of GWAS results</td>
<td><a href="http://genome.sph.umich.edu/wiki/METAL_Program">http://genome.sph.umich.edu/wiki/METAL_Program</a></td>
<td>Willer, 2010</td>
</tr>
<tr>
<td>Polyphen (ver.2.1.0)</td>
<td>Prediction of damaging missense mutations</td>
<td><a href="http://genetics.bwh.harvard.edu/pph2/index.shtml">http://genetics.bwh.harvard.edu/pph2/index.shtml</a></td>
<td>Adzhubei, 2010</td>
</tr>
<tr>
<td>PLINK (ver.1.1)</td>
<td>Genome-wide association analysis package</td>
<td><a href="http://pngu.mgh.harvard.edu/~purcell/plink/">http://pngu.mgh.harvard.edu/~purcell/plink/</a></td>
<td>Purcel, 2007</td>
</tr>
<tr>
<td>ProbABEL (ver.1.0.3)</td>
<td>Genome-wide association analysis package for imputed data</td>
<td><a href="http://www.genabel.org/packages/ProbABEL">http://www.genabel.org/packages/ProbABEL</a></td>
<td>Aulchenko, 2010</td>
</tr>
<tr>
<td>QUANTO (ver.1.2.4)</td>
<td>Power and sample size calculation in the presence of gene x environment interactions</td>
<td><a href="http://hydra.usc.edu/GxE/">http://hydra.usc.edu/GxE/</a></td>
<td>Gauderman, 2006</td>
</tr>
<tr>
<td>Snipper (ver.1.2)</td>
<td>SNPs annotation including nearby genes and expression quantitative loci</td>
<td><a href="http://csg.sph.umich.edu/boehnke/sniper/">http://csg.sph.umich.edu/boehnke/sniper/</a></td>
<td>Welch, 2010</td>
</tr>
</tbody>
</table>
E. REFERENCES


32. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 2010;6:e1000888.


CHAPTER 4

A. TITLE AND AUTHORS

MANUSCRIPT #1: EXPLORING THE GENETIC BASIS OF CHRONIC PERIODONTITIS: A GENOME-WIDE APPROACH

B. MANUSCRIPT #1 SUMMARY

Chronic periodontitis (CP) is a common-complex oral disease that affects the majority of the adult population and is a major cause of tooth loss. Although a substantial genetic component of CP is theorized, to-date, no whole-genome association (GWA) analyses have been performed. We conducted a GWA analysis among 4610 white participants of the Atherosclerosis in Communities Study (mean age of 62 years). Traits of interest were the three-level CDC/AAP periodontitis classification (severe—17%, moderate—43%, healthy—40%), and the continuous trait of “extent” of attachment loss (EAL; proportion of sites exhibiting ≥3mm attachment loss). Genotyping was performed with the Affymetrix 6.0 platform and imputation to 2.5 million markers was based on HapMap II-CEU. We used logistic genetic models for the examination of the “severe” and “moderate” CP, and a linear model for the EAL trait using a genome-wide significance threshold of $P<5\times10^{-8}$. No genome-wide significant association signals were noted. However, we found suggestive evidence of association ($P<5\times10^{-6}$) for seven loci including $\text{NIN, NPY, WNT5A}$ for severe, $\text{NCR2, EMR1}$ for moderate, and $\text{TBX18, ETS1, DYNC2H1, TTC26}$ and $\text{ZC3HAV1}$ for EAL. These genome-wide association results from a large well-defined cohort provide information on multiple candidate regions for interrogation in future genetic studies of CP.

**Keywords:** periodontitis, genome-wide association studies; oral health; dentistry; genetics;
C. INTRODUCTION

Chronic periodontitis (CP) is a common-complex disease of the oral cavity that is characterized by an inflammatory response to commensal and pathogenic oral bacteria (1). This relatively common chronic inflammatory disease is found in about 20% of the adult US population, manifests with gingival pocket formation and clinical attachment loss (CAL), and results in gradual destruction of periodontal tissues and tooth-supporting alveolar bone. CP is considered the main cause of tooth loss among most adult populations worldwide (2). Moreover, a growing body of evidence has linked the disease with increased risk for systemic conditions including coronary heart disease (CHD) (3), pregnancy outcomes (4), poor diabetes control (5), and other conditions.

There are more than 450 species identified in the human microbiome (6), and although harboring of periodontal pathogens is virtually universal, only a small proportion of individuals develop the severe form of the disease. Risk factors for CP have been well-studied and include smoking and diabetes mellitus (DM). In addition, age, race and obesity have also been shown to be important risk indicators (7). A genetic component of CP risk was supported by early reports of familial aggregation of severe forms of the disease (8), as well as twin studies (9), but the magnitude of risk conferred by genetics and the role of specific genes has been under debate.

Recent candidate-gene studies for CP have focused on genes related to host immunity and inflammatory response, such as cytokines, cell-surface receptors, chemokines, enzymes and antigen recognition. Most of these studies have examined polymorphisms in the interleukin (IL)-1, IL-6, Fc gamma receptor (FcyR), tumor necrosis factor alpha (TNFα),
human vitamin D receptor (VDR), cluster of differentiation (CD)-14, matrix metalloproteinase (MMP)-1, toll-like receptor (TLR), cyclo-omega-2 (COX-2), and C-reactive protein (CRP) gene coding regions (10).

A recent genome-wide association (GWA) study of generalized aggressive periodontitis (gAgP) among a sample of whites of European descent identified associations with a susceptibility locus on 9p34.3 intronic to the glycosyltransferase 6 domain containing 1 (GLT6D1) gene, as well as a shared susceptibility locus on 9p21.3 for both gAgP and CHD (11, 12). However, gAgP is a rare form of periodontitis, found in less than 1% of adults, and is a distinct entity from CP. To-date no GWA exploration has been performed for CP. To add to the knowledge base of the genetic etiology of CP, this study aims to investigate genetic risk loci for CP using a GWA approach in the context of a well-defined cohort.

D. MATERIALS AND METHODS

We conducted a GWA study among 4610 white participants of the Atherosclerosis Risk In Communities (ARIC) longitudinal cohort investigation (13). The ARIC is a longitudinal investigation of atherosclerosis, CVD risk factors and outcomes, which included a complete oral-dental examination between 1996 and 1998. As part of the Dental ARIC ancillary study, participants underwent complete periodontal examinations that recorded the number of missing teeth, probing depth, attachment loss and bleeding upon probing measurements at six sites per tooth, including third molars. Clinical examiners were trained and calibrated against a standard examiner, with corresponding kappas indicating excellent to outstanding level of agreement (14).
For the GWA analyses we used two traits, the Centers of Disease Control (CDC) and American Academy of Periodontology (AAP) consensus three-level classification and second, a continuous “extent” score of disease severity. The CDC/AAP classification uses CAL and PD criteria to define three CP categories as, healthy-mild, moderate and severe (15) (Supplemental Table 1). The “extent” of disease score was defined as the proportion of measured sites that exhibited CAL equal or greater than 3mm (16). The rationale for examining the continuous trait is supported by the fact that attachment loss is a non-reversible marker of periodontal destruction, whereas the CDC/AAP taxonomy includes a PD criterion, which is potentially reversible. Additional covariates that we used for descriptive and exploratory ‘adjusted’ genetic models were smoking (measured as a three-level ordinal categorical variable: never, former and current), and diabetic status (fasting glucose levels of ≥126 mg/dL, non-fasting of ≥200 mg/dL, or pharmacological treatment for diabetes).

Genotyping was based on the Affymetrix Genome-Wide Human SNP Array 6.0 chip which offers 906,600 SNP markers. The platform offers 906,600 markers for SNPs. Following rigorous quality control procedures, imputation to 2.5million markers was performed using 669,450 SNPs and MACH v1.0.16 (http://www.sph.umich.edu/csg/abecasis/MaCH/index.html), based on HapMap Phase II CEU build 36. Comprehensive descriptions of genotyping and imputation, quality control and population stratification procedures, are presented in the Appendix.

Two analytical endpoints were considered for the main effects analysis of the present project: CDC/AAP CP disease classification and “extent” of disease (attachment loss). Two contrasts were considered for the first trait: moderate vs. mild/healthy and severe vs. mild/healthy. The rationale for considering these contrasts is that severe and moderate CP are
considered distinct forms rather than variable expressions or “natural progression” stages of the disease. These analyses relied upon logistic regression models where allele effects were considered multiplicative (log-additive). To investigate the continuous trait of attachment loss, the proportion of sites exhibiting ≥3mm CAL was Z-score transformed into a normal curve equivalent variable, and a linear genetic model was used. All models included age, sex, examination center and ten principal components from the EIGENSTRAT analysis as covariates. A correction for multiple comparisons was employed assuming 1 million independent tests resulting in a threshold of genome-wide statistical significance of P<5x10^{-8}.

Although variables such as smoking and DM are not likely confounders of the association between genetic polymorphisms and risk of CP, we developed a series of models adjusting for these variables, as a sensitivity analysis. For this step, the results of which are presented in the supplemental material, we used an arbitrary criterion of >10% change-in-estimate which is often used in confounding evaluation in epidemiologic studies (17). We used this criterion to inspect for “important” changes in estimate for the prioritized SNPs, and explored whether additional ones emerged below the P<5x10^{-6} threshold upon the resulting variance reduction.

All genetic analyses were performed with the ProbABEL software (18). Post-analysis procedures included the generation of quantile-quantile (Q-Q) and Manhattan plots. A threshold of P<5x10^{-6} was set for prioritizing SNPs for further investigation. SNPs were annotated using WGAViewer ver.1.26l (19) and Snipper ver. 1.2 (http://csg.sph.umich.edu/boehnke/snipper/), and regions were viewed using LocusZoom ver.1.1 (20) and Haploview ver.4.2 (21). We used additional online resources of the National Center for Biotechnology Information (NCBI- http://www.ncbi.nlm.nih.gov/). Reporting of
genes was based on the “HUGO Gene Nomenclature” naming convention (www.genenames.org).

E. RESULTS

The descriptive characteristics of the Dental ARIC cohort participants that were included in this analysis are presented in Table 4.1. Participants had a mean age of 62 years, with a balanced sex distribution. Twelve percent were current smokers, and eleven percent had DM. In terms of CDC/AAP periodontal diagnoses, these were severe—17%, moderate—43%, and healthy—40%. In the main analyses we found no genome-wide significant association signals. Of the total 2178777 examined SNPs, 26 had a P<5x10^-6 and thus were prioritized for further investigation (Figures 4.1-3). Lambda variance inflation factors for the three traits were low, and ranged between 1.003 for attachment loss to 1.024 for severe CP (Figures 4.10-12). The prioritized SNPs marked three loci as associated with moderate (6p21.1, 19p13.3 and 10p15), three with severe CP (14.21, 7p15, and 3p21), and four loci with the continuous trait of attachment loss (6q15, 11q24, 11q22, and 7q34).

A comprehensive annotation of the prioritized SNPs, as well as corresponding allele frequencies by trait is presented in Table 4.2. Visualizations of the corresponding loci, along with nearby genes and recombination rates are presented in Figure 4.4-9 and Supplementary Figures 4.13-18. For severe CP, the strongest (with respect to P-value) association in the 14q21 locus was produced by rs12883458, intronic to NIN. The minor C allele showed a 4% enrichment among severe CP patients, and was associated with an OR=1.89, P=3.5x10^-7. In the 7p15 locus, the common allele of rs2521634 (47Kb from NPY) produced an OR=1.47, P=1.6x10^-6. Similarly, the common G allele of rs11925054 in the 3p21 locus, adjacent to
WNT5A and ERC2 produced the strongest signal in the region, showing 4% enrichment among severe CP patients. With regard to moderate CP, the SNP with the lowest P-value in 6p21.1 was rs7762544 (OR=1.41; P=1.1x10^{-7}). This variant is 61Kb from NCR2, and its minor (risk) allele showed 5% enrichment among moderate CP patients compared to healthy participants. Rs3826782 which is intronic to EMR1 and 30Kb from VAV1 provided the strongest signal in that locus with OR=2.00 and P=4.0x10^{-6}.

With regard to attachment loss, several SNPs in LD with the top SNP rs17792917 (177Kb from TBX18; P=1.8x10^{-6}) were found in the 6q15 locus. Individuals with 2 copies of the rare (risk) T allele had more than 10% higher mean extent of attachment loss (22.7% vs. 20.2%) compared to those with CC genotype. The 11q24 locus where rs10790919 (P=1.9x10^{-6}) provided the strongest signal was in an intergenic region (671Kb from ETS1 and 787Kb from KIRREL3). Rs7120142 provided the lowest P-value in the 11q22 locus (P=3.9x10^{-6}), and was located 67Kb from DYNC2H1 and 360Kb from PDGFD. Carriers of 2 risk allele copies (rare allele T) had a substantially higher mean extent of attachment loss (25.0) compared to those with no risk allele copy (20.4). The 7q34 locus is a gene-rich area, where rs10500130 emerged as the top SNP (P=4.6x10^{-6}), with the common C allele being the one associated with higher extent of attachment loss score. Rs10500130 was 3Kb from TTC26 and 21Kb from ZC3HAV1. Rs1537415 which was the “top hit” in a GWA of AgP (11) did not show an important association (P=0.5) with the any of our examined traits.

F. DISCUSSION

This manuscript presents results of the first genome-wide exploration for loci associated with the risk of CP. The study is limited by the sample size for a GWA, of about 4,000
subjects and the absence of a replication sample. The study strengths include a well-defined cohort with a detailed phenotypical characterization of CP, using full mouth periodontal examinations and the latest commonly accepted taxonomy of CP. The CDC/AAP classification has been used in epidemiologic studies and surveillance. Offenbacher and colleagues (22) introduced a refined CP classification characterizing the disease’s biological (versus clinical-only) phenotype that may be more useful when exploring genetic effects of a trait with underlying heterogeneity. Nonetheless, the results of the present investigation provide a wealth of new information on potential candidate genes and mechanistic pathway analyses that will require further exploration, replication and validation in future studies. In general, it can be anticipated that interrogations of the genetic etiology of CP will identify markers associated with defense molecules and pathways, as is the case with numerous candidate-gene approaches (10).

Several promising loci and candidate genes were identified in the present analysis. *NIN* is a gene known to encode a protein that plays a role in centrosomal microtubule organization and anchoring, which have recently been recognized as elements of the T cell cytolytic response (23). Interestingly, a recent candidate-gene study by Olson and colleagues reported two of the “top hits” for severe CP in 14q21 of the present study (rs12893300 and rs1004832) as associated with breast cancer risk (24). *NPY* is a gene encoding a neuropeptide that is widely expressed in the central nervous system and has been suggested to function as an anxiolytic peptide that helps explain inter-individual variation in trait anxiety and resiliency to stress (25); a common risk factor for many diseases, including CP. Moreover, experimental evidence supports a role of WNT5A in the activation of MMP-2 and the regulation of inflammatory cytokine genes of macrophages (26). *NCR2* is the gene in the
region with strongest signal for moderate CP. It appears to have an important role in both normal and pathological innate immune responses, and is selectively expressed by Interleukin 2-activated natural killer cells (27). EMR1, VAV1 and CELF2 have also been implicated in immune functions such as eosinophilic inflammation, T-cell and B-cell development and activation (28).

With regard to the “extent” of disease trait, apart from two genes are adjacent to the locus 6q15 (TBX18 at 177kb and KIAA1009 at 300Kb), the top SNP is 1Kb upstream of the novel processed transcript RP1-90L14.1-001. Other promising loci and potential candidate genes include PDGFD in 11q22, TTC26 and ZC3HAV1 at 7q34. Wagsater and colleagues (29) reported that PDGFs play important roles in atherosclerosis by stimulating matrix metalloproteinase (MMP) activity and influencing monocyte migration. ZC3HAV1 encodes a zinc finger protein that is thought to prevent infection by retroviruses, possibly by directly interacting with viral RNA (30).

The lack of an overlap of identified loci for the three examined traits is not surprising, and in fact verifies the rationale of examining these three phenotypes separately. Moderate and severe CP are considered largely distinct entities and include in their case definitions a reversible clinical marker (probing depth), while “extent” attachment loss is a cumulative and non-reversible marker of periodontal destruction. From a statistical standpoint, the lack of overlap is not surprising because small stochastic variations can have a big impact on the tails of the test statistic distribution. However, because these traits share a common pathogenetic underpinning, and in our analyses we used the same controls for both contrasts, some overlap in GWA signals should be anticipated. Explorations at lower P-value thresholds may reveal more “good signals” and common risk loci. As an example, in our
exploratory smoking and diabetes-adjusted analyses 4q21.3 emerged as a common risk locus for severe CP and attachment loss and may be a reflection of this overlap (Figure 4.19).

It must be acknowledged that although GWAS have provided invaluable new information on the genetic basis for many disease and health outcomes, a large component of the disease variation remains to-date unexplained (31). The inability of GWAS to detect very small effect sizes and interrogate rare polymorphisms, gene x gene interactions and epigenetics are additional layers of unaddressed complexity. In our study, although no SNP reached the level of genome-wide statistical significance, several loci were nominally associated with substantial effect sizes and, if replicated, may offer promising avenues for further investigation and mechanistic studies.

Further investigations providing replication of these findings and additional examination of specific associations may lead to an improved understanding of the pathogenesis of the disease, as well as novel preventive and therapeutic approaches.

MANUSCRIPT #1- REPORTING OF FUNDING SOURCES

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C; and by a grant from the National Institute of Environmental Health Sciences (P30ES010126). The authors thank the staff and participants of the ARIC
study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The authors declare no potential conflicts of interest with regard to the authorship and publication of this article.
Table 4.1 Distribution in the total sample and bivariate associations of periodontal diagnosis (CDC/AAP disease classification) and periodontal attachment loss (extent score: percent of sites with $\geq 3$mm attachment loss) with sex, smoking and diabetic status among the white participants of the Dental ARIC study (n=4610).

<table>
<thead>
<tr>
<th></th>
<th>Total sample</th>
<th>CDC/AAP Periodontitis classification</th>
<th>Attachment loss ($\geq 3$mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (n, column %)</td>
<td>Health-mild (n, row %)</td>
<td>Moderate (n, row %)</td>
</tr>
<tr>
<td>Total sample</td>
<td>4610 (100)</td>
<td>1864 (40)</td>
<td>1961 (43)</td>
</tr>
<tr>
<td>Sex $^{3,4}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>2415 (52)</td>
<td>1197 (50)</td>
<td>939 (39)</td>
</tr>
<tr>
<td>Males</td>
<td>2195 (48)</td>
<td>667 (30)</td>
<td>1022 (47)</td>
</tr>
<tr>
<td>Smoking status $^{3,5}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>2104 (47)</td>
<td>1055 (50)</td>
<td>817 (39)</td>
</tr>
<tr>
<td>Former smoker</td>
<td>1876 (42)</td>
<td>633 (34)</td>
<td>870 (46)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>526 (12)</td>
<td>147 (28)</td>
<td>221 (42)</td>
</tr>
<tr>
<td>Diabetic status $^{3,4}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy</td>
<td>4077 (89)</td>
<td>1704 (42)</td>
<td>1704 (42)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>527 (11)</td>
<td>158 (30)</td>
<td>254 (48)</td>
</tr>
</tbody>
</table>

1: Third molars were included in the calculation of the extent of attachment loss trait
2: Rounded to the closest integer
3: Chi-square test of equivalence between strata of periodontitis P<0.05
4: Median test of stratum-specific attachment loss estimates P<0.05
5: Kruskal-Wallis test of stratum-specific attachment loss or covariate estimates P<0.05
Table 4.2. Genome-wide association analysis results of the CDC/AAP chronic periodontitis (CP) classification traits (severe CP vs. healthy and moderate CP vs. healthy) and extent of attachment loss trait (EAL-proportion of sites exhibiting attachment loss ≥3mm), among the white participants of the Dental ARIC study (n=4610). Single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF-HapMap II CEU) of ≥5% and associated P<5x10^{-6}. The SNP with the lowest P-value per locus is presented; additional prioritized SNPs in each locus are presented in the footnote, along with corresponding R² (based on 1000 genomes pilot 1 release) with the top SNPs.

<table>
<thead>
<tr>
<th>Chr.</th>
<th>SNP</th>
<th>Position Build36</th>
<th>ca¹</th>
<th>nca²</th>
<th>Risk allele frequency (HapMap II-CEU)</th>
<th>Closest gene(s) and position or distance</th>
<th>P value (beta)</th>
<th>Risk allele frequency disease/healthy</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14q21</td>
<td>rs12883458</td>
<td>50349129</td>
<td>C</td>
<td>T</td>
<td>[C] 0.104</td>
<td>NIN (Intronic)</td>
<td>3.5x10^{-7} (0.64)</td>
<td>0.13/0.09</td>
<td>1.89 (1.48, 2.41)</td>
</tr>
<tr>
<td>7p15</td>
<td>rs2521634</td>
<td>24344565</td>
<td>A</td>
<td>G</td>
<td>[G] 0.754</td>
<td>NPY (47Kb)</td>
<td>1.6x10^{-6} (0.39)</td>
<td>0.80/0.74</td>
<td>1.47 (1.25, 1.73)</td>
</tr>
<tr>
<td>3p21</td>
<td>rs11925054</td>
<td>55365926</td>
<td>G</td>
<td>T</td>
<td>[G] 0.865</td>
<td>WNT5A (109Kb); ERC2 (151Kb)</td>
<td>6.5x10^{-7} (0.53)</td>
<td>0.90/0.86</td>
<td>1.69 (1.37, 2.10)</td>
</tr>
<tr>
<td>Moderate CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6p21.1</td>
<td>rs7762544</td>
<td>41487293</td>
<td>A</td>
<td>G</td>
<td>[G] 0.184</td>
<td>NCR2 (61Kb)</td>
<td>1.1x10^{-7} (0.34)</td>
<td>0.21/0.16</td>
<td>1.41 (1.24, 1.60)</td>
</tr>
<tr>
<td>19p13.3</td>
<td>rs3826782</td>
<td>6838736</td>
<td>A</td>
<td>G</td>
<td>[A] 0.070</td>
<td>EMR1 (Intronic); VAV1 (30Kb)</td>
<td>4.0x10^{-6} (0.69)</td>
<td>0.05/0.04</td>
<td>2.00 (1.48, 2.70)</td>
</tr>
<tr>
<td>10p15</td>
<td>rs12260727</td>
<td>10378335</td>
<td>A</td>
<td>G</td>
<td>[G] 0.846</td>
<td>CELF2 (709Kb)</td>
<td>6.0x10^{-7} (0.43)</td>
<td>0.89/0.85</td>
<td>1.54 (1.30, 1.82)</td>
</tr>
<tr>
<td>Extent of attachment loss (≥3mm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6q15</td>
<td>rs17792917</td>
<td>85323684</td>
<td>C</td>
<td>T</td>
<td>[T] 0.242</td>
<td>TRX18 (177Kb); KIAA1009 (330Kb)</td>
<td>1.8x10^{-6} (1.92)</td>
<td>20.2</td>
<td>22.4</td>
</tr>
<tr>
<td>11q24</td>
<td>rs10790919</td>
<td>127162281</td>
<td>A</td>
<td>G</td>
<td>[A] 0.788</td>
<td>ETS1 (671Kb); KIRREL3 (787Kb)</td>
<td>1.9x10^{-6} (2.04)</td>
<td>17.1</td>
<td>19.9</td>
</tr>
<tr>
<td>11q22</td>
<td>rs7120142</td>
<td>102922991</td>
<td>C</td>
<td>T</td>
<td>[T] 0.197</td>
<td>DYN2CH1 (67Kb); PDGF (360Kb)</td>
<td>3.9x10^{-6} (1.98)</td>
<td>20.4</td>
<td>22.0</td>
</tr>
<tr>
<td>7q34</td>
<td>rs10500130</td>
<td>138466453</td>
<td>A</td>
<td>C</td>
<td>[C] 0.889</td>
<td>TTC26 (3Kb); ZC3HAV1 (21Kb)</td>
<td>4.6x10^{-6} (2.56)</td>
<td>18.8</td>
<td>19.0</td>
</tr>
</tbody>
</table>

¹: coded allele
²: non-coded allele
3: confidence limits
4: Additional SNPs in locus with P<5x10^{-6}: rs1004832 (R²=1.00), rs8009874 (R²=0.84), rs12893300 (R²=0.3)
5: Additional SNP in locus with P<5x10^{-6}: rs11771124 (R²=1.00)
6: Additional SNP in locus with P<5x10^{-6}: rs503022 (R²=0.52)
7: Additional SNPs in locus with P<5x10^{-6}: rs9357360 (R²=0.89), rs1853406 (R²=0.89), rs1535582 (R²=0.33)
8: Additional SNP in locus with P<5x10^{-6}: rs12610529 (R²=0.79)
9: Assuming a log-additive linear genetic model
10: CL, confidence limits
11: Additional SNPs in locus with P<5x10^{-6}: rs9791329 (R²=0.58), rs11756650 (R²=0.96), rs7741380 (R²=0.60), rs4510639 (R²=0.67)
12: Additional SNPs in locus with P<5x10^{-6}: rs6590279 (R²=0.87), rs10893747 (R²=0.87)
13: Additional SNP in locus with P<5x10^{-6}: rs4440990 (R²=1.00)
H. FIGURES

Figure 4.1 Manhattan plot of the genome-wide association analysis results for severe chronic periodontitis among the 4610 white dental ARIC participants.
Figure 4.2 Manhattan plot of the genome-wide association analysis results for moderate chronic periodontitis among the 4610 white dental ARIC participants.
**Figure 4.3** Manhattan plot of the genome-wide association analysis results for extent of attachment loss (proportion of sites exhibiting $\geq 3\text{mm}$ attachment loss) among the 4610 white dental ARIC participants.
Figure 4.4 Visualization of the 14q21 locus that was marked by rs12883458 ($P=3.5 \times 10^{-7}$) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
Figure 4.5 Visualization of the 19p13.3 locus that was marked by rs3826782 ($P=4.0\times10^{-6}$) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
**Figure 4.6** Visualization of the 6p21.1 locus that was marked by rs7762544 ($P=1.1\times10^{-7}$) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
**Figure 4.7** Visualization of the 7p15 locus that was marked by rs2521634 (P=1.6x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
Figure 4.8 Visualization of the 6q15 locus that was marked by rs17792917 (P=1.8x10^{-6}) for extent of attachment loss (CAL; sites exhibiting CAL ≥3 mm) among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
**Figure 4.9** Visualization of the 7q34 locus that was marked by rs10500130 (P=4.6x10^{-6}) for extent of attachment loss (CAL; sites exhibiting CAL ≥3 mm) among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
I. REFERENCES


SUPPLEMENTAL MATERIALS AND METHODS

Genotyping and imputation

In the study population, DNA was extracted from blood samples drawn from an antecubital vein into tubes containing serum separator gel. Blood samples were analyzed at a central ARIC laboratory in Houston, TX. Genotyping was performed with the Affymetrix Genome-Wide Human SNP Array 6.0 chip. The platform offers 906,600 markers for SNPs. The rigorous quality control procedures included initial blind duplicate genotyping and identification/flagging of SNPs with kappa<0.95 and reconciliation of unintentional duplicate samples (17 duplicates and one triplicate). Imputation to 2.5million markers was performed using 669,450 SNPs and MACH v1.0.16 (http://www.sph.umich.edu/csg/abecasis/MaCH/index.html), based on HapMap Phase II CEU build 36. The SNPs used for imputation were selected from 839,048 autosomal SNPs restricted to those with minor allele frequency (MAF)>0.01 (129,543 excluded), Hardy-Weinberg equilibrium (HWE) P>10^{-5} (12,432 excluded) and call rate >95% (1,693 excluded). We used the following SNP exclusion criteria for further analyses: quality score <0.8 and missing data rate >10% after imputation, and MAF of <5%.

Population stratification

To obtain estimates of relatedness and population stratification a subset of 85,947 “high quality” SNPs was selected. These SNPs met the following criteria among self-reported whites: MAF ≥0.1, call rate >99.5%, HWE P≥10^{-3}, autosomal, with annotation in the platform annotation file, not labeled “AFFX” or “chromosome 0”, and not monomorphic.
Using these SNPs identity-by-state (IBS) allele sharing distance (DST values) were computed using PLINK, as such: DST = IBS distance (IBS_2 + 0.5*IBS_1) / (n SNP pairs). First degree relative status was assigned to pairs of individuals with DST ≥ 0.8 and second degree relatives were considered those with 0.763 ≤ DST < 0.8. Among the white ARIC participants there were 380 pairs of first degree and 207 pairs of second degree relatives identified. To minimize exclusions, related pairs were broken by iterative selection of individuals with most relatives using a custom program.

Population stratification was further evaluated with principal component (PC) analysis using the EIGENSTRAT method (1). The above chosen set of SNPs was used for the computation of ten principal components. Genetic outliers were considered those that were further than 8 standard deviations (SD) away from any of ten PCs over ten runs of PC computation. Based on DST and PC criteria there were 716 subjects flagged from removal from the analysis (206 as genetic outliers based on PCs and 16 based on average DST values (“too little IBS sharing” with the rest of the sample), 351 first degree relatives and 143 second degree relatives. All but ten second degree relatives (whose relatives were excluded as genetic outliers) were re-entered in the dataset and were assigned PCs. After exclusion of 364 individuals (4%) there were 9349 whites who were included in the GWA analysis and of those, 4610 had periodontal phenotype data available as Dental ARIC participants.

SUPPLEMENTAL RESULTS

Adjusted genetic models
Adjustment for smoking and diabetic status resulted in small (<10%) and predominantly upwards changes in odds ratios of the prioritized SNPs for the two disease classification traits (Supplemental Table S2). Noteworthy, the effect estimate of rs7762544 associated with moderate CP increased by 6% for a genome-wide significant adjusted OR=1.44 (P=3.8x10^{-8}). Larger and exclusively downwards changes in effect size estimates were noted for the prioritized SNPs in the domain of attachment loss. These estimate attenuations ranged from 11% for rs10500130 to 23% for rs7120142. Four additional loci emerged below the P<5x10^{-6} threshold in the adjusted analyses, one of those common for severe CP and extent of attachment loss: moderate CP—[rs11615037, OR=1.33; P_{adj}=4.4x10^{-6}, in a region between CPM (98Kb) and CPSF6 (178Kb)]; severe CP—(rs10493998, OR=1.68; P_{adj}=2.2x10^{-6}, 90Kb from COL11A); extent of CP—(rs8006336, P_{adj}=1.4x10^{-6} adjacent to RPS29 and PPIL5). Rs17006135, was intronic to SCD5, approximately 3Kb from the exon boundary, marking the 4q21.3 locus, and was associated with both severe CP (OR=1.82; P_{adj}=4.4x10^{-6}) and increased extent of attachment loss (P_{adj}=8.2x10^{-7}).
K. SUPPLEMENTAL TABLES

Table 4.3 Clinical case definitions of Periodontitis proposed by the CDC working group for use in population-based surveillance of Periodontitis\(^1\) (2)

<table>
<thead>
<tr>
<th>Disease Category</th>
<th>Clinical Attachment Level</th>
<th>Probing Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Severe periodontitis</strong></td>
<td>(\geq 2) interproximal sites(^2) with CAL (\geq 6)mm</td>
<td>(\geq 1) interproximal site with PD (\geq 5)mm</td>
</tr>
<tr>
<td><strong>Moderate periodontitis</strong></td>
<td>(\geq 2) interproximal sites(^2) with CAL (\geq 4)mm</td>
<td>(\geq 2) interproximal sites(^2) with PD (\geq 5)mm</td>
</tr>
<tr>
<td><strong>No or mild periodontitis</strong></td>
<td>Neither “moderate” nor “severe” periodontitis</td>
<td></td>
</tr>
</tbody>
</table>

1: third molars are excluded

2: not on the same tooth
Table 4.4 Changes in estimate of top SNPs for the three chronic periodontitis traits after adjustment for smoking (never, former, current) and diabetic status (binary) among the white participants of the Dental ARIC study.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Chr.</th>
<th>SNP</th>
<th>RAF(^1) Disease/Healthy</th>
<th>Crude estimates</th>
<th>Adjusted(^2) estimates</th>
<th>Change in estimate(^3) (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate vs. Healthy</td>
<td>6p21.1</td>
<td>rs7762544</td>
<td>0.21/0.16</td>
<td>1.41 (1.24, 1.60)</td>
<td>1.1x10(^{-7})</td>
<td>1.44 (1.26, 1.63)</td>
</tr>
<tr>
<td></td>
<td>19p13.3</td>
<td>rs3826782</td>
<td>0.05/0.04</td>
<td>2.00 (1.48, 2.70)</td>
<td>4.0x10(^{-6})</td>
<td>2.09 (1.54, 2.84)</td>
</tr>
<tr>
<td></td>
<td>10p15</td>
<td>rs12260727</td>
<td>0.15/0.11</td>
<td>1.54 (1.30, 1.82)</td>
<td>6.0x10(^{-7})</td>
<td>1.51 (1.27, 1.79)</td>
</tr>
<tr>
<td>Severe vs. Healthy</td>
<td>14q21</td>
<td>rs12883458</td>
<td>0.13/0.09</td>
<td>1.89 (1.48, 2.41)</td>
<td>3.5x10(^{-7})</td>
<td>1.88 (1.46, 2.43)</td>
</tr>
<tr>
<td></td>
<td>7p15</td>
<td>rs2521634</td>
<td>0.80/0.74</td>
<td>1.47 (1.25, 1.73)</td>
<td>1.6x10(^{-6})</td>
<td>1.43 (1.21, 1.69)</td>
</tr>
<tr>
<td></td>
<td>3p21</td>
<td>rs11925054</td>
<td>0.90/0.86</td>
<td>1.69 (1.37, 2.10)</td>
<td>6.5x10(^{-7})</td>
<td>1.78 (1.42, 2.23)</td>
</tr>
<tr>
<td>Extent of attachment loss</td>
<td>6q15</td>
<td>rs17792917</td>
<td>1.92 (1.13, 2.71)</td>
<td>1.8x10(^{-6})</td>
<td></td>
<td>1.61 (0.84, 2.38)</td>
</tr>
<tr>
<td></td>
<td>11q24</td>
<td>rs10790919</td>
<td>2.04 (1.20, 2.88)</td>
<td>1.9x10(^{-6})</td>
<td></td>
<td>1.78 (0.96, 2.59)</td>
</tr>
<tr>
<td></td>
<td>11q22</td>
<td>rs7120142</td>
<td>1.98 (1.14, 2.83)</td>
<td>3.9x10(^{-6})</td>
<td></td>
<td>1.54 (0.71, 2.36)</td>
</tr>
<tr>
<td></td>
<td>7q34</td>
<td>rs10500130</td>
<td>2.56 (1.47, 3.66)</td>
<td>4.6x10(^{-6})</td>
<td></td>
<td>2.28 (1.21, 3.34)</td>
</tr>
</tbody>
</table>

1: Risk allele frequency
2: Adjusted for smoking (three-level categorical where 0: never, 1: former, 2: current smoker) and diabetic status (dichotomous variable where 0: healthy and 1: diabetic)
3: Calculated as follows: change-in-estimate = [(beta\(_{adjusted}\) - beta\(_{crude}\)) / beta\(_{crude}\)] *100
4: OR (odds ratios) are presented for the two binary periodontitis classification traits and beta coefficients for the continuous trait of attachment loss
5: Confidence limits
Table 4.5 Distribution of genotype (based on number of minor allele copies of SNPs with $P<5\times10^{-6}$) by phenotypic classification (three disease traits and mean “extent” of disease) among the Dental ARIC participants.

<table>
<thead>
<tr>
<th>Minor allele copies (n)</th>
<th>Chr. SNP</th>
<th>Healthy (%)</th>
<th>Moderate chronic periodontitis (%)</th>
<th>Severe chronic periodontitis (%)</th>
<th>Mean extent of attachment loss (≥3mm) and (95% confidence limits)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0 (%) 1 (%) 2 (%)</td>
<td>0 (%) 1 (%) 2 (%)</td>
<td>0 (%) 1 (%) 2 (%)</td>
<td>0 (%) 1 (%) 2 (%)</td>
</tr>
<tr>
<td>6p21.1</td>
<td>rs7762544$^{1,4}$</td>
<td>69.6 27.9 2.5</td>
<td>62.3 33.9 3.9</td>
<td>68.1 29.3 2.6</td>
<td>20.9 (20.1, 21.6) 21.5 (20.4, 22.7) 22.4 (18.6, 26.2)</td>
</tr>
<tr>
<td>19p13.3</td>
<td>rs3826782$^{1,4}$</td>
<td>93.0 6.9 0.1</td>
<td>90.2 9.6 0.3</td>
<td>91.7 8.3 0.0</td>
<td>21.1 (20.4, 21.7) 22.5 (20.3, 24.7) 16.1 (9.1, 23.1)</td>
</tr>
<tr>
<td>10p15</td>
<td>rs12260727$^1$</td>
<td>72.4 25.5 2.1</td>
<td>78.3 20.6 1.1</td>
<td>76.3 22.0 1.7</td>
<td>21.5 (20.8, 22.3) 20.0 (18.7, 21.3) 17.1 (12.6, 21.6)</td>
</tr>
<tr>
<td>14q21</td>
<td>rs12883458$^{2,4}$</td>
<td>82.6 17.1 0.3</td>
<td>80.2 18.9 0.9</td>
<td>74.8 23.6 1.7</td>
<td>20.7 (20.0, 21.4) 22.6 (21.2, 24.1) 25.6 (19.1, 32.1)</td>
</tr>
<tr>
<td>7p15</td>
<td>rs2521634$^{2,4}$</td>
<td>54.1 39.2 6.8</td>
<td>55.4 38.8 5.8</td>
<td>63.8 32.4 3.8</td>
<td>21.6 (20.8, 22.4) 20.8 (19.8, 21.8) 18.6 (16.2, 21.1)</td>
</tr>
<tr>
<td>3p21</td>
<td>rs11925054$^2$</td>
<td>74.0 24.0 2.0</td>
<td>74.5 23.8 1.7</td>
<td>81.0 18.0 1.0</td>
<td>21.7 (21.0, 22.4) 19.6 (18.4, 20.8) 16.8 (13.2, 20.5)</td>
</tr>
<tr>
<td>6q15</td>
<td>rs17792917$^{3,4}$</td>
<td>61.9 33.0 5.1</td>
<td>56.9 37.6 5.5</td>
<td>53.6 39.1 7.3</td>
<td>20.2 (19.4, 21.0) 22.4 (21.3, 23.4) 22.7 (20.0, 25.4)</td>
</tr>
<tr>
<td>11q24</td>
<td>rs10790919$^3$</td>
<td>60.2 35.4 4.4</td>
<td>63.0 33.0 3.9</td>
<td>68.7 28.8 2.5</td>
<td>22.0 (21.2, 22.8) 19.9 (18.9, 20.9) 17.1 (14.8, 19.5)</td>
</tr>
<tr>
<td>11q22</td>
<td>rs7120142$^{3,4}$</td>
<td>66.7 30.5 2.8</td>
<td>61.6 34.3 4.1</td>
<td>62.9 33.3 3.8</td>
<td>20.4 (19.7, 21.2) 22.0 (20.9, 23.1) 25.0 (21.3, 28.7)</td>
</tr>
<tr>
<td>7q34</td>
<td>rs10500130$^3$</td>
<td>79.0 19.7 1.3</td>
<td>81.6 17.5 0.9</td>
<td>82.9 15.9 1.2</td>
<td>18.8 (12.6, 25.0) 19.0 (17.7, 20.4) 21.6 (20.9, 22.3)</td>
</tr>
</tbody>
</table>

1: Associated $P<5\times10^{-6}$ for moderate CP
2: Associated $P<5\times10^{-6}$ for severe CP
3: Associated $P<5\times10^{-6}$ for extent of attachment loss
4: The minor allele is the ‘risk allele’ for the corresponding trait
Table 4.6 Reported genes’ symbols [HUGO Gene Nomenclature Committee (HGNC) approved symbols], names and chromosomal locations.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCR2</td>
<td>Natural cytotoxicity triggering receptor 2</td>
<td>6p21.1</td>
</tr>
<tr>
<td>EMR1</td>
<td>Egf-like module containing, mucin-like, hormone receptor-like 1</td>
<td>19p13.3</td>
</tr>
<tr>
<td>VAV1</td>
<td>Vav 1 guanine nucleotide exchange factor</td>
<td>19p13.2</td>
</tr>
<tr>
<td>CELF2</td>
<td>CUGBP, Elav-like family member 2</td>
<td>10p13</td>
</tr>
<tr>
<td>NIN</td>
<td>Ninein (GSK3B interacting protein)</td>
<td>14q21-q22</td>
</tr>
<tr>
<td>NPY</td>
<td>Neuropeptide Y</td>
<td>7p15.3</td>
</tr>
<tr>
<td>WNT5A</td>
<td>Wingless-type MMTV integration site family, member 5A</td>
<td>3p21-p14</td>
</tr>
<tr>
<td>ERC2</td>
<td>ELKS/RAB6-interacting/CAST family member 2</td>
<td>3p14.3</td>
</tr>
<tr>
<td>TBX18</td>
<td>T-box 1</td>
<td>6q14.1-q15</td>
</tr>
<tr>
<td>KIAA1009</td>
<td>KIAA1009</td>
<td>14.3</td>
</tr>
<tr>
<td>ETS1</td>
<td>V-ets erythroblastosis virus E26 oncogene homolog 1 (avian)</td>
<td>11q23.3</td>
</tr>
<tr>
<td>KIRREL3</td>
<td>Kin of IRRE like 3 (Drosophila)</td>
<td>11q24</td>
</tr>
<tr>
<td>DYNC2H1</td>
<td>Dynein, cytoplasmic 2, heavy chain 1</td>
<td>11q21-q22.1</td>
</tr>
<tr>
<td>PDGFD</td>
<td>Platelet derived growth factor D</td>
<td>11q22.3</td>
</tr>
<tr>
<td>TTC26</td>
<td>Tetramericopeptide repeat domain 26</td>
<td>7q34</td>
</tr>
<tr>
<td>ZC3HAV1</td>
<td>Zinc finger CCCH-type, antiviral 1</td>
<td>7q34</td>
</tr>
<tr>
<td>CPM*</td>
<td>Carboxypeptidase M</td>
<td>12q15</td>
</tr>
<tr>
<td>CPSF6*</td>
<td>Cleavage and polyadenylation specific factor 6, 68kDa</td>
<td>12q15</td>
</tr>
<tr>
<td>COL11A1*</td>
<td>Collagen, type XI, alpha 1</td>
<td>1p21</td>
</tr>
<tr>
<td>RNPC3*</td>
<td>RNA-binding region (RNP1, RRM) containing 3</td>
<td>1p21.1</td>
</tr>
<tr>
<td>SCD5*</td>
<td>Stearoyl-CoA desaturase 5</td>
<td>4q21.3</td>
</tr>
<tr>
<td>SEC31A*</td>
<td>SEC31 homolog A (S. cerevisiae)</td>
<td>4q21</td>
</tr>
<tr>
<td>RPS29*</td>
<td>Ribosomal protein S29</td>
<td>14q21.3</td>
</tr>
<tr>
<td>LRR1*</td>
<td>Leucine rich repeat protein 1 (PPIL5)</td>
<td>14q21.3</td>
</tr>
</tbody>
</table>

*Genes adjacent to loci that emerged in adjusted analyses.
Table 4.7 Reported SNPs address according to the 1000 genomes (1000G) convention and proxy SNPs (R^2>0.8 where available) based on 1000G pilot 1 release. All SNPs were imputed based on HapMap II CEU with the exception of rs2521634 with was directly genotyped on the Affymetrix 6.0 platform.

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>1000G</th>
<th>Proxy SNPs (R^2&gt;0.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7762544</td>
<td>6:41379315</td>
<td>rs1853406, rs9357360</td>
</tr>
<tr>
<td>rs3826782</td>
<td>19:6887736</td>
<td>rs12610529</td>
</tr>
<tr>
<td>rs12260727</td>
<td>10:10338329</td>
<td>rs12258450, rs7919833</td>
</tr>
<tr>
<td>rs12883458</td>
<td>14:51279379</td>
<td>rs1004832</td>
</tr>
<tr>
<td>rs2521634</td>
<td>7:24378040</td>
<td>rs11771124, rs10487606, rs10487605</td>
</tr>
<tr>
<td>rs11925054</td>
<td>3:55390886</td>
<td>rs503022^1</td>
</tr>
<tr>
<td>rs17792917</td>
<td>6:85266965</td>
<td>rs16874800</td>
</tr>
<tr>
<td>rs10790919</td>
<td>11:127657071</td>
<td>rs6590279</td>
</tr>
<tr>
<td>rs7120142</td>
<td>11:103417781</td>
<td>rs4440990</td>
</tr>
<tr>
<td>rs10500130</td>
<td>7:138815913</td>
<td>rs12673905^2, rs12056143^3</td>
</tr>
</tbody>
</table>

^1: R^2=0.40; ^2: R^2=0.46; ^3: R^2=0.43
Table 4.8 Reported SNPs as expression quantitative loci (eQTLs) and their association with gene expression (lymphoblastoid cell line-based expression) in populations of European descent with $P<10^{-4}$ in the Scan database (http://www.scandb.org).

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Locus</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3826782</td>
<td>GPR113</td>
<td>G protein-coupled receptor 113</td>
<td>2p23.3</td>
<td>$2\times10^{-5}$</td>
</tr>
<tr>
<td>rs11925054</td>
<td>TMEM5</td>
<td>transmembrane protein 5</td>
<td>12q14.2</td>
<td>$3\times10^{-5}$</td>
</tr>
<tr>
<td>SLC20A2</td>
<td>Rab9 effector protein with kelch motifs</td>
<td>8p12-p11</td>
<td>$8\times10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>RABEPK</td>
<td></td>
<td></td>
<td>9q33.3</td>
<td>$9\times10^{-5}$</td>
</tr>
<tr>
<td>rs17792917</td>
<td>ULK4</td>
<td>unc-51-like kinase 4</td>
<td>3p22.1</td>
<td>$1\times10^{-4}$</td>
</tr>
<tr>
<td>DUSP4</td>
<td></td>
<td>dual specificity phosphatase 4</td>
<td>8p12-p11</td>
<td>$1\times10^{-4}$</td>
</tr>
<tr>
<td>CLCF1</td>
<td></td>
<td>cardiotrophin-like cytokine factor 1</td>
<td>11q13.3</td>
<td>$3\times10^{-5}$</td>
</tr>
<tr>
<td>CLYBL</td>
<td></td>
<td>citrate lyase beta like</td>
<td>13q32</td>
<td>$2\times10^{-5}$</td>
</tr>
<tr>
<td>TSC22D1</td>
<td>TSC22 domain family, member 1</td>
<td>13q14</td>
<td>$9\times10^{-6}$</td>
<td></td>
</tr>
<tr>
<td>TRAK1</td>
<td>trafficking protein, kinesin binding 1</td>
<td>3p22.1</td>
<td>$1\times10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>DNMBP</td>
<td></td>
<td>dynamin binding protein</td>
<td>10q24.2</td>
<td>$3\times10^{-5}$</td>
</tr>
<tr>
<td>TNFRSF19</td>
<td>tumor necrosis factor receptor superfamily, member 19</td>
<td>13q12.11-q12.3</td>
<td>$1\times10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>TBC1D8</td>
<td>TBC1 domain family, member 8</td>
<td>2q11.2</td>
<td>$5\times10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>MNX1</td>
<td>motor neuron and pancreas homeobox 1</td>
<td>7q36</td>
<td>$3\times10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>SLC43A2</td>
<td>solute carrier family 43, member 2</td>
<td>17p13.3</td>
<td>$8\times10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>GOLGA8A</td>
<td>golgin A8 family, member A</td>
<td>15q11.2</td>
<td>$7\times10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>GOLGA8B</td>
<td>golgin A8 family, member B</td>
<td>15q14</td>
<td>$7\times10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>PLCL2</td>
<td>phospholipase C-like</td>
<td>3p24.3</td>
<td>$5\times10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>rs11615037</td>
<td>CENTA1/ ADAP1 ArfGAP with dual PH domains 1</td>
<td>7p22.3</td>
<td>$1\times10^{-4}$</td>
<td></td>
</tr>
</tbody>
</table>
L. SUPPLEMENTAL FIGURES

Figure 4.10 Quantile-quantile plot of genome-wide association analysis results of severe chronic periodontitis among the 4610 white Dental ARIC participants.
Figure 4.11 Quantile-quantile plot of genome-wide association analysis results of moderate chronic periodontitis among the 4610 white Dental ARIC participants.
Figure 4.12 Quantile-quantile plot of genome-wide association analysis results of the extent of attachment loss trait (proportion of sites exhibiting ≥3 mm attachment loss) among the 4610 white Dental ARIC participants.
Figure 4.13 Visualization of the 3p21 locus that was marked by rs11925054 (P=6.5x10^{-7}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
Figure 4.14 Visualization of the 4q21.3 locus that was marked by rs17006135 (P=1.8x10^{-5} and P=4.4x10^{-6} after adjustment for smoking and diabetic status) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
Figure 4.15 Visualization of the 1p21.1 locus that was marked by rs10493998 (P=9.2x10^{-6} and P=2.2x10^{-6} after adjustment for smoking and diabetic status) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
Figure 4.16 Visualization of the 10p15 locus that was marked by rs12260727
(P=6.0x10^{-7}) for moderate chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
**Figure 4.17** Visualization of the 12q15 locus that was marked by rs11615037 (P=2.6x10\(^{-5}\) and P=4.4x10\(^{-6}\) after adjustment for smoking and diabetic status) for moderate chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
Figure 4.18 Visualization of the 11q22 locus that was marked by rs7120142 ($P=3.9 \times 10^{-6}$) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
Figure 4.19 Venn diagram representing genes adjacent to identified SNPs for the three chronic periodontitis traits. Genes with asterisk are associated with SNPs that emerged below the P<5x10^{-6} threshold in genetic models adjusted for smoking and diabetic status.
M. MANUSCRIPT #1- SUPPLEMENTAL MATERIAL REFERENCES


CHAPTER 5

A. TITLE AND AUTHORS

MANUSCRIPT #2. Genome-wide association study of periodontal pathogen colonization

B. MANUSCRIPT #2 SUMMARY

Periodontitis is characterized by an oral biofilm pathological shift that contributes to cascade of events leading to periodontal destruction. Host factors modulate the establishment of a dysbiotic oral microbiome, but to-date limited evidence exists with regard to associated genetic risk loci. We conducted a genome-wide association analysis (GWA) among 1020 white participants enrolled in the Atherosclerosis in Communities Study cohort, for whom quantification of eight periodontal pathogens was performed using a “checkerboard” DNA hybridization technique. We examined three traits: “high red” and “high orange” bacterial complex, and “high” Aggregatibacter actinomycetemcomitans (Aa) colonization, with “high” being defined as the top quintile of each trait distribution. Genotyping was performed using the Affymetrix 6.0 platform. Imputation to 2.5million markers was based on HapMap II-CEU and a multiple-test correction was applied (threshold of P<5x10^{-8}). We detected no genome-wide significant signals. However, thirteen loci including KCNK1, FBXO38, UHRF2, IL33, RUNX2, TRPS1, CAMTA1 and VAMP3 provided suggestive evidence (P<5x10^{-6}) of association with the measured traits. Further investigations providing replication of these findings and examination of specific associations may lead to an improved understanding of the complex nature of host-biofilm and -bacteria interactions that characterizes states of health and disease.
C. INTRODUCTION

Despite advances in the science and practice of dentistry, periodontal diseases continue to affect adult populations worldwide. The understanding of the pathogenesis of CP has evolved through advances in biology, biochemistry, microbiology, immunology and genetics. Study of the composition and complexity of the oral ecology has recently re-emerged as a focus of research, with investigations of the oral microbiome gaining increased attention (1). To date, over 450 species have been identified in the human microbiome, and its diversity is greater than initially theorized (2, 3). In fact, while harboring of periodontal pathogens is virtually universal, only a small only proportion of individuals develop severe forms of CP.

It is a common ground that an oral ecosystem in harmonious symbiosis with its host will likely be associated with health, whereas departure from this balance is characteristic of disease (4). Specific bacterial species that are implicated in CP have been identified, and in general are commensal and include gram-negative anaerobes. The degree or severity of the host response, which is a determinant of periodontal tissue destruction, has been found to be modulated by several local, systemic, and genetic factors. In this respect, the role of smoking and diabetes, as well as several single nucleotide polymorphisms (SNPs) in immune response-related genes such as interleukin (IL)-1, have been shown to alter the host response and impact the phenotype and clinical course of periodontitis (4,5).

There is a large body of literature in candidate-gene studies investigating the genetics of host inflammatory response, inflammatory mediators and cytokines (5). One recent study found a positive association between an IL-6 polymorphism and harboring of Aggregatibacter actinomycetemcomitans (Aa), but that report was based on a small sample of
forty patients (6). No study to our knowledge has carried out a whole-genome investigation for genetic markers of host colonization with periodontal bacteria. To this end, and to add to the knowledge base of the genetic component of periodontitis, the aim of this study was to investigate susceptibility loci for colonization with sub-gingival pathogenic periodontal bacteria using a GWA analysis approach.

D. MATERIALS AND METHODS

Detailed descriptions of the study population, genotyping and imputation, quality control and population stratification procedures are included in the online Appendix. In brief, we conducted a GWA study among 1020 white participants of the Atherosclerosis Risk In Communities (ARIC) longitudinal cohort investigation (7). While ARIC is a study of atherosclerosis, CVD risk factors and outcomes, a complete oral-dental examination took place between 1996 and 1998 during the fourth ARIC visit. As part of the Dental ARIC, apart from a complete clinical examination, sub-gingival microbial plaque samples were collected for a subset of participants (8).

The “checkerboard” DNA-DNA hybridization method (9) was used on plaque samples to measure the extent of sub-gingival colonization with eight periodontal pathogens: *Prevotella intermedia* [American Type Culture Collection (ATCC) 25611], *Campylobacter rectus* (ATCC 33238), *Fusobacterium nucleatum* (ATCC 10953), and *Prevotella nigrisens* (ATCC 33563) (belonging to the “orange” complex); *Porphyromonas gingivalis* (ATCC 33277), *Tannerella forsythia* (ATCC 43037) and *Treponema denticola* (ATCC 35404) (“red” complex); and *Aggregatibacter actinomycetemcomitans* (ATCC 43718). In that method, bacterial levels are expressed as counts relative to established microbial standards. Three dichotomous traits of “high” colonization with “red” complex, “orange” complex, and
Aggregatibacter actinomycetemcomitans were considered for analytical purposes. The two composite phenotypes were derived by the summation of bacterial count for each species belonging to the “red” (n=3) and “orange” (n=4) group, as described above. Because Porphyromonas gingivalis is considered the major periodontal pathogen implicated in periodontitis in adult populations, we explored for additional risk loci using its “high” colonization phenotype defined as above, as a separate trait, in exploratory analyses that we present in the Appendix. Various approaches in defining the bacterial colonization profiles have been previously used, including summations of the absolute microbial counts (10), tertile-categorization (11), and five-level categorization of log10-transformed counts (12). For the present investigation, we defined a “high” colonization trait as the top quintile (20%) of each trait’s distribution. The rationale for the selection of this phenotype is based on the fact that the “checkerboard” semi-quantitative method has a lower detection threshold of ~10^4 and reduced precision in the lower end of the distribution, whereas a individuals with “high” bacterial colonization profile may be those with reduced or impaired host and at high risk for periodontal tissue destruction.

Participants’ DNA was extracted from blood samples drawn from an antecubital vein, and genotyping was performed with the Affymetrix Genome-Wide Human SNP Array 6.0 chip. The platform offers 906,600 markers for SNPs. Following rigorous quality control procedures, imputation to 2.5million markers was performed using 669,450 SNPs and MACH v1.0.16 (http://www.sph.umich.edu/csg/abecasis/MaCH/index.html), based on HapMap Phase II CEU build 36. For analytical purposes, the dichotomous traits of “high colonization” phenotypes were entered in three logistic regression models assuming multiplicative (log-additive) allelic effects. The models included age, sex, examination center
and ten principal components from the population stratification analysis as covariates. A correction for multiple comparisons assuming 1 million independent tests was applied and a threshold of genome-wide statistical significance was set at $P<5\times10^{-8}$. An additional arbitrary threshold of $P<5\times10^{-6}$ was set to prioritize and investigate loci with suggestive evidence of association. All GWA analyses were performed with the ProbABEL software (13). SNP annotations were performed with WGAViewer ver.1.26l (14), Snipper ver.1.2 (http://csg.sph.umich.edu/boehnke/snipper/), and loci visualizations with LocusZoom ver.1.1 (15) and Haplovie ver.4.2 (16). Reporting of genes was based on the “HUGO Gene Nomenclature” naming convention (www.genenames.org).

E. RESULTS

The sample’s descriptive information is presented in Table 1. The participants had mean age of 63 years and approximately even gender distribution. “High” bacterial colonization profiles were found in all groups of periodontal diagnosis according to the CDC classification; however, the prevalence of “high red” colonization cases was more than double among participants with severe periodontitis compared to those with mild or no disease. Similar, but less pronounced associations were noted for “orange” complex and Aa. The bacterial counts of “high colonization cases” were [median (interquartile range); range] “red” complex bacteria—[64725 (108516); 21943-3894605], “orange” complex bacteria—[304155 (483373); 111926-1.5x10^7], and Aggregatibacter actinomycetemcomitans—[17309 (20034); 7520-410713].

Of the 2178777 examined SNPs, none had $P<5\times10^{-8}$. However, 53 had a $P<5\times10^{-6}$ and thus were prioritized for further investigation. Lambda inflation factors for three traits
were: “red”—1.040, “orange”—1.045 and $Aa$—1.032. The corresponding Q-Q plots are presented in Figure S1 in the Supplemental Material. Upon inspection of the prioritized SNPs, there were five loci that emerged for “red”, three loci for “orange” and five loci for $Aa$ colonization (Figure 1). Of those, one locus on 1q42 was shared for the “red” complex and $Aa$. Graphical representations of the genomic areas adjacent to six of these loci are presented in Figure 2.

The strongest signal with regard to “red” complex colonization (Table 1) was produced by rs11800854 in the 1q42 locus [$P=2.8\times10^{-7}$; $OR=12.3$ (95% CL=3.7, 41.3); MAF in HapMap-CEU (MAF-CEU): 0.067] in the promoter region (30Kb upstream) of $KCNK1$ and adjacent to $KIAA1804$. The common [G] allele showed 3% enrichment among “high colonization cases” for both “red” complex and $Aa$. Another locus in chromosome 1p22 was marked by rs12032672 ($P=9.6\times10^{-7}$), ~500Kb upstream of $PKN2$. Rs10043775, in LD with multiple markers in the 5q33 locus, represents a missense change in the $FBXO38$ gene (resulting in [Pro]→[Arg] substitution, predicted as ‘benign’ by PolyPhen-2) and provided the strongest signal in that locus ($P=2.4\times10^{-6}$), also adjacent to $HTR4$. A high LD area in 9p24 including the $UHRF2$, $GLDC$, $TPD52L3$ and $IL33$ genes is marked by rs16924631 [intronic to $UHRF2$; $P=3.2\times10^{-6}$ $OR=2.29$ (95% CL=1.61, 3.24); MAF-CEU: 0.275], of which the [C] risk allele showed almost 10% enrichment among “cases”. An intronic variant (rs10010758) of the $TBC1D1$ gene, adjacent (24Kb) to $PTTG2$ provided the strongest signal in the 4p14 locus [$P=3.7\times10^{-6}$; $OR=1.91$ (95% CL=1.25, 2.21)]. In the Appendix we present results of the exploratory analysis for $Pg$ “high” colonization, which revealed three loci with $P<5\times10^{-6}$ including $OTOF$, C2orf70, CIB4, DAB2IP, TLL11 and AKNRD3.
The common allele of rs1932040 showed 9% enrichment and provided the strongest association signal with “high orange” bacterial colonization [P=1.3x10^{-6}, OR=2.47 (95% CL=1.67, 3.65)], marking an intergenic area between RUNX2 and CLIC5 on the 6p21.1 locus. A low recombination area on 8q23, adjacent to TSPS1 (1.3Mb) and CSMD3 (672Kb) is marked by multiple alleles, of which rs9942773 provided the strongest signal (P=1.9x10^{-6}) and 10% enrichment among “cases”. A variant intronic to CAMTA1 [rs1616122; P=4.9x10^{-6}; OR=1.85 (95% CL=1.41, 2.42)] marks the 1p36.2 locus.

The 1q42 locus that was identified for “red” bacteria also provided the third strongest association signal for Aa [rs11800854; P=4.0x10^{-6}; OR=8.12 (95% CL=2.73, 24.11)]. The common [T] allele of rs11621969 was also nominally associated with high Aa colonization (P=9.4x10^{-7}) and was adjacent to FOS and JPD2 in 14q24. The rare [G] allele of rs1970525 was more than twice as prevalent among “cases” (0.118 versus 0.054 among “non-cases”), provided the strongest signal in the 10q23 locus [P=3.8x10^{-6}; OR=2.89 (95% CL=1.85, 4.52); MAF-CEU: 0.045] , and represents a nonsense-mediated decay transcript variant in the GRID1 gene. Rs9287989 is adjacent to KIAA1715 (30Kb) and EVX2 (227Kb) and marks the 2q31 locus (P=4.4x10^{-6}). An intronic variant of ODZ2 (rs6885116) provided the strongest signal in the locus 5q35 [P=1.4x10^{-6}; OR=2.57 (95% CL=1.76, 3.74); MAF-CEU: 0.084], showing 9% enrichment among “cases”. Rs1800795, that was found to be significantly associated with Aa colonization in the recent study of Nibali and colleagues (6) did not show any important association in this GWA scan [P=0.34, OR=1.12 (95% CL=0.88, 1.43)].
F. DISCUSSION

This study is the first report of a genome-wide association analysis investigating risk loci for colonization with pathogenic periodontal bacteria. Although limited by the sample size, this investigation explores a novel phenotype and benefits from a comprehensive quantitative phenotypical characterization. Upon replication or validation, these findings have the potential to unveil pathways and mechanisms that direct the host’s symbiosis with healthy microflora that if altered may predispose for states of disease. In fact, several of the risk loci identified in this study may offer promising leads for further exploration and mechanistic studies.

The 1q42 locus and rs11800854, which emerged as a common risk marker for both “red” complex and Aa colonization is in the promoter region of KCNK1, a gene that encodes a potassium channel protein and has been linked to cardiac outcomes (17). The prioritized SNPs at the 9p24 locus marked an area with low recombination rate that includes the UHRF2, TPD52L3 and IL33 genes. IL33, as other members of the interleukin family, has important roles in immunity and inflammation and has been suggested to function as an “alarmin”, alerting the immune system to endogenous trauma such as physical stress or infection (18).

RUNX2, a gene in the 6p21.1 locus that emerged due to the association of rs1932040 with “orange” complex colonization encodes a transcription factor that is essential for osteoblastic differentiation and skeletal morphogenesis. Mutations in this gene have been associated with the cleidocranial disorder syndrome, which has multiple and severe oral manifestations (19). Experimental evidence shows that RUNX2 is involved in the inhibition of MMP-13
expression, which appears to be involved in periodontitis, as well as osteoblastic and osteoclastic activity (20). Another locus that was identified as associated with “orange” colonization was 1p36.2 with the “top” SNP rs1616122 being intronic to the CAMTA1 gene. CAMTA1 has been reported as a candidate gene for type II diabetes risk (21). VAMP3 is adjacent (310Kb upstream) to the marked locus and may also be plausible gene candidate, as it has been shown to regulate podosome organization in macrophages, and thus mediate their adhesion, spreading and migration (22).

With regard to association results with Aa “high” colonization, the strongest signal was produced by the 14q24 locus, where the FOS and JDP2 genes are located. FOS was recently reported as part of a novel mechanism of RANKL expression in T cells (23), which may constitute an important link with immune responses, as FOS levels have been shown to increase by lippopolysaccharide infusion in vivo (24). JDP2 (85Kb downstream of the top SNP of the 14q24 locus) is a transcription factor that has been associated with the maintenance of Epstein-Barr virus latency (25), which has been implicated in the pathogenesis of chronic periodontitis via inhibition of oral bacteria-induced macrophage activation and phagocytosis (26).

Socransky (27) and other investigators have described how organisms such as Pg, Tannerella forsythia, Treponema denticola, Aa and other species found in dental plaque, organize themselves in complex communities collectively called “biofilm”, and interact with each other and with the host to result in different states of health and disease. While the presence of a highly organized biofilm appears a unique feature of dental plaque-induced diseases including caries and periodontitis, several other chronic conditions such as tuberculosis, rheumatic fever, syphilis, gastrointestinal ulcers and Lyme disease share a
bacterial colonization etiologic component (28). Consequently, an increased understanding of the genetic underpinning of interactions between the host and exogenous or symbiosing bacterial communities has the potential to advance the state of knowledge in periodontitis, but also other chronic inflammatory diseases (29).

The findings of the present investigation, if replicated, have the potential to add to the knowledge base of oral microbiome, host-biofilm, host-bacteria interactions, and more. We acknowledge that the study of periodontal pathogen colonization phenotypes is rare and this limits the replication options of our findings. However, the consideration of specific microbial factors as a distinct exposure in investigations of periodontal, oral and systemic health is consistent with the paradigm of “periodontal medicine” and may provide novel insight in the oral-systemic diseases connection. Although the prevention and treatment of periodontitis is an obvious goal, the links of CP with other systemic conditions and the “common theme” of pathogenic ecological shift in other diseases, provide opportunities for even greater impact.

MANUSCRIPT #2 REPORTING OF FUNDING SOURCES

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C; and by a grant from the National Institute of Environmental
Health Sciences (P30ES010126). The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The authors declare no potential conflicts of interest with regard to the authorship and publication of this article.
### G. TABLES

**Table 5.1** Distribution of subjects’ demographic characteristics and bacterial “high colonization” (defined as the top quintile of the distribution) profile (with “Red”, “Orange” complex, and *Aggregatibacter actinomycetemcomitans*) overall, and stratified by periodontal diagnosis (CDC/AAP classification) among the Dental ARIC study participants (n=1020)

<table>
<thead>
<tr>
<th></th>
<th>Total n (row %)</th>
<th>Healthy/mild n (column %)</th>
<th>Moderate n (column %)</th>
<th>Severe n (column %)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (column %)</td>
<td>416 (41)</td>
<td>415 (41)</td>
<td>189 (19)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>478 (47)</td>
<td>246 (59)</td>
<td>178 (43)</td>
<td>54 (29)</td>
</tr>
<tr>
<td>Males</td>
<td>542 (53)</td>
<td>170 (41)</td>
<td>237 (57)</td>
<td>135 (71)</td>
</tr>
<tr>
<td><strong>Age</strong> (years; mean, standard deviation)</td>
<td>63.2 (5.7)</td>
<td>62.3 (5.5)</td>
<td>63.6 (5.7)</td>
<td>64.0 (5.9)</td>
</tr>
<tr>
<td><strong>“High” bacterial colonization</strong> (n, % of column)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Red” complex</td>
<td>203 (20)</td>
<td>58 (14)</td>
<td>84 (20)</td>
<td>61 (32)</td>
</tr>
<tr>
<td>“Orange” complex</td>
<td>201 (20)</td>
<td>73 (18)</td>
<td>72 (17)</td>
<td>56 (30)</td>
</tr>
<tr>
<td>A. actinomycetemcomitans</td>
<td>204 (20)</td>
<td>75 (18)</td>
<td>81 (20)</td>
<td>48 (25)</td>
</tr>
</tbody>
</table>
Table 5.2 Genome-wide association analysis results of the high colonization traits (highest quintile of the distribution versus the other four; quantified with DNA-DNA “checkerboard” hybridization) for “Red”, “Orange” complex and Aggregatibacter actinomycetemcomitans, among the white participants of the Dental ARIC study (n=1020). Single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF-HapMap II CEU) of ≥5% and associated P<5x10^-6. The SNP with the lowest P-value per locus is presented; additional prioritized SNPs in each locus are presented in the footnote, along with corresponding R^2 (based on 1000 genomes pilot 1 release) with the top SNPs.

<table>
<thead>
<tr>
<th>Chr.</th>
<th>SNP Position</th>
<th>MAF (HapMap II-CEU)</th>
<th>Closest gene(s) and position or distance</th>
<th>Risk allele frequency low/ high colonization</th>
<th>P value</th>
<th>Odds ratio (95% CL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q42</td>
<td>rs11800854</td>
<td>231786607</td>
<td>A G [A] 0.068</td>
<td>KCNK1 (30Kb)</td>
<td>[G] 0.947/0.978</td>
<td>2.8x10^-7</td>
</tr>
<tr>
<td>1p22</td>
<td>rs12032672</td>
<td>88398224</td>
<td>A C [C] 0.350</td>
<td>PKN2 (524Kb)</td>
<td>[C] 0.332/0.446</td>
<td>9.6x10^-7</td>
</tr>
<tr>
<td>5q33</td>
<td>rs10043775</td>
<td>147785313</td>
<td>C T [C] 0.274</td>
<td>FBXO38 (missense change)</td>
<td>[T] 0.703/0.791</td>
<td>2.4x10^-6</td>
</tr>
<tr>
<td>5p24</td>
<td>rs16924631</td>
<td>6476308</td>
<td>C G [C] 0.142</td>
<td>HTR4</td>
<td>[C] 0.122/0.219</td>
<td>3.2x10^-6</td>
</tr>
<tr>
<td>4p14</td>
<td>rs10010758</td>
<td>37614913</td>
<td>C T [C] 0.275</td>
<td>UHRF2 (non-coding transcript variant)</td>
<td>[C] 0.291/0.384</td>
<td>3.7x10^-6</td>
</tr>
</tbody>
</table>

| 6p21.1 | rs19320400 | 45804766 | A G [A] 0.142 | CLIC5 (169Kb)/ RUNX2 (178Kb) | [G] 0.808/0.896 | 1.3x10^-6 | 2.47 (1.67, 3.65) |
| 8q23 | rs9942773 | 115190203 | A C [C] 0.283 | CSMD3 (672Kb)/ TRPS1 (1.3Mb) | [A] 0.703/0.803 | 1.9x10^-6 | 2.07 (1.51, 2.82) |
| 1p36.2 | rs1616122 | 7444172 | C T [C] 0.482 | CAMTA1 (intron variant)/ VAMP3 (310Kb) | [T] 0.506/0.624 | 4.9x10^-6 | 1.85 (1.41, 2.42) |

**“Orange” complex**

| 14q24 | rs11621969 | 74883781 | C T [C] 0.167 | FOS (65Kb)/ JDP2 (85Kb) | [T] 0.789/0.885 | 9.4x10^-7 | 2.46 (1.68, 3.62) |
| 10q23 | rs1970525 | 87624904 | C G [G] 0.045 | GRID1 (NMD12 transcript variant) | [G] 0.054/0.118 | 3.8x10^-6 | 2.89 (1.85, 4.52) |
| 1q42 | rs11800854 | 231786607 | A G [A] 0.067 | KCNK1 (30Kb)/ KIAA1904 (199Kb) | [G] 0.947/0.978 | 4.0x10^-6 | 8.12 (2.73, 24.11) |
| 2q31 | rs9287989 | 176425987 | C T [T] 0.433 | KIAA1715 (73Kb)/ EVX2 (227Kb)/ EXTLP2 (310Kb) | [C] 0.484/0.605 | 4.4x10^-6 | 1.80 (1.39, 2.33) |
| 1q24 | rs11800854 | 231786607 | A G [A] 0.067 | KCNK1 (30Kb)/ KIAA1904 (199Kb) | [G] 0.947/0.978 | 4.0x10^-6 | 8.12 (2.73, 24.11) |
| 5q35 | rs6885116 | 167576123 | A G [G] 0.084 | WWCI (76Kb) | [G] 0.078/0.169 | 1.4x10^-6 | 2.57 (1.76, 3.74) |

1: coded allele
2: non-coded allele
3: Additional SNPs in locus with P<5x10^-6: rs6682365 (R^2=1.00)
4: Additional SNP in locus with P<5x10^-6: rs10068216 (R^2=1.00), rs10072051 (R^2=1.00), rs10044061 (R^2=1.00), rs349707 (R^2=1.00), rs10477376 (R^2=1.00), rs9325095 (R^2=1.00), rs10041283 (R^2=1.00), rs9325097 (R^2=0.87), rs3734120 (R^2=1.00), rs4574333 (R^2=0.87), rs4274967 (R^2=1.00), rs4274968 (R^2=1.00), rs6884076 (R^2=1.00), rs9325098 (R^2=1.00)
5: Additional SNP in locus with P<5x10^-6: rs11795355 (R^2=1.00), rs7860000 (R^2=1.00), rs10975603 (R^2=0.93), rs1692462 (R^2=0.93), rs16924624 (R^2=1.00), rs10975605 (R^2=0.93), rs10115883 (R^2=0.93), rs10122116 (R^2=0.93)
6: T>G – Ser>Pro, 35b from the exon boundary
7: R^2=0.29 with rs6811863 which is a missense change in PTTG2: G>C – [Arg]>[Pro]
8: Additional SNPs in locus with P<5x10^-6: rs12525547 (R^2=0.93), rs9349326 (R^2=0.93), rs16873698 (R^2=0.93)
9: Additional SNP in locus with P<5x10^-6: rs10089040 (R^2=1.00), rs9942776 (R^2=1.00), rs10086149 (R^2=1.00), rs7845243 (R^2=0.87), rs10105817 (R^2=1.00), rs7006291 (R^2=1.00), rs1779159 (R^2=1.00), rs11783996 (R^2=1.00), rs10098056 (R^2=1.00), rs7018200 (R^2=0.92)
10: Additional SNPs in locus with P<5x10^-6: rs4325261 (R^2=1.00)
11: Additional SNP in locus with P<5x10^-6: rs6682365 (R^2=1.00)
12: nonsense-mediated decay
13: EXTLP2 is a pseudogene
H. FIGURES

Figure 5.1 Manhattan plot of the GWAS results (-log10 P-values of the ~2.5 million examined SNPs arranged by chromosome) for “high red” complex bacterial colonization among the 1020 white participants of the Dental ARIC Study cohort.
Figure 5.2 Manhattan plot of the GWAS results (-log10 P-values of the ~2.5 million examined SNPs arranged by chromosome) for “high orange” complex bacterial colonization among the 1020 white participants of the Dental ARIC Study cohort.
**Figure 5.3** Manhattan plot of the GWAS results (−log10 P-values of the ~2.5 million examined SNPs arranged by chromosome) for “high” *Aggregatibacter actinomycetemcomitans* bacterial colonization among the 1020 white participants of the Dental ARIC Study cohort.
Figure 5.4 Visualization of the 1p22 locus that was marked by rs12032672 (P=9.6x10^{-7}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
**Figure 5.5** Visualization of the 5q33 locus that was marked by rs10043775 ($P=2.4 \times 10^{-6}$) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
**Figure 5.6** Visualization of the 9p24 locus that was marked by rs16924631 ($P=3.2 \times 10^{-6}$) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
**Figure 5.7** Visualization of the 4p14 locus that was marked by rs10010758 (P=3.7x10^{-6}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
**Figure 5.8** Visualization of the 1p36.2 locus that was marked by rs1616122 (P=4.9x10^{-6}) for “high orange” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
Figure 5.9 Visualization of the 1q42 locus that was marked by rs11800854 (P=4.0x10^-6) for “high” Aggregatibacter actinomycetemcomitans colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
I. REFERENCES


Supplemental Materials and Methods

Study Population

The study sample consisted of participants of the Dental ARIC, an ancillary study of the Atherosclerosis Risk In Communities (ARIC) longitudinal cohort investigation (1). ARIC is a study of atherosclerosis, CVD risk factors and outcomes. The Dental ARIC, an ancillary study was undertaken between 1996 and 1998 during the ARIC visit 4. For the purposes of Dental ARIC a complete oral-dental examination was performed among 6979 subjects, whereas collection of gingival crevicular fluid and sub-gingival microbial plaque samples was undertaken for a subset of participants (2). Specifically, of the 6979 Dental ARIC participants, 1450 had microbial plaque samples collected (2-4). For the purposes of the present GWA study, we considered white subjects who had provided microbial plaque samples (n=1236) and were successfully matched with genotype data (n=1020). Additional information that was collected as part of ARIC visit 4 and was used for data presentation and analysis included the following variables: sex, age, CP diagnosis [Centers for Disease Control (CDC) and American Academy of Periodontology (AAP) three-level classification: health/mild, moderate, and severe disease] (5), smoking status (never, former, current), and DM (healthy or DM).

Genotyping and imputation

In the study population, DNA was extracted from blood samples drawn from an antecubital vein into tubes containing serum separator gel. Blood samples were analyzed at a central ARIC laboratory in Houston, TX. Genotyping was performed with the Affymetrix
Genome-Wide Human SNP Array 6.0 chip. The platform offers 906600 markers for SNPs. The rigorous quality control procedures included initial blind duplicate genotyping and identification/flagging of SNPs with kappa<0.95 and reconciliation of unintentional duplicate samples (17 duplicates and one triplicate). Imputation to 2.5 million markers was performed using 669450 SNPs and MACH v1.0.16 (http://www.sph.umich.edu/csg/abecasis/MaCH/index.html), based on HapMap Phase II CEU build 36. The selected SNPs were selected from 839048 autosomal SNPs restricted to those with minor allele frequency (MAF)>0.01 (129543 excluded), Hardy-Weinberg equilibrium (HWE) P>10^{-5} (12432 excluded) and call rate >95% (1693 excluded). After imputation, SNPs with a quality score <0.8 and missing data rate >10% were identified and flagged for removal, while only those with MAF of ≥5% were considered for analytical purposes.

**Population stratification**

To obtain estimates of relatedness and population stratification a subset of 85,947 “high quality” SNPs was selected. These SNPs met the following criteria among self-reported whites: MAF ≥0.1, call rate >99.5%, HWE P≥10^{-3}, autosomal, with annotation in the platform annotation file, not labeled “AFFX” or “chromosome 0”, and not monomorphic. Using these SNPs identity-by-state (IBS) allele sharing distance (DST values) were computed using PLINK, as such: DST = IBS distance (IBS^2 + 0.5*IBS_1) / (n SNP pairs). First degree relative status was assigned to pairs of individuals with DST≥0.8 and second degree relatives were considered those with 0.763 ≤DST< 0.8. Among the Caucasian ARIC participants there were 380 pairs of first degree and 207 pairs of second degree relatives identified. To minimize exclusions, related pairs were broken by iterative selection of individuals with most relatives using a custom program.
Population stratification was further evaluated with principal component (PC) analysis using the EIGENSTRAT program (6). The above chosen set of SNPs was used for the computation of ten principal components. Genetic outliers were considered those that were further than 8 standard deviations (SD) away from any of ten PCs over ten runs of PC computation. Based on DST and PC criteria there were 716 subjects flagged from removal from the analysis (206 as genetic outliers based on PCs and 16 based on average DST values (“too little IBS sharing” with the rest of the sample), 351 first degree relatives and 143 second degree relatives. All but ten second degree relatives (whose relatives were excluded as genetic outliers) were re-entered in the dataset and were assigned PCs. After exclusion of 364 individuals (4%) there were 9349 Caucasians who were included in the GWA analysis and of those, 1020 had periodontal microbiological data available as Dental ARIC participants.

SUPPLEMENTAL RESULTS REGARDING *PORPHYROMONAS GINGIVALIS*

We detected no genome-wide significant association signals with regard to *Pg* “high” colonization. However, six SNPs marking three loci emerged below the $P<5\times10^{-6}$ threshold and were prioritized for further investigation. Of those, rs10760187 in 9q33.2 provided the lowest $P$-value [$P=4.5\times10^{-7}$; OR=$2.07$ ($95\%$ CL=$1.55$, $2.76$)], with the ‘risk’ [C] allele (HapMap II CEU MAF [T]: 0.514) showing 12% enrichment among subjects with “high” colonization. This SNP is in an intergenic area between *DAB2IP* (18Kb) and *TTLL11* (18Kb) and in LD with several variants intronic to both genes. Rs1011108 [$P=2.0\times10^{-6}$; OR=$1.79$ ($95\%$ CL=$1.40$, 2.27); MAF-CEU: [T] 0.374] is in a gene-dense locus at 2p23.3, 1Kb from *OTOF*, 3Kb from *C2Orf70* and 18Kb from *CIB4*. In fact, the second prioritized SNP in this locus (rs1275992) is intronic to *CIB4* and in linkage disequilibrium ($R^2=0.58$, $D’=0.94$) with
rs13002673 which represents a missense change in \textit{C2Orf70} resulting to a [Gln]\(\rightarrow\)[His] substitution, predicted as ‘benign’ by PolyPhen-2. The third locus that was associated with \textit{Pg} “high” colonization was marked by rs1360573 at 10p11.21 [\(P=1.5\times10^{-6}\); OR=2.75 (95% CL=1.78, 4.26)]. The common “risk” allele [A] showed a 10% enrichment, and the closest gene is \textit{ANKRD30A} (529Kb).

\textbf{SUPPLEMENTAL DISCUSSION REGARDING \textit{PORPHYROMONAS GINGIVALIS}}

The examination of \textit{Pg} colonization as a distinct trait is supported by the major role that this pathogen has in chronic periodontitis (7, 8). These exploratory analysis results, above and beyond those reported for “red” complex bacteria, can be considered as additional candidate regions that may be implicated in increased colonization with \textit{Pg} and thus, higher risk for chronic periodontitis.

With regard to genes in the 2p23.3 locus, \textit{CIB4 (KIP4)} is known as calcium and integrin binding family member 4. Mutations in the \textit{OTOF} (otoferlin) gene have been associated with non-syndromic types of deafness (9). No information is available on the role of \textit{C2Orf70}. Interestingly, the 9q33.3 locus that was marked by rs10760187 in our GWAS was also identified as associated with abdominal aortic aneurysm in a Northern European GWAS (10). This genomic area has also been reported to harbor important susceptibility variants associated with cleft lip/palate (11). One candidate gene in the 9q33.3 locus, \textit{DAB2IP}, is also known as DAB2-interacting protein or \textit{AIP1}. This gene is member of a family of regulators of extracellular stimuli that serve to maintain the homeostasis of cellular functions (12) and has been shown to be inactivated by methylation in prostate and breast cancer (13). \textit{TTLL11}, another candidate gene in this locus, encodes a member of a large
family of proteins involved in the catalytic ligation of amino acids to tubulins or other substrates (14). Finally, the closest gene to the top SNP in the 21q22.3 locus was ANKRD30A, or ankyrin repeat domain-containing protein 30A. Reports have linked this gene and its protein (NY-BR-1) as a differentiation marker of the mammary gland, and thus as a potential diagnostic and immunotherapeutic aid for breast carcinomas (15, 16) with cancer. Moreover, a recent study found an association of a 3’UTR SNP of ANKRD30A with the human serum metabolic profile (17).
**K. SUPPLEMENTAL TABLE**

**Table 5.3 Porphyromonas gingivalis “high” colonization (quantified with DNA “checkerboard” hybridization) GWA results among the white participants of the Dental ARIC study (n=1020).** Single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF-HapMap II CEU) of ≥5% and P<5x10^{-6}. The SNP with the lowest P-value per locus is presented; additional prioritized SNPs in each locus are presented in the footnote, along with corresponding R^2 (based on 1000 genomes pilot 1 release) with the top SNPs.

<table>
<thead>
<tr>
<th>Chr.</th>
<th>SNP</th>
<th>Position</th>
<th>ca^1</th>
<th>nca^2</th>
<th>Risk allele frequency (HapMap II-CEU)</th>
<th>Closest gene(s) and position or distance</th>
<th>Risk allele frequency low/high colonization</th>
<th>P value</th>
<th>Odds ratio (95% CL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2p23.3</td>
<td>rs1011108^4</td>
<td>26636125</td>
<td>C</td>
<td>T</td>
<td>[T] 0.374</td>
<td>OTOF (1Kb); C2orf70 (3Kb); CIB4 (21Kb)</td>
<td>0.341/0.485</td>
<td>2.0x10^{-6}</td>
<td>1.79 (1.40, 2.27)</td>
</tr>
<tr>
<td>9q33.2</td>
<td>rs10760187^5</td>
<td>123605641</td>
<td>C</td>
<td>T</td>
<td>[C] 0.514</td>
<td>DAB2IP (18Kb); TTLL11 (18Kb)</td>
<td>0.498/0.618</td>
<td>4.5x10^{-7}</td>
<td>2.07 (1.55, 2.76)</td>
</tr>
<tr>
<td>10p11.21</td>
<td>rs1360573</td>
<td>36925927</td>
<td>A</td>
<td>G</td>
<td>[A] 0.759</td>
<td>ANKRD30A (529 Kb); FZD8 (956 Kb)</td>
<td>0.752/0.853</td>
<td>1.5x10^{-6}</td>
<td>2.75 (1.78, 4.26)</td>
</tr>
</tbody>
</table>

1: coded allele
2: non-coded allele
3: Confidence limits
4: Additional SNP in locus with P<5x10^{-6}: rs1275992 (R^2=0.33), intronic to CIB4 and in linkage disequilibrium (R^2=0.58, D’0.94) with rs13002673 which represents a missense change in C2orf70 resulting to a [Gln]\[His] substitution
5: Additional SNPs in locus with P<5x10^{-6}: rs7849478 (R^2=0.63), rs10985387 (R^2=0.67)
Table 5.4 Reported SNPs as expression quantitative loci (eQTLs) and their association with gene expression (lymphoblastoid cell line-based expression) in populations of European descent with $P < 10^{-4}$ in the Scan database (http://www.scandb.org).

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Locus</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10043775</td>
<td>KIAA0515</td>
<td>proline-rich coiled-coil 2B</td>
<td>9q34.13</td>
<td>$6 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>(PRRC2B)</td>
<td>coiled-coil domain containing 57</td>
<td>17q25.3</td>
<td>$5 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs1932040</td>
<td>MRPL15</td>
<td>mitochondrial ribosomal protein L15</td>
<td>8q11.2-13</td>
<td>$9 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>EAF1</td>
<td>ELL associated factor 1</td>
<td>3p25.1</td>
<td>$3 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>TAF5L</td>
<td>TAF5-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor, 65kDa</td>
<td>1q42.11-42.3</td>
<td>$8 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs16924631</td>
<td>MED15</td>
<td>mediator complex subunit 15</td>
<td>22q11.2</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>PSMB7</td>
<td>proteasome (prosome, macropain) subunit, beta type, 7</td>
<td>9q34.11-34.12</td>
<td>$2 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>MVD</td>
<td>mevalonate (diphospho) decarboxylase</td>
<td>16q24.3</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>MAN2A2</td>
<td>mannosidase, alpha, class 2A, member 2</td>
<td>15q25</td>
<td>$9 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>DNMBP</td>
<td>dynamin binding protein</td>
<td>10q24.3</td>
<td>$6 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>PQLC1</td>
<td>PQ loop repeat containing 1</td>
<td>20q13.1</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>PRKCE</td>
<td>protein kinase C, epsilon</td>
<td>2p21</td>
<td>$3 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>NEK6</td>
<td>NIMA (never in mitosis gene a)-related kinase 6</td>
<td>3p26</td>
<td>$2 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>PREX1</td>
<td>phosphatidylinositol-3,4,5-trisphosphate-dependent Rac exchange factor 1</td>
<td>3p26.2-22</td>
<td>$4 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs1011108</td>
<td>TMEM2</td>
<td>transmembrane protein 2</td>
<td>9q34.13</td>
<td>$3 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>LRRK1</td>
<td>leucine-rich repeat kinase 1</td>
<td>15q26.3</td>
<td>$10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>A4GALT</td>
<td>alpha 1,4-galactosyltransferase</td>
<td>22q13.2</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>MTTP</td>
<td>microsomal triglyceride transfer protein</td>
<td>4q24</td>
<td>$5 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>MYO6</td>
<td>myosin VI</td>
<td>6q14.1</td>
<td>$2 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>ASPH</td>
<td>aspartate beta-hydroxylase</td>
<td>8q12.1</td>
<td>$6 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>CNKSR3</td>
<td>CNKSR family member 3</td>
<td>6q25.2</td>
<td>$9 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>F13A1</td>
<td>coagulation factor XIII, A1 polypeptide</td>
<td>6p24.2-26</td>
<td>$4 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>RASSF6</td>
<td>Ras association (RalGDS/AF-6) domain family member 6</td>
<td>4q21.1</td>
<td>$5 \times 10^{-6}$</td>
</tr>
<tr>
<td>LOC100129069</td>
<td>LOC100129069</td>
<td>hypothetical protein LOC100129069</td>
<td>11q24.2</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>HNF4G</td>
<td>hepatocyte nuclear factor 4, gamma</td>
<td>8q21-22</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>MLSTD1</td>
<td>fatty acyl CoA reductase 2</td>
<td>2p11.23</td>
<td>$10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>CHL1</td>
<td>cell adhesion molecule with homology to L1CAM (close homolog of L1)</td>
<td>3p26</td>
<td>$6 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>PIP3-E</td>
<td>interaction protein for cytohesin exchange</td>
<td>6q25.2</td>
<td>$10^{-5}$</td>
<td></td>
</tr>
</tbody>
</table>

137
<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Chromosome</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNF1B</td>
<td>HNF1 homeobox B</td>
<td>17q12</td>
<td>3x10^{-5}</td>
</tr>
<tr>
<td>SERPINB1</td>
<td>serpin peptidase inhibitor, clade B (ovalbumin), member 1</td>
<td>6p25</td>
<td>7x10^{-5}</td>
</tr>
<tr>
<td>ESAM</td>
<td>ribosomal protein L6 pseudogene</td>
<td>12q21.3</td>
<td>10^{-4}</td>
</tr>
<tr>
<td>IL17RB</td>
<td>interleukin 17 receptor B</td>
<td>3p21.1</td>
<td>10^{-4}</td>
</tr>
<tr>
<td>CCDC64</td>
<td>coiled-coil domain containing 64</td>
<td>12q24.2</td>
<td>10^{-4}</td>
</tr>
<tr>
<td>C14orf105</td>
<td>chromosome 14 open reading frame 105</td>
<td>14q22.2</td>
<td>2x10^{-5}</td>
</tr>
<tr>
<td>C1orf161</td>
<td>mab-21-like 3</td>
<td>1p13.1</td>
<td>2x10^{-5}</td>
</tr>
<tr>
<td>MMAA</td>
<td>methylmalonic aciduria (cobalamin deficiency) cblA type</td>
<td>4q31.1</td>
<td>5x10^{-5}</td>
</tr>
<tr>
<td>C10orf137</td>
<td>chromosome 10 open reading frame 137</td>
<td>10q26.2</td>
<td>6x10^{-5}</td>
</tr>
<tr>
<td>ZNF503</td>
<td>zinc finger protein 503</td>
<td>10q22.3</td>
<td>5x10^{-5}</td>
</tr>
<tr>
<td>ZNF44</td>
<td>zinc finger protein 44</td>
<td>19p13.2</td>
<td>10^{-4}</td>
</tr>
<tr>
<td>CXorf21</td>
<td>chromosome X open reading frame 21</td>
<td>Xp21.3</td>
<td>6x10^{-5}</td>
</tr>
<tr>
<td>PIR</td>
<td>pirin (iron-binding nuclear protein)</td>
<td>Xp22.31</td>
<td>2x10^{-5}</td>
</tr>
<tr>
<td>MGC13057</td>
<td>chromosome 2 open reading frame 88</td>
<td>2q32.2</td>
<td>4x10^{-5}</td>
</tr>
<tr>
<td>SERPINB9</td>
<td>serpin peptidase inhibitor, clade B (ovalbumin), member 9</td>
<td>6p25</td>
<td>2x10^{-5}</td>
</tr>
<tr>
<td>CCDC4</td>
<td>BEN domain containing 4</td>
<td>4p13</td>
<td>3x10^{-5}</td>
</tr>
<tr>
<td>SYNE2</td>
<td>spectrin repeat containing, nuclear envelope 2</td>
<td>14q22.1-q22.3</td>
<td>8x10^{-5}</td>
</tr>
<tr>
<td>BHLHB5</td>
<td>basic helix-loop-helix family, member e22</td>
<td>8q12.1</td>
<td>10^{-4}</td>
</tr>
<tr>
<td>LRRK1</td>
<td>leucine-rich repeat kinase 1</td>
<td>15q26.3</td>
<td>4x10^{-6}</td>
</tr>
<tr>
<td>GTF3C6</td>
<td>general transcription factor IIIC, polypeptide 6, alpha 35kDa</td>
<td>6q21</td>
<td>7x10^{-5}</td>
</tr>
<tr>
<td>PLA1A</td>
<td>phospholipase A1 member A</td>
<td>3q13.13-q13.2</td>
<td>2x10^{-5}</td>
</tr>
<tr>
<td>WDR91</td>
<td>WD repeat domain 91</td>
<td>7q33</td>
<td>6x10^{-5}</td>
</tr>
<tr>
<td>LIN7A</td>
<td>lin-7 homolog A</td>
<td>12q21</td>
<td>9x10^{-6}</td>
</tr>
<tr>
<td>MAGEF1</td>
<td>melanoma antigen family F, 1</td>
<td>3q13</td>
<td>10^{-4}</td>
</tr>
</tbody>
</table>
SUPPLEMENTAL FIGURES

Figure 5.10 Quantile-quantile plot of genome-wide association analysis results of “high red” complex bacterial colonization, among the 1020 Dental ARIC participants.
Figure 5.11 Quantile-quantile plot of genome-wide association analysis results of “high orange” complex bacterial colonization, among the 1020 Dental ARIC participants.
Figure 5.12 Quantile-quantile plot of genome-wide association analysis results of “high” Aggregatibacter actinomycetemcomitans colonization, among the 1020 Dental ARIC participants.
Figure 5.13 Visualization of the 2p23.3 locus that was marked by rs1011108 (P=2.0x10^{-6}) for “high” Porphyromonas gingivalis colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
**Figure 5.14** Visualization of the 9q33.2 locus that was marked by rs10760187 (P=4.5x10^{-7}) for “high” *Porphyromonas gingivalis* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
**Figure 5.15** Visualization of the 10p11.21 locus that was marked by rs1360573 ($P=1.5 \times 10^{-6}$) for “high” *Porphyromonas gingivalis* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort. The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
SUPPLEMENTAL REFERENCES


A. Summary of findings

In our genome-wide analysis of chronic periodontitis (CP) in a well-defined cohort of approximately 4600 white subjects we detected no genome-wide significant signals. However, we found suggestive evidence of association (P<5x10^{-6}) for CP with markers at ten loci including NPY, NIN, and WNT5A for severe CP, NCR2 and EMR1 for moderate CP, and TBX18, ETS1, DYN2H1, TTC26 and ZC3HAV1 for the “extent” of attachment loss (EAL) trait. Several of these loci include candidate genes that are related to immune function and regulation, as well as neurological function. Noteworthy, four loci marked by rs3826782, rs11925054, rs17792917, rs11615037 have been identified as expression quantitative trait loci (eQTL) on lymphoblastoid cell lines (LCL). Sensitivity analyses including smoking and diabetes-adjusted genetic models did not show any material impact on these results. However, in these adjusted analyses five additional loci including SCD5, RPS29, PPIL5, COL11A1, CPM and CPSF6 emerged below the P<5x10^{-6} threshold. The 4q21.3 (SCD5) locus was shared for severe CP and EAL, and the 6p21.1 (NCR2) locus reached genome-wide significance for moderate CP (rs7762544; P_{adjusted}= 3.8x10^{-8}). The NPY (7p15) locus was replicated for severe CP in an independent sample of whites of European descent. We performed inverse-variance weighted meta-analysis for three loci that
revealed no evidence of effect heterogeneity between the two studies. Pooled odds ratios for these three loci were: 7p15 (NPY, rs2521634)—OR=1.49 (95% CI=1.28-1.73), P=3.5x10^{-7}; 19p13.3 (EMR1, rs3826782)—OR=2.01 (95% CI=1.52-2.65), P=8.2x10^{-7}; 6p21.1 (NCR2, rs7762544)—OR=1.40 (95% CI=1.24-1.59), P=7.5x10^{-8}.

Our exploratory analyses of SNP interactions with sex, smoking and diabetes detected several interaction effects below an empirical P<10^{-5} threshold: 20 SNPs with sex, 17 with diabetic status, and 12 with the three-level smoking categorization (never, former, current).

In the GWA scan in a subset of approximately 1000 subjects for which periodontal microbiological data were available we detected associations with markers at thirteen loci including KCNK1, FBXO38, UHRF2, IL33, RUNX2, TRPS1, CAMTA1 and VAMP3 providing suggestive signals of association (P<5x10^{-6}) with “high” bacterial colonization with “red” complex, “orange” complex, and Aggregatibacter actinomycetemcomitans (Aa). Exploratory analyses revealed three additional loci including OTOF, C2Orf70, CIB4, DAB2IP, TTLL11 and AKNRD3 associated with Porphyromonas gingivalis (Pg) “high” colonization. Four of these loci marked by rs10043775, rs1932040, rs16924631 and rs1011108 were identified to be eQTLs on LCL. Two missense changes that were identified for “high” bacterial colonization traits (rs10043775, 5q33, FBXO38 gene [Pro]→[Arg] and rs13002673 (in LD with the prioritized rs1011108), 2p23.3, C2Orf70 gene [Gln]→[His]) were predicted to be ‘benign’.

These genome-wide scan results provide new information on multiple candidate regions for interrogation in genetic studies of CP. Further investigations
providing replication of these findings and examination of specific associations may lead to an improved understanding of the pathogenesis of CP as well as the complex nature of host-biofilm and bacteria interactions that characterizes states of health and disease.

B. Strengths

1. Genome-wide association analysis approach

The opportunity that GWAS provide for investigators to examine the association of currently more than a million DNA markers with a given phenotype without being limited by prior knowledge, known pathways or mechanistic hypotheses in their marker selection, is unique. It has been argued that most associations detected by GWAS represent new discoveries. This “unbiased” or “agnostic” (1) approach in examining the genetic component of various phenotypes facilitates identification of novel loci and thus, potential insights into new pathways or mechanisms of biological processes. The present study is the first report of a GWA analysis of CP to-date, and provides suggestive evidence of association for several genomic loci that may be promising leads for further genetic investigations.

2. Phenotype characterization

The value and quality of the Dental ARIC clinical examinations protocol is truly unique. It is known that clinical-periodontal examination data are accurate and reproducible when they are performed by trained and calibrated dental examiners, as was the case with Dental ARIC study. Recent reports (2) stress that national estimates
of CP prevalence may severely underestimate the true prevalence due to disease misclassification in the NHANES III and IV. The next NHANES oral-dental examinations will be performed for the first time following the “full-mouth six site per tooth” protocol used in the Dental ARIC study. As outlined in the introduction, numerous definitions and quantitative measures exist to capture the extent and severity of periodontitis (3), with each been subject to varying degrees of bias (measurement error, bias due to loss of severely affected teeth, and more). The use of the most widely accepted and consensus CDC/AAP classification of CP, in combination with a clinical examinations protocol of highest standard is a major strength of the proposed investigation. We considered a continuous and non-reversible measure of periodontal destruction as an additional strength.

The investigation of microbiological colonization phenotypes in addition to CP is offering additional insights into the factors facilitating or antagonizing the cascade of pathological oral ecological shifts that are characteristic of the disease. The oral microbial load of Gram-negative and anaerobic periodontal pathogens may be an independent risk factor for, or modifier of several systemic conditions (4-7), and its exploration adds value to this study. While periodontitis is a clinical manifestation of exaggerated immune response to predominantly commensal oral bacteria, investigating genetic factors associated with colonization and proliferation of these causative bacterial agents is a fundamental question.
3. Covariates and sensitivity analysis

Because population stratification ("systematic" differences in ancestry between trait categories) can result in spurious GWA results, appropriate adjustment for population substructure was undertaken in all analyses, using 10 EIGENSTRAT-derived principal components. Inspection of the quantile-quantile plots did not reveal any "early departure" of the observed versus expected P-values. Additionally, genomic inflation (lambda) factors for most analyses were generally low, below 1.030. Our analyses also benefitted from the detailed measurement and consideration in sensitivity analyses of the well-documented risk factors or disease modifiers, smoking and diabetic status. Although no confounding is expected in the association between SNPs and the examined traits, the consideration of smoking history in the sensitivity analysis provides the opportunity to rule out the identification of risk loci for this risk factor, which has been shown to confer epigenetic changes (8, 9). Moreover, the inclusion of the diabetic status as a covariate in the sensitivity analysis is based on the fact that diabetes is associated with altered host response (10, 11) and increased susceptibility to bacterial infections (12). A sensitive definition of diabetes was used in the ARIC study, including fasting and non-fasting serum glucose levels, and pharmacological treatment for the disease.

C. Limitations

1. Phenotype characterization

As noted previously, the use of the CDC/AAP definition of CP represents the most widely accepted standard for classifying the disease. While our clinical data and
case-definition used for CP are of the highest standards possible, the disease itself entails some unique methodological challenges. Although periodontitis is a defined as a disease at the “person- or oral cavity-level”, it is actually characterized by multiple-attacks on the periodontal attachment of individual teeth (13). As it progresses, periodontitis leads to gradual periodontal tissue destruction, alveolar bone loss, and eventually tooth loss. For example, in a cross-sectional study of periodontal status in relation to age and tooth type it was found that incisor teeth were the most frequently affected by advanced disease, whereas molar teeth were most frequently missing (14). In “periodontal maintenance” studies, it has similarly been found that molar teeth are the most frequently extracted due to periodontal reasons (15). This constitutes a form of “censoring” of the most affected “diseased units”, the teeth. In this respect, CP can be a self-limiting disease on a tooth-level. This is also the case on a patient-level, as (considering the extreme scenario) an edentulous individual is not at risk to develop periodontitis. A consequence of this phenomenon is the observation that the prevalence, as well as extent and severity of CP in epidemiologic studies are likely biased downwards from tooth loss attributable to the disease. It has been recognized that variation in the criteria and case definitions used for the classification of CP has been a source of inconsistencies in the dental literature (16). Also, as outlined in the introduction, more quantitative measures exist to capture the extent and severity of periodontitis compared to “crude” classifications, such as the “extent” of attachment loss, which was employed in this study. It will be worthwhile for future studies to consider juxtaposing the present investigation’s findings to additional sensitivity
analyses using extent scores (e.g. $\geq$3mm CAL) as continuous variables for the GWA analyses, while adjusting for tooth loss.

Another consideration is that periodontitis, when measured by clinical measures of PD and (mainly) CAL, essentially represents the cumulative history of periodontal tissue destruction and not necessarily active disease or infection. Capturing this cumulative impact on periodontal tissues may be beneficial when considering longitudinal effects or progression patterns of periodontitis that may be biologically admixed. While more biologically-relevant disease classifications have been developed, our independent consideration of the association of host genetic variants with the microbial biofilm composition represents an improvement over the study of CP diagnoses alone. Despite being limited by small sample size for a GWAS (approximately 1000 individuals with microbial level counts), we consider this approach as the first exploration of loci associated with the composition of the oral microbiome. However, we acknowledge that this discovery step is limited by the very low power to detect modest effects. Similarly, our GWAS was underpowered to detect interaction effects in this additional exploratory step.

2. Genome-wide analysis approach

Whole-genome sequencing for population-based studies is financially and resource-prohibitive with today’s state of technology. For this purpose, imputation methods for non-typed markers have been developed, and imputation has become a standard step in GWA studies (17). Imputing does introduce some uncertainty in the non-typed allele frequencies, but publicly available programs for GWA analyses such as
probABEL (18) allow the use of “dosage files” including imputed SNPs. As denser mapping of the human genome is realized, opportunities to use a 1000 genomes-based (19) imputation of the ARIC genotyped data (up to 10 million markers) will be an important improvement to the current approach.

A frequent criticism of the GWAS concerns the current state of very conservative criteria of (genome-wide) statistical significance that are necessary to reduce the false-positive rate (1). The traditional Bonferroni correction is believed to be ultra-conservative due to the fact that SNPs are to some degree correlated, as the ~2.5 million association tests performed in imputed datasets are not truly independent from each other. This stringency may allow only for relatively strong associations to be discovered, while much “good signal” obtained by SNPs with modest or small effects goes without attention. This is compounded by the fact that many common complex traits are thought to be influenced by a large numbers of low-penetrance, small effect variants. Park et al. recently suggested that if these large numbers of low-penetrance variants is identified from sufficiently powered GWAS, together they could explain 20% or more of the known heritability of such traits (20). On the other hand, another concern with GWAS findings is related to the “winner’s curse” (21) which becomes evident when identified loci are not being replicated in external samples or are not reproduced in subsequent studies. For these reasons, the interpretation of GWAS findings is now proposed to become subject to different criteria (compared to “traditional” epidemiologic evidence) to assess their inferential potential. Based on these proposed “interim guidelines” (22), genome-wide evidence should be weighed in terms of amount of evidence, extent of replication and
protection from bias and thus classified in a three-level categorization as strong, moderate or weak evidence.

Moreover, it is becoming increasingly evident that the variation detected by GWAS is only one dimension of the genetic basis of health and disease (23). For instance, the role of copy number polymorphisms (24), rare alleles (25), epistatic effects (26) and that of the non-coding regions (27, 28) in the human genome are thought to be parts of the missing piece of the “missing heritability” (25). For example, Dickson and colleagues suggested that “synthetic genome-wide associations” can be detected from combinations of rare variants (29). In a recent review, Panagiotou and colleagues summarized the published evidence of genome-wide association for rare (<5%) variants and found that there are still few genome-wide significant rare variants, but suggested that a possible confluence of rare and common variants in the same loci is likely to be depended on the different ancestry references used (30). Bansal and colleagues, in a recent review provided recommendations with regard to analytical approaches that could potentially aid in the discovery of “rare variant” effects (31).

Moreover, it is common ground that biological processes are best represented by “cycles” and the action of biological messengers is understood on “pathways”. Therefore, it may be overly simplistic to consider independent effects of SNPs when it is recognized that usually multi-molecular pathways, feedback loops and regulating mechanisms are functioning in tandem. For this reason, any opportunity to jointly evaluate the effect of multiple “functionally relevant” SNPs is appealing and is becoming more feasible with advances in GWA methods (32, 33).
Further, the study of epigenetic phenomena has opened a whole new chapter in research of the pathogenesis and therapeutics, by a joint consideration of genotype and environmental influences (34) that is not routinely considered in GWAS. Feinberg recently suggested that the consideration of epigenomics, the genome-scale study of epigenetics in the function of our genomic make-up will likely be equally important to our understanding of the relation between function and organ anatomy in humans (35).

To place this in perspective of studies of CP, it has been shown that bacteria involved in periodontal infection are capable of inducing such epigenetic DNA alterations (36).

3. Potential biases

As explained in detail in the two previous sections, the phenotype characterization and limitations inherent to GWAS are two major potential sources of bias for this investigation. With regard to the CP phenotype, it is possible that the CDC classification is not sensitive or quantitative enough to capture the entire spectrum of disease expression. For example, using the CDC trait may be an inappropriate phenotype if one considers CP equally likely to develop due to a hyper-inflammatory host reaction or an immune-compromised state. Most likely, there are different genetic influences for these two types of host-response profile. To address this limitation, the consideration of biologically-refined classifications of CP such as the BGI may be indicated.

Another phenotype-related source of bias is the definition of the “high” bacterial colonization traits for our Specific Aim 2. It is known that the semi-quantitative method of DNA-DNA “checkerboard” hybridization has a lower count
detection threshold of $10^3$ to $10^4$. On the other hand, experts agree that CP risk is “frankly elevated” at high count levels, in the range of $10^5$ to $10^6$. In our study we arbitrarily defined “high colonization cases” those who were in the highest quintile of the distribution for each bacterial trait. Using this approach we were able to retain an “adequate” proportion of 20\% of subjects as cases, representing a “pragmatic” approach that allowed an analytically feasible contrast- defining a higher threshold would result in fewer “high colonization cases” and perhaps reduced power to detect GWA signals. Future studies should, however, explore the feasibility of different definitions of the “high colonization” traits. Moreover, although our summation of different bacteria into the “red” and “orange” complex groups is based solid biological and microbiological ground, it could still represent an over-simplification, because different micro-organisms possess different pathogenic and host interaction profiles. Therefore, it will be worthwhile for investigators to examine individual pathogen colonization profiles using a GWA approach.

With regard to the GWAS methodology, in brief, apart from our modest sample size for this type of investigation, there are three inherent biases. First, it is very likely that many “true association” signals are below the stringent threshold set for genome-wide significance or suggestive evidence of association. Second, due the low power of the study, it is also likely that the identified associations are overestimates of the true effect sizes of these loci. Third, our analysis of only common variants restricts our inference to common-only polymorphisms and ignores possibly rare (<5\%) but large effect size SNPs. To address these biases future studies
need to be conducted replicating our findings, investigate for novel ones, and explore the role of rare variants.

Tooth loss due to CP is also another important source of bias. An example would be a patient with severe CP that lost all his teeth and in this investigation was classified as edentulous and was excluded from the dental ARIC analytical sample; or another patient who had his “most severely” affected teeth extracted few weeks before the clinical examination, and was subsequently diagnosed with moderate or no CP instead of severe CP. An analytical approach quantifying this source of bias would include sensitivity analyses that either adjust the genetic models for tooth loss, or impute ‘likely CP diagnoses’ based on certain patterns of tooth loss that are found among CP patients.

D. Potential impact and public health significance

1. Genetic discovery and applications

The significance of any new discovery in health research is measured by its clinical relevance, potential for prevention and public health context (37). The unraveling of the genetic component of human traits and diseases has been challenged in terms of its potential for immediate public health benefits (38). The contrary can be argued, for instance, for pharmaceutical research, were new drugs or vaccines may have immediate and dramatic global public health benefits (39). In this respect, the central role of epidemiology as the “quarterback” of steering translational research into application and public health practice has been increasingly emphasized (40). Other authors stress the fact that still not enough attention or efforts are invested in
translating novel findings in a manner that can have meaningful impacts in public health (41). The findings of this study have the potential to offer insights into the pathogenesis of CP, providing possible therapeutic and preventive avenues, but also benefit the study of other pathological conditions that have a host-microbiome interaction or immunological underpinning.

Despite the accelerated rates of gene discovery for common complex diseases, valid applications of GWAS findings in personalized health care are not yet evident (42, 43). Influences by the emerging “personalized medicine” market may subject individuals to unnecessary costs or even health risks (44); in this regard, the clinicians’ and investigators’ role in providing accurate and evidence-based counseling about the potential and the limitations of genetic testing, is critical. However, regardless of immediate translational potential, discovery of gene variants associated with health or disease can lead to the discovery of previously unknown biological mechanisms and pathways. Intervening on the genetic makeup is not applicable to-date; however, development of drugs that could interact on these pathways is an important opportunity treatment and even prevention (vaccine development) of disease.

2. Therapeutic avenues for periodontitis

Despite the great strides in our understanding of the biology of periodontitis during the last decades, the therapeutic means available for CP have been long unchanged. Efforts in the prevention of CP and periodontal diseases in general have a strong behavioral component (45, 46), because meticulous oral hygiene, oral self-care
and dental attendance are the cornerstones of oral disease prevention. Periodontal therapy however, is needed for individuals who develop the disease, so that the supra- and sub-gingival microbial load can be reduced, the inflammatory response controlled, and the patient can maintain a periodontal health status via preventive efforts and a maintenance-recall protocol. Specifically, the maintenance of a microbial colonization profile that is host-compatible has been proposed as a “microbial treatment goal” for periodontal therapy (47)

The use of mechanical means for the microbial load reduction (conventional or surgical periodontal therapy based on root scaling and planning) remains unchallenged for the treatment of periodontitis (48), while adjunctive antibiotic therapy is becoming more frequent in severe or refractory cases of periodontitis (49, 50). The patterns of colonization by and proliferation of periodontal pathogens has been suggested to have a genetic (host) component (51), and this question was indeed interrogated in our study. Beyond that, the current stage of knowledge allows clinicians and investigators to expect that response to periodontal treatment varies greatly between individuals who harbor the same pathogenic microflora (52). The fact that non-uniform and often unsustainable results are to be expected by CP patients after treatment is a frequent observation (53), which has invited the exploration of individual susceptibility or treatment resistance biological markers.

In medicine, the microbiological component and etiology of disease is well-established and common sense. In dentistry, in spite of advances in oral microbiology and the increasing popularity of antimicrobial applications, the unexplored potential is immense (49). This is supported by a substantial variation in quantitative and
qualitative aspects of periodontal ecology between patients (54), and recent findings that link subgingival bacterial colonization profiles with gingival tissue gene expression (55). The incorporation of oral microbiology applications for the prevention and therapy of periodontitis, consistent with the “periodontal medicine” model, is likely to increase in the future (56-58).

The introduction of antimicrobials (59-61) or anti-inflammatory drugs (61, 62) has showed varying but generally promising results. These reports indicate that combinations of conventional and pharmaceutical treatment regimes are likely to produce the most predictable treatment outcomes in selected cases. However, it would be optimal if individual characteristics of host response could be incorporated in diagnosis and therapy to optimize treatment responses. For example, identification from the present GWA of susceptibility loci and potential functional genetic variants can be used to determine whether adjunctive antimicrobial or anti-inflammatory therapy would be indicated for biological subtypes of periodontitis (63).

Improvements in prevention of periodontitis would have significant and multi-level implications in the oral health and related quality of life of a substantial proportion of the adult population, and may be more important as demographic effects increase the life expectancy and mean age worldwide (64). Prevention of CP will contribute to the reduction of edentulousness rates (65), which constitutes a frequent form of functional and psychosocial impairment. Such advances in prevention could indeed stem from a better understanding of the disease pathogenesis, progression, and treatment or from the development of new preventive strategies, such as the envisioned vaccine against periodontitis (66). Moreover, it will reduce the
disease-related somatic and psychological distress and co-morbidities, as well as the financial and time loss impact to systems and individuals, due to its treatment or sequelae. Additionally, prevention of periodontitis has substantial public health implications when its links with systemic health are considered (48). One major and direct benefit would be, for instance, improved glycaemic control among patients with type 2 diabetes (67).

E. Replication of GWA results and relevance of identified loci

As noted above, replication of results of genome-wide association studies has virtually become a standard requirement for reporting (22, 38). Although successful replication of GWA finding provides some level of additional “confidence” that the results do not represent spurious associations, non-replication does not always imply a “false alarm”. The “winner’s curse” phenomenon, representing novel findings’ failure to replicate in subsequent studies, as well as other methodological features of GWA studies such as stringent significance thresholds may be responsible for overestimation of the true underlying genetic effects (68). For this reason, GWAS are considered essentially “discovery attempts” (69, 70) and their results are subject to very conservative statistical criteria, as well as replication in independent samples.

1. Replication of prioritized SNPs of chronic periodontitis

The Health ABC dental cohort

Replication of the current study’s findings was sought via collaboration with investigators of the Health Aging and Body Composition (ABC) study. The Health
ABC is a prospective study of well-functioning adults of ages 70-79 at baseline. White participants were recruited from a random sample of Medicare beneficiaries in the Pittsburgh, Pennsylvania, and Memphis, Tennessee. Eligibility criteria included age between 70 and 79, self-reported no difficulty of walking one-quarter mile or climbing 10 steps without resting, no difficulty performing basic activities of daily living, no reported need of an assistive walking device, and no active treatment for cancer (71). As part of the study year 2 and 3 follow-up clinical visits (1998-2000), a total of 1,133 white and African American participants were eligible and received complete dental and periodontal examinations. About two-thirds of this ancillary dental study’s participants were white. Among whites, 21% of men and 14% of women were diabetic. With regard to smoking history, proportions of current/former smokers among men and women were 5%/61%, and 5%/32%, respectively. Replication was based on this subset of approximately 750 Health ABC participants (~400 men and ~350 women) that were subsequently genotyped using the Illumina Human1M Duo platform.

**Replication of the dental ARIC GWA findings**

Replication was sought for our prioritized SNPs based on the genome-wide association P<5x10^-6 threshold in ‘crude’ or ‘adjusted’ analyses. Fourteen SNPs met this threshold and therefore, effect estimates and P-values based derived from the Health ABC cohort GWAS were obtained. For these SNPs, the same coded and non-coded alleles were genotyped in both GWAS (Table A15). These estimates are presented in Tables A16 and A17. A same direction effect and a conventional P<0.05
threshold were set as for evidence of replication. Moreover, we took advantage of the opportunity to perform a meta-analysis of the reported effect estimates, as an additional approach in summarizing the effect of the reported risk loci (72). Meta-analytical pooled odds ratios and P-values were obtained using the METAL program (73), using an inverse-variance weighting method (74). Summary estimates (pooled odds ratio) were obtained for SNPs that did not depart substantially from effect homogeneity between the two cohorts (defined as homogeneity $X^2 \ P<0.2$).

The summary of the replication findings is presented in the Appendix, in Tables A16-A18. Noteworthy, only one locus associated with severe CP ($7p15$, rs2521634) met the $P<0.05$ and same effect direction replication criteria. Additionally, one locus associated with moderate CP ($19p13.3$, rs3826782) neared replication ($P=0.06$) and another ($6p21.1$, rs7762544) showed a same direction effect with no evidence of heterogeneity. Therefore, summary estimates based on inverse-variance weighted meta-analysis are presented in Table A18. Pooled odds ratios and associated summary P-values for these three loci were: $7p15$ (rs2521634)—$\text{OR}=1.49$ (95% CI=1.28-1.73), $P=3.5\times10^{-7}$; $19p13.3$ (rs3826782)—$\text{OR}=2.01$ (95% CI=1.52-2.65), $P=8.2\times10^{-7}$; $6p21.1$ (rs7762544)—$\text{OR}=1.40$ (95% CI=1.24-1.59), $P=7.5\times10^{-8}$.

Interestingly, one of these SNPs prioritized for severe CP (rs2521634) provided a replication signal also for moderate CP in Health ABC, whereas one moderate CP-associated SNP in ARIC (rs3826782) provided a replication signal for severe CP in Health ABC. It is possible that these risk loci indeed confer risk for both types of CP. However, this observation warrants more investigation to clarify whether systematic differences between the two study samples may be responsible for this difference (i.e.
tooth loss and retention, or periodontal treatment levels that may result to different phenotypical expression). An alternative explanation may be that these polymorphisms actually confer protection against periodontal pathology, and therefore may be associated with “periodontal health” when studying the two CP traits.

**Relevance of identified and replicated loci**

Presentation and discussion of all prioritized loci for the CP, EAL and “high” bacterial colonization traits are presented in the discussion sections of and supplemental material of manuscripts 1 and 2 (Sections 4.F, 4.J, 5.F and 5.J). This section offers a more detailed presentation of one replicated locus for severe CP, at 7p15.

**The NPY (7p15) locus**

This locus was marked by rs2521634, which was the top SNP in the region with \( P=1.6 \times 10^{-6} \) and located 47Kb downstream of \( NPY \). The latter is a gene encoding a neuropeptide that is widely expressed in the central nervous system and has been suggested to function as an anxiolytic peptide that helps explain inter-individual variation in trait anxiety and resiliency to stress (75). Several reports have associated a Leu→Pro substitution in the \( NPY \) gene with significantly elevated serum triglycerides and lipids among Northern European populations (76, 77). The SNP responsible for this missense change (rs16139, predicted as ‘benign’ by PolyPhen-2)
was 10Kb upstream of rs2521634 and not in linkage disequilibrium with it (D’=1.00; r²=0.01).

An early immunohistological Swedish study among by Barr-Agholme and colleagues observed higher neuropeptide Y levels and sensory hyper-innervation in the gingival tissue of Down syndrome (DS) patients, attributing this finding to the gingival inflammatory reaction rather that DS itself (78). A 2004 critical review by Lundy and Linden (79) emphasized the role of the nervous system in the pathophysiology of peripheral inflammation, suggesting a neurogenic inflammatory component for periodontitis. These authors suggested that neurogenic inflammation is in fact a protective mechanism, with NPY having a pivotal role in the cascade of relevant chemical activities.

A recent case-control study by Lundy and colleagues (80) provided experimental evidence consistent with this hypothesis. Elevated NPY levels were found in the gingival crevicular fluid of periodontitis compared to healthy sites. Additional evidence supports a role of the sympathetic nervous system and NPY in the modulation of periodontal inflammation and bone resorption (81, 82). Interestingly, inhibition of nerve growth factor (NGF), which is known to up-regulate neuropeptides in sensory neurons, via systemic anti-NGF was recently shown to be associated with reduced Interleukin 1beta and bone resorption in a rat periodontitis model (83). In sum, combined with our GWAS results, there is accumulating evidence pointing towards the role of the nervous system and neuropeptide messengers (with NPY being an obvious candidate one) across the axis of a “neurological inflammatory reflex” in the cascade of events that lead to periodontitis.
F. Future research directions

1. Improvement over candidate-gene approaches

Candidate-gene studies have been limited in their ability to identify risk loci or genes for common complex diseases. In general, candidate-gene approaches have had better results when studying “highly penetrant” diseases, where a “disease-gene” may exist. Successful candidate-gene studies are typically based on strong prior knowledge or mechanistic theory, as is the case of inflammatory mediator gene markers and cytokines in the context of inflammatory diseases, such as ulcerative colitis (84) and CP (85). In the case of CP in particular, such studies have focused on inflammatory mediator molecules and receptors such as the interleukin gene cluster. Despite the obvious relevance and replication of this finding across many candidate-gene studies, it should be expected that multiple genes are implicated in the pathogenesis of CP, each conferring likely small to modest excess risk. Although previously reported candidate-gene study targets were not marked by strong signals in our GWAS we explored several loci that contain such candidate genes, including $IL1A$, $IL1RN$, $IL6$, $TLR4$, $MMP2$, $GLT6D1$, $FCGR3A$, and $VDR$ with visualizations that are presented in Figures A30-A33. Moreover, we did identify few loci that were contained genes related to previous candidate-gene reports. For example, the 9p24 locus was identified for “high red” complex bacterial colonization and included $IL33$. Additionally, one locus on chromosome 19 that was identified for severe CP contained a tumor necrosis family ($TNFSF14$) and a complement ($C3$) gene, which are well-documented immune- and periodontitis-related genes (86-90). One of the
prioritized loci on chromosome 6 (rs17792917) for the “extent” of attachment loss trait has been reported as an eQTL for another TNF-family gene on chromosome 13 (TNFRSF19), which has been reported as a susceptibility locus for lung cancer (91) and vascular dementia (92). Finally, the risk locus on 2p23.3 that emerged for Pg “high” colonization is an eQTL for multiple genes including IL17RB on chromosome 3. We anticipate that future candidate-gene studies will embark in testing and potentially replicating our GWA findings.

2. Beyond genome-wide association analyses

The analytical methods of GWAS data are constantly evolving, and may allow a future extension of the present analysis via implementation of novel methodologies. For example, despite the initial “agnostic” scan, there is an opportunity to prioritize candidate SNPs for analysis in the GWA context, when for instance epistatic effects are hypothesized or suspected (32, 93). Other examples of novel but not well-established approaches include attempts to link prior knowledge (of validated polymorphisms) with analytical strategies (94), use of pathway-clustering analysis of genetic signals (95), efficient approaches to explore gene-gene and gene-environment interactions (96, 97), and the incorporation of expression quantitative trait loci (eQTL) with GWAS results (98, 99).

It is becoming increasingly understood that major advances in our understanding of the genetic underpinning of human health and disease can be realized by efficient comprehensive, efficient and inexpensive interrogations of genomes, transcriptomes and interactomes (100). Next-generation sequencing studies (101-103) have recently
emerged and are promising in this respect, with exome sequencing investigations providing a high rate of novel findings (104, 105). A particular strength of whole-exome sequencing studies is that they offer opportunities for both disease gene discovery as well as clinical diagnosis, as protein coding genes harbor the vast majority of large-effect disease-related mutations (106).

Finally, several additional explorations of the data presented in this dissertation may offer novel insights in the study of CP. The consideration of gene expression modulation by proximal (cis) or distal (trans) eQTLs, as noted above (98, 99) can provide information that may be functionally relevant. For example, one of the three replicated loci, 19p13.3 (marked by rs3826782) has been reported as an eQTL for GPR113 (G-protein coupled receptor 113) on chromosome 2p23.3, which has functions on the neuropeptide signaling pathway as well as signal transducer activity; pathways relevant to the other prioritized locus on 7p15, where neuropeptide Y (NPY) is located. Moreover, GPR113 has been found to be particularly expressed intra-orally, in the taste receptor cells (107).

Moreover, examining larger sets of prioritized loci in ‘pathway analysis’ networks may reveal patterns of association that are otherwise difficult to detect by ‘single-locus’ explorations. Examples of such exploratory visualizations of functional and physical interactions between genes that were identified from our analyses are presented in the Appendix Figures A28 and A29. Although these networks arguably contain a lot of “noise”, empirical inspection of represented functional pathways and physical interactions can be informative. In this network of interactions (figure A28), some of these well-represented pathways were collagen, protein phosphatase 2A
binding, extracellular matrix, positive regulation of transcription from RNA polymerase II promoter, epithelial to mesenchymal transition, platelet activation, Wnt receptor signaling pathway, cellular response to growth factor stimulus, regulation of fibroblast proliferation and positive regulation of kinase activity.

3. Future studies

As explained in the previous sections, the present study can be considered as a ‘fist look’ into the genetic component of chronic periodontitis using a genome-wide association approach. Within the limitations of the analyses presented here, this work offers new information about candidate loci associated with CP. There is a large number of follow-up research and different avenues that can and should be followed after the conduct of a GWAS (108, 109). Future studies in this field are needed to replicate and validate these associations, discover new ones, elucidate their functional role and mechanistic pathways, determine their potential for diagnostic and therapeutic applications, and more.

To be more specific, future GWAS examining the three-level CP classification are warranted to replicate the associations with loci that our study reported. Although GWAS of bacterial colonization profiles are harder to realize, replication of the high periodontal pathogen colonization findings will add to our understanding of the contribution of genetic factors to host-microbiome interactions. Because GWAS are generally under-powered to detect most small to modest genetic effects, future studies of equal size may discover new loci that did not meet the suggestive association threshold in our investigation. However, larger GWAS or
pooling of existing cohorts may be necessary to enable the detection of small effects and GxE or GxG interactions. Factors such as sex, smoking and diabetes should be strong candidates as ‘environmental exposures’ for these interaction analyses. We emphasize the need and importance of conducting GWAS in populations other than whites of European descent, as evidence of the genetic component of CP among populations of African, Asian and American descent are lacking. Such studies in diverse populations can provide insights about loci that were in fact identified in our study population of whites, or discover novel ones.

Identifying the causal variant(s) in the loci that we report is also an important step that future studies should address (110). From our GWAS results it cannot be inferred that the SNPs that were identified tagging the prioritized loci are the causal variants. It is possible that non-genotyped and non-imputed, perhaps rare (<5%) variants are the risk-conferring ones. Fine-mapping represents one strategy that can be implemented in future studies seeking to identify the risk variants and causal alleles in these or other candidate loci for CP. However, some opportunities to interrogate the contribution of rare variants will emerge with the imputation of our existing genotype data using the ‘1000 genomes’ haplotypes, a strategy that will offer approximately 10 million markers for GWAS, including a substantial proportion of rare ones. It is not unrealistic to anticipate that these strategies will eventually be superseded by whole genome sequencing, which will become progressively less costly and time-consuming in the next decade.

Subsequent experimental or association studies linking GWAS data with gene expression information can also provide information about the functional relevance
and candidate mechanistic pathways that the prioritized genes operate on. Experimental studies “knocking-out” the new candidate genes in animal models will be needed to clarify the role of GWAS discoveries. The use of microarrays to interrogate gene expression can also provide tissue-specific insights (111). As more data become available with regard to gene expression regulation by quantitative trait loci on multiple tissues, future studies may be able to annotate our findings to such cis- or trans-acting eQTLs (112).

Future studies examining traits other than the three-level CDC disease taxonomy that we employed may offer additional insights into the genetic component of CP. Disease indicators or “proxies” are less valid in the ascertainment of cases, but they may be cost-effective and more feasible than gold-standard clinical examinations. It may be possible for indicators such as number of remaining teeth, radiographically measured alveolar bone loss, and self-reported diagnosis of CP to be obtained for population-based samples or existing cohorts with genotype data, thus enabling more GWAS of CP.

Furthermore, genome-wide interrogations of endophenotypes, which represent physiologic indicators, biochemical assays, or other “intermediate” traits relevant to the disease under study (113, 114) may offer additional opportunities to examine the genetic influences for CP. These traits have the benefit of being ‘biologically proximal’ to the hypothesized genetic effects, whereas the phenotype expression of complex and polygenic diseases is likely to be the result of many genetic and non-genetic, measured, un-measured and unobservable factors. The bacterial colonization profile that we examined as part of our Specific Aim 2 can be considered as an
endophenotype of CP. Others may include in vivo quantified inflammatory markers such as gingival crevicular fluid Interleukin levels, serum levels of antibodies to periodontal pathogens, markers of oxidative stress, and more.

Extending the concept of the study of intermediate CP phenotypes, it can be argued that the identified or the still undiscovered risk loci for CP confer excess risk for conditions other than CP. Certain molecular functions and pathways may be common in the pathogenesis of a number of diseases. This is indicative of pleiotropy, which is the phenomenon of a single gene influencing multiple traits (115). Future studies can investigate common risk loci for CP and other, frequently associated conditions, such as diabetes and cardiovascular disease. Moreover, simultaneous examinations of a wide array of phenotypes for a given genotype, in the context of phenome-wide (116, 117) association studies (PheGWAS) are efficient strategies that are gaining popularity and may include CP in the future.

Finally, incorporating the current GWAS findings into epidemiologic and clinical investigations will be an important “research translation” step. One can envisage how novel candidate genes can be considered in investigations of CP risk or treatment response. Genes or promoter regions in the candidate loci that we report, such as NPY, could be evaluated as CP risk indicators or prognostic markers CP treatment response, alone or in combination with genetic markers that are already being tested as such (i.e. the interleukin family gene cluster). The identification of a group of genetic markers that have a good combined positive predictive value could in the future be used in the design of genetic tests that would inform clinicians and patients of their CP risk or their likelihood to respond favorably to various therapeutic
options. Such investigations could evaluate for example whether CP patients who are carriers of certain genotype of \textit{NIN, EMR1, NPY, FBXO38, TBC1D1, CAMTA1} or \textit{GRID1} have better therapeutic results when treated with conventional-surgical, antimicrobial, or anti-inflammatory periodontal treatment, or combinations of these.
References


80. Lundy FT, El Karim IA, Linden GJ. Neuropeptide Y (NPY) and NPY Y1 receptor in periodontal health and disease. Arch Oral Biol 2009;54:258-62.


116. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the
feasibility of a phenome-wide scan to discover gene-disease associations.
Bioinformatics 2010;26:1205-10.

wide association studies (PheWAS) for exploration of novel genotype-phenotype
## APPENDIX

**Table A1.** Distribution in the total sample and bivariate associations of periodontal diagnosis (CDC classification) and periodontal attachment loss (extent score: percent of sites with $\geq 3$mm attachment loss) with brushing and flossing frequency, dental attendance reason and frequency, and education level among the Caucasian participants of the Dental ARIC study, for whom clinical and genetic data were matched after exclusions (n=4,610)

<table>
<thead>
<tr>
<th>Total sample</th>
<th>CDC Periodontitis classification</th>
<th>Attachment loss ($\geq 3$mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (n, column %)</td>
<td>Health-mild (n, row %)</td>
</tr>
<tr>
<td></td>
<td>4610 (100)</td>
<td>1864 (40)</td>
</tr>
<tr>
<td><strong>Brushing frequency</strong>&lt;sup&gt;1,2&lt;/sup&gt; (yesterday)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One time or none</td>
<td>1350 (29)</td>
<td>482 (36)</td>
</tr>
<tr>
<td>Two or more</td>
<td>3239 (71)</td>
<td>1374 (42)</td>
</tr>
<tr>
<td><strong>Flossed last week</strong>&lt;sup&gt;1,2&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did not floss</td>
<td>1464 (32)</td>
<td>506 (35)</td>
</tr>
<tr>
<td>Flossed one or more times</td>
<td>3124 (68)</td>
<td>1349 (43)</td>
</tr>
<tr>
<td><strong>Dental visit reason</strong>&lt;sup&gt;1,2&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular visits</td>
<td>3809 (83)</td>
<td>1609 (42)</td>
</tr>
<tr>
<td>Only when problem</td>
<td>780 (17)</td>
<td>247 (32)</td>
</tr>
<tr>
<td><strong>Last dental visit</strong>&lt;sup&gt;1,3&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6 months ago</td>
<td>3174 (69)</td>
<td>1312 (41)</td>
</tr>
<tr>
<td>6-24 months</td>
<td>1031 (22)</td>
<td>427 (41)</td>
</tr>
<tr>
<td>More than 2 years ago</td>
<td>381 (8)</td>
<td>117 (31)</td>
</tr>
<tr>
<td><strong>Education level</strong>&lt;sup&gt;1,3&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic (0-11 years)</td>
<td>433 (9)</td>
<td>136 (31)</td>
</tr>
<tr>
<td>Intermediate (12-15 years)</td>
<td>2102 (56)</td>
<td>838 (40)</td>
</tr>
<tr>
<td>Advanced (16+ years)</td>
<td>2069 (45)</td>
<td>886 (43)</td>
</tr>
</tbody>
</table>

*third molars were included; †rounded to the closest integer; ‡Chi-square test of equivalence between strata of periodontitis P<0.05; §Median test of stratum-specific attachment loss estimates P<0.05; ¶Kruskal-Wallis test of stratum-specific attachment loss or covariate estimates P<0.05
**Table A2.** Distribution in the total sample and bivariate associations of periodontal diagnosis (CDC classification) and periodontal attachment loss (extent score: percent of sites with ≥3mm attachment loss) with quintiles of age and BMI among the Caucasian participants of the Dental ARIC study, for whom clinical and genetic data were matched after exclusions (n=4,610)

<table>
<thead>
<tr>
<th>CDC Periodontitis classification</th>
<th>Attachment loss (≥3mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (n; median)</td>
</tr>
<tr>
<td><strong>Total sample</strong></td>
<td>4610</td>
</tr>
<tr>
<td><strong>Age&lt;sup&gt;1,2&lt;/sup&gt; (years, quintiles; mean)</strong></td>
<td></td>
</tr>
<tr>
<td>Q1 (55.5)</td>
<td>1054; 56</td>
</tr>
<tr>
<td>Q2 (59.5)</td>
<td>1026; 60</td>
</tr>
<tr>
<td>Q3 (63.0)</td>
<td>724; 63</td>
</tr>
<tr>
<td>Q4 (66.5)</td>
<td>890; 66</td>
</tr>
<tr>
<td>Q5 (70.9)</td>
<td>916; 71</td>
</tr>
<tr>
<td><strong>BMI&lt;sup&gt;1,2&lt;/sup&gt; (kg/cm&lt;sup&gt;2&lt;/sup&gt;, quintiles; mean)</strong></td>
<td></td>
</tr>
<tr>
<td>Q1 (22.1)</td>
<td>922; 22.5</td>
</tr>
<tr>
<td>Q2 (25.3)</td>
<td>922; 25.3</td>
</tr>
<tr>
<td>Q3 (27.5)</td>
<td>919; 27.5</td>
</tr>
<tr>
<td>Q4 (30.1)</td>
<td>923; 30.0</td>
</tr>
<tr>
<td>Q5 (35.6)</td>
<td>918; 34.7</td>
</tr>
</tbody>
</table>

<sup>*</sup>third molars were included; †rounded to the closest integer; <sup>1</sup>Chi-square test of equivalence between strata of periodontitis P<0.05; <sup>2</sup>Kruskal-Wallis test of stratum-specific attachment loss or covariate estimates P<0.05
Table A3. Distribution and bivariate associations of attachment loss (quintile-categorical and three-level classification of extent score) with periodontal diagnosis (CDC classification) among the Caucasian participants of the Dental ARIC study, for whom clinical and genetic data were matched after exclusions (n=4,610)

<table>
<thead>
<tr>
<th>Attachment loss</th>
<th>Total (n, column %)</th>
<th>Health-mild (n, row %)</th>
<th>Moderate (n, row %)</th>
<th>Severe (n, row %)</th>
<th>Extent score* mean(SD); median†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total sample</strong></td>
<td>4610</td>
<td>1864 (40)</td>
<td>1961 (43)</td>
<td>785 (17)</td>
<td>21.1 (21.2); 14</td>
</tr>
<tr>
<td><strong>Quintile-categorical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>922</td>
<td>892 (97)</td>
<td>29 (3)</td>
<td>1 (0)</td>
<td>2.3 (1.5); 2</td>
</tr>
<tr>
<td>Q2</td>
<td>929</td>
<td>634 (68)</td>
<td>291 (31)</td>
<td>4 (0)</td>
<td>7.6 (1.6); 8</td>
</tr>
<tr>
<td>Q3</td>
<td>918</td>
<td>262 (29)</td>
<td>611 (67)</td>
<td>45 (5)</td>
<td>14.3 (2.3); 14</td>
</tr>
<tr>
<td>Q4</td>
<td>944</td>
<td>61 (6)</td>
<td>622 (70)</td>
<td>221 (23)</td>
<td>25.1 (4.4); 25</td>
</tr>
<tr>
<td>Q5</td>
<td>897</td>
<td>15 (2)</td>
<td>368 (41)</td>
<td>514 (57)</td>
<td>57.3 (19.2); 52</td>
</tr>
<tr>
<td><strong>3-level classification</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None-mild (&lt;10%)</td>
<td>1760 (38)</td>
<td>1480 (84)</td>
<td>276 (16)</td>
<td>4 (0)</td>
<td>4.7 (2.9); 5</td>
</tr>
<tr>
<td>Moderate (10-&lt;30%)</td>
<td>1769 (38)</td>
<td>365 (21)</td>
<td>1197 (68)</td>
<td>207 (12)</td>
<td>18.0 (5.5); 17</td>
</tr>
<tr>
<td>Severe (≥30%)</td>
<td>1081 (23)</td>
<td>19 (2)</td>
<td>488 (45)</td>
<td>574 (53)</td>
<td>52.9 (20.0); 46</td>
</tr>
</tbody>
</table>

*third molars were included; †rounded to the closest integer
Table A4. Median (interquartile range, IQR) of sub-gingival microbial counts of eight periodontal pathogens obtained via DNA checkerboard immunoblotting among a subset of the Dental ARIC participants (n=1020)

<table>
<thead>
<tr>
<th></th>
<th>P gingivalis</th>
<th>T forsythia</th>
<th>T denticola</th>
<th>P intermedia</th>
<th>P nigricens</th>
<th>C rectus</th>
<th>F nucleatum</th>
<th>A actinom.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total sample (n)</strong></td>
<td>1020</td>
<td>1020</td>
<td>1017</td>
<td>1018</td>
<td>1020</td>
<td>1018</td>
<td>1012</td>
<td>1020</td>
</tr>
<tr>
<td></td>
<td>945 (3061)</td>
<td>1314 (4368)</td>
<td>2665 (9273)</td>
<td>4270 (15432)</td>
<td>3185 (12758)</td>
<td>2524 (19676)</td>
<td>5605 (32326)</td>
<td>2044 (5592)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>806 (2716)</td>
<td>1248 (4695)</td>
<td>2788 (9922)</td>
<td>3579 (15930)</td>
<td>3179 (12834)</td>
<td>2260 (10252)</td>
<td>5994 (30347)</td>
<td>1891 (5298)</td>
</tr>
<tr>
<td>Males</td>
<td>1162 (3341)</td>
<td>1435 (4188)</td>
<td>2537 (8968)</td>
<td>5074 (15009)</td>
<td>3191 (12683)</td>
<td>2736 (11230)</td>
<td>5355 (32393)</td>
<td>2257 (5762)</td>
</tr>
<tr>
<td><strong>Smoking status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>987 (2881)</td>
<td>1184 (4777)</td>
<td>2647 (9895)</td>
<td>4164 (18039)</td>
<td>3146 (14186)</td>
<td>2389 (10985)</td>
<td>5502 (39732)</td>
<td>2129 (6476)</td>
</tr>
<tr>
<td>Past-light</td>
<td>1351 (3927)</td>
<td>1558 (4268)</td>
<td>3252 (8968)</td>
<td>5204 (15480)</td>
<td>3514 (13292)</td>
<td>3230 (13005)</td>
<td>7088 (33977)</td>
<td>2138 (5906)</td>
</tr>
<tr>
<td>Past-heavy</td>
<td>725 (2470)</td>
<td>1116 (3732)</td>
<td>1869 (8043)</td>
<td>3200 (11721)</td>
<td>2567 (9438)</td>
<td>1827 (6974)</td>
<td>3811 (23385)</td>
<td>1651 (4309)</td>
</tr>
<tr>
<td>Current-light</td>
<td>2358 (2935)</td>
<td>2671 (5696)</td>
<td>1718 (13022)</td>
<td>5417 (13724)</td>
<td>3216 (13207)</td>
<td>3375 (8652)</td>
<td>6696 (35865)</td>
<td>2491 (5933)</td>
</tr>
<tr>
<td>Current-heavy</td>
<td>1003 (3082)</td>
<td>1318 (6286)</td>
<td>3503 (11592)</td>
<td>6858 (15881)</td>
<td>2726 (13957)</td>
<td>3649 (10491)</td>
<td>6950 (29101)</td>
<td>2250 (4306)</td>
</tr>
<tr>
<td><strong>Diabetic status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy</td>
<td>897 (2996)</td>
<td>1242 (4079)</td>
<td>2574 (9149)</td>
<td>4118 (14895)</td>
<td>2924 (12601)</td>
<td>2499 (10467)</td>
<td>5433 (28535)</td>
<td>2035 (5539)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>1305 (3229)</td>
<td>1849 (6770)</td>
<td>2980 (10907)</td>
<td>6031 (17418)</td>
<td>4595 (15681)</td>
<td>2707 (16178)</td>
<td>6853 (42203)</td>
<td>2398 (6474)</td>
</tr>
<tr>
<td><strong>Tooth loss</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-9 lost teeth</td>
<td>862 (2658)</td>
<td>955 (3280)</td>
<td>2154 (8016)</td>
<td>3466 (13724)</td>
<td>2718 (10288)</td>
<td>1978 (8310)</td>
<td>4621 (26557)</td>
<td>1810 (4818)</td>
</tr>
<tr>
<td>10-18 lost teeth</td>
<td>783 (3195)</td>
<td>1445 (5519)</td>
<td>2249 (10392)</td>
<td>3320 (16363)</td>
<td>3407 (14219)</td>
<td>2679 (12632)</td>
<td>5518 (32747)</td>
<td>1997 (5031)</td>
</tr>
<tr>
<td>19-32 lost teeth</td>
<td>2247 (5848)</td>
<td>3556 (9088)</td>
<td>6192 (19980)</td>
<td>10486 (32331)</td>
<td>6021 (25088)</td>
<td>4947 (25396)</td>
<td>14486 (48825)</td>
<td>4185 (8360)</td>
</tr>
<tr>
<td><strong>No. teeth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rho (95% CL)</td>
<td>-0.20</td>
<td>-0.14</td>
<td>-0.11</td>
<td>-0.10</td>
<td>-0.08</td>
<td>-0.12</td>
<td>-0.10</td>
<td>-0.13</td>
</tr>
<tr>
<td>Age</td>
<td>0.01</td>
<td>0.07</td>
<td>0.02</td>
<td>0.01</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>BMI</td>
<td>-0.05, 0.07</td>
<td>0.01, 0.13</td>
<td>-0.03, 0.08</td>
<td>-0.06, 0.07</td>
<td>-0.02, 0.10</td>
<td>-0.02, 0.10</td>
<td>-0.03, 0.10</td>
<td>-0.03, 0.09</td>
</tr>
<tr>
<td>rho (95% CL)</td>
<td>-0.05, 0.07</td>
<td>-0.06, 0.07</td>
<td>-0.04, 0.08</td>
<td>-0.07, 0.06</td>
<td>-0.03, 0.10</td>
<td>-0.05, 0.07</td>
<td>-0.06, 0.07</td>
<td>-0.03, 0.09</td>
</tr>
</tbody>
</table>
Table A5. Median (interquartile range, IQR) of sub-gingival microbial counts of eight periodontal pathogens obtained via DNA checkerboard immunoblotting among a subset of the Dental ARIC participants (n=1020)

<table>
<thead>
<tr>
<th></th>
<th>P. gingivalis</th>
<th>T. forsythia</th>
<th>T. denticola</th>
<th>P. intermedia</th>
<th>P. nigрисcens</th>
<th>C. rectus</th>
<th>F. nucleatum</th>
<th>A. actinom.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total sample (n)</strong></td>
<td>1020</td>
<td>1020</td>
<td>1017</td>
<td>1018</td>
<td>1020</td>
<td>1018</td>
<td>1012</td>
<td>1020</td>
</tr>
<tr>
<td></td>
<td>945 (3061)</td>
<td>1314 (4368)</td>
<td>2665 (9273)</td>
<td>4270 (15432)</td>
<td>3185 (12758)</td>
<td>2524 (19676)</td>
<td>5605 (32326)</td>
<td>2044 (5592)</td>
</tr>
<tr>
<td><strong>Brushing frequency (yesterday)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One time or none</td>
<td>996 (3798)</td>
<td>1506 (5244)</td>
<td>3129 (10393)</td>
<td>4948 (17741)</td>
<td>3075 (18176)</td>
<td>2736 (12907)</td>
<td>7447 (40283)</td>
<td>2584 (6912)</td>
</tr>
<tr>
<td>Two or more</td>
<td>904 (2838)</td>
<td>1277 (4109)</td>
<td>2330 (9112)</td>
<td>3977 (14675)</td>
<td>3216 (10557)</td>
<td>2324 (8879)</td>
<td>5048 (26681)</td>
<td>1882 (5004)</td>
</tr>
<tr>
<td><strong>Flossed last week</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did not floss</td>
<td>1386 (3905)</td>
<td>1734 (6439)</td>
<td>3294 (11529)</td>
<td>6185 (21216)</td>
<td>4208 (18039)</td>
<td>3030 (14578)</td>
<td>7252 (42203)</td>
<td>2779 (6603)</td>
</tr>
<tr>
<td>Flossed 1+ times</td>
<td>780 (2657)</td>
<td>1094 (3588)</td>
<td>2302 (8314)</td>
<td>3582 (13698)</td>
<td>2711 (10246)</td>
<td>2218 (8797)</td>
<td>4830 (26433)</td>
<td>1750 (5004)</td>
</tr>
<tr>
<td><strong>Dental visit reason</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular visits</td>
<td>876 (2817)</td>
<td>1143 (3581)</td>
<td>2203 (8226)</td>
<td>3733 (13696)</td>
<td>2827 (10342)</td>
<td>2246 (8377)</td>
<td>4811 (24388)</td>
<td>1876 (4852)</td>
</tr>
<tr>
<td>Only when problem</td>
<td>1409 (5391)</td>
<td>2900 (10448)</td>
<td>4624 (19327)</td>
<td>8204 (32715)</td>
<td>6128 (24793)</td>
<td>4328 (26746)</td>
<td>10653 (73484)</td>
<td>3012 (8944)</td>
</tr>
<tr>
<td><strong>Last dental visit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6 months ago</td>
<td>908 (2880)</td>
<td>1164 (3573)</td>
<td>2183 (7841)</td>
<td>3624 (13697)</td>
<td>2637 (10246)</td>
<td>2231 (8310)</td>
<td>5048 (25670)</td>
<td>1822 (4816)</td>
</tr>
<tr>
<td>6-24 months</td>
<td>666 (3086)</td>
<td>1295 (4939)</td>
<td>3355 (11840)</td>
<td>4713 (16502)</td>
<td>4662 (14860)</td>
<td>2618 (13005)</td>
<td>6080 (36723)</td>
<td>2204 (6916)</td>
</tr>
<tr>
<td>2+ years ago</td>
<td>1633 (5947)</td>
<td>3497 (12599)</td>
<td>4875 (21366)</td>
<td>10334 (37253)</td>
<td>6654 (26833)</td>
<td>4112 (28367)</td>
<td>10245 (110673)</td>
<td>4208 (10449)</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic (0-11 years)</td>
<td>1719 (5961)</td>
<td>2680 (8607)</td>
<td>3616 (16033)</td>
<td>9248 (34555)</td>
<td>4911 (21757)</td>
<td>5424 (27103)</td>
<td>15088 (77493)</td>
<td>3945 (10016)</td>
</tr>
<tr>
<td>Intermediate (12-15 yrs)</td>
<td>859 (2887)</td>
<td>1535 (4914)</td>
<td>3255 (9823)</td>
<td>3980 (14096)</td>
<td>3385 (12601)</td>
<td>2630 (10467)</td>
<td>6272 (38108)</td>
<td>2267 (5525)</td>
</tr>
<tr>
<td>Advanced (16+ yrs)</td>
<td>910 (2990)</td>
<td>959 (3383)</td>
<td>1973 (8226)</td>
<td>3948 (13950)</td>
<td>2711 (9889)</td>
<td>2048 (8032)</td>
<td>3966 (21940)</td>
<td>1536 (4710)</td>
</tr>
</tbody>
</table>
**Table A6.** Median (interquartile range, IQR) values of sub-gingival microbial counts of the “red” and “orange” complex pathogens, and total microbial load obtained via DNA checkerboard immunoblotting, in the total sample, stratified by sex, smoking and diabetic status and their association with age and BMI among the Caucasian participants of the Dental ARIC study, for whom clinical and genetic data were matched after exclusions (n=1020)

<table>
<thead>
<tr>
<th></th>
<th>“Red” complex</th>
<th>“Orange” complex</th>
<th>Total microbial load</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total sample</strong></td>
<td>5676 (13629)</td>
<td>22668 (60169)</td>
<td>31640 (76343)</td>
</tr>
<tr>
<td>n</td>
<td>1017</td>
<td>1009</td>
<td>1006</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>5684 (13150)</td>
<td>22579 (60294)</td>
<td>31594 (73568)</td>
</tr>
<tr>
<td>Males</td>
<td>5676 (14061)</td>
<td>22731 (60252)</td>
<td>32084 (82237)</td>
</tr>
<tr>
<td><strong>Smoking status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>5604 (13100)</td>
<td>23437 (73000)</td>
<td>31186 (96592)</td>
</tr>
<tr>
<td>Ever-light</td>
<td>6848 (12977)</td>
<td>25544 (69412)</td>
<td>35654 (83731)</td>
</tr>
<tr>
<td>Ever- heavy</td>
<td>4591 (10733)</td>
<td>17750 (41055)</td>
<td>26584 (58326)</td>
</tr>
<tr>
<td>Current-light</td>
<td>5540 (16016)</td>
<td>23990 (46456)</td>
<td>42370 (48097)</td>
</tr>
<tr>
<td>Current- heavy</td>
<td>6500 (19215)</td>
<td>22071 (54349)</td>
<td>32580 (61111)</td>
</tr>
<tr>
<td><strong>Diabetic status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy</td>
<td>5539 (13169)</td>
<td>21728 (56399)</td>
<td>30766 (73756)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>7246 (20397)</td>
<td>26616 (84427)</td>
<td>34893 (107047)</td>
</tr>
<tr>
<td><strong>Tooth loss</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-9 lost teeth</td>
<td>4867 (12089)</td>
<td>18727 (48253)</td>
<td>27334 (63495)</td>
</tr>
<tr>
<td>10-18 lost teeth</td>
<td>5770 (16539)</td>
<td>23813 (71346)</td>
<td>32668 (83626)</td>
</tr>
<tr>
<td>19-32 lost teeth</td>
<td>11044 (25744)</td>
<td>34946 (143873)</td>
<td>49798 (222836)</td>
</tr>
<tr>
<td><strong>Age</strong> (rho, 95% CL)</td>
<td>0.05 (-0.01, 0.11)</td>
<td>0.02 (-0.04, 0.08)</td>
<td>0.03 (-0.04, 0.09)</td>
</tr>
<tr>
<td><strong>BMI</strong> (rho, 95% CL)</td>
<td>0.01 (-0.05, 0.08)</td>
<td>0.01 (-0.05, 0.06)</td>
<td>0.01 (-0.06, 0.07)</td>
</tr>
<tr>
<td><strong>Number of natural teeth</strong> (rho, 95% CL)</td>
<td>-0.14 (-0.20, -0.08)</td>
<td>-0.13 (-0.20, -0.07)</td>
<td>-0.15 (-0.21, -0.09)</td>
</tr>
</tbody>
</table>

1Kruskal-Wallis test of stratum-specific bacterial count estimates P<0.05
Table A7. Median (interquartile range, IQR) values of sub-gingival microbial counts of the “red” and “orange” complex pathogens, and total microbial load obtained via DNA checkerboard immunoblotting, in the total sample, stratified by brushing and flossing frequency, dental attendance frequency and reason, and education level among the Caucasian participants of the Dental ARIC study, for whom clinical and genetic data were matched after exclusions (n=1020)

<table>
<thead>
<tr>
<th></th>
<th>“Red” complex</th>
<th>“Orange” complex</th>
<th>Total microbial load</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total sample</strong></td>
<td>5676 (13629)</td>
<td>22668 (60169)</td>
<td>31640 (76343)</td>
</tr>
<tr>
<td><strong>n</strong></td>
<td>1017</td>
<td>1009</td>
<td>1006</td>
</tr>
<tr>
<td><strong>Brushing frequency</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(yesterday)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One time or none</td>
<td>6352 (17930)</td>
<td>25038 (80044)</td>
<td>33245 (101539)</td>
</tr>
<tr>
<td>Two or more</td>
<td>5518 (12998)</td>
<td>20464 (49853)</td>
<td>28904 (69526)</td>
</tr>
<tr>
<td><strong>Flossed last week</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did not floss</td>
<td>7244 (19794)</td>
<td>25883 (89482)</td>
<td>33900 (127185)</td>
</tr>
<tr>
<td>Flossed one or more times</td>
<td>5444 (12521)</td>
<td>20006 (48350)</td>
<td>28182 (64442)</td>
</tr>
<tr>
<td><strong>Dental visit reason</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular visits</td>
<td>5397 (12535)</td>
<td>20283 (45756)</td>
<td>29129 (61763)</td>
</tr>
<tr>
<td>Only when problem</td>
<td>9289 (30032)</td>
<td>33660 (150830)</td>
<td>49084 (233372)</td>
</tr>
<tr>
<td><strong>Last dental visit</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6 months ago</td>
<td>5338 (12387)</td>
<td>20822 (45828)</td>
<td>28888 (62424)</td>
</tr>
<tr>
<td>6-24 months</td>
<td>6251 (18396)</td>
<td>21441 (69189)</td>
<td>31233 (131809)</td>
</tr>
<tr>
<td>More than 2 years ago</td>
<td>9525 (39649)</td>
<td>40110 (176897)</td>
<td>54380 (367130)</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic (0-11 years)</td>
<td>8709 (29505)</td>
<td>33695 (158940)</td>
<td>49798 (220499)</td>
</tr>
<tr>
<td>Intermediate (12-15 years)</td>
<td>6356 (13756)</td>
<td>23504 (65055)</td>
<td>32477 (82767)</td>
</tr>
<tr>
<td>Advanced (16+ years)</td>
<td>4935 (12199)</td>
<td>19120 (43498)</td>
<td>27929 (58540)</td>
</tr>
</tbody>
</table>

\(^1\)Median test of stratum-specific bacterial count estimates P<0.05; ^2Kruskal-Wallis test of stratum-specific bacterial count estimates P<0.05
Table A8. Sample power calculation for SNP main effect among the ARIC cohort study whites. Assumptions are: MAF 0.1-0.5, estimated SNP effects of 1.1-1.6, and outcome prevalence 0.2 (severe periodontitis).

<table>
<thead>
<tr>
<th>Frequency</th>
<th>R</th>
<th>Power</th>
<th>Gene</th>
<th>P</th>
<th>σ_0=0</th>
<th>$σ_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100000</td>
<td>1.1000</td>
<td>0.1737</td>
<td>0.1356890</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.200000</td>
<td>0.5206</td>
<td>0.1356890</td>
<td>0.1356890</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.300000</td>
<td>0.3332</td>
<td>0.1356890</td>
<td>0.1356890</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.400000</td>
<td>0.6664</td>
<td>0.1356890</td>
<td>0.1356890</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.500000</td>
<td>0.9597</td>
<td>0.1356890</td>
<td>0.1356890</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.600000</td>
<td>0.9907</td>
<td>0.1356890</td>
<td>0.1356890</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.200000</td>
<td>1.1000</td>
<td>0.2810</td>
<td>0.1438333</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.200000</td>
<td>0.7687</td>
<td>0.1438333</td>
<td>0.1438333</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.300000</td>
<td>0.9713</td>
<td>0.1438333</td>
<td>0.1438333</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.400000</td>
<td>0.9087</td>
<td>0.1438333</td>
<td>0.1438333</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.500000</td>
<td>0.9999</td>
<td>0.1438333</td>
<td>0.1438333</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.600000</td>
<td>0.9999</td>
<td>0.1438333</td>
<td>0.1438333</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.300000</td>
<td>1.1000</td>
<td>0.3499</td>
<td>0.1496830</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.200000</td>
<td>0.8570</td>
<td>0.1496830</td>
<td>0.1496830</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.300000</td>
<td>0.9931</td>
<td>0.1496830</td>
<td>0.1496830</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.400000</td>
<td>0.9939</td>
<td>0.1496830</td>
<td>0.1496830</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.500000</td>
<td>0.9999</td>
<td>0.1496830</td>
<td>0.1496830</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.600000</td>
<td>0.9999</td>
<td>0.1496830</td>
<td>0.1496830</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.400000</td>
<td>1.1000</td>
<td>0.3883</td>
<td>0.1578799</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.200000</td>
<td>0.8948</td>
<td>0.1578799</td>
<td>0.1578799</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.300000</td>
<td>0.9962</td>
<td>0.1578799</td>
<td>0.1578799</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.400000</td>
<td>0.9969</td>
<td>0.1578799</td>
<td>0.1578799</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.500000</td>
<td>0.9999</td>
<td>0.1578799</td>
<td>0.1578799</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.600000</td>
<td>0.9999</td>
<td>0.1578799</td>
<td>0.1578799</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.500000</td>
<td>1.1000</td>
<td>0.4061</td>
<td>0.1649860</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.200000</td>
<td>0.9023</td>
<td>0.1649860</td>
<td>0.1649860</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.300000</td>
<td>0.9967</td>
<td>0.1649860</td>
<td>0.1649860</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.400000</td>
<td>0.9969</td>
<td>0.1649860</td>
<td>0.1649860</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.500000</td>
<td>0.9999</td>
<td>0.1649860</td>
<td>0.1649860</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.600000</td>
<td>0.9999</td>
<td>0.1649860</td>
<td>0.1649860</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table A9 Sample power calculation for SNP main effect among the ARIC cohort study whites. Assumptions are: MAF 0.1-0.5, estimated SNP effects of 1.2-1.6, and outcome prevalence 0.2 (highest quintile of bacterial colonization phenotypes).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Null</th>
<th>Full</th>
<th>Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>$\delta=0$</td>
<td>$\rho_0$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency</th>
<th>$R_0$</th>
<th>$P_0$</th>
<th>Gene</th>
<th>$P_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100000</td>
<td>1.2000</td>
<td>0.1715</td>
<td>0.193948</td>
<td></td>
</tr>
<tr>
<td>1.3000</td>
<td>0.3101</td>
<td>0.191161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4000</td>
<td>0.4741</td>
<td>0.188514</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5000</td>
<td>0.6240</td>
<td>0.185998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6000</td>
<td>0.7669</td>
<td>0.183800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.200000</td>
<td>1.2000</td>
<td>0.2652</td>
<td>0.188098</td>
<td></td>
</tr>
<tr>
<td>1.3000</td>
<td>0.4880</td>
<td>0.182747</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4000</td>
<td>0.7022</td>
<td>0.177741</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5000</td>
<td>0.9547</td>
<td>0.173045</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6000</td>
<td>0.9397</td>
<td>0.168030</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.300000</td>
<td>1.2000</td>
<td>0.3279</td>
<td>0.182444</td>
<td></td>
</tr>
<tr>
<td>1.3000</td>
<td>0.5899</td>
<td>0.174743</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4000</td>
<td>0.8634</td>
<td>0.167645</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5000</td>
<td>0.9248</td>
<td>0.161082</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6000</td>
<td>0.9765</td>
<td>0.154935</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.400000</td>
<td>1.2000</td>
<td>0.3619</td>
<td>0.176931</td>
<td></td>
</tr>
<tr>
<td>1.3000</td>
<td>0.6382</td>
<td>0.167134</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4000</td>
<td>0.8431</td>
<td>0.158185</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5000</td>
<td>0.9468</td>
<td>0.150049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6000</td>
<td>0.9853</td>
<td>0.142599</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.500000</td>
<td>1.2000</td>
<td>0.3699</td>
<td>0.171704</td>
<td></td>
</tr>
<tr>
<td>1.3000</td>
<td>0.6468</td>
<td>0.160002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4000</td>
<td>0.8482</td>
<td>0.149355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5000</td>
<td>0.9485</td>
<td>0.139884</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6000</td>
<td>0.9856</td>
<td>0.131343</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table A10. Sample power calculation for SNP*environment interaction among Caucasians. Assumptions are: MAF 0.1-0.5, estimated SNP effect of 1.4, environmental factor prevalence 0.1 (current heavy smoking), interaction effect 1.2-2.0 and outcome prevalence 0.2 (severe periodontitis).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Null</th>
<th>Full</th>
<th>Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>$R_{0E} = 0$</td>
<td>$R_0$, $R_{0E}$</td>
<td>$R_0$, $R_{0E}$</td>
</tr>
<tr>
<td>Gene</td>
<td>$R_{G} = 0$</td>
<td>$R_G$</td>
<td>$R_G$</td>
</tr>
<tr>
<td>Environment</td>
<td>$R_{E} = 0$</td>
<td>$R_E$</td>
<td>$R_E$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency</th>
<th>$R_{0E}$</th>
<th>Interaction</th>
<th>Gene</th>
<th>Environment</th>
<th>$P_0$</th>
<th>$R_0$</th>
<th>$R_E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100000</td>
<td>1.2000</td>
<td>0.0989</td>
<td>0.9980</td>
<td>0.0633</td>
<td>0.187326</td>
<td>1.4289</td>
<td>1.0453</td>
</tr>
<tr>
<td>1.4000</td>
<td>0.2217</td>
<td>0.9880</td>
<td>0.0968</td>
<td>0.187202</td>
<td>1.4516</td>
<td>1.0874</td>
<td></td>
</tr>
<tr>
<td>1.8000</td>
<td>0.4020</td>
<td>0.9995</td>
<td>0.1513</td>
<td>0.186834</td>
<td>1.4742</td>
<td>1.1285</td>
<td></td>
</tr>
<tr>
<td>2.0000</td>
<td>0.5752</td>
<td>0.9995</td>
<td>0.2159</td>
<td>0.186117</td>
<td>1.4949</td>
<td>1.1627</td>
<td></td>
</tr>
<tr>
<td>2.0000</td>
<td>0.7247</td>
<td>0.9971</td>
<td>0.2877</td>
<td>0.185645</td>
<td>1.5138</td>
<td>1.1964</td>
<td></td>
</tr>
<tr>
<td>0.200000</td>
<td>1.2000</td>
<td>0.1343</td>
<td>0.9995</td>
<td>0.1019</td>
<td>0.176444</td>
<td>1.4271</td>
<td>1.0902</td>
</tr>
<tr>
<td>1.4000</td>
<td>0.3520</td>
<td>0.9998</td>
<td>0.2431</td>
<td>0.175265</td>
<td>1.4521</td>
<td>1.1760</td>
<td></td>
</tr>
<tr>
<td>1.6000</td>
<td>0.8059</td>
<td>0.9999</td>
<td>0.4058</td>
<td>0.174195</td>
<td>1.4751</td>
<td>1.2571</td>
<td></td>
</tr>
<tr>
<td>1.8000</td>
<td>0.9002</td>
<td>0.9999</td>
<td>0.7261</td>
<td>0.173222</td>
<td>1.4959</td>
<td>1.3337</td>
<td></td>
</tr>
<tr>
<td>2.0000</td>
<td>0.9119</td>
<td>0.9999</td>
<td>0.7759</td>
<td>0.172336</td>
<td>1.5149</td>
<td>1.4057</td>
<td></td>
</tr>
<tr>
<td>0.300000</td>
<td>1.2000</td>
<td>0.1572</td>
<td>0.9999</td>
<td>0.1647</td>
<td>0.165816</td>
<td>1.4273</td>
<td>1.1347</td>
</tr>
<tr>
<td>1.4000</td>
<td>0.4273</td>
<td>0.9999</td>
<td>0.4505</td>
<td>0.164145</td>
<td>1.4528</td>
<td>1.2650</td>
<td></td>
</tr>
<tr>
<td>1.6000</td>
<td>0.7073</td>
<td>0.9999</td>
<td>0.7623</td>
<td>0.162028</td>
<td>1.4762</td>
<td>1.3629</td>
<td></td>
</tr>
<tr>
<td>1.8000</td>
<td>0.8826</td>
<td>0.9999</td>
<td>0.9173</td>
<td>0.161251</td>
<td>1.4975</td>
<td>1.5147</td>
<td></td>
</tr>
<tr>
<td>0.400000</td>
<td>1.2000</td>
<td>0.1638</td>
<td>0.9999</td>
<td>0.2421</td>
<td>0.155900</td>
<td>1.4275</td>
<td>1.1789</td>
</tr>
<tr>
<td>1.6000</td>
<td>0.4640</td>
<td>0.9999</td>
<td>0.6410</td>
<td>0.153795</td>
<td>1.4536</td>
<td>1.3579</td>
<td></td>
</tr>
<tr>
<td>2.0000</td>
<td>0.7517</td>
<td>0.9999</td>
<td>0.9331</td>
<td>0.151881</td>
<td>1.4777</td>
<td>1.5348</td>
<td></td>
</tr>
<tr>
<td>0.500000</td>
<td>1.2000</td>
<td>0.1696</td>
<td>0.9999</td>
<td>0.2600</td>
<td>0.140557</td>
<td>1.4278</td>
<td>1.2288</td>
</tr>
<tr>
<td>2.0000</td>
<td>0.7592</td>
<td>0.9999</td>
<td>0.9861</td>
<td>0.141902</td>
<td>1.4794</td>
<td>1.6836</td>
<td></td>
</tr>
<tr>
<td>1.8000</td>
<td>0.9184</td>
<td>0.9999</td>
<td>0.9997</td>
<td>0.139845</td>
<td>1.5021</td>
<td>1.9148</td>
<td></td>
</tr>
<tr>
<td>2.0000</td>
<td>0.9776</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.137964</td>
<td>1.5226</td>
<td>2.1434</td>
<td></td>
</tr>
</tbody>
</table>
Table A11 Sample power calculation for SNP*environment interaction among Caucasians. Assumptions are: MAF 0.1-0.5, estimated SNP effect of 1.4, environmental factor prevalence 0.5 (ever smoker), interaction effect 1.2-2.0 and outcome prevalence 0.2 (severe periodontitis).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Null</th>
<th>Full</th>
<th>Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>$R_{AE}=0$</td>
<td>$R_{AE}R_{D}R_{E}$</td>
<td>$R_{D}R_{E}$</td>
</tr>
<tr>
<td>Gene</td>
<td>$R_{G}=0$</td>
<td>$R_{G}$</td>
<td>___</td>
</tr>
<tr>
<td>Environment</td>
<td>$R_{E}=0$</td>
<td>$R_{E}$</td>
<td>___</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency</th>
<th>$R_{AE}$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10000</td>
<td>1.2000</td>
<td>0.1059</td>
</tr>
<tr>
<td>0.20000</td>
<td>1.2000</td>
<td>0.1059</td>
</tr>
<tr>
<td>0.30000</td>
<td>1.2000</td>
<td>0.1059</td>
</tr>
<tr>
<td>0.40000</td>
<td>1.2000</td>
<td>0.1059</td>
</tr>
<tr>
<td>0.50000</td>
<td>1.2000</td>
<td>0.1059</td>
</tr>
</tbody>
</table>

Disease model: Summary parameters
- $P_0 = 0.185102$
- $P_{AE} = 0.200000$
- $P_{D} = 1.4000$
- $P_{E} = 1.6000$
- $P_{AE} = 1.2000$

Parameter: Log-additive
Mode of inheritance: 0.1000 to 0.5000 by 0.1000
Allele frequency: 0.10000, 2-sided
Binary environmental factor: 0.50000
Outcome prevalence: 0.20000
Environmental factor prevalence: 0.50000

Outcome: Unmatched case-control (1:4)
Design: Gene-environment interaction
Sample size: 800 cases, 4 controls per case are required
Significance: 0.050000, 2-sided
Gene

<table>
<thead>
<tr>
<th>$P_0$</th>
<th>$P_{AE}$</th>
<th>$P_{D}$</th>
<th>$P_{E}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1059</td>
<td>0.1059</td>
<td>1.5369</td>
<td>1.0455</td>
</tr>
<tr>
<td>0.1059</td>
<td>0.1059</td>
<td>1.5369</td>
<td>1.0455</td>
</tr>
<tr>
<td>0.1059</td>
<td>0.1059</td>
<td>1.5369</td>
<td>1.0455</td>
</tr>
<tr>
<td>0.1059</td>
<td>0.1059</td>
<td>1.5369</td>
<td>1.0455</td>
</tr>
<tr>
<td>0.1059</td>
<td>0.1059</td>
<td>1.5369</td>
<td>1.0455</td>
</tr>
</tbody>
</table>

(*Indicates calculated value)
Table A12 Sample power calculation for SNP main effect among African Americans. Assumptions are: MAF 0.1-0.5, estimated SNP effects of 1.3-1.7 and outcome prevalence 0.2 (severe periodontitis).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Null</th>
<th>Full</th>
<th>Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>β_0 = 0</td>
<td>β_G</td>
<td>——</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency</th>
<th>R_0</th>
<th>Power</th>
<th>Gene</th>
<th>P_0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100000</td>
<td>1.3000</td>
<td>0.3101</td>
<td>0.191161</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.4000</td>
<td>0.4741</td>
<td>0.188514</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5000</td>
<td>0.6340</td>
<td>0.185998</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.6000</td>
<td>0.7669</td>
<td>0.183600</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.7000</td>
<td>0.8632</td>
<td>0.181314</td>
<td></td>
</tr>
<tr>
<td>0.200000</td>
<td>1.3000</td>
<td>0.4880</td>
<td>0.182747</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.4000</td>
<td>0.7022</td>
<td>0.177741</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5000</td>
<td>0.8547</td>
<td>0.173045</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.6000</td>
<td>0.9397</td>
<td>0.168630</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.7000</td>
<td>0.9783</td>
<td>0.164471</td>
<td></td>
</tr>
<tr>
<td>0.300000</td>
<td>1.3000</td>
<td>0.5899</td>
<td>0.174743</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.4000</td>
<td>0.6034</td>
<td>0.167645</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5000</td>
<td>0.9248</td>
<td>0.161082</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.6000</td>
<td>0.9765</td>
<td>0.154995</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.7000</td>
<td>0.9938</td>
<td>0.149335</td>
<td></td>
</tr>
<tr>
<td>0.400000</td>
<td>1.3000</td>
<td>0.6382</td>
<td>0.167134</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.4000</td>
<td>0.8431</td>
<td>0.158195</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5000</td>
<td>0.9468</td>
<td>0.155049</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.6000</td>
<td>0.9653</td>
<td>0.142599</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.7000</td>
<td>0.9966</td>
<td>0.135762</td>
<td></td>
</tr>
<tr>
<td>0.500000</td>
<td>1.3000</td>
<td>0.6468</td>
<td>0.159902</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.4000</td>
<td>0.8482</td>
<td>0.149355</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5000</td>
<td>0.9485</td>
<td>0.139884</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.6000</td>
<td>0.9856</td>
<td>0.131343</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.7000</td>
<td>0.9966</td>
<td>0.123611</td>
<td></td>
</tr>
<tr>
<td>Chr</td>
<td>SNP</td>
<td>Position Build36</td>
<td>Nearby gene (location)</td>
<td>P value</td>
</tr>
<tr>
<td>-----</td>
<td>--------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>7</td>
<td>rs1615939</td>
<td>152391164</td>
<td>ACTRI3B (205Kb)</td>
<td>1.4x10^{-6}</td>
</tr>
<tr>
<td>1</td>
<td>rs1551110</td>
<td>105334285</td>
<td>LOC100129138 (913Kb)</td>
<td>1.5x10^{-6}</td>
</tr>
<tr>
<td>4</td>
<td>rs1001956</td>
<td>148166054</td>
<td>TTC19 (78Kb)</td>
<td>2.0x10^{-6}</td>
</tr>
<tr>
<td>8</td>
<td>rs1215581</td>
<td>9610089</td>
<td>TNKS (intronic)</td>
<td>2.3x10^{-6}</td>
</tr>
<tr>
<td>4</td>
<td>rs1372486</td>
<td>46864735</td>
<td>GABRB1 (intronic)</td>
<td>2.5x10^{-6}</td>
</tr>
<tr>
<td>12</td>
<td>rs4761973</td>
<td>50166536</td>
<td>SLC4A8</td>
<td>2.6x10^{-6}</td>
</tr>
<tr>
<td>1</td>
<td>rs4659467</td>
<td>234987803</td>
<td>ACTN2 (intronic)</td>
<td>3.0x10^{-6}</td>
</tr>
<tr>
<td>13</td>
<td>rs1085126</td>
<td>110767630</td>
<td>C13orf16 (13Kb); ARHGEF7 (12Kb)</td>
<td>3.3x10^{-6}</td>
</tr>
<tr>
<td>8</td>
<td>rs2063794</td>
<td>78581773</td>
<td>PEX2 (506Kb)</td>
<td>3.6x10^{-6}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table A14. Genome-wide association analysis results of the CDC/AAP chronic periodontitis (CP) classification traits (severe CP vs. healthy and moderate CP vs. healthy) and extent of attachment loss trait (EAL-proportion of sites exhibiting attachment loss ≥3mm), among the white participants of the Dental ARIC study (n=4610). Supplemental list of single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF-HapMap II CEU) of ≥5% and that emerged below the P<5x10^{-6} in the “sensitivity” analysis that included logistic regression models adjusted for diabetes (binary variable) and smoking (never, former, current).

<table>
<thead>
<tr>
<th>Chr.</th>
<th>SNP</th>
<th>Position Build36</th>
<th>CA1</th>
<th>CA2</th>
<th>chr1 SNP</th>
<th>Position Build36</th>
<th>CA1</th>
<th>CA2</th>
<th>Chr. SNP</th>
<th>Position Build36</th>
<th>CA1</th>
<th>CA2</th>
<th>Chr. SNP</th>
<th>Position Build36</th>
<th>CA1</th>
<th>CA2</th>
<th>Chr. SNP</th>
<th>Position Build36</th>
<th>CA1</th>
<th>CA2</th>
<th>Chr. SNP</th>
<th>Position Build36</th>
<th>CA1</th>
<th>CA2</th>
<th>Chr. SNP</th>
<th>Position Build36</th>
<th>CA1</th>
<th>CA2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1: coded allele
2: non-coded allele
3: based on logistic regression models that were adjusted for diabetes (binary) and smoking (never, former, current)
4: Confidence interval
Table A15. Coded and non-coded alleles of prioritized SNPs for severe and moderate CP on the Affymetrix 6.0 (ARIC GWAS) and the Illumina Human 1M Duo (ver.3) (Health ABC GWAS) platforms. Pending verification by the Health ABC collaborating group.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Affymetrix 6.0</th>
<th>Illumina Human1M Duo Beadchip</th>
<th>Different strand (yes/no)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coded allele</td>
<td>Non-coded allele</td>
<td>Coded allele</td>
</tr>
<tr>
<td>rs12883458</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>rs2521634</td>
<td>A</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>rs11925054</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>rs10493998</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>rs17006135</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>rs7762544</td>
<td>A</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>rs3826782</td>
<td>A</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>rs12260727</td>
<td>A</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>rs11615037</td>
<td>C</td>
<td>G</td>
<td>C</td>
</tr>
</tbody>
</table>
Table A16. Genome-wide association replication results [SNPs prioritized based on a P<5x10^-6 threshold in the dental ARIC (n=4610)] of severe and moderate CP in the Health ABC cohort. Reported estimates correspond to the ARIC cohort ‘risk’ allele effect on severe and moderate CP and EAL in the Health ABC cohort.

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>Severe CP</th>
<th>Moderate CP</th>
<th>EAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P-value</td>
<td>OR (95% CI)</td>
<td>P-value</td>
</tr>
<tr>
<td>Severe CP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12883458</td>
<td>0.25</td>
<td>0.65 (0.31, 1.34)</td>
<td>0.71</td>
</tr>
<tr>
<td>rs2521634</td>
<td>0.05</td>
<td>1.65 (1.01, 2.69)</td>
<td>0.03</td>
</tr>
<tr>
<td>rs11925054</td>
<td>0.06</td>
<td>0.51 (0.25, 1.04)</td>
<td>0.05</td>
</tr>
<tr>
<td>rs10493998</td>
<td>0.89</td>
<td>1.06 (0.52, 2.15)</td>
<td>0.34</td>
</tr>
<tr>
<td>rs17006135</td>
<td>0.03</td>
<td>0.43 (0.20, 0.90)</td>
<td>0.32</td>
</tr>
<tr>
<td>Moderate CP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7762544</td>
<td>0.34</td>
<td>1.30 (0.76, 2.20)</td>
<td><strong>0.28</strong></td>
</tr>
<tr>
<td>rs3826782</td>
<td>0.01</td>
<td>4.18 (1.66, 10.50)</td>
<td><strong>0.06</strong></td>
</tr>
<tr>
<td>rs12260727</td>
<td>0.29</td>
<td>1.39 (0.74, 2.60)</td>
<td><strong>0.80</strong></td>
</tr>
<tr>
<td>rs11615037</td>
<td>0.37</td>
<td>0.78 (0.46, 1.32)</td>
<td><strong>0.77</strong></td>
</tr>
</tbody>
</table>

1 opposite direction effect; verification of coded SNP in Health ABC GWAS is pending
2 SNPs that were prioritized based on the P<5x10^-6 criterion in exploratory smoking- and diabetes-adjusted analysis
3 rs17006135 marked a shared risk locus for severe CP and EAL in the Dental ARIC GWAS adjusted exploratory analysis
Table A17. Replication of GWAS results (P-values and effect estimates) of prioritized SNPs of the ARIC genome-wide association analysis (crude and exploratory smoking and diabetes-adjusted analyses) in the Health ABC cohort. Reported effect estimates correspond to the ‘risk’ allele in the ARIC cohort.

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>ARIC GWAS P-value</th>
<th>OR (95% CI)</th>
<th>ARIC GWAS adjusted P-value</th>
<th>OR (95% CI)</th>
<th>Health ABC P-value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12883458</td>
<td>3.5x10^{-7}</td>
<td>1.89 (1.48, 2.41)</td>
<td>1.1x10^{-6}</td>
<td>1.88 (1.46, 2.43)</td>
<td>0.25</td>
<td>0.65 (0.31, 1.34)</td>
</tr>
<tr>
<td>rs2521634</td>
<td>1.6x10^{-6}</td>
<td>1.47 (1.25, 1.73)</td>
<td>2.6x10^{-5}</td>
<td>1.43 (1.21, 1.69)</td>
<td>0.05</td>
<td>1.65 (1.01, 2.69)</td>
</tr>
<tr>
<td>rs11925054</td>
<td>6.5x10^{-7}</td>
<td>1.69 (1.37, 2.10)</td>
<td>2.6x10^{-7}</td>
<td>1.78 (1.42, 2.23)</td>
<td>0.06</td>
<td>0.51 (0.25, 1.04)</td>
</tr>
<tr>
<td>rs10493998</td>
<td>9.2x10^{-6}</td>
<td>1.59 (1.30, 1.95)</td>
<td>2.2x10^{-6}</td>
<td>1.68 (1.36, 2.08)</td>
<td>0.89</td>
<td>1.06 (0.52, 2.15)</td>
</tr>
<tr>
<td>rs17006135</td>
<td>1.8x10^{-5}</td>
<td>1.70 (1.34, 2.16)</td>
<td>4.4x10^{-6}</td>
<td>1.82 (1.41, 2.35)</td>
<td>0.03</td>
<td>0.43 (0.20, 0.90)</td>
</tr>
<tr>
<td>Moderate CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7762544</td>
<td>1.1x10^{-7}</td>
<td>1.41 (1.24, 1.60)</td>
<td>3.8x10^{-8}</td>
<td>1.44 (1.26, 1.63)</td>
<td>0.28</td>
<td>1.32 (0.81, 2.16)</td>
</tr>
<tr>
<td>rs3826782</td>
<td>4.0x10^{-6}</td>
<td>2.00 (1.48, 2.70)</td>
<td>1.4x10^{-6}</td>
<td>2.09 (1.54, 2.84)</td>
<td>0.06</td>
<td>2.08 (0.99, 4.37)</td>
</tr>
<tr>
<td>rs12260727</td>
<td>6.0x10^{-7}</td>
<td>1.54 (1.30, 1.82)</td>
<td>3.0x10^{-6}</td>
<td>1.51 (1.27, 1.79)</td>
<td>0.80</td>
<td>0.93 (0.55, 1.58)</td>
</tr>
<tr>
<td>rs11615037</td>
<td>2.6x10^{-5}</td>
<td>1.29 (1.15, 1.46)</td>
<td>4.4x10^{-6}</td>
<td>1.33 (1.18, 1.51)</td>
<td>0.77</td>
<td>0.93 (0.61, 1.44)</td>
</tr>
</tbody>
</table>

1 opposite direction effect; verification of coded SNP in Health ABC GWAS is pending
2 SNPs that were prioritized based on the P<5x10^{-6} criterion in exploratory smoking- and diabetes-adjusted analysis
3 rs17006135 marked a shared risk locus for severe CP and EAL in the Dental ARIC GWAS adjusted exploratory analysis
Table A18. Meta-analysis of GWAS results (meta-analysis P-values and effect estimates based on inverse variance weighting) of prioritized SNPs for severe and moderate CP in the ARIC and the Health ABC studies. Reported effect estimates correspond to the ‘risk’ allele in the ARIC cohort. Pooled estimates based on inverse-variance meta-analysis are presented for results that did not demonstrate substantial heterogeneity (homogeneity $X^2$ $P<0.2$ criterion).

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>Pooled estimates</th>
<th>Diff. direction</th>
<th>$I^2$</th>
<th>$X^2$</th>
<th>$P&lt;0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Severe CP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12883458</td>
<td>$3.5\times10^{-7}$</td>
<td>1.89 (1.48, 2.41)</td>
<td>0.25</td>
<td>0.65 (0.31, 1.34)</td>
<td>*</td>
<td>86.6</td>
<td>7.5</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2521634</td>
<td>$1.6\times10^{-6}$</td>
<td>1.47 (1.25, 1.73)</td>
<td>0.05</td>
<td>1.65 (1.01, 2.69)</td>
<td>$3.5\times10^{-7}$</td>
<td>1.49 (1.28, 1.73)</td>
<td>0</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11925054</td>
<td>$6.5\times10^{-7}$</td>
<td>1.69 (1.37, 2.10)</td>
<td>0.06</td>
<td>0.51 (0.25, 1.04)</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>90.1</td>
<td>10.1</td>
<td>*</td>
</tr>
<tr>
<td>rs10493998</td>
<td>$9.2\times10^{-6}$</td>
<td>1.59 (1.30, 1.95)</td>
<td>0.89</td>
<td>1.06 (0.52, 2.15)</td>
<td></td>
<td></td>
<td></td>
<td>48.7</td>
<td>2.0</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>rs17006135</td>
<td>$1.8\times10^{-5}$</td>
<td>1.70 (1.34, 2.16)</td>
<td>0.03</td>
<td>0.43 (0.20, 0.90)</td>
<td>*</td>
<td>91.6</td>
<td>11.9</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Moderate CP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7762544</td>
<td>$1.1\times10^{-5}$</td>
<td>1.41 (1.24, 1.60)</td>
<td>0.28</td>
<td>1.32 (0.81, 2.16)</td>
<td>$7.5\times10^{-8}$</td>
<td>1.40 (1.24, 1.59)</td>
<td>0</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3826782</td>
<td>$4.0\times10^{-6}$</td>
<td>2.00 (1.48, 2.70)</td>
<td>0.06</td>
<td>2.08 (0.99, 4.37)</td>
<td>$8.2\times10^{-7}$</td>
<td>2.01 (1.52, 2.65)</td>
<td>0</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12260727</td>
<td>$6.0\times10^{-7}$</td>
<td>1.54 (1.30, 1.82)</td>
<td>0.80</td>
<td>0.43 (0.20, 0.90)</td>
<td>*</td>
<td>67.7</td>
<td>3.1</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11615037</td>
<td>$2.6\times10^{-5}$</td>
<td>1.29 (1.15, 1.46)</td>
<td>0.77</td>
<td>0.93 (0.61, 1.44)</td>
<td>*</td>
<td>51.4</td>
<td>2.6</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 SNPs that were prioritized based on the $P<5\times10^{-6}$ criterion in exploratory smoking- and diabetes-adjusted analysis in the ARIC cohort
2 Asterisks (*) correspond to effects of different direction in the ARIC and Health ABC cohorts, and homogeneity $X^2$ $P$-values of less than 0.2
FIGURES

Figure A1. Power calculation for severe CP among whites for MAF: 0.1-0.5.
Figure A2. Power calculation for the “high” bacterial colonization trait among whites for MAF: 0.1-0.5.
Figure A3. Power calculation for GxE interaction effects for severe CP, assuming ‘exposure’ prevalence 0.5 and MAF: 0.1-0.5 among the dental ARIC whites.
Figure A4. Power calculation for GxE interaction effects for severe CP, assuming ‘exposure’ prevalence 0.1 and MAF: 0.1-0.5 among the dental ARIC whites.
Figure A5. Density plot of average DST among the ARIC cohort study whites zoomed in the regions of DST cut-offs for related individuals (first degree DST $\geq 0.8$, second degree $0.745 \leq$ DST $< 0.8$). Red dots represent those 380 pairs of first degree and 207 pairs of second degree relatives.
**Figure A6.** Density plot of average DST among the ARIC cohort study whites. Red dots represent 16 genetic outliers based on average DST values.
Figure A.7 Scatter plot matrix of the ten principal components generated with the EIGENSTRAT program for the white ARIC cohort study genotyped participants.
Figure A.8 Catalogue of published GWA studies (n=951) through June 2011.
Figure A.9 Visualization of the 14q21 locus that was marked by rs12883458 (P=3.5x10^{-7}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.10 Visualization of the 7p15 locus that was marked by rs2521634 ($P=1.6\times10^{-6}$) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.11 Visualization of the 3p21 locus that was marked by rs11925054 (P=6.5x10^{-7}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
**Figure A.12** Visualization of the 6p21.1 locus that was marked by rs7762544 (P=1.1x10^{-7}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.13 Visualization of the 19p13.3 locus that was marked by rs3826782 (P=4.0x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.14 Visualization of the 10p15 locus that was marked by rs12260727 ($P=6.0 \times 10^{-7}$) for moderate chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.15 Visualization of the 6q15 locus that was marked by rs17792917 (P=1.8x10^{-6}) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.16 Visualization of the 4q21.3 locus that was marked by rs17006135 ($P=4.4\times10^{-6}$ after adjustment for smoking and diabetic status) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.17 Visualization of the 11q22 locus that was marked by rs7120142 ($P=3.9 \times 10^{-6}$) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.18 Visualization of the 7q34 locus that was marked by rs10500130 ($P=4.6 \times 10^{-6}$) for severe chronic periodontitis among the 4610 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
**Figure A.19** Visualization of the 1q42 locus that was marked by rs11800854 ($P=4.0 \times 10^{-6}$) for “high” *Aggregatibacter actinomycetemcomitans* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.20 Visualization of the 1p22 locus that was marked by rs12032672 (P=9.6x10^{-7}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.21 Visualization of the 5q33 locus that was marked by rs10043775 (P=2.4x10^{-6}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.22 Visualization of the 9p24 locus that was marked by rs16924631 (P=3.2×10^{-6}) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.23 Visualization of the 4p14 locus that was marked by rs10010758 ($P=3.7 \times 10^{-6}$) for “high red” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.24 Visualization of the 1p36.2 locus that was marked by rs1616122 (P=4.9x10^{-6}) for “high orange” complex bacterial colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.25 Visualization of the 1q42 locus that was marked by rs11800854 ($P=4.0\times10^{-6}$) for “high” Aggregatibacter actinomycetemcomitans colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
Figure A.26 Visualization of the 9q33.2 locus that was marked by rs10760187 (P=4.5x10^{-7}) for “high” *Porphyromonas gingivalis* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haplovie (based on 1000 genomes pilot 1 release haplotype data).
Figure A.27 Visualization of the 10p11.21 locus that was marked by rs1360573 (P=1.5x10^{-6}) for “high” *Porphyromonas gingivalis* colonization among the subset of 1020 white participants of the Dental Atherosclerosis in Communities Study cohort combined with a linkage disequilibrium plot generated with Haploview (based on 1000 genomes pilot 1 release haplotype data).
**Figure A.28** Predicted network of associations [physical interactions (grey line), co-expression (dark blue line), pathway (light blue line), shared protein domain (green line)] among genes of prioritized loci for the CP and the “high” bacterial traits in the Dental ARIC cohort. Grey lines represent co-expression and blue lines represent physical interactions.
Figure A.29 Predicted network of associations [physical interactions (grey line), co-expression (dark blue line), pathway (light blue line), shared protein domain (green line)], generated with the GeneMANIA module) among genes of prioritized loci for the CP and the “high” bacterial traits in the Dental ARIC cohort, along with genes that have been previously reported to be associated with periodontitis.
Figure A.30 Visualization of GWA results for severe CP in the ARIC cohort for loci that have been previously associated with periodontitis (IL1A, IL1RN, TLR4, MMP2, GLT6D1, GCGR3A). The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
Figure A.31 Visualization of GWA results for moderate CP in the ARIC cohort for loci that have been previously associated with periodontitis (IL1A, IL1RN, TLR4, MMP2, GLT6D1, GCGR3A). The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
Figure A.32 Visualization of GWA results for the extent of attachment loss trait (proportion of sites exhibiting ≥3mm attachment loss) CP in the ARIC cohort for loci that have been previously associated with periodontitis (IL1A, IL1RN, TLR4, MMP2, GLT6D1, GCGR3A). The vertical axis corresponds to each marker’s associated –log10 P-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
**Figure A.33** Visualization of GWAS results in the *IL6* (Interleukin 6) and *VDR* (vitamin D receptor) loci in the ARIC cohort for severe (left column) and moderate CP (right column). The vertical axis corresponds to each marker’s associated $-\log_{10} P$-value. The overlaid recombination rate plot and the color-coded pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU (human genome 18, build 36).
Figure A.34 Visualization of the 6p21.1 locus (chr6:41,279,315-41,479,315; flanking 100Kb of rs7762544, which was prioritized for moderate CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters.
Figure A.35 Visualization of the 19p13.3 locus (chr19:6,787,736-6,987,736; flanking 100Kb of rs3826782, which was prioritized for moderate CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters.
**Figure A.36** Visualization of the 10p15 locus (chr10:10,238,329-10,438,329; flanking 100Kb of rs12260727, which was prioritized for moderate CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters.
Figure A.37 Visualization of the 14q21 locus (chr14:51,179,379-51,379,379; flanking 100Kb of rs12883458, which was prioritized for severe CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters.
Figure A.38 Visualization of the 7p15 locus (chr7:24,278,040-24,478,040; flanking 100Kb of rs2521634, which was prioritized for severe CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters.
Figure A.39 Visualization of the 3p21 locus (chr3:55,290,886-55,490,886; flanking 100Kb of rs11925054, which was prioritized for severe CP) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters.
Figure A.40 Visualization of the 6q15 locus (chr6:85,166,965-85,366,965; flanking 100Kb of rs17792917, which was prioritized for the “extent” of attachment loss trait) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters.
Figure A.41 Visualization of the 11q24 locus (chr11:127,557,071-127,757,071; flanking 100Kb of rs10790919, which was prioritized for the “extent” of attachment loss trait) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters.
**Figure A.42** Visualization of the 11q22 locus (chr11:103,317,781-103,517,781; flanking 100Kb of rs7120142, which was prioritized for the “extent” of attachment loss trait) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters.
Figure A.43 Visualization of the 7q34 locus (chr7:138,715,913-138,915,913; flanking 100Kb of rs10500130, which was prioritized for the “extent” of attachment loss trait) using the ENCODE/USCF browser and displaying enhancer/promoter associated regions and DNaseI hypersensitivity clusters.
REFERENCES

Chapter 1


42. Koromantzos PA, Makrilakis K, Dereka X, Katsilambros N, Vrotsos IA, Madianos PN. A randomized, controlled trial on the effect of non-surgical periodontal therapy


100. Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. JAMA 2001;286:327-34.


Chapter 2


Chapter 3


201. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 2010;6:e1000888.


Chapter 4


Chapter 4 appendix


Chapter 5


Chapter 5 appendix


Chapter 6


267


