RESEARCH REPORTS
Clinical

K. Rhodin[1], K. Divaris[1,2], K.E. North[2,3], S.P. Barros[4], K. Moss[5], J.D. Beck[5], and S. Offenbacher[4]*

[1]Department of Pediatric Dentistry, School of Dentistry, University of North Carolina, Chapel Hill, NC, USA; [2]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA; [3]Carolina Center for Genome Sciences, Chapel Hill, NC, USA; [4]Department of Periodontology, School of Dentistry, University of North Carolina, Chapel Hill, NC, USA; and [5]Department of Dental Ecology, School of Dentistry, University of North Carolina, Chapel Hill, NC, USA; *corresponding author, Steven_Offenbacher@unc.edu

J Dent Res 93(9):882-890, 2014

Chronic Periodontitis Genome-wide Association Studies: Gene-centric and Gene Set Enrichment Analyses

ABSTRACT

Recent genome-wide association studies (GWAS) of chronic periodontitis (CP) offer rich data sources for the investigation of candidate genes, functional elements, and pathways. We used GWAS data of CP (n = 4,504) and periodontal pathogen colonization (n = 1,020) from a cohort of adult Americans of European descent participating in the Atherosclerosis Risk in Communities study and employed a MAGENTA approach (*i.e.*, meta-analysis gene set enrichment of variant associations) to obtain gene-centric and gene set association results corrected for gene size, number of single-nucleotide polymorphisms, and local linkage disequilibrium characteristics based on the human genome build 18 (National Center for Biotechnology Information build 36). We used the Gene Ontology, Ingenuity, KEGG, Panther, Reactome, and Biocarta databases for gene set enrichment analyses. Six genes showed evidence of statistically significant association: 4 with severe CP (*NIN,* $p = 1.6 \times 10^{-7}$; *ABHD12B,* $p = 3.6 \times 10^{-7}$; *WHAMM,* $p = 1.7 \times 10^{-6}$; *AP3B2,* $p = 2.2 \times 10^{-6}$) and 2 with high periodontal pathogen colonization (red complex–*KCNK1,* $p = 3.4 \times 10^{-7}$; *Porphyromonas gingivalis–DAB2IP,* $p = 1.0 \times 10^{-6}$). Top-ranked genes for moderate CP were *HGD* ($p = 1.4 \times 10^{-5}$), *ZNF675* ($p = 1.5 \times 10^{-5}$), *TNFRSF10C* ($p = 2.0 \times 10^{-5}$), and *EMR1* ($p = 2.0 \times 10^{-5}$). Loci containing *NIN, EMR1, KCNK1*, and *DAB2IP* had showed suggestive evidence of association in the earlier single-nucleotide polymorphism–based analyses, whereas *WHAMM* and *AP2B2* emerged as novel candidates. The top gene sets included severe CP ("endoplasmic reticulum membrane," "cytochrome P450," "microsome," and "oxidation reduction") and moderate CP ("regulation of gene expression," "zinc ion binding," "BMP signaling pathway," and "ruffle"). Gene-centric analyses offer a promising avenue for efficient interrogation of large-scale GWAS data. These results highlight genes in previously identified loci and new candidate genes and pathways possibly associated with CP, which will need to be validated via replication and mechanistic studies.

KEY WORDS: periodontal disease, genetics, genomics, bacteria, inflammation, pathway analysis.

INTRODUCTION

Chronic periodontitis (CP) is a common multifactorial oral disease entailing a dysbiotic oral microbial shift and a deregulated host inflammatory response that contribute to progressive periodontal tissue destruction (Berezow and Darveau, 2011; Hajishengallis, 2014). Its clinical manifestations include gingival pocket formation and clinical attachment loss, and it is considered a major cause of tooth loss among most adult populations. According to the most recent epidemiologic data, during 2009-2010 approximately 47% of adults aged 30 years and older in the United States had CP, with 30% having moderate and 8.5% severe periodontitis (Thornton-Evans *et al.*, 2013).

DOI: 10.1177/0022034514544506

Received April 27, 2014; Last revision July 1, 2014; Accepted July 2, 2014

A supplemental appendix to this article is published electronically only at http://jdr.sagepub.com/supplemental.

Behavioral, lifestyle, and systemic risk factors for CP have been well studied and include, among others, smoking, diabetes, obesity, osteoporosis, and vitamin D deficiency (Genco and Borgnakke, 2013).

The investigation of genetic factors in CP has been based on early twin and family studies (Michalowicz *et al.*, 1991), candidate gene studies (Laine *et al.*, 2012), and, most recently, genome-wide association studies (GWAS; Schaefer *et al.*, 2010; Divaris *et al.*, 2013; Teumer *et al.*, 2013). This body of literature has substantiated the genetic underpinning of CP, with heritability estimates ranging from 0.38 for clinical attachment loss among monozygotic twins reared apart (Michalowicz *et al.*, 1991) to 0.07 and 0.13 for moderate and severe CP variance explained by GWAS single-nucleotide polymorphisms (SNPs; Divaris *et al*., 2013). Two additional GWAS examined periodontal disease-related traits, including increased periodontal probing depth (Shaffer *et al.*, 2014) and high periodontal pathogen colonization (Divaris *et al.*, 2012).

Due to their agnostic nature and broad genomic coverage, the latest GWAS have offered novel insights into possible genetic influences and regulators of CP, including several promising candidate loci and genes. However, to the best of our knowledge, no genome-wide statistically significant loci have been reported to date for CP or related traits. Challenges common to all GWAS include typically low power to detect small genetic effects, a likely small representation of truly causal or rare variants, and the nonconsideration of prior biological or mechanistic information (McCarthy *et al.*, 2008). The lack of conclusive GWAS findings for CP can additionally be attributed to relatively small to moderate sample sizes and the inability of currently applied phenotypic case definitions to fully capture disease expression patterns (Divaris *et al.*, 2013; Vaithilingam *et al.*, 2014). Moreover, no GWAS of CP to date have included a formal analytical consideration of SNP genomic context and linkage disequilibrium (LD) structure. The implications of the latter can be significant, as gene size, the number of SNPs available per gene, and the local LD structure have been shown to function as confounders of reported association results (Segrè *et al.*, 2010). Accordingly, the purpose of this study was to improve on previous GWAS by conducting a gene-centric and gene set reanalysis of 2 recently conducted GWAS, for CP and periodontal pathogen colonization. The motivation for considering 2 GWAS traits in this report—the clinically determined disease classification and high levels of periodontal pathogen colonization—was to capture a wider spectrum of periodontal disease expression compared to clinical disease alone.

MATERIALS & METHODS

Studies and Participants

We used data from 2 previously reported GWAS conducted among European American participants of the Atherosclerosis Risk in Communities (ARIC) study (ARIC Investigators, 1989), examining CP (n = 4,504; Divaris *et al.*, 2013) and periodontal pathogen colonization (n = 1,020; Divaris *et al.*, 2012). The ARIC study is a prospective cohort investigation of atherosclerosis, cardiovascular disease risk factors, and outcomes, and it initially enrolled 15,792 community-based adults in 4 U.S. communities in 1987 to 1989. As part of the Dental ARIC ancillary study (Beck *et al.*, 2001), a subset of approximately 6,000 participants underwent comprehensive oral-dental examinations between 1996 and 1998. These examinations recorded the number of missing teeth, probing depth, attachment loss, and bleeding upon probing measurements at 6 sites per tooth, including third molars, and it collected gingival crevicular fluid and subgingival microbial plaque samples. The study's 5 clinical examiners were trained and calibrated against a gold standard, with corresponding kappas indicating excellent to outstanding level of agreement (Beck *et al.*, 2001). At the time of the examination, Dental ARIC participants had a mean age of 62 years (range, 53-74) and a balanced sex distribution. Twelve percent were current smokers, and 11% had diabetes mellitus (Table 1).

Genome-wide Association Data

Genotyping, imputation, quality control, and GWAS analysis steps have been described in previous reports (Divaris *et al.*, 2012; Divaris *et al.*, 2013). Genotyping in ARIC was carried out with the Affymetrix Genome-wide Human SNP Array 6.0 chip, and imputation to 2.5 million markers for the European American sample was based on the HapMap II–CEU panel build 36. SNP exclusion criteria were missing data rate >10%, minor allele frequency <5%, and imputation quality score <0.8. Adjustment for population stratification was based on 10 ancestry principal components generated with the EIGENSTRAT program. Analyses were conducted with ProbABEL and a conventional $p < 5 \times 10^{-8}$ genome-wide significance threshold.

The primary traits examined in the GWAS of CP were severe disease *vs.* mild/healthy and moderate disease *vs.* mild/healthy, based on case definition criteria per the Centers for Disease Control and Prevention and the American Academy of Periodontology (Page and Eke, 2007). These periodontal diagnoses in the analytical sample were as follows: severe, 17%; moderate, 43%; and mild/healthy, 40%. Analyses were based on logistic regression models assuming log-additive genetic effects, adjusting for examination center, age, sex, and 10 ancestry principal components.

The traits examined in the GWAS of periodontal pathogen colonization were "high" colonization with bacteria of the red complex, orange complex, *Aggregatibacter actinomycetemcomitans* (*Aa*), and *Porphyromonas gingivalis* (*Pg*). High colonization was defined as being in the top quintile of semiquantitative traits derived via "checkerboard" DNA-DNA hybridization method (Socransky *et al.*, 1994). As in the GWAS of CP, analyses were based on logistic regression models assuming multiplicative genetic effects, adjusting for examination center, age, sex, and 10 ancestry principal components.

Analytical Approach

Gene-centric Association Testing

We used GWAS results sets, including SNP identifiers (reference SNP ID), genomic position (base pair based on the human genome build 36), and GWAS-derived p values for the 6 aforementioned traits (severe CP, moderate CP, and high colonization

Table 1. Genome-wide Association Studies of Chronic Periodontitis and Periodontal Pathogen Colonization

		Periodontitis Classification[b]		
Genome-wide Association Studies	Total[a]	Healthy-Mild	Moderate	Severe
Chronic periodontitis[c]				
Entire sample	4,504 (100)	1,823 (40)	1,920 (43)	761 (17)
Sex				
Females	2,362 (52)	1,173 (50)	923 (39)	266 (11)
Males	2,142 (48)	650 (30)	997 (47)	495 (23)
Smoking status				
Never smoker	2,048 (47)	1,025 (50)	800 (39)	223 (11)
Former smoker	1,840 (42)	626 (34)	853 (46)	361 (20)
Current smoker	514 (12)	143 (28)	216 (42)	155 (30)
Diabetes status				
No	3,984 (89)	1,670 (42)	1,665 (42)	649 (16)
Yes	514 (11)	151 (29)	252 (49)	111 (22)
Periodontal pathogen colonization[d]				
Entire sample	1,020 (100)	416 (41)	415 (41)	189 (19)
Sex				
Females	478 (47)	246 (59)	178 (43)	54 (29)
Males	542 (53)	170 (41)	235 (57)	135 (71)
Smoking status				
Never smoker	410 (41)	204 (50)	151 (37)	55 (30)
Former smoker	468 (47)	166 (40)	211 (52)	91 (50)
Current smoker	121 (12)	42 (10)	42 (10)	37 (20)
Diabetes status				
No	875 (86)	379 (91)	353 (85)	143 (76)
Yes	144 (14)	37 (9)	62 (15)	45 (24)

Distribution of periodontal diagnoses (per Centers for Disease Control and Prevention and the American Academy of Periodontology disease classification), sex, smoking, and diabetes status among the European American participants of the Dental Atherosclerosis Risk in Communities study included in the genome-wide association studies of chronic periodontitis (and periodontal pathogen colonization [B]).
[a]Values in No. (%) per column.
[b]Values in No. (%) per row.
[c]*n* = 4,504 (Divaris *et al.*, 2013).
[d]*n* = 1,020 (Divaris *et al.*, 2012).

for red complex, orange complex, *Aa*, and *Pg*). We employed a MAGENTA approach (*i.e.*, meta-analysis gene set enrichment of variant associations; Segrè *et al.*, 2010) to obtain gene-centric association results, corrected for gene size, number of SNPs in region, and local LD characteristics. On the basis of evidence derived from simulation and real-data applications (Segrè *et al.*, 2010), we initially defined gene regions that extended from 110 Kb upstream to 40 Kb downstream of each gene's most extreme exon boundaries. To examine the influence of these "gene region" definition criteria on our results, we repeated the MAGENTA process using 2 additional gene region boundary definitions: ±50 Kb and ±300 Kb, flanking each gene's most extreme exon boundary.

A multiple testing–corrected gene-centric *p* value threshold for testing 18,307 genes was calculated as 2.7×10^{-6}. We used LocusZoom 1.1 (Pruim *et al.*, 2010) to visualize the genes and loci of interest. We additionally examined genes that were previously implicated in CP (Laine *et al.*, 2012) or were prioritized via bioinformatics tools (Zhan *et al.*, 2014). A comprehensive list of all genes' association results was also generated and has been made available at the http://genomewide.net online repository with links provided in the results section. Reporting of genes was based on HUGO Gene Nomenclature (http://www.genenames.org).

Gene Set and Pathway Enrichment Testing

The gene-centric association results were carried forward into gene set and pathway enrichment analyses (*i.e.*, gene set enrichment analysis [GSEA]), based on a library containing Gene Ontology, Ingenuity, KEGG, Panther, Reactome, and Biocarta gene sets and pathways. A detailed presentation of these databases is presented in the Appendix. For this step, we excluded genes falling into the 3.4-Mb human leucocyte antigen region of chromosome 6 due to their high LD structure, resulting in 3,225 gene sets. As in recent studies employing MAGENTA-based GSEA (Bønnelykke *et al.*, 2013), corrected gene-centric *p* values were ranked in ascending order, and the number of genes in each set or pathway above a predefined significance threshold was calculated. Because CP- and periodontitis-related traits are considered polygenic traits, we used a 75th percentile cutoff per the developer's recommendations; this cutoff provided the optimal power to detect weak genetic associations in analyses reported by Segrè *et al.* (2010). To generate a GSEA *p* value for each gene set or pathway, the number of observed genes above

the 75th threshold was compared with that generated from 10,000 randomly generated gene sets of identical size. Gene sets and pathways below the multiple testing–corrected p value criterion of 0.05/3,225 gene sets (1.6×10^{-5}) were deemed statistically significant. We additionally conducted and report (at the http://genomewide.net repository) GSEA results using the 2 alternative gene boundary definitions (flanking ±50- and ±300-Kb regions), as well as p values calculated at the 95th percentile threshold.

RESULTS

Single Marker–based GWAS Findings

No genome-wide significant association signals were detected in the previously reported single marker–based GWAS of CP; however, 6 loci showed suggestive evidence of association ($p < 5 \times 10^{-6}$), including *NIN, NPY, WNT5A* for severe CP and *NCR2, EMR1*, 10p15 for moderate CP (Divaris *et al.*, 2013). *NPY, EMR1*, and *NCR2* had concordant effect size and direction in an independent sample of 656 adult European American participants of the Health, Aging, and Body Composition Study (Weyant *et al.*, 2004), which were included in a joint meta-analysis. The full results of the discovery GWAS of CP among the European American Dental ARIC participants have been made publicly available at http://genomewide.net/public/aric/dental/periodontitis/CDC/cdc/2vs0_full.txt and http://genomewide.net/public/aric/dental/periodontitis/CDC/cdc1vs0_full.txt, respectively. Similarly, no genome-wide significant association signals emerged in the discovery GWAS analyses of periodontal pathogen colonization (Divaris *et al.*, 2012). However, 16 loci provided suggestive ($P < 5 \times 10^{-6}$) evidence of association, including *KCNK1, FBXO38, UHRF2, IL33, RUNX2, TRPS1, CAMTA1*/*VAMP3, OTOF*, and *DAB2IP*. The list of association results for the top 10,000 SNPs for each trait has also been made publicly available at http://genomewide.net/public/aric/dental/infectogenomics/topSNPs.xls.

Gene-centric Association Analysis Results

Of 18,307 genes tested, 6 showed genome-wide statistically significant evidence of association with our dental phenotypes—4 with severe CP, 1 for red complex high colonization, and 1 for *Pg* high colonization (Table 2):

Severe CP: *NIN* ($p = 1.6 \times 10^{-7}$), *ABHD12B* ($p = 3.6 \times 10^{-7}$), *WHAMM* ($p = 1.7 \times 10^{-6}$), *AP3B2* ($p = 2.2 \times 10^{-6}$)
Red complex: *KCNK1* ($p = 3.4 \times 10^{-7}$)
Pg: *DAB2IP* ($p = 1.0 \times 10^{-6}$)

Among the 6 significant genes, 2 pairs were adjacent. The 4 loci including these genes are presented in Figure 1. Additional top-ranked genes ($p < 10^{-4}$) for moderate CP were *HGD* ($p = 1.4 \times 10^{-5}$), *ZNF675* ($p = 1.5 \times 10^{-5}$), *TNFRSF10C* ($p = 2.0 \times 10^{-5}$), and *EMR1* ($p = 2.0 \times 10^{-5}$). With the exception of *VDR* (severe CP: $p = 0.09$, moderate CP: $p = 0.03$), no other previously reported candidate-study genes showed any evidence of association. Three genes that were prioritized in the Zhan *et al.* (2014) report showed nominal evidence of association with severe CP: *TNFSF14* ($p = 3.1 \times 10^{-2}$), *GADD45B* ($p = 4.6 \times 10^{-2}$), and *VAV1* ($p = 4.1 \times 10^{-2}$). Noteworthy, *VAV1* had been prioritized by the parent single-marker GWAS for its association with moderate CP, and its gene-centric association in the present analysis had a p value of 4.4×10^{-5}.

Little variation in the patterns of association of genome-wide significant genes was noted upon examination of alternative gene boundary definitions: *NIN*/*ABHD12B, KCNK1*, and *DAB2IP* remained genome-wide significantly associated with disease and bacterial traits, whereas *WHAMM* (*WHDC1*)/*AP3B2* remained strongly associated severe CP ($p < 10^{-5}$) in ±300-Kb gene boundary analyses. Additional genes in the already highlighted loci emerged with the use of the 300-Kb flanking gene region definition. The full list of gene-centric association results are available at http://genomewide.net/public/aric/dental/gene-centric/CP_Bact_Genes.xls, whereas the additional exploratory analysis results are presented at http://genomewide.net/public/aric/dental/gene-centric/CP_Bact_Genes_50Kb.xls (±50-Kb gene region definition) and http://genomewide.net/public/aric/dental/gene-centric/CP_Bact_Genes_300Kb.xls (±300-Kb gene region definition).

Gene Set and Pathway Enrichment Analysis Results

None of the 3,225 gene sets tested met statistical significance criteria after correction for multiple testing (Table 3). "Endoplasmic reticulum membrane" was the gene set with the lowest GSEA p value (2.2×10^{-5}) for the severe CP trait. Other high-ranked sets were "cytochrome P450" ($p = 6.3 \times 10^{-5}$) for severe CP and "regulation of gene expression" for moderate CP. With regard to periodontal pathogen colonization "circadian clock system" was the top gene set for red complex, "G alpha Z signaling events" for orange complex, "KEGG mismatch repair" for *Aa*, and "protein binding" for *Pg*. Of the 6 genome-wide significant genes, 4 were members of prioritized gene sets (Table 3): *NIN* and *AP3B2* are members of "protein binding" (red complex); *WHAMM* and *AP3B2* are members of "Golgi apparatus" (*Pg*); and *DAB2IP* is a member of "intracellular" (*Pg*).

The pathway rankings were substantially altered when we carried forward to GSEA results derived from ±50- and ±300-Kb gene-centric association analyses. The large protein binding gene set (>1,000 members) that was the top one in the main analyses for *Pg* colonization was significantly enriched ($p = 9.9 \times 10^{-6}$) when we utilized the ±300-Kb gene region definition. Other associations, including "microsome" and "oxidation reduction" (severe CP) and BMP signaling pathway (moderate CP), remained relatively robust to alternative gene boundary definitions. The full list of GSEA results is available at http://genomewide.net/public/aric/dental/gene-centric/CP_Bact_GeneSets.xls, whereas the additional results based on ±50- and ±300-Kb gene region definitions are presented at http://genomewide.net/public/aric/dental/gene-centric/CP_Bact_GeneSets_50Kb.xls and http://genomewide.net/public/aric/dental/gene-centric/CP_Bact_GeneSet_300Kb.xls, respectively.

Table 2. Top Genes for Periodontal Traits as Determined by Gene-centric Genome-wide Analyses

Rank	Gene	*p* Value	Chromosome	Gene Size, Kb	SNPs in region
		Severe chronic periodontitis *vs.* mild/healthy			
1	***NIN***	1.6×10^{-7}	14	111	277
2	***ABHD12B***	3.6×10^{-7}	14	33	210
3	***WHAMM***	1.7×10^{-6}	15	26	117
4	***AP3B2***	2.2×10^{-6}	15	51	121
5	*CPEB1*	3.4×10^{-6}	15	105	132
		Moderate chronic periodontitis *vs.* mild/healthy			
1	*HGD*	1.4×10^{-5}	3	54	56
2	*ZNF675*	1.5×10^{-5}	19	34	35
3	*TNFRSF10C*	2.0×10^{-5}	8	15	100
4	*EMR1*	2.0×10^{-5}	19	53	232
5	*TNFRSF10B*	2.6×10^{-5}	8	49	121
		Red complex high colonization			
1	***KCNK1***	3.4×10^{-7}	1	58	243
2	*UHRF2*	4.0×10^{-6}	9	94	210
3	*FBXO38*	4.8×10^{-6}	5	59	164
4	*ARGLU1*	1.8×10^{-5}	13	25	160
5	*HTR4*	2.0×10^{-5}	5	203	307
		Orange complex high colonization			
1	*WDR59*	5.0×10^{-6}	16	112	230
2	*FAM82A*	1.8×10^{-5}	2	117	341
3	*C9orf3*	2.4×10^{-5}	9	360	454
4	*CYP1B1*	2.6×10^{-5}	2	9	241
5	*EXOSC5*	2.8×10^{-5}	19	11	100
		Aggregatibacter actinomycetemcomitans high colonization			
1	*JDP2*	5.5×10^{-6}	14	38	214
2	*AAMP*	1.3×10^{-5}	2	6	65
3	*VIL1*	1.4×10^{-5}	2	30	75
4	*CTDSP1*	1.6×10^{-5}	2	6	62
5	*C2orf62*	1.7×10^{-5}	2	11	68
		Porphyromonas gingivalis high colonization			
1	***DAB2IP***	1.0×10^{-6}	9	218	304
2	*CIB4*	7.3×10^{-6}	2	60	148
3	*OTOF*	9.3×10^{-6}	2	101	252
4	*LOC339778*	1.1×10^{-5}	2	17	207
5	*ANGEL2*	1.5×10^{-5}	1	24	68

Periodontal traits: severe and moderate periodontitis and high colonization (top quintile) with bacteria of the red complex, orange complex, *Aa*, and *Pg*. Corrected for gene size, single-nucleotide polymorphisms (SNPs) in region, and local disequilibrium structure. Based on European American participants of the Dental Atherosclerosis Risk in Communities study. Genome-wide statistical significance treshold for testing of 18,307 genes (NCBI36/hg18) = 2.7×10^{-6} (genes meeting this criterion in bold).

DISCUSSION

In this article, we present a gene-centric reanalysis of 2 recently conducted single marker–based GWAS and offer novel insights into candidate genes, possible gene sets, and pathways underlying CP. Our results highlight 4 previously implicated genes, as well as additional candidates and pathways for consideration. Of note, 4 of 6 genome-wide significant genes and the lowest GSEA *p* values for gene sets were found for association with severe CP. This is unsurprising because genetic influences may be more pronounced and easier to detect for severe phenotypes, such as severe CP, which is also known to be more heritable than milder forms of the disease. Upon replication and mechanistic confirmation, these findings can enhance our understanding of the disease pathogenesis and aid in the development of new risk assessment and preventive and therapeutic modalities.

Several promising loci and candidate genes were identified or confirmed in the present analysis, as shown in Figure 2. Of those showing statistically significant evidence of association, *NIN* and *ABHD12B* (severe CP), *KCNK1* (red complex) and *DAB2IP* (*Pg*) were in loci that had showed suggestive evidence of association in the previous single marker–based analyses (Divaris *et al.*, 2013; Divaris *et al.*, 2012); *WHAMM* and *AP3B2* emerged as novel associations. Although a formal discussion of these genes as novel candidates for periodontitis would require replication in independent cohorts, a brief presentation of their

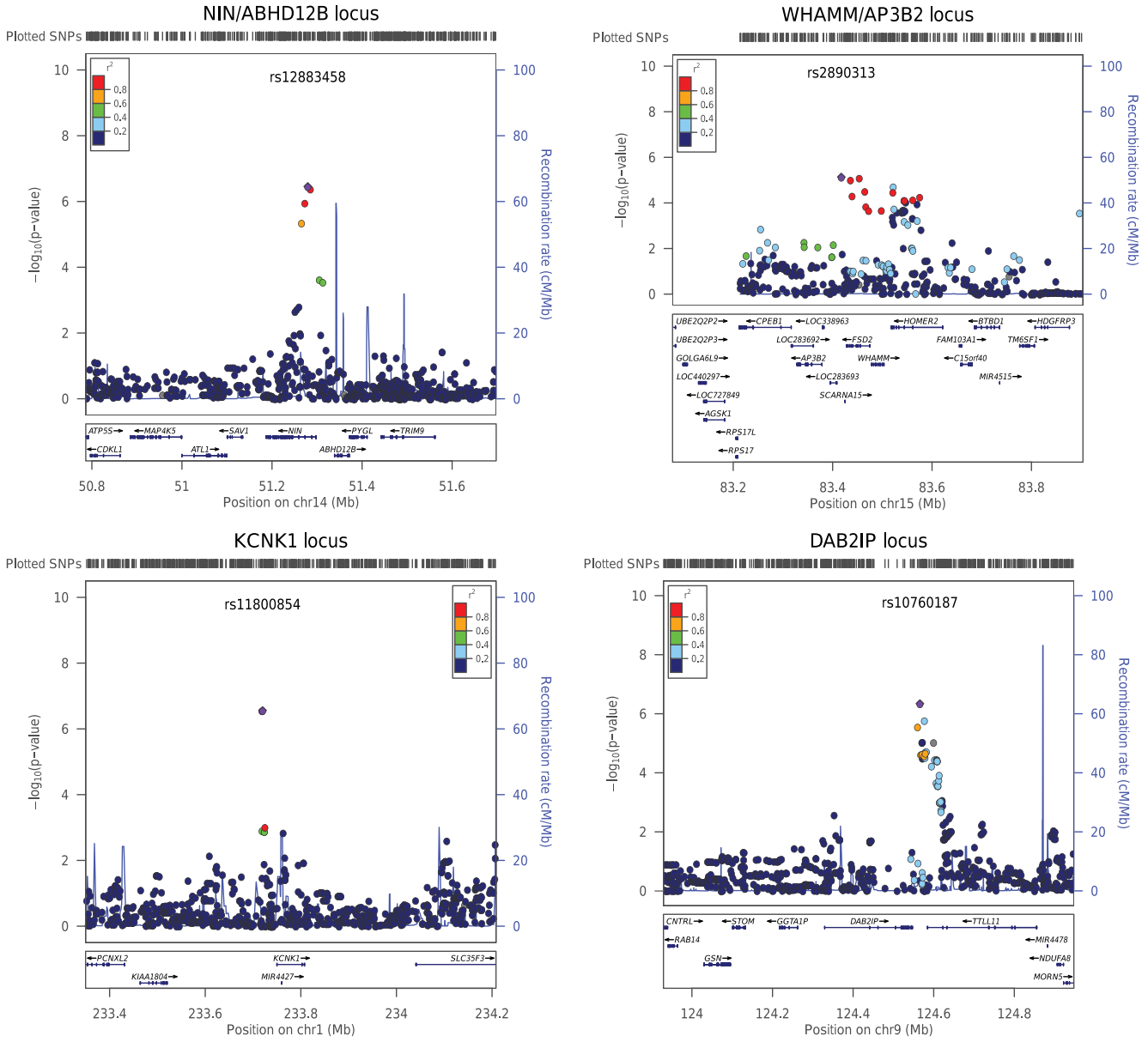

Figure 1. Visualization of the 4 loci that contained genome-wide significant genes, as determined by genome-wide gene-centric analyses of chronic periodontitis and periodontal pathogen colonization, among the European American participants of the Dental Atherosclerosis Risk in Communities Study cohort (n = 4,504). Four identified genes (*NIN, ABHD12B, WHAMM, AP3B2*) showed genome-wide evidence of association with severe chronic periodontitis, *KCNK1* with red complex high colonization, and *DAB2IP* with *Pg* high colonization. The vertical axis corresponds to each marker's associated –log10 *p* value. The overlaid recombination rate plot and the pairwise linkage disequilibrium values with index single-nucleotide polymorphisms were calculated per HapMap II–CEU.

putative biological relevance can be informative. *NIN* encodes a centrosomal microtubule organization and anchoring protein and was recently implicated in breast cancer risk in a case-control study of invasive breast cancer (Olson *et al.*, 2011). Importantly, the centrosome is known to play a pivotal role in the polarization and lytic granule delivery potential of cytotoxic T lymphocytes (Jenkins *et al.*, 2009), providing a possible link with pathogenesis of CP. Although the function of *ABHD12B* remains to be elucidated, this gene was involved in 1 of 4 significant gene-gene interactions reported as associated with longitudinal changes in ventricle size in the Alzheimer's Disease Neuroimaging Initiative cohort (Koran *et al.*, 2014). *WHAMM* functions at the interface of the microtubule and actin cytoskeletons and thus appears to be important for the regulation of the cell membrane tabulation (Campellone *et al.*, 2008). *AP3B2* encodes a subunit of the adaptor protein 3 complex, which appears to have neuron-specific functions, including neurotransmitter release via involvement in the formation of neurosecretory

Table 3. Top Gene Sets and Pathways for Severe and Moderate Periodontitis as Determined by Gene Set Enrichment Analysis

Rank	DB	Gene Set	Size (No. of Genes)	> 75th Percentile		*p* Value
				Expected	Observed	
		Severe CP *vs.* mild/healthy				
1	GO	Endoplasmic reticulum membrane	531	118	156	2.2×10^{-5}
2	GO	Cytochrome P450 arranged by substrate type	49	11	23	6.3×10^{-5}
3	GO	Microsome	232	51	75	2.0×10^{-4}
4	GO	Oxidation reduction	537	116	148	2.8×10^{-4}
5	GO	Phase 1 functionalization of compounds	67	14	26	3.0×10^{-4}
		Moderate CP *vs.* mild/healthy				
1	GO	Regulation of gene expression	34	8	19	1.0×10^{-4}
2	RE	Zinc ion binding	1,988	377	422	5.0×10^{-4}
3	GO	BMP signaling pathway	42	10	19	1.0×10^{-3}
4	GO	Ruffle	54	13	23	1.1×10^{-3}
5	RE	Stem cell maintenance	11	3	8	1.1×10^{-3}
		Red complex high colonization				
1	PA	Circadian clock system	9	2	8	1.0×10^{-4}
2	RE	Signaling by NGF	215	50	70	6.0×10^{-4}
3	GO	Protein binding	6,648	1,179	1,194	7.0×10^{-4}
4	GO	B cell differentiation	34	8	17	1.3×10^{-3}
5	PM	Gap junction	19	4	9	1.3×10^{-3}
		Orange complex high colonization				
1	RE	G alpha Z signaling events	14	4	10	1.0×10^{-4}
2	PB	Protein ADP ribosylation	10	3	8	2.0×10^{-4}
3	GO	Female pregnancy	66	16	27	6.0×10^{-4}
4	RE	Adenylate cyclase activating pathway	11	3	8	8.0×10^{-4}
5	GO	Endoplasmic reticulum membrane	531	118	148	9.0×10^{-4}
		Aggregatibacter actinomycetemcomitans high colonization				
1	KE	KEGG mismatch repair	23	6	14	1.0×10^{-4}
2	PB	Extracellular matrix protein-mediated signaling	62	15	26	9.0×10^{-4}
3	GO	Synapse	246	56	75	1.1×10^{-3}
4	GO	Intra-Golgi vesicle-mediated transport	16	4	10	1.3×10^{-3}
5	GO	Mismatched DNA binding	15	3	8	1.3×10^{-3}
		Porphyromonas gingivalis high colonization				
1	GO	Protein binding	6,648	1,172	1,188	1.0×10^{-4}
2	GO	Filopodium	25	6	14	4.0×10^{-4}
3	GO	Golgi apparatus	809	182	214	9.0×10^{-4}
4	BC	CARM ER pathway	35	8	16	1.3×10^{-3}
5	GO	Intracellular	1,911	356	392	1.4×10^{-3}

BC, Biocarta; GO, Gene Ontology; KE, KEGG; PA, Panther; PB, Panther-Biological Process; RE, Reactome.
Per the 75th percentile cutoff among the 4,504 European American participants of the Dental Atherosclerosis Risk in Communities study (Divaris *et al.*, 2013), based on NCBI36/hg18. Genome-wide statistical significance treshold for testing of 3,225 gene sets was 1.6×10^{-5}.

vesicles (Grabner *et al.*, 2006). Finally, *ERM1* is involved in the host immune response and has recently been recognized as a highly specific marker for eosinophils in humans (Hamann *et al.*, 2007).

Gene set and pathway analyses provided an additional layer of information on possible pathways and processes involved in periodontal health and disease. While no gene set displayed genome-wide significant associations with the examined periodontal traits, several high-ranked associations merit being highlighted. The finding of "endoplasmic reticulum membrane" as the top gene set associated with severe periodontitis parallels recent emerging data on the importance of endoplasmic reticulum stress in periodontitis (Domon *et al.*, 2009; Lee *et al.*, 2012). "Circadian clock rhythm" was the top enriched pathway for high colonization with red complex pathogens and has gained recent attention in oral health research (Papagerakis *et al.*, 2014). Interestingly, Keller *et al.* (2009) recently reported that on the molecular level, >8% of the macrophage transcriptome oscillates in a circadian fashion, including many important regulators for pathogen recognition and cytokine secretion. Moreover, experimental evidence (Barros *et al.*, 2010) supports the possible association of BMP signaling with periodontal disease traits that we identified in this report (moderate CP and orange complex colonization).

Our genome-wide gene-centric analysis approach has some limitations. Most important, it relies on important assumptions

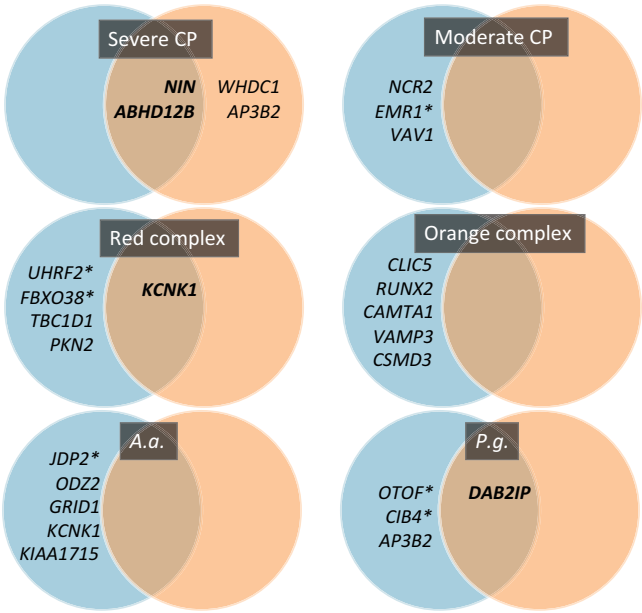

Figure 2. Venn diagrams of genes *in loci* prioritized in the single-marker genome-wide association studies of chronic periodontitis (CP; severe *vs.* mild/healthy and moderate *vs.* mild/healthy; Divaris *et al.*, 2013) and high periodontal pathogen colonization (red complex, orange complex, *Aggregatibacter actinomycetemcomitans* [*Aa*], and *Porphyromonas gingivalis* [*Pg*]; Divaris *et al.*, 2012] based on a $p < 5 \times 10^{-6}$ criterion (blue circles) and showing genome-wide statistically significant evidence of association ($p < 2.7 \times 10^{-6}$) in gene-centric analyses for the same traits (brown circles). Genes denoted with asterisks (*) were prioritized in single-marker genome-wide association studies and were among the top 5 in gene-centric analyses without meeting genome-wide statistical significance criteria.

relative to gene mapping, boundaries, and regions upon which SNPs-based signals are clustered on the basis of the NCBI36 human genome build. For our main analyses, we considered a 100-Kb upstream/40-Kb downstream gene boundary definition, which has been shown as the most efficient and has been employed in recent genome-wide gene-centric investigations. Our top gene-centric association findings were relatively robust to changes in this specification; however, it is likely that regions of varying length are functionally important for different genes (Qu and Fang, 2013). The fact that gene set and pathway enrichment analysis results were largely negative and showed substantial variation depending on gene boundary definitions warrants further investigation. It is likely that a more efficient and less multiple testing–penalized GSEA approach may involve testing a smaller number of gene sets compared to all available gene sets and pathways available to date that we examined in this study.

In this report, we demonstrate the potential utility of gene-centric analyses in the context of GWAS. We adjusted SNP-based association results for genomic context, SNP density, and LD structure to enable gene-based inferences. This step reduces the multiple-testing burden that is inherent in GWAS, and it facilitates downstream gene set and pathway analyses, as well as across-cohort and transethnic gene-based meta-analyses. Nevertheless, our study is limited by the fact that it is based on a single cohort and no replication is available. However, to enable future replication studies, as more GWAS of periodontal traits are undertaken, we have made publicly available our full results of gene-centric association and GSEA and thus provide a rich data source for hypothesis generation, replication, and mechanistic studies. The nonoverlap with previous GWAS loci reviewed by Vaithilingam *et al.* (2014) is likely due to the different disease traits employed in these GWAS (*e.g.*, aggressive periodontitis by Schaefer *et al.*, 2010).

Although a recent development in dental research, GWAS have advanced our understanding of genetic contributions to the risk and pathophysiology of CP by highlighting suggestive genetic loci that may be implicated in periodontitis. Additional studies based on large population samples and high-quality phenotypes are warranted to add to the knowledge base created by the initial GWAS of CP. Pooling of existing data and studying biologically refined and incident (*vs.* prevalent) periodontal outcomes may offer additional insights. Gene-centric analyses, as demonstrated in this study, offer a promising avenue for efficient interrogation of such large-scale GWAS data and may aid in the discovery of additional genetic associations, above and beyond those highlighted by single marker–based discovery GWAS.

ACKNOWLEDGMENTS

The Atherosclerosis Risk in Communities Study was carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HSN268201100009C, HHSN268201100010C, HHSN268201100011C, HSN268201100012C, R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute contract U01HG004402; National Institutes of Health contract HHSN268200625226C; National Institute of Environmental Health Sciences grant P30ES010126; and National Institute of Dental and Craniofacial Research grants R01DE11551 and R01DE021418. Infrastructure was partly supported by grant UL1RR025005, a component of the National Institutes of Health and the National Institutes of Health Roadmap for Medical Research. The authors thank the staff and participants of the Atherosclerosis Risk in Communities study for their important contributions and Dr. Ayellet V. Segrè for her support and guidance with respect to analyses involving MAGENTA. The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.

REFERENCES

Barros SP, Wirojchanasak S, Barrow DA, Panagakos FS, Devizio W, Offenbacher S (2010). Triclosan inhibition of acute and chronic inflammatory gene pathways. *J Clin Periodontol* 37:412-418.

Beck JD, Elter JR, Heiss G, Couper D, Mauriello SM, Offenbacher S (2001). Relationship of periodontal disease to carotid artery intima-media wall thickness: the atherosclerosis risk in communities (ARIC) study. *Arterioscler Thromb Vasc Biol* 21:1816-1822.

Berezow AB, Darveau RP (2011). Microbial shift and periodontitis. *Periodontol* 2000 55:36-47.

Bønnelykke K, Matheson MC, Pers TH, Granell R, Strachan DP, Alves AC, *et al.*; Australian Asthma Genetics Consortium; Early Genetics and

Lifecourse Epidemiology Consortium (2013). Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. *Nat Genet* 45:902-906.

Campellone KG, Webb NJ, Znameroski EA, Welch MD (2008). WHAMM is an Arp2/3 complex activator that binds microtubules and functions in ER to Golgi transport. *Cell* 134:148-161.

Divaris K, Monda KL, North KE, Olshan AF, Lange EM, Moss K, *et al*. (2012). Genome-wide association study of periodontal pathogen colonization. *J Dent Res* 91(7):21S-28S.

Divaris K, Monda KL, North KE, Olshan AF, Reynolds LM, Hsueh WC, *et al*. (2013). Exploring the genetic basis of chronic periodontitis: a genome-wide association study. *Hum Mol Genet* 22:2312-2324.

Domon H, Takahashi N, Honda T, Nakajima T, Tabeta K, Abiko Y, *et al*. (2009). Up-regulation of the endoplasmic reticulum stress-response in periodontal disease. *Clin Chim Acta* 401:134-140.

Genco RJ, Borgnakke WS (2013). Risk factors for periodontal disease. *Periodontol* 2000 62:59-94.

Grabner CP, Price SD, Lysakowski A, Cahill AL, Fox AP (2006). Regulation of large dense-core vesicle volume and neurotransmitter content mediated by adaptor protein 3. *Proc Natl Acad Sci U S A* 103:10035-10040.

Hajishengallis G (2014). Immunomicrobial pathogenesis of periodontitis: keystones, pathobionts, and host response. *Trends Immunol* 35:3-11.

Hamann J, Koning N, Pouwels W, Ulfman LH, van Eijk M, Stacey M, *et al*. (2007). EMR1, the human homolog of F4/80, is an eosinophil-specific receptor. *Eur J Immunol* 37:2797-2802.

Jenkins MR, Tsun A, Stinchcombe JC, Griffiths GM (2009). The strength of T cell receptor signal controls the polarization of cytotoxic machinery to the immunological synapse. *Immunity* 31:621-631.

Keller M, Mazuch J, Abraham U, Eom GD, Herzog ED, Volk HD, *et al*. (2009). A circadian clock in macrophages controls inflammatory immune responses. *Proc Natl Acad Sci U S A* 106:21407-21412.

Koran ME, Hohman TJ, Meda SA, Thornton-Wells TA (2014). Genetic interactions within inositol-related pathways are associated with longitudinal changes in ventricle size. *J Alzheimers Dis* 38:145-154.

Laine ML, Crielaard W, Loos BG (2012). Genetic susceptibility to periodontitis. *Periodontol* 2000 58:37-68.

Lee SI, Kang KL, Shin SI, Herr Y, Lee YM, Kim EC (2012). Endoplasmic reticulum stress modulates nicotine-induced extracellular matrix degradation in human periodontal ligament cells. *J Periodontal Res* 47:299-308.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, *et al*. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9:356-369.

Michalowicz BS, Aeppli D, Virag JG, Klump DG, Hinrichs JE, Segal NL, *et al*. (1991). Periodontal findings in adult twins. *J Periodontol* 62:293-299.

Olson JE, Wang X, Pankratz VS, Fredericksen ZS, Vachon CM, Vierkant RA, *et al*. (2011). Centrosome-related genes, genetic variation, and risk of breast cancer. *Breast Cancer Res Treat* 125:221-228.

Page RC, Eke PI (2007). Case definitions for use in population-based surveillance of periodontitis. *J Periodontol* 78:1387S-1399S.

Papagerakis S, Zheng L, Schnell S, Sartor MA, Somers E, Marder W, *et al*. (2014). The circadian clock in oral health and diseases. *J Dent Res* 93:27-35.

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, *et al*. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26:2336-2337.

Qu H, Fang X (2013). A brief review on the Human Encyclopedia of DNA Elements (ENCODE) project. *Genomics Proteomics Bioinformatics* 11:135-141.

Schaefer AS, Richter GM, Nothnagel M, Manke T, Dommisch H, Jacobs G, *et al*. (2010). A genome-wide association study identifies GLT6D1 as a susceptibility locus for periodontitis. *Hum Mol Genet* 19:553-562.

Segrè AV; DIAGRAM Consortium; MAGIC investigators; Groop L, Mootha VK, Daly MJ, Altshuler D (2010). Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* 6:e1001058.

Shaffer JR, Polk DE, Wang X, Feingold E, Weeks DE, Lee MK, *et al*. (2014). Genome-wide association study of periodontal health measured by probing depth in adults ages 18-49 years. *G3 (Bethesda)* 4:307-314.

Socransky SS, Smith C, Martin L, Paster BJ, Dewhirst FE, Levin AE (1994). "Checkerboard" DNA-DNA hybridization. *Biotechniques* 17:788-792.

Teumer A, Holtfreter B, Völker U, Petersmann A, Nauck M, Biffar R, *et al*. (2013). Genome-wide association study of chronic periodontitis in a general German population. *J Clin Periodontol* 40:977-985.

The ARIC Investigators (1989). The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol* 129:687-702.

Thornton-Evans G, Eke P, Wei L, Palmer A, Moeti R, Hutchins S, *et al*.; Centers for Disease Control and Prevention (2013). Periodontitis among adults aged ≥30 years: United States, 2009-2010. *MMWR Surveill Summ* 62(Suppl 3):129-135.

Vaithilingam RD, Safii SH, Baharuddin NA, Ng CC, Cheong SC, Bartold PM, *et al*. (2014). Moving into a new era of periodontal genetic studies: relevance of large case-control samples using severe phenotypes for genome-wide association studies. *J Periodontal Res*. [Epub ahead of print 2/17/2014]

Weyant RJ, Newman AB, Kritchevsky SB, Bretz WA, Corby PM, Ren D, *et al*. (2004) Periodontal disease and weight loss in older adults. *J Am Geriatr Soc* 52:547-553.

Zhan Y, Zhang R, Lv H, Song X, Xu X, Chai L, *et al*. (2014). Prioritization of candidate genes for periodontitis using multiple computational tools. *J Periodontol* 85:1059-1069.