OXFORD

Human Molecular Genetics, 2016, Vol. 25, No. 10 2113–2129

doi: 10.1093/hmg/ddw069
Advance Access Publication Date: 8 March 2016
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Genome-wide association study of biologically informed periodontal complex traits offers novel insights into the genetic basis of periodontal disease

Steven Offenbacher[1,*], Kimon Divaris[2,4], Silvana P. Barros[1], Kevin L. Moss[3], Julie T. Marchesan[1], Thiago Morelli[1], Shaoping Zhang[1], Steven Kim[1], Lu Sun[1], James D. Beck[3], Matthias Laudes[5], Matthias Munz[6,7], Arne S. Schaefer[6] and Kari E. North[4]

[1]Department of Periodontology, [2]Department of Pediatric Dentistry and [3]Department of Dental Ecology, UNC School of Dentistry, Chapel Hill, NC, USA, [4]Department of Epidemiology, UNC Gillings School of Global Public Health, Chapel Hill, NC, USA, [5]Clinic of Internal Medicine I, University Clinic Schleswig-Holstein, Kiel, Germany, [6]Department of Periodontology, Institute of Dental, Oral and Maxillary Medicine, Charité–University Medicine Berlin, Berlin, Germany and [7]Institute of Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany

*To whom correspondence should be addressed at: Center for Oral and Systemic Diseases, Department of Periodontology, 3501F Koury Oral Health Sciences Building, 385 S. Columbia Street, Chapel Hill, NC 27599-7455, USA. Tel: +1 9195373205; Fax: +1 9198431798; Email: steven_offenbacher@unc.edu

Abstract

Genome-wide association studies (GWAS) of chronic periodontitis (CP) defined by clinical criteria alone have had modest success to-date. Here, we refine the CP phenotype by supplementing clinical data with biological intermediates of microbial burden (levels of eight periodontal pathogens) and local inflammatory response (gingival crevicular fluid IL-1β) and derive periodontal complex traits (PCTs) via principal component analysis. PCTs were carried forward to GWAS (~2.5 million markers) to identify PCT-associated loci among 975 European American adult participants of the Dental ARIC study. We sought to validate these findings for CP in the larger ARIC cohort ($n = 821$ participants with severe CP, 2031—moderate CP, 1914—healthy/mild disease) and an independent German sample including 717 aggressive periodontitis cases and 4210 controls. We identified six PCTs with distinct microbial community/IL-1β structures, although with overlapping clinical presentations. PCT1 was characterized by a uniformly high pathogen load, whereas PCT3 and PCT5 were dominated by *Aggregatibacter actinomycetemcomitans* and *Porphyromonas gingivalis*, respectively. We detected genome-wide significant signals for PCT1 (*CLEC19A, TRA, GGTA2P, TM9SF2, IFI16, RBMS3*), PCT4 (*HPVC1*) and PCT5 (*SLC15A4, PKP2, SNRPN*). Overall, the highlighted loci included genes associated with immune response and epithelial barrier function. With the exception of associations of *BEGAIN* with severe and *UBE3D* with moderate CP, no other loci were associated with CP in ARIC or aggressive periodontitis in the German sample. Although not associated with current clinically determined periodontal disease taxonomies, upon replication and mechanistic validation these candidate loci may highlight dysbiotic microbial community structures and altered inflammatory/immune responses underlying biological sub-types of CP.

Received: June 4, 2015. **Revised:** February 19, 2016. **Accepted:** February 26, 2016

Introduction

Periodontal disease is a condition that is associated with an exaggerated inflammatory response to the oral biofilm. In the absence of adequate oral hygiene, disease initiates as inflamed gingiva (gingivitis), but in susceptible individuals it can progress to periodontitis that is characterized by local periodontal tissue destruction with resorption of alveolar bone, loss of supporting ligament and tooth loss. Recent studies have shown that periodontal conditions are quite prevalent in humans with approximately 90% of the US adult population exhibiting at least some form of gingivitis and 47% of the population with periodontitis; with the severe form of periodontitis affecting 18% of those individuals over the age of 65 (1,2). Epidemiological studies indicate that poor oral hygiene is associated with more severe disease, but chronic plaque accrual is not as strong a predictor of disease in the population as other factors such as smoking or diabetes (3–5).

Although the oral microbes are believed to be essential to the causal pathway, they are not sufficient. Studies in twins suggest a strong heritability estimate for adult periodontal disease (i.e. ~50%) and a polygenic predisposition is presumed (6,7). Recent analyses have indicated that periodontitis is associated with a shift in the normal human commensal oral microbiome resulting in different microbial colonization patterns that represent distinct, but stable, dysbiotic communities associated with disease (8,9). For example, aggressive forms of periodontal disease have been associated with a biofilm composition which is dominated by an emergent periodontal pathogen, such as *Porphyromonas gingivalis* (10) or *Aggregatibacter actinomycetemcomitans* (11). These dysbiotic microbial states are often accompanied with altered inflammatory signatures and more severe clinical disease. More prevalent forms of chronic periodontal disease have lower levels of these two microorganisms with clustered combination of pathogens including *Treponema denticola (Td), Tannerella forsythia (Tf), Campylobacter rectus (Cr)* and *Prevotella nigrescens (Pn)* (12). Viral contribution to dysbiosis also seems to be relevant. During periods of disease exacerbation, there are reported concomitant increases in the levels of tissue viral expression including Herpes and Cytomegalovirus species (13).

Recent genome-wide association studies (GWAS) of chronic periodontitis (CP) by our and other groups have highlighted loci potentially associated with clinically derived disease definitions (14–17) and high levels of specific periodontal pathogens (15,18) among populations of European descent. No single-marker association in these reports met strict genome-wide significance criteria; however, four loci (*NIN/ABHD12B*, *WHAMM/AP3B2*, *KCNK1* and *DAB2IP*) met gene-centric statistical significance criteria, whereas *NPY* was independently reported by two different investigations (14,19). Taken together, these results underscore the need to continue exploring for the genetic basis of CP using larger or pooled samples and high-quality phenotypes.

Conceptually, genetic variants that are causally related to the clinical manifestation of disease are likely to be associated with altered biological intermediates or 'endophenotypes' which have the potential to modify host barrier function, inflammatory responses and microbial colonization patterns. In this model, we suggest that periodontal disease is actually a group of distinct conditions with similar and overlapping clinical presentations. Each of these conditions is influenced by human genetic variation that results in a non-protective inflammatory response that interacts with the normal commensal biofilm to induce dysbiotic microbial shifts and clinical disease. We hypothesize that human genetic variation influences specific dysbiotic shifts in the oral flora and altered inflammatory signature for each individual that could provide insight into identifying the genetic influences on disease expression. There is precedent for this concept. For example, Elinav *et al.* (20) have demonstrated that mutations in the NLRP6 inflammasome complex regulate both the composition of the colonic microbial ecology, as well as the cytokine and chemokine inflammatory pathways associated with colitis. In this investigation, we define the disease phenotype as complex traits using principal component analyses (PCA) to identify genetic determinants, as described by Suo *et al.* (21) and others (22,23). Our PCA models create distinct periodontal complex traits (PCTs) based upon subgingival microbial composition data and biomarkers of the tissue inflammatory response (IL-1β) to interrogate in GWA analysis. The intent is to identify loci related to the biological underpinning and pathogenesis of CP and not strictly to serve as population markers for clinical CP.

Results

Defining PCTs

The demographic, risk factor and clinical characteristics of the study participants are presented in Supplementary Material, Table S1. We used PCA and biological intermediate (microbial and inflammatory) characteristics to define PCTs for interrogation in GWA analyses, using trait analyses methods described previously (24,25). PCA was carried out among 975 participants including microbial (eight pathogens), inflammatory mediator [gingival crevicular fluid (GCF)-IL1®] and clinical disease classification (CDC/AAP chronic periodontitis classification). Using the three-level CDC/AAP classification there were 394 (40%) subjects with periodontal health, which includes subjects with gingivitis, 389 (39.9%) subjects with entry-level disease (mild–moderate periodontitis) and 192 (19.7%) subjects with severe periodontal disease. Participants had a mean age of 63 years, were approximately equally distributed by sex, 13% had a history of smoking and 14% had diabetes mellitus.

The eigenvalues (loadings) for the microbial and inflammatory variables for the first six PCTs (PCT1–PCT6) are presented in Supplementary Material, Table S2. These six PCTs are shown selecting those that account for >5% of the variance. The three-level CDC/AAP disease classification is included as a control variable. Variables with eigenvalues of >0.4 or <–0.4 are generally considered as significantly contributing to the principal component (26,27); the values for each microbial and inflammatory variable are also shown. Each PCT defines a specific microbial community structure with varying levels of IL-1β. The first principal component, which defines the dominant PCT1 explains 46.5% of the variance (Supplementary Material, Table S2). PCT2 through PCT5 each explain about 10% of the variance each with the first six complex traits together accounting for 90.2% of the total variance. The remaining five principal components each explain <5% of the variance, are not associated with clinical disease and have no significantly associated GWAS loci. However, PCT7 through PCT11 define five different microbial community structures that are consistent with health states (data not shown).

Figure 1 shows the principal component profile illustrating correlations with the constitutive parameters. PCT1 has a microbial community structure that has a high correlation and positive loading with all pathogens (Supplementary Material, Table S2). Since these organisms are historically associated with the microbial clusters originally identified by Socransky *et al.* (28), we are referring to this first PCT as the Socransky Trait. PCT2 is positively associated with severe periodontal disease and has weak positive loading of *P. gingivalis* and IL-1β. PCT3 is characterized by a high

positive loading of GCF-IL-1β and *A. actinomycetemcomitans* (*Aa* trait) (Fig. 1). PCT4 is associated with a mixed-infection community with high positive loading of GCF-IL1®, weak positive loadings of *Pn*, *Pi* and *Td* and low levels of *Aa*. PCT5 has the highest loading of *P. gingivalis* across all PCTs with an eigenvalue of 0.882 (*Pg* trait) followed by *A. actinomycetemcomitans* (0.134) (Supplementary Material, Table S2) and the correlation with *P. gingivalis* can be seen in Figure 1. Notably, no microbial or inflammatory variables are associated with PCT6, but this trait is positively correlated with CDC/AAP entry and severe clinical categories. Complex traits PCT7–PCT11 are associated with health and have microbial community structures dominated by low counts of other organisms including *P. nigrescens*, *T. denticola*, *T. forsythia* and *P. intermedia* with lower clinical plaque scores (data not shown).

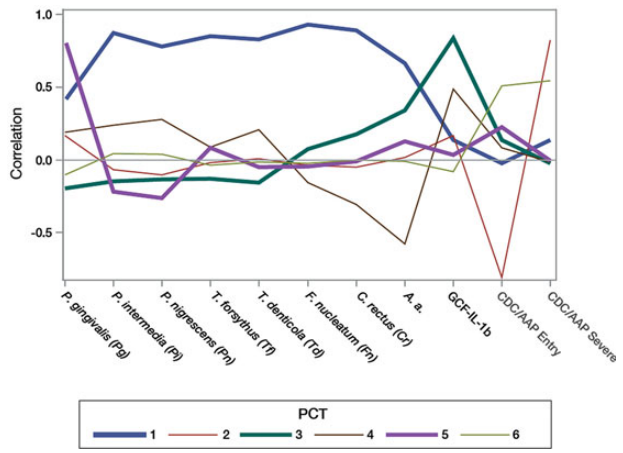

Figure 1. Principal component trait (PCT) pattern profiles.

Associations of PCTs with clinical signs of CP (probing depth, interproximal attachment level, bleeding on probing, gingival index, tooth loss and plaque scores) are presented in Table 1. The clinical signs of periodontal disease, expressed as continuous variables (extent scores), are significantly associated with the first six PCTs, as are categorical classifications of disease states. Further analyses of the subject-level risk factors [sex (female = 0, male = 1), age, Body mass Index (BMI), smoking (pack/year) and diabetes (yes = 1, no = 0)] are presented for each PCT in Table 1. Importantly, the data demonstrate that the PCTs reflect overlapping clinical presentations and emphasize that the current disease classifications that are based upon clinical signs display remarkable microbial and inflammatory heterogeneity. Of note, PCT6 is the trait most strongly associated with clinically determined disease, highest plaque levels and highest association with smoking and diabetes.

Table 2 presents the levels of organisms, concentration of GCF-IL-1β distribution of CDC/AAP disease categories, total microbial counts and clinical signs including number of teeth, measures of interproximal attachment loss (iAL), probing depth (PD), bleeding on probing (BOP), plaque index scores (PI) and gingival inflammation scores (GI) stratified by quartiles (Q1–Q4) of each PCT (1–6). Overall, there is an increase in prevalence in entry and severe disease with increasing quartiles of PCT and concomitant increases in clinical signs of disease, although differences in patterns can be seen. These clinical changes are accompanied by a dysbiotic shift among the PCTs. For example, within PCT1, comparing total counts in PCT1–Q1 versus PCT1–Q4 there is a 54-fold increase in total counts and a 79-fold increase in *P. gingivalis*. In contrast, *F. nucleatum* demonstrates a 1288-fold increase from Q1 to Q4 in PCT1. Similar significant, but perhaps less dramatic, dysbiotic

Table 1. Correlation matrix (Pearson correlation coefficients and *P*-values; bold-face indicates statistically significant correlations) of PCTs with clinical signs of disease and subject characteristics known to modify disease risk

Clinical measures and risk factor characteristics	PCT1 'Socransky trait'	PCT2	PCT3 '*Aa* trait'	PCT4	PCT5 '*Pg* trait'	PCT6
EPDGE4	**0.120**	**0.416**	**0.211**	**0.079**	**0.117**	**0.492**
	0.0002	**<0.0001**	**<0.0001**	**0.0133**	**0.0002**	**<0.0001**
EALGE3I	**0.156**	**0.314**	**0.169**	**0.072**	**0.123**	**0.603**
	<0.0001	**<0.0001**	**<0.0001**	**0.02**	**0.0001**	**<0.0001**
EBL	**0.198**	**0.230**	**0.198**	**0.093**	**0.099**	**0.276**
	<0.0001	**<0.0001**	**<0.0001**	**0.0037**	**0.0019**	**<0.0001**
EGIGE2	**0.231**	**0.164**	**0.167**	**0.014**	**0.078**	**0.128**
	<0.0001	**<0.0001**	**<0.0001**	**0.690**	**0.0242**	**0.0002**
NTEETH	**−0.106**	0.016	**−0.074**	−0.043	0.008	−0.046
	0.0009	0.625	**0.0217**	0.178	0.807	0.151
EPQGE1	**0.137**	**0.131**	**0.149**	**0.074**	**0.081**	**0.240**
	<0.0001	**<0.0001**	**<0.0001**	**0.0264**	**0.0144**	**<0.0001**
Sex	0.021	**0.067**	0.021	0.004	0.032	**0.205**
	0.513	**0.0368**	0.515	0.895	0.319	**<0.0001**
Age	0.039	0.024	−0.044	0.032	0.035	**0.127**
	0.225	0.450	0.170	0.326	0.280	**<0.0001**
BMI	0.044	0.032	**0.065**	0.015	0.023	**0.067**
	0.168	0.323	**0.0426**	0.644	0.465	**0.0368**
Smoking	0.003	**0.078**	0.027	0.014	0.021	**0.230**
	0.935	**0.0158**	0.410	0.677	0.512	**<0.0001**
Diabetes	0.043	**0.070**	0.059	0.024	−0.007	**0.144**
	0.183	**0.0292**	0.0661	0.453	0.838	**<0.0001**

EPDGE4, extent probing depth ≥4 mm; EALGE3I, extent interproximal clinical attachment loss ≥3 mm; EBL, extent bleeding score; EGIGE2, extent gingival index >1; Sex (female = 0, male = 1); Age (years); BMI, body mass index; Smoking (pack/years); Diabetes (yes = 1, no = 0).

Table 2. Association of microbiological (log-levels of eight periodontal pathogens), inflammatory [log-level of GCF-Interleukin (IL)-1β] and clinical (CDC/AAP chronic periodontitis classification, number of teeth and 'extent' scores) parameters with PCT (defined as quartile-categorical variable) among the sample of 975 European American participants of the Dental ARIC study

	Q1 Mean (SE) or *n* (%)	Q2 Mean (SE) or *n* (%)	Q3 Mean (SE) or *n* (%)	Q4 Mean (SE) or *n* (%)	*P*-value[a]
PCT1 (Socransky trait)					
P. gingivalis (Pg)	1.47 (0.10)	1.87 (0.10)	2.24 (0.10)	3.37 (0.10)	<0.0001
P. intermedia (Pi)	1.66 (0.12)	2.24 (0.12)	2.58 (0.12)	4.04 (0.12)	<0.0001
P. nigrescens (Pn)	1.70 (0.11)	2.26 (0.11)	2.57 (0.11)	4.21 (0.11)	<0.0001
T. forsythus (Tf)	1.30 (0.09)	2.06 (0.09)	2.44 (0.09)	4.01 (0.09)	<0.0001
T. denticola (Td)	1.32 (0.10)	2.25 (0.10)	2.63 (0.10)	4.18 (0.10)	<0.0001
F. nucleatum (Fn)	1.59 (0.12)	2.43 (0.12)	2.79 (0.12)	4.70 (0.12)	<0.0001
C. rectus (Cr)	1.21 (0.09)	2.36 (0.09)	2.84 (0.09)	4.40 (0.09)	<0.0001
A. a.	1.33 (0.09)	2.49 (0.09)	2.69 (0.09)	3.97 (0.09)	<0.0001
GCF-IL-1β	1.93 (0.02)	2.07 (0.02)	2.13 (0.02)	2.15 (0.02)	<0.0001
CDC/AAP Health	101 (25.6)	139 (35.3)	77 (19.5)	77 (19.5)	
Entry	142 (36.5)	105 (27.0)	71 (18.3)	71 (18.3)	
Severe	0 (0.0)	1 (0.5)	96 (50.0)	95 (49.5)	<0.0001
Log total counts	3.80 (0.03)	4.38 (0.03)	4.55 (0.03)	5.53 (0.03)	<0.0001
Number of teeth	23.9 (0.43)	22.9 (0.43)	21.8 (0.43)	21.3 (0.43)	<0.0001
EALGE3i	16.2 (1.39)	14.3 (1.38)	28.1 (1.38)	30.6 (1.39)	<0.0001
EPDGE4	4.01 (0.61)	4.11 (0.61)	9.52 (0.61)	11.1 (0.61)	<0.0001
EBL	15.9 (1.38)	19.6 (1.37)	29.3 (1.37)	37.3 (1.38)	<0.0001
EPQGE1	26.3 (2.11)	29.4 (2.12)	41.7 (2.15)	54.0 (2.18)	<0.0001
EGIGE1	15.2 (2.04)	15.4 (2.08)	27.6 (2.19)	36.8 (2.26)	<0.0001
PCT2					
P. gingivalis (Pg)	2.14 (0.10)	2.35 (0.10)	1.94 (0.10)	2.52 (0.10)	0.0008
P. intermedia (Pi)	2.70 (0.13)	2.56 (0.13)	2.47 (0.13)	2.79 (0.13)	0.32
P. nigrescens (Pn)	2.76 (0.12)	2.80 (0.12)	2.30 (0.12)	2.89 (0.12)	0.004
T. forsythus (Tf)	2.41 (0.11)	2.51 (0.11)	2.16 (0.11)	2.73 (0.11)	0.002
T. denticola (Td)	2.58 (0.12)	2.73 (0.12)	2.07 (0.12)	3.01 (0.12)	<0.0001
F. nucleatum (Fn)	2.95 (0.14)	3.01 (0.13)	2.41 (0.14)	3.15 (0.14)	0.0006
C. rectus (Cr)	2.73 (0.12)	2.85 (0.12)	2.25 (0.12)	2.98 (0.12)	<0.0001
A. a.	2.68 (0.11)	2.69 (0.11)	2.32 (0.11)	2.78 (0.11)	0.01
GCF-IL-1β	1.88 (0.02)	2.06 (0.02)	2.11 (0.02)	2.23 (0.02)	<0.0001
CDC/AAP Health	0 (0.0)	100 (25.4)	244 (61.9)	50 (12.7)	
Entry	243 (62.5)	145 (37.3)	0 (0.0)	1 (0.3)	
Severe	0 (0.0)	0 (0.0)	0 (0.0)	192 (100)	<0.0001
Log total counts	4.59 (0.05)	4.59 (0.05)	4.33 (0.05)	4.75 (0.05)	<0.0001
Number of teeth	21.9 (0.43)	22.4 (0.43)	23.7 (0.43)	21.9 (0.43)	0.007
EALGE3i	25.0 (1.20)	17.0 (1.20)	5.93 (1.20)	41.3 (1.20)	<0.0001
EPDGE4	5.99 (0.55)	5.20 (0.54)	1.53 (0.54)	16.0 (0.55)	<0.0001
EBL	23.5 (1.40)	23.2 (1.39)	17.3 (1.39)	37.9 (1.40)	<0.0001
EPQGE1	38.7 (2.16)	36.8 (2.15)	24.5 (2.18)	50.6 (2.21)	<0.0001
EGIGE1	20.4 (2.13)	21.2 (2.14)	16.6 (2.15)	35.4 (2.28)	<0.0001
PCT3 [*Aa* Trait]					
P. gingivalis (Pg)	2.37 (0.11)	2.26 (0.10)	1.97 (0.10)	2.35 (0.11)	0.02
P. intermedia (Pi)	2.66 (0.13)	2.44 (0.13)	2.66 (0.13)	2.75 (0.13)	0.34
P. nigrescens (Pn)	2.55 (0.13)	2.61 (0.12)	2.67 (0.12)	2.91 (0.18)	0.18
T. forsythus (Tf)	2.58 (0.11)	2.32 (0.11)	2.32 (0.11)	2.59 (0.11)	0.13
T. denticola (Td)	2.47 (0.12)	2.45 (0.12)	2.50 (0.12)	2.97 (0.12)	0.004
F. nucleatum (Fn)	2.77 (0.14)	2.62 (0.14)	2.94 (0.14)	3.19 (0.14)	0.02
C. rectus (Cr)	2.49 (0.12)	2.43 (0.12)	2.82 (0.12)	3.07 (0.12)	0.0002
A. a.	2.46 (0.11)	2.54 (0.11)	2.56 (0.11)	2.91 (0.11)	0.02
GCF-IL-1β	1.71 (0.02)	1.96 (0.02)	2.18 (0.02)	2.42 (0.02)	<0.0001
CDC/AAP Health	150 (38.1)	101 (25.6)	79 (20.1)	64 (16.2)	
Entry	21 (5.4)	102 (26.2)	133 (34.2)	133 (34.2)	
Severe	72 (37.5)	42 (21.9)	32 (16.7)	46 (24.0)	<0.0001
Log total counts	4.52 (0.05)	4.52 (0.05)	4.49 (0.05)	4.73 (0.05)	0.004
Number of teeth	22.5 (0.44)	22.5 (0.43)	22.7 (0.44)	22.2 (0.44)	0.88
EALGE3i	20.0 (1.45)	20.9 (1.45)	22.1 (1.45)	26.1 (1.45)	0.02
EPDGE4	5.94 (0.64)	6.41 (0.63)	6.28 (0.63)	10.1 (0.64)	<0.0001
EBL	19.2 (1.44)	23.6 (1.43)	25.5 (1.44)	33.6 (1.44)	<0.0001
EPQGE1	30.0 (2.19)	35.6 (2.25)	39.6 (2.23)	45.4 (2.24)	<0.0001
EGIGE1	20.3 (2.16)	18.5 (2.21)	22.9 (2.20)	30.7 (2.24)	0.0006

Table continues

Table 2. Continued

	Q1 Mean (SE) or *n* (%)	Q2 Mean (SE) or *n* (%)	Q3 Mean (SE) or *n* (%)	Q4 Mean (SE) or *n* (%)	*P*-value[a]
PCT4					
P. gingivalis (Pg)	2.67 (0.10)	2.17 (0.10)	1.84 (0.10)	2.26 (0.10)	<0.0001
P. intermedia (Pi)	3.24 (0.13)	2.31 (0.13)	2.47 (0.13)	2.50 (0.13)	<0.0001
P. nigrescens (Pn)	3.22 (0.12)	2.37 (0.12)	2.43 (0.12)	2.72 (0.12)	<0.0001
T. forsythus (Tf)	3.21 (0.11)	2.34 (0.11)	1.94 (0.11)	2.32 (0.11)	<0.0001
T. denticola (Td)	3.26 (0.12)	2.29 (0.12)	2.13 (0.12)	2.71 (0.12)	<0.0001
F. nucleatum (Fn)	3.81 (0.13)	2.50 (0.13)	2.47 (0.13)	2.75 (0.13)	<0.0001
C. rectus (Cr)	3.58 (0.11)	2.21 (0.11)	2.44 (0.11)	2.59 (0.11)	<0.0001
A. a.	3.65 (0.10)	2.36 (0.10)	2.22 (0.10)	2.24 (0.10)	<0.0001
GCF-IL-1β	1.78 (0.02)	1.92 (0.02)	2.13 (0.02)	2.44 (0.02)	<0.0001
CDC/AAP Health	131 (33.3)	118 (30.0)	74 (18.8)	71 (18.0)	
Entry	46 (11.8)	77 (19.8)	139 (35.7)	127 (32.7)	
Severe	66 (34.4)	49 (25.5)	32 (16.7)	45 (23.4)	<0.0001
Log total counts	5.01 (0.05)	4.36 (0.05)	4.38 (0.05)	4.51 (0.05)	<0.0001
Number of teeth	22.1 (0.44)	23.0 (0.43)	22.8 (0.43)	22.0 (0.44)	0.23
EALGE3i	20.2 (1.45)	20.7 (1.45)	22.2 (1.45)	26.0 (1.45)	0.02
EPDGE4	6.43 (0.64)	5.68 (0.64)	7.01 (0.64)	9.59 (0.64)	0.0001
EBL	24.0 (1.45)	20.3 (1.45)	25.9 (1.45)	31.7 (1.45)	<0.0001
EPQGE1	37.8 (2.23)	32.1 (2.24)	36.3 (2.23)	44.2 (2.27)	0.002
EGIGE1	23.7 (2.21)	18.7 (2.23)	21.1 (2.17)	28.7 (2.22)	0.01
PCT5 [*Pg* Trait]					
P. gingivalis (Pg)	1.28 (0.09)	2.59 (0.09)	1.80 (0.09)	3.28 (0.09)	<0.0001
P. intermedia (Pi)	2.87 (0.13)	2.36 (0.13)	2.57 (0.13)	2.73 (0.13)	0.03
P. nigrescens (Pn)	2.70 (0.12)	2.36 (0.12)	2.67 (0.12)	3.01 (0.12)	0.004
T. forsythus (Tf)	2.23 (0.11)	2.27 (0.11)	2.31 (0.11)	2.91 (0.11)	<0.0001
T. denticola (Td)	2.51 (0.12)	2.31 (0.12)	2.64 (0.12)	2.92 (0.12)	0.003
F. nucleatum (Fn)	2.79 (0.14)	2.58 (0.14)	2.81 (0.14)	3.32 (0.14)	0.001
C. rectus (Cr)	2.51 (0.12)	2.39 (0.12)	2.74 (0.12)	3.16 (0.12)	<0.0001
A. a.	2.27 (0.10)	2.60 (0.10)	2.44 (0.10)	3.16 (0.10)	<0.0001
GCF-IL-1β	2.00 (0.02)	2.12 (0.02)	2.04 (0.02)	2.11 (0.02)	0.002
CDC/AAP Health	165 (41.9)	170 (43.2)	47 (11.9)	12 (3.1)	
Entry	6 (1.5)	1 (0.3)	163 (41.9)	219 (56.3)	
Severe	72 (37.5)	73 (38.0)	35 (18.2)	12 (6.3)	<0.0001
Log total counts	4.54 (0.05)	4.39 (0.05)	4.52 (0.05)	4.80 (0.05)	<0.0001
Number of teeth	23.3 (0.44)	22.6 (0.43)	22.1 (0.43)	21.8 (0.43)	0.07
EALGE3i	19.0 (1.44)	18.3 (1.44)	24.7 (1.44)	27.1 (1.44)	<0.0001
EPDGE4	6.66 (0.64)	6.91 (0.64)	7.52 (0.64)	7.61 (0.64)	0.67
EBL	23.2 (1.47)	23.5 (1.47)	26.8 (1.46)	28.5 (1.47)	0.03
EPQGE1	34.0 (2.25)	33.7 (2.25)	38.2 (2.21)	44.5 (2.26)	0.002
EGIGE1	21.3 (2.24)	20.5 (2.26)	23.4 (2.14)	26.8 (2.24)	0.20
PCT6					
P. gingivalis (Pg)	2.19 (0.11)	2.41 (0.11)	2.10 (0.11)	2.25 (0.11)	0.19
P. intermedia (Pi)	2.72 (0.13)	2.70 (0.13)	2.42 (0.13)	2.67 (0.13)	0.31
P. nigrescens (Pn)	2.61 (0.12)	2.76 (0.12)	2.63 (0.12)	2.74 (0.12)	0.76
T. forsythus (Tf)	2.41 (0.11)	2.56 (0.11)	2.22 (0.11)	2.61 (0.11)	0.054
T. denticola (Td)	2.57 (0.12)	2.69 (0.12)	2.54 (0.12)	2.59 (0.12)	0.83
F. nucleatum (Fn)	2.81 (0.14)	3.03 (0.14)	2.79 (0.14)	2.88 (0.14)	0.56
C. rectus (Cr)	2.67 (0.12)	2.75 (0.12)	2.67 (0.12)	2.72 (0.12)	0.95
A. a.	2.56 (0.11)	2.77 (0.11)	2.49 (0.11)	2.65 (0.11)	0.26
GCF-IL-1β	2.25 (0.02)	1.98 (0.02)	2.08 (0.02)	1.97 (0.02)	<0.0001
CDC/AAP Health	240 (60.9)	154 (39.1)	0 (0.0)	0 (0.0)	
Entry	1 (0.3)	82 (21.1)	240 (61.7)	66 (17.0)	
Severe	2 (1.0)	9 (4.7)	3 (1.6)	178 (92.7)	<0.0001
Log total counts	4.54 (0.05)	4.65 (0.05)	4.45 (0.05)	4.61 (0.05)	0.05
Number of teeth	22.9 (0.44)	22.6 (0.43)	22.3 (0.44)	22.2 (0.43)	0.70
EALGE3i	6.84 (1.18)	14.7 (1.18)	25.0 (1.18)	42.5 (1.18)	<0.0001
EPDGE4	1.94 (0.55)	4.30 (0.55)	6.95 (0.55)	15.5 (0.55)	<0.0001
EBL	19.3 (1.43)	22.5 (1.43)	25.8 (1.43)	34.4 (1.43)	<0.0001
EPQGE1	28.8 (2.23)	35.6 (2.20)	40.2 (2.20)	46.0 (2.26)	<0.0001
EGIGE1	18.6 (2.21)	24.1 (2.18)	22.2 (2.16)	27.5 (2.31)	0.04

Aa, Aggregatibacter actinomycetemcomitans; EALGE3i, extent interproximal attachment loss ≥3 mm; EPDGE4, extent probing depth ≥4 mm; EBL, extent bleeding on probing; EPQGE1, extent plaque score ≥1; EGIGE1, extent gingival index score ≥1.

Log total counts represent the sum of the specific pathogens identified by whole chromosomal DNA probes for the individual tested species.

[a]Derived from ANOVA for continuous and χ^2 tests for categorical variables.

shifts illustrate the magnitude of the changes associated with these traits. For example, there is a 2.8-fold increase in *Aa* in PCT3 and a 100-fold increase in *Pg* in PC5 comparing Q1 versus Q4, respectively.

Identification of candidate gene loci associated with complex traits

After applying the exclusion criteria outlined in the Materials and Methods section, there remained 2135235 SNPs that were included in the GWA analysis. Q–Q plots were generated for each PCT and the lambda coefficients for all six traits ranged between 0.994 and 1.031, suggesting virtually no genomic inflation after adjustment for global ancestry. There were 69 genome-wide statistically significant SNPs that were below the *P*-value threshold 5.0×10^{-8}, marking 10 loci. There were six loci associated with the Socransky trait (PCT1), two loci for the *Aa* trait PCT3, one locus for PCT4 and three loci associated with the *Pg* trait (PCT5). Manhattan plots are presented in Figure 2A–C. Two SNPs within the *Aa* trait (PCT3) met genome-wide significance criteria: rs4074082 (*C1QTNF7*; MAF = 0.059, $P = 2.2 \times 10^{-8}$) and rs9772881 (*FLJ43860*; MAF = 0.10, $P = 3.1 \times 10^{-8}$). No genome-wide significant SNPs were identified for PCT2 or PCT6, which are both associated with smoking and diabetes, as shown in Table 1. Consistent with the current recommendations for reporting GWAS findings, we have made available the entire set of single-marker association results for the six PCTS listed in ascending *P*-value order at: http://genomewide.net/public/aric/dental/complex_traits/pct#.txt, where # ranges between 1 and 6. The annotations for the significant loci for the Socransky trait, the *Aa* trait, PCT4 and the *Pg* trait appear in Table 3. The lowest imputation score was 0.946.

Three loci which emerged as promising candidates are illustrated using LocusZoom in Figure 3A–C. Within the Socransky trait the SNP with the lowest *P*-value (3.0×10^{-10}) is rs1156327 (MAF = 0.192), at 16q11.2 and in moderate proximity to *CLEC19A* (C-type lectin domain family 19 member A) (Fig. 3A). This locus includes an additional coding, non-synonymous SNP with $P = 9.7 \times 10^{-8}$: rs179196 ($r^2 = 0.34$). This missense SNP (MAF = 0.21) represents a T–G base change resulting an amino acid change of S→A in protein amino acid position 16 which is located in the signal peptide region. CLEC19A is a glycoprotein which can be membrane-tethered or secreted and believed to be involved in carbohydrate recognition functioning as a pattern recognition receptor (CLR, C-lectin receptor) for immune signaling, as other

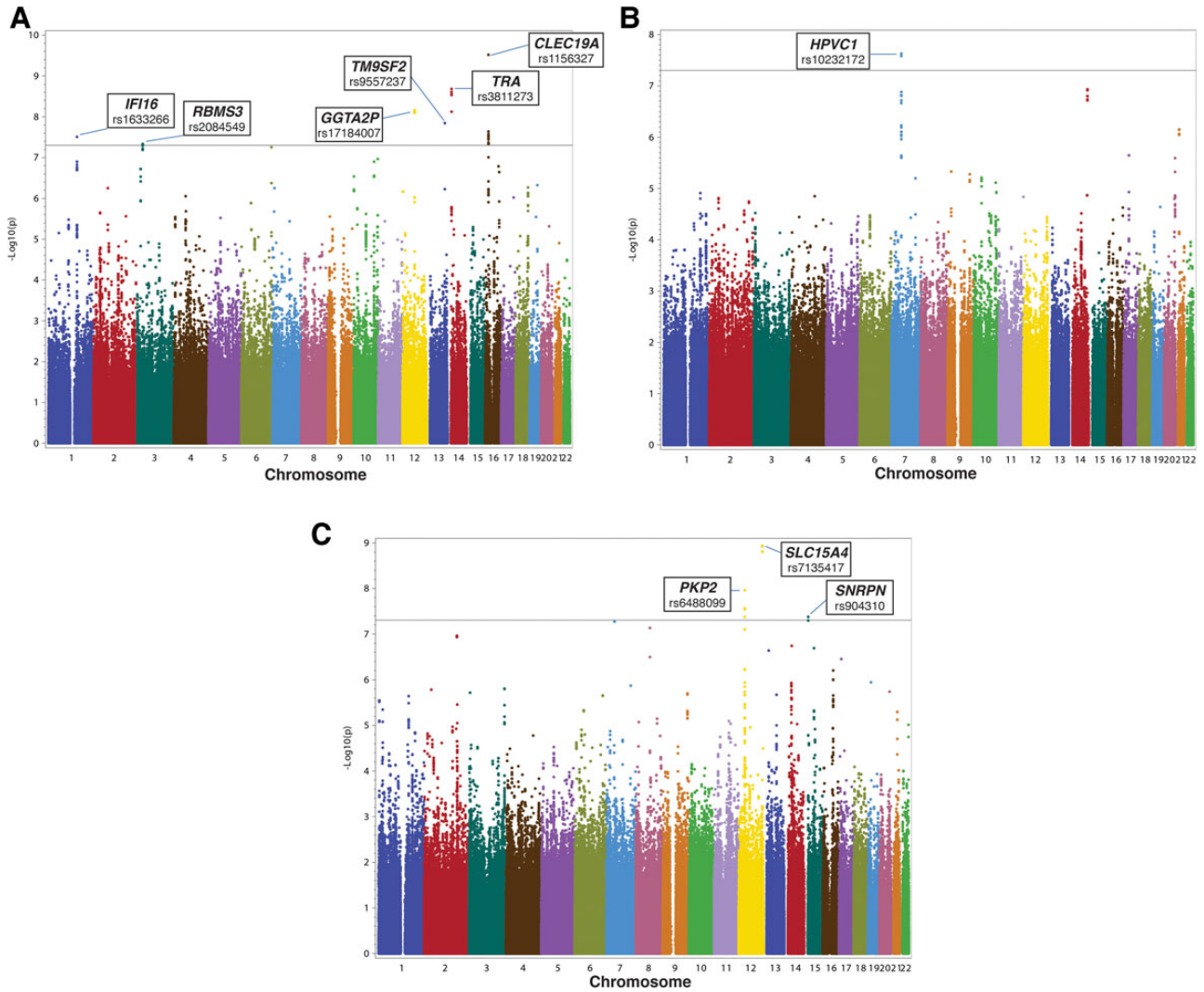

Figure 2. (**A**) Periodontal Complex Trait 1 (Socransky Trait) Manhattan plot. (**B**) Periodontal Complex Trait 4 Manhattan plot. (**C**) Periodontal Complex Trait 5 (Pg Trait) Manhattan plot.

Table 3. Genome-wide association analysis results of the six PCTs among the European American participants of the Dental ARIC study cohort ($n = 975$)

Locus	SNP	Position Bld37.4	Imputed	Major allele	MAF[a] (1000 genomes)	MAF (Cohort)	Closest gene and position or distance (Kb)	i-value[b]	Beta (SE)
PCT1 (Socransky Trait)									
16q11.2[c]	rs1156327	193448524	Y	C	[T] 0.192	[T] 0.095	*CLEC19A* (33)	3.0×10^{-10}	1.45 (0.23)
14q21	rs3811273	22715158	Y	A	[G] 0.149	[G] 0.149	*TRA* (65)	2.1×10^{-9}	1.22 (0.20)
12q14	rs17184007	67594578	N	T	[C] 0.050	[C] 0.118	*GGTA2P* (65)	6.9×10^{-9}	1.35 (0.23)
13q32.3	rs9557237	100110956	Y	G	[C] 0.178	[C] 0.113	*TM9SF2* (43)	1.4×10^{-8}	1.33 (0.24)
1q12[d]	rs1633266	159005977	Y	T	[C] 0.264	[C] 0.226	*IFI16* (19)	3.1×10^{-8}	0.93 (0.17)
3q12	rs17718700	29055028	Y	T	[C] 0.034	[C] 0.127	*RBMS3* (268)	4.6×10^{-8}	1.22 (0.22)
PCT3 (*Aa* Trait)									
4p15.33	rs4074082	152322	Y	C	[T] 0.059	[T] 0.126	*C1QTNF7* (108)	2.2×10^{-8}	0.55 (0.10)
8q24.3	rs9772881		Y	A	[G] 0.100	[G] 0.188	*TSNARE* (290)	3.1×10^{-8}	0.46 (0.08)
PCT4									
7q21.1	rs10232172	54182781	Y	G	[T] 0.147	[T] 0.139	*HPVC1* (86)	2.3×10^{-8}	−0.45 (0.08)
PCT5 (*Pg* Trait)									
12q14	rs7135417	128176235	Y	C	[T] 0.055	[T] 0.106	*SLC15A4* (1102)	1.2×10^{-9}	0.44 (0.07)
11q14	rs6488099	33035483	Y	A	[G] 0.125	[G] 0.316	*PKP2* (intronic)	1.1×10^{-8}	0.29 (0.05)
15q24	rs904310	30439524	Y	G	[A] 0.420	[A] 0.133	*SNRPN* (intronic)	4.2×10^{-8}	0.39 (0.07)

Single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF-HapMap II CEU) of ≥5% and associated $P < 5 \times 10^{-8}$; the SNP with the lowest *P*-value per locus is presented. MAF estimates and annotations are based upon 1000 Genomes. Footnotes indicate additional missense (coding, non-synonymous) SNPs in linkage disequilibrium with the index SNP.

[a]Minor allele frequency.

[b]Based on linear regression including terms for age, sex, study center and ancestry (10 first principal components).

[c]Additional coding, non-synonymous SNP within locus: rs179196 ($r^2 = 0.34$; $P = 9.7 \times 10^{-8}$).

[d]Additional coding, non-synonymous SNPs within locus rs6940 ($r^2 = 1.00$; $P = 2.0 \times 10^{-7}$); rs1057028 ($r^2 = 0.70$; $P = 4.5 \times 10^{-6}$) and rs1057027 ($r^2 = 0.70$; $P = 4.5 \times 10^{-6}$).

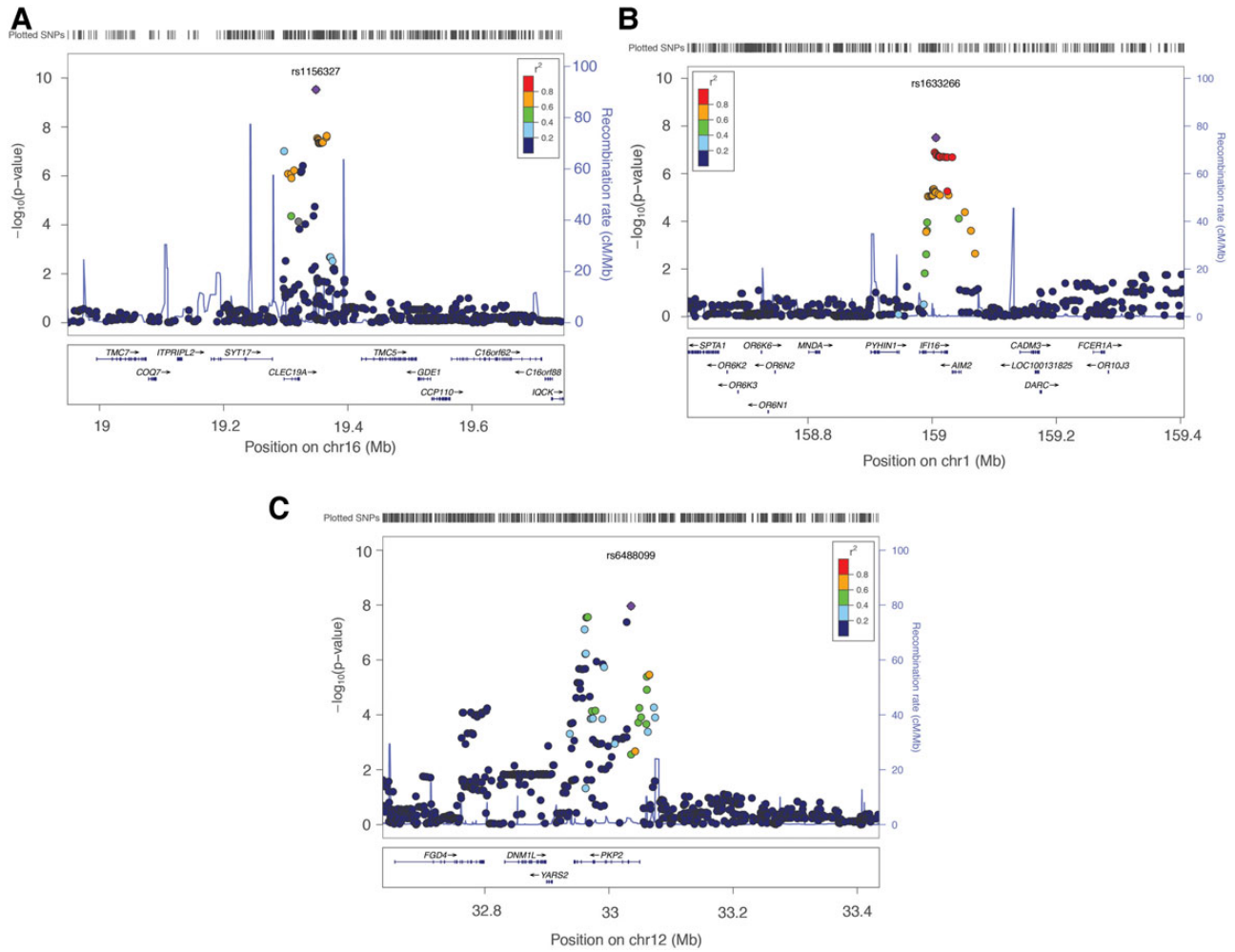

Figure 3. (**A**) PCT1 locus CLEC19A. (**B**) PCT1 locus IFI16. (**C**) PCT1 locus PKP2.

members of the CLR protein family. Also associated with the Socransky Trait (PCT1) is *IFI16* (Gamma interferon inducible protein 16, rs1633266, $P = 3.8 \times 10^{-8}$, MAF = 0.26). *IFI16* encodes a protein that contains two HIN-200 domain and functions as a sensor of intracellular DNA that stimulates interferon expression and inflammasome activation in response to viral challenge (29). The *IFI16* locus appears in Figure 3B and is marked by rs1633266 which is in linkage disequilibrium with three additional missense SNPs: rs6940 ($r^2 = 1.00$; $P = 2.0 \times 10^{-7}$), rs1057028 ($r^2 = 0.70$; $P = 4.5 \times 10^{-6}$) and rs1057027 ($r^2 = 0.70$; $P = 4.5 \times 10^{-6}$). Of those, rs6940 and rs1057028 are predicted to result in a potentially/probably damaging alteration in the protein structure. Rs6940 shows a dual amino acid substitution (T–S) at positions 723 and 779 within the 785-amino acid protein (30) while rs1057028 indicates a Y→N at position 413 and rs1057027 represents a potentially benign substitution (R→S at position 409). The *Pg* Trait (PCT5) showed a significant association with a chromosome 12 locus marked by rs6488099 ($P = 1.1 \times 10^{-8}$), which is intronic to *PKP2* (plakophilin 2, Fig. 3C). PKP2 is a structural protein present within desmosomal structures of basal epithelial cells and within the myocardium.

Identification of candidate genes associated with complex traits

We carried forward the single-marker results to genome-wide gene-centric association analysis for the first six complex traits using MAGENTA (15), correcting for gene size (including a 110 kb upstream and the 40 kb downstream region, as employed in previous investigations) (15,31), the number of SNPs within the gene region, and local linkage disequilibrium characteristics (15). Gene-centric analyses involve a smaller number of tests of association (i.e. genes versus SNPs) and aggregate multiple signals per locus, thus offering gains in power to detect candidate genes. We queried 18 307 genes and used a Bonferroni multiple testing-corrected *P*-value threshold of $0.05/18307 = 2.7 \times 10^{-6}$. Eighty genes in 37 loci were found to be significantly associated with the complex traits as shown in Table 4. Significant gene-centric associations were also found for 5 of the 10 loci identified in Table 3: Socransky trait—*CLEC19A*, *GGTA2P* and *IFI16*; *Aa* trait—*C1QTNF7*; *Pg* trait—*PKP2* and *SNRPN*. The full list of gene-centric association testing results is available at http://genomewide.net/public/aric/dental/gene-centric/CP_PCT_Genes.xls.

Validation of loci for clinically defined disease taxonomy and severity

We examined the association of all previously highlighted loci with clinically defined disease in the larger ARIC cohort ($n = 4766$) and an independent German sample including 717 aggressive periodontitis (AgP) cases and 4210 controls generated by Schaefer *et al.* and previously reported (32–36). Overall, we found little evidence of validation for clinically defined CP (moderate versus health and severe

Table 4. Results of genome-wide gene-centric association analyses for the six PCTs

PCT	Chr	Gene *P*-value	Symbol	Entrez Gene Name	Top SNP in region
1	16	8.00E-12	*TMC5*	Transmembrane Channel-Like 5	rs1156327
1	12	1.47E-09	*CAND1*	Cullin Associated and Neddylation-Dissociated 1	rs17184007
1	1	4.70E-09	*AIM2*	Absent in melanoma 2	rs1633266
1	13	1.03E-08	*TM9SF2*	Transmembrane 9 Superfamily Member 2	rs9557237
1	1	1.16E-08	*IFI16*	Interferon, Gamma-Inducible Protein 16	rs1633266
1	16	2.35E-08	*MLKL*	Mixed Lineage Kinase Domain-Like	rs2240246
1	16	2.46E-08	*RFWD3*	Ring Finger and WD Repeat Domain 3	rs2240246
1	10	5.48E-08	*SMC3*	Structural maintenance of chromosomes 3 (bamacan)	rs11195183
1	1	8.89E-08	*CADM3*	Cell Adhesion Molecule 3	rs1633266
1	19	9.19E-08	*PLEKHG2*	Pleckstrin Homology Domain Containing, Family G (With RhoGef Domain) Member 2	rs3786949
1	19	9.78E-08	*SUPT5H*	Suppressor Of Ty 5 Homolog (*S. cerevisiae*)	rs3786949
1	10	1.13E-07	*DUSP5*	Dual specificity phosphatase 5	rs11195183
1	19	1.43E-07	*IL28B*	Interferon, Lambda 3	rs3786949
1	19	1.47E-07	*MED29*	Mediator Complex Subunit 29	rs3786949
1	19	1.49E-07	*GMFG*	Glia maturation factor, gamma	rs3786949
1	19	1.70E-07	*ZFP36*	ZFP36 Ring Finger Protein	rs3786949
1	19	2.19E-07	*SAMD4B*	sterile alpha motif domain containing 4B	rs3786949
1	19	3.18E-07	*IFNL1*	Interferon, lambda 1 (IL29)[a]	rs3786949
1	19	3.40E-07	*LRFN1*	Leucine rich repeat and fibronectin type III domain containing 1	rs3786949
1	16	3.77E-07	*FA2H*	Fatty acid 2-hydroxylase	rs2240246
1	16	5.75E-07	*SYT17*	Synaptotagmin I	rs1156327
1	18	1.05E-06	*CBLN2*	Cerebellin 2 precursor	rs1432076
2	12	5.51E-07	ANKLE2	Ankyrin Repeat and LEM Domain Containing 2	rs10781645
2	1	8.48E-07	DMAP1	DNA Methyltransferase 1 Associated Protein 1	rs4660770
2	1	8.94E-07	KLF17	Kruppel-like factor 17[a]	rs4660770
2	12	1.48E-06	GOLGA3	Golgin A3[a]	rs10781645
2	12	1.61E-06	POLE	Polymerase (DNA Directed), Epsilon, Catalytic Subunit	rs10781645
2	1	1.80E-06	SLC6A9	Solute Carrier Family 6 (Neurotransmitter Transporter, Glycine), Member 9	rs4660770
3	4	1.53E-08	C1QTNF7	C1q and tumor necrosis factor-related protein 7	rs4074082
3	2	4.74E-08	DGKD	Diacylglycerol kinase, delta 130 kDa	rs7568534
3	9	7.55E-08	ZNF483	Zinc Finger Protein 483	rs10980953
3	9	1.05E-07	KIAA0368	KIAA0368	rs10980953
3	9	1.60E-07	DNAJC25	DnaJ (Hsp40) Homolog, Subfamily C, Member 25	rs10980953
3	9	2.04E-07	DNAJC25-GNG10	DNAJC25-GNG10 readthrough	rs10980953
3	9	2.36E-07	GNG10	Guanine Nucleotide Binding Protein (G Protein), Gamma 10	rs10980953
3	9	2.43E-07	PTGR1	Prostaglandin reductase 1	rs10980953
3	17	2.90E-07	ANKRD40	Ankyrin Repeat Domain 40	rs4148411
3	9	5.47E-07	IFNA14	Interferon, Alpha 14	rs13340713
3	17	5.99E-07	CROP	LUC7-Like 3 (*S. cerevisiae*)	rs4148411
3	6	8.28E-07	DAAM2/KIF6	Dishevelled Associated Activator of Morphogenesis 2/Kinesin Family Member 6	rs9471152
3	9	8.83E-07	IFNA17	Interferon, alpha 17	rs13340713
3	10	9.57E-07	INPP5F	inositol polyphosphate-5-phosphatase F	rs17680667
3	10	1.06E-06	DUPD1	Dual specificity phosphatase DUPD1[a]	rs2395126
3	17	1.12E-06	ABCC3	ABCC3 protein	rs4148411
3	9	1.18E-06	IFNA16	Interferon, Alpha 16	rs13340713
3	10	1.24E-06	SAMD8	Sterile Alpha Motif Domain Containing 8	rs2395126
3	9	1.27E-07	IFNA7	Interferon, Alpha 7	rs13340713
3	6	1.36E-06	REV3L/TRAF3IP2	REV3-Like, Polymerase (DNA Directed), Zeta, Catalytic Subunit/TRAF3 Interacting Protein 2[a]	rs13196377
3	9	1.38E-06	IFNA10	Interferon, Alpha 10	rs13340713
3	17	1.98E-06	CACNA1G	Voltage-dependent T-type calcium channel subunit alpha-1G	rs757415
3	10	2.21E-06	SEC23IP	SEC23 Interacting Protein	rs17680667
3	10	2.35E-06	MCMBP	Minichromosome Maintenance Complex Binding Protein	rs17680667
3	10	2.57E-06	KAT6B	K(Lysine) Acetyltransferase 6B	rs2395126
3	9	2.59E-06	IFNA21	Interferon, Alpha 21	rs13340713
4	14	7.43E-09	*BEGAIN*	Brain-enriched guanylate kinase-associated	rs7144384
4	14	1.23E-07	*DLK1*	Delta-like 1 homolog (*Drosophila*)	rs7144384
5	7	1.29E-09	MRPL32	Mitochondrial Ribosomal Protein L32	rs2288354
5	7	3.27E-09	C7orf25	Chromosome 7 Open Reading Frame 25	rs2288354

Table continues

Table 4. Continued

PCT	Chr	Gene P-value	Symbol	Entrez Gene Name	Top SNP in region
5	15	3.44E-09	SNURF	SNRPN Upstream Reading Frame	rs904310
5	12	3.65E-09	PKP2	Plakophilin 2	rs6488099
5	12	5.17E-09	YARS2	Tyrosyl-TRNA Synthetase 2, Mitochondrial	rs6488099
5	7	6.66E-09	PSMA2	Proteasome (Prosome, Macropain) Subunit, Alpha Type, 2	rs2288354
5	15	5.18E-08	SNRPN	Small Nuclear Ribonucleoprotein Polypeptide N	rs904310
5	14	3.82E-07	SEC23A	Sec23 homolog A (*S. cerevisiae*)	rs11622736
5	14	4.41E-07	PNN	Pinin, desmosome-associated protein	rs11622736
5	14	5.10E-07	MIA2	Melanoma Inhibitory Activity 2	rs11622736
5	14	5.14E-07	TRAPPC6B	Trafficking Protein Particle Complex 6B	rs11622736
5	14	5.23E-07	GEMIN2	Gem (Nuclear Organelle) Associated Protein 2	rs11622736
5	2	6.24E-07	DHX57/MORN2	DEAH (Asp–Glu–Ala–Asp/His) Box Polypeptide 57/MORN Repeat Containing 2[a]	rs3099950
5	19	7.36E-07	COLGALT1/ UNC13A	Collagen Beta(1-O)Galactosyltransferase 1/Unc-13 Homolog A (*C. Elegans*)	rs3746200
5	13	9.12E-07	TEX26-ASI	Testis expressed 26	rs12877861
5	13	9.38E-07	MEDAG	Mesenteric Estrogen-Dependent Adipogenesis	rs12877861
5	14	1.04E-06	CTAGE5	CTAGE Family, Member 5	rs11622736
5	9	1.18E-06	POMT1/UCK1	Protein-O-Mannosyltransferase 1/Uridine–Cytidine Kinase 1[a]	rs2282010
5	16	1.23E-06	KIFC3	Kinesin family member C3	rs12922275
5	1	1.33E-06	HTR1D	5-hydroxytryptamine (serotonin) receptor 1D, G protein-coupled	rs16828047
5	9	1.43E-06	PRRC2B	Proline-rich coiled-coil 2B	rs2282010
5	1	1.60E-06	LUZP1	Leucine zipper protein 1	rs9426689
5	16	2.10E-06	KATNB1	Katanin p80 (WD repeat containing) subunit B 1	rs12922275
6	6	1.45E-06	UBE3D	Ubiquitin protein ligase E3D	rs10484711

Statistically significant genes are shown for all six PCTs identifying chromosomal location, gene symbol and HUGO Gene Name and top SNP within region (lowest *P*-value). *P*-values reflect the value for the entire gene, rather than the best SNP.

[a]Coding, non-synonymous SNPs are within loci for *IFNL1*, *GOLGA3*, *KLF17*, *DUPD1*, *TRAF3IP2*, *MORN2* and *UCK1*.

versus health); the data shown in Supplementary Material, Table S3 are derived from gene-centric analyses reported by Rhodin *et al.* (15) and illustrate that with the exception of associations of *BEGAIN* with severe CP ($P = 4.3 \times 10^{-4}$) and *UBE3D* ($P = 1.3 \times 10^{-4}$) with moderate CP, no other loci were associated with CP in ARIC after a multiple testing correction. Two additional loci met nominal statistical significance criteria (*KLF17*—$P = 3.0 \times 10^{-2}$ with moderate CP and *DLK1*—$P = 2.1 \times 10^{-3}$ with severe CP). Similarly, no association was found with AgP (Supplementary Material, Table S4) in the German sample. No SNP surpassed the multiple testing-corrected significance threshold and in fact 27 of 37 loci showed discordant directionality (binomial test two-tailed $P = 0.008$).

The nearly complete lack of replication for clinical disease taxonomies raises the possibility that these PCTs may have little relevance to clinically defined disease taxonomies, which represent conglomerates of different pathogenic effector mechanisms. On the other hand, it is also possible that the identified PCT loci may be associated with clinical disease only in the context of specific microbial community/inflammatory patterns, which reflect biological sub-types of clinical periodontitis. Some evidence in support of this notion is illustrated in Table 2, where the PCTs (quartile-categorical) are generally associated with more severe disease, as defined by clinical criteria. For this reason, we tested the putative candidate gene-centric associations (obtained from Tables 3 and 4) with clinical disease severity in the context of each respective PCT (microbial and inflammatory pattern) via orthogonal analyses (adjusting for the other PCTs). Since we only tested 80 genes in this manner, we used a Bonferroni multiple comparison-corrected statistical significance threshold of $P = 6.2 \times 10^{-4}$.

Three genes (*GOLGA3*, *KLF17* and *IPP5F*) were found to be significantly associated with clinical disease severity in the context of their PCT phenotypic environment, while six other loci approached significance (Table 5). However, these results should be considered with caution. First, these associations were less pronounced (null, after correction for multiple testing) if considered outside the context of the biologically informed trait. As shown in Supplementary Material, Table S5, adjustment 'endophenotype context' (i.e. the other PCTs) resulted in mostly attenuated estimates of association (betas) for the lead SNPs in the nine highlighted loci (five out of nine SNPs had smaller estimates of association) or had no material impact (three out of nine SNPs with small increases in beta, 11% or less). A substantial increase in the magnitude of association was noted for only one SNP (rs4660770; *KLF17* locus) which was not associated with the index PCT in unadjusted analysis ($P = 0.5$). Moreover, the inclusion of all other PCTs in the models might raise issues of false positives due to collinearity. To rule out this possibility, we performed a formal test of covariate collinearity by computing the variance inflation factor (VIF) for the models used to produce Table 5. The seven models tested had VIF values ranging between 1.02 and 1.18 for the PCTs entered into each model as covariates, with VIF >10 considered indicative of serious collinearity. Based on this diagnostic information, we suggest that there is little evidence of bias due to collinearity (37). Nonetheless, these results must be interpreted with caution, and for this reason we present both adjusted and unadjusted (for the other PCTs) single-marker association results for these loci in Supplementary Material, Table S5.

Discussion

We used PCA enriched with biologically informed periodontal phenotypes to identify six PCTs that are characterized by specific

Table 5. Results of gene-centric association analyses of PCTs with a continuous measure of disease severity (extent interproximal attachment loss ≥3 mm) derived from gene-centric (MAGENTA) orthogonal analysis (adjusting for all other PCTs except the index one) among the Dental ARIC discovery sample ($n = 975$)

PCT	Chr.	Lead SNP	Gene(s)	Gene-centric *P*-value
1	19	rs3786949[a]	*GMFG*	2.3×10^{-3}
			SAMD4B	2.7×10^{-3}
			IFNL1 (IL29)	3.3×10^{-3}
			LRFN1	3.5×10^{-3}
2	12	rs10781645	*GOLGA3*	6.0×10^{-5}[b]
	1	rs4660770	*KLF17*	9.0×10^{-5}[b]
3	10	rs17680667	*INPP5F*	1.4×10^{-4}[b]
	17	rs4148411	*ABCC3*	6.7×10^{-3}
		rs757415	*CACNA1G*	8.9×10^{-3}
5	1	rs9426689	*LUZP1*	7.3×10^{-3}
		rs16828047	*HTR1D*	8.4×10^{-3}
6	6	rs10484711	*UBE2CBP*	6.6×10^{-3}

[a]Lead SNP in the chromosome 19 locus that includes *GMFG*, *SAMD4B*, *INFL1(IL29)* and *LRFN1*.

[b]Denotes gene-centric statistically significant association (the Bonferroni multiple testing-corrected *P*-value threshold for testing of 80 gene associations with the continuous measure of disease severity was $0.05/80 = 6.2 \times 10^{-4}$).

dysbiotic microbial community structures and inflammatory mediator levels. The GWAS results reported here were generated among 975 genotyped with known subgingival microbial plaque composition and gingival inflammatory response (IL-1β levels) and provide evidence in support of several novel candidate loci. The Dental ARIC population currently represents one of the largest reported samples of community-dwelling subjects with both full-mouth periodontal clinical examinations and genotype data. We were able to define six microbial/IL1β community structures as PCTs that are associated with clinical disease and identified 10 loci and 80 candidate genes possibly associated with these complex traits. With the exceptions of *BEGAIN* (severe CP) and *UBE3D* (moderate CP) these were not associated with clinically defined disease neither in the larger cohort of 4910 Dental ARIC participants nor in an independent German data set of aggressive periodontitis of 4916 participants (717 cases). The nearly complete lack of replication in the clinically defined disease cohorts raises the possibility that these PCTs may be genetically tractable endophenotypes that nevertheless have little relevance to disease defined with clinical criteria alone. This suggests that the six PCTs, although having overlapping clinical presentations, may actually reflect six different conditions with distinct genetic risk profiles that may be discoverable only in the context of specific patterns of microbial dysbiosis and inflammatory response. One of the consequences of identifying genes for common complex traits is that as sample sizes increase to boost power, so too does phenotypic heterogeneity (36). The approach considered here may allow for the identification of meaningful subgroups of disease pathogenesis, highlighting the role of microbial dysbiosis and inflammatory response. These loci may help identify new pathways of pathogenesis which can facilitate the identification of complex periodontal traits for the study of genetic susceptibility.

Genes in loci associated with the Socransky (PCT1), the *Aa* (PCT3) and the *Pg* traits (PCT5) suggest a common biological underpinning that controls the epithelial, connective tissue and innate immune response (Fig. 4). These loci, which were highlighted by several missense variants, appear to enhance disease susceptibility in the presence of a dysbiotic microbial structure. PCT1 explains the greatest variance and has the microbial community structure associated with the most common form of periodontal disease. We refer to this as the Socransky trait, since he first described the emergence of this microbial pattern of colonization in chronic adult periodontitis against a background of attenuated commensals that are associated with health (28). *CLEC19A* has an unknown function, but belongs to a family of carbohydrate recognition receptors and provides an interesting new candidate for further study. Interestingly, also within PCT1 two candidate genes, *IFI16* and *AIM2*, are both involved with the intracellular processing of foreign DNA. IFI16 recognizes foreign intracellular DNA and activates type I interferon expression via the STING pathway and complexes with AIM2 to form an inflammasome. AIM2 is also capable of binding foreign DNA and the complex is a caspase-1/inflammasome activator, triggering a local IL-1β response. This suggests the possibility that polymorphisms in these genes may result in an altered processing of invasive intracellular oral pathogens, including viruses and possibly organisms such as *Pg*, *Cr* and *Aa*, which are associated with the Socransky trait. Dendritic, synaptic or neural proteins also emerge as a molecular theme for genetic variation in the Socransky trait with significant loci identified including *CBLN2*, *GMFG*, *LRFN1*, *FA2H* and *SAMD4*. The theme of neural regulation of periodontal inflammation has been previously reported from whole transcriptome analyses (38). 'Nervous system-signaling' has also been reported as the most enriched pathway in the chronic periodontitis GWAS reported by Divaris *et al.* (14). Interestingly, two genes with missense variants emerge as potential candidates for PCT2- *GOLGA3* and *KLF17* (a member of the KLF transcriptional factor family that regulates epithelial–mesenchymal transition). The *Aa* trait *(PCT3)* showed a significant association with additional genes including the adipokine *C1QTNF7*, another type I interferon *IFNA14* and *TRAF3IP2*. The TRAF3IP2 protein is involved in IL-17 signaling and mucosal immunity serving as an adaptor protein for the IL-17 receptor. Genetic variants of *TRAF3IP2* have been associated with epithelial dysfunction in psoriasis and inflammatory bowel disease (39,40). Variants of the neuronal proteins BEGAIN and DLK1 are associated with PCT4. Genes of several epithelial structural proteins including components of desmosomes and tight junctions including *PKP2*, *PNN* and *KIFC3* were found to be significantly associated with the *Pg* trait (PCT5).

The *Aa* (PCT3) and *Pg* (PCT5) traits identified in this investigation have characteristics which are remarkably consistent with the literature reports of various subsets of individuals with aggressive forms of periodontal disease. The former is associated with several genes that point to potential abnormalities in the innate immune response and neutrophil function. Abnormal neutrophil function that impairs chemotaxis, phagocytosis and/or killing is a well-established risk factor or co-morbidity associated with less common, but aggressive forms of periodontal disease (41,42). For example, one specific aggressive form of severe disease is a periodontal syndrome (early onset periodontitis) that displays advanced progression and tooth loss in young individuals (<35 years), and neutrophil chemotactic defects with abnormal neutrophil DKG (diacylglycerol kinase) and IP3 (inositol triphosphate) levels (43), as discussed in Supplementary Material, Discussion. The *Aa* trait (PCT3) is associated with genetic variants of diacylglycerol kinase and inositol polyphosphate phosphatase, which are critically involved with regulating neutrophil function. The data from the current investigation are based upon an older population, for which we cannot assign an age of disease onset, and therefore the findings do not support

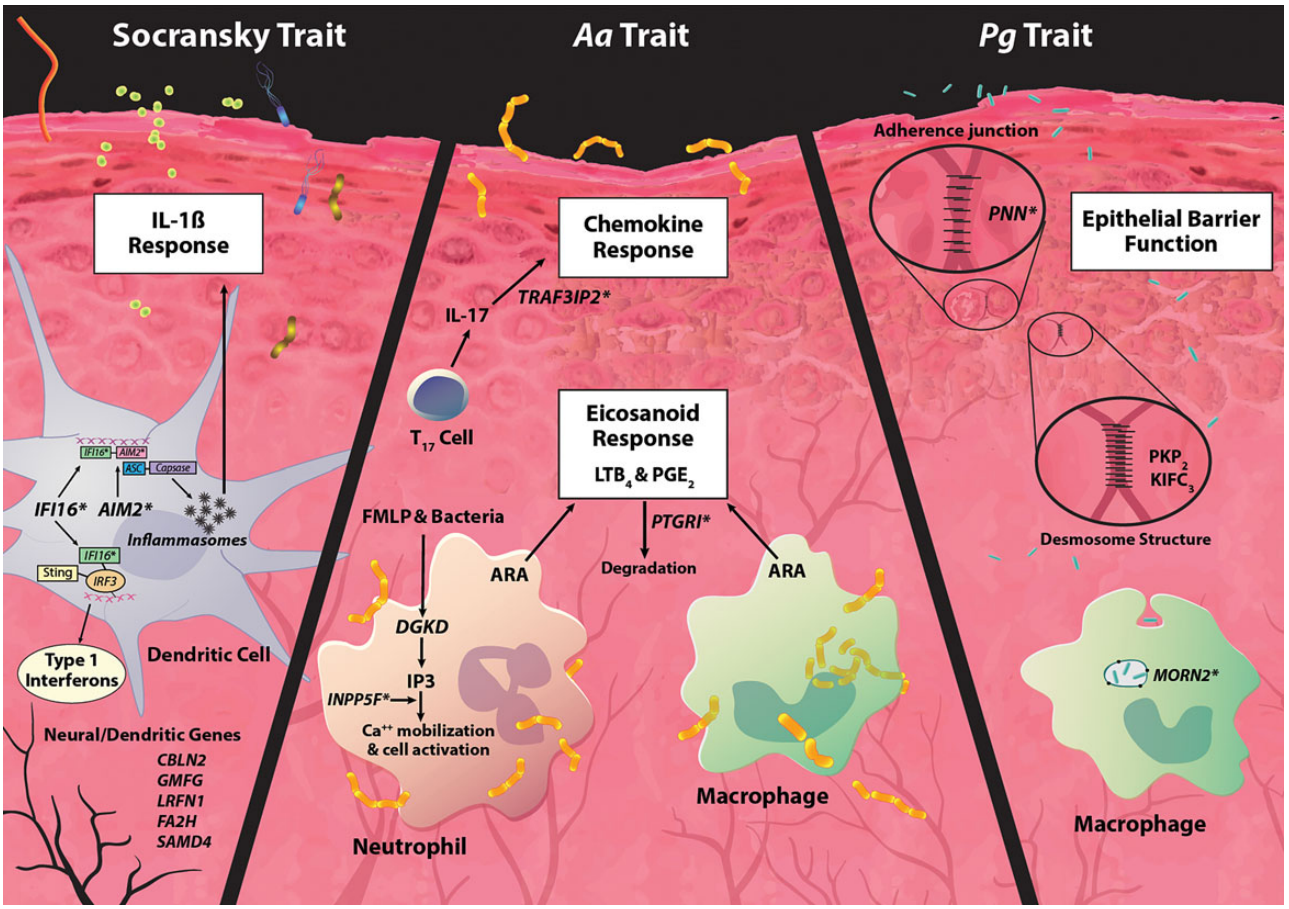

Figure 4. Model of genetic and microbial patterns of periodontal disease.

the genetic basis of this syndrome that affects younger individuals. However, they do strengthen the role of these genetic variants as potential modifiers of neutrophil function and disease expression in the presence of *Aa*. Further supportive evidence of the role of genetic variants associated with the innate immune response is provided by the association with genes that regulate prostanoid and leukotriene metabolism. A locus including prostaglandin reductase 1, which metabolizes eicosanoids, was associated with the *Aa* trait (PCT3). There are several reports of abnormally elevated eicosanoids [prostaglandin E_2 and leukotriene B4 (LTB_4)] within the GCF and in tissue biopsies in aggressive and early-onset periodontitis, when compared with the more common, less severe chronic periodontitis or in periodontal health (44–46). Abnormalities in LTB_4 levels are known to modify risk for *Mycobacterium tuberculosis* infection in the lung (47,48), suggesting a key role for leukotriene metabolism in microbial resistance at mucosal surfaces. The potential for abnormal IL-17 protection at the mucosal surface with missense variants of the IL-17 adaptor protein TRAFIP3 leading to altered chemokine response further supports the role of the potential of genetic modulation of the immune response in *Aa* dominated dysbiosis (Fig. 4).

The second important insight is offered in the *Pg* trait (PCT5), which is associated with the strong emergence of *P. gingivalis*, a well-established causal organism for severe chronic periodontitis. We found associations of this with several intronic SNPs of *PKP2*, *PNN* and *KIFC3* that are all associated with epithelial cell–cell adhesion, and include components of desmosomes and tight junctions. It is not unreasonable to suggest that any functional impairment in the epithelial boundary adjacent to the omnipresent subgingival biofilm would have the potential to increase the risk for disease. Interestingly, the *MORN2* locus contains a missense SNP (rs3099950) that is predicted to be a damaging E > K variant (Polyphen-2 score 1.0). MORN2 is a phagosome protein of macrophages that is involved with autophagy and promotes the recruitment of LC3, an autophagy protein also involved in phagocytosis, to *M. tuberculosis*-containing phagosomes and subsequent maturation to degradative phagolysosomes (49). Thus, MORN2 structural variants would appear to be promising candidates for further study. Clearly, these findings alone do not suffice as proof of the genetic basis of CP, as further replication, experiments and functional genomics will be needed to understand the role of these candidate loci and variants. However, the findings not only highlight important candidate genes for further investigation, but emphasize the utility of exploring GWAS in the context of the existing microbial and inflammatory signatures.

For decades we have recognized the tremendous heterogeneity of disease among patients with periodontitis in terms of individual subject's expression of severity, prognosis and response to therapy. These new findings provide a logical sub-classification of disease based upon genetic and microbial-inflammatory signatures that warrants further validation. One key limitation

of this study is that there are no additional data sets currently available that include clinical, microbial and genotype data to enable replication or meta-analyses. It is clear that these loci do not independently discriminate between clinical health and disease and have no diagnostic utility in this context. Nonetheless, the present investigation demonstrates the potential utility of constructing a biologically informed phenotype to identify candidate genes associated with pathogenic pathways in disease as reflected by microbial dysbiosis and inflammatory host response. A second limitation of this study is that the microbial characterization profiling was done using whole-genomic, high stringency probes—an approach that is not an unbiased 16s sequencing analysis that would include non-cultivable species. Although the lack of identification of non-cultivable species represents an important limitation of this investigation and an opportunity for future investigation, three points suggest that this method does not negatively impact our current findings. First, the six distinct microbial community structures that are associated with disease have microbial patterns that are consistent with previous studies that have performed 16s sequencing methods; including studies that employ multiple intraoral microbial sampling sites in each subject and among subjects with health, gingivitis, chronic periodontitis and aggressive periodontitis (50,51). Secondly, the non-cultivable pathogens associated with disease identified to-date tend to be associated with other organisms which are identified in the present investigation. For example, *Fretibacter fastidiosum* that belongs to the family *Synergistetes* and has only recently been cultivated (52) has been associated with disease and with high levels of *Pg* (53). Thus in the present study, we cannot exclude the possibility that identifying *Pg* as a dominant organism of the microbial community structure, as in PCT5, is actually a reflection of high levels of *Fretibacter* or some other non-identified pathogenic organism that co-exists within the dysbiotic community identified by these marker organisms. Similarly, certain PCTs have low levels of targeted bacteria and high IL-1β suggesting that perhaps the inflammation and attachment loss is attributable to microbes that were not within our panel, such as non-cultivable bacteria or viruses. This might suggest that some of the complex traits such as PCT2, PCT3 and PCT6 that are associated with periodontal disease may have incomplete phenotypic characterization to achieve statistical significance in single-marker analyses. It is also interesting that smoking and diabetes that are the two major risk factors for periodontal disease aggregate with PCT2 and especially PCT6. Finally, the identification and quantification of the microbes was performed using whole chromosomal DNA probes that do not identify bacterial at a strain level and do not allow the assessment of virulence of putative pathogens. Many oral pathogens possess virulence traits that enable them to selectively emerge in the plaque by inhibiting other microbes or preventing host clearance. For example, *Aa* or *Pg* that are associated with respective PCTs might have different bacteriocin and host-stealth properties than those *Aa* and *Pg* associated with the Socransky trait (PCT1) or other PCTs, enabling them to emerge within the subgingival biofilm.

In conclusion, the current investigation has provided new data that suggest a genetic basis of periodontal disease defined by microbial community structure and inflammatory characteristics. In addition, our results highlight loci and candidate genes for replication, functional analyses and further characterization as potential risk markers for subsets of individuals with this common disease, which is associated with microbial dysbiosis at the tooth–oral mucosal interface.

Materials and Methods

Study population

Subjects were selected from four US communities to participate in the Dental Atherosclerosis Risk in Communities (Dental ARIC) Study (54–56). For this investigation, genotype and clinical phenotype data were obtained from 4910 Northern European descendants in the Dental ARIC cohort. GCF levels of IL-1β, measured at four gingival sampling areas were assayed independently to create a mean GCF-IL1® level for each subject (55). The composition of the oral biofilm was quantitated by measuring levels of eight periodontal pathogens by microbe-specific DNA probes on 975 randomly selected EA individuals from the Dental ARIC cohort (54).

Clinical measurement and definition of periodontal disease

The Dental ARIC ancillary study participants were seen at the fourth visit of the ARIC Study to collect full-mouth periodontal data and biological samples as described elsewhere (54,56). Briefly, all sites on all teeth including third molars were examined using trained and calibrated examiners with >90% agreement (48). Measurements included number of missing teeth, gingival index, plaque index, probing depths, attachment loss measurements and bleeding upon probing. Biological samples were collected to include serum, GCF and subgingival plaque, as previously reported. For this investigation, the plaque samples were analyzed for levels of key periodontal pathogens and the GCF processed for IL-1β and subgingival plaque, as described subsequently. We used the Centers of Disease Control (CDC) and American Academy of Periodontology (AAP) three-level classification. The CDC/AAP classification uses interproximal attachment loss and probing depths to define disease categories as healthy, entry and severe (57). Finally, we used extent scores for the clinical parameters as continuous variables to describe the level of disease and to investigate correlations with principal components.

Quantification of periodontal organisms in plaque samples by DNA checkerboard analyses

Levels of bacteria within plaque samples were determined as previously described (55). One plaque sample was used from each subject, sampling the subgingival mesio-buccal site of the maxillary right first molar, and assayed by DNA whole chromosomal checkerboard for the eight periodontal pathogens. Levels of organism were expressed as counts using known microbial standards. Periodontal pathogen variables included counts of each of the following organisms [*P. gingivalis (Pg), Prevotella intermedia (Pi), T. denticola (Td), T. forsythia (Tf, formerly Bacteroides forsythus), C. rectus (Cr), Fusobacterium nucleatum (Fn), A. actinomycetemcomitans (Aa, formerly Actinobacillus actinomycetemcomitans), P. nigrescens (Pn)*]. Total counts reflect a summation of these targeted pathogens for each subject.

Measurements of IL-1β in GCF

GCF was collected from the mesio-buccal region of each first molar or, if missing, an alternate site as previously described in detail (58). Four GCF strips were eluted and analyzed separately for each subject. IL-1β was evaluated by enzyme-linked immune-absorbent assay (ELISA) according to the manufacturer's instructions as described (59). GCF analyte concentration data were pooled to provide a patient mean value in ng/mL.

Creating PCTs using PCA

PCA was performed to create PCTs that included subject-level periodontal disease status (CDC three-level definition), microbial composition (counts of eight periodontal pathogens) and mean GCF levels of IL-1β. The PROC PRINCOMP procedure was used to derive the PCTs using SAS v9.4 (SAS Institute, Cary, NC); correlation matrices with the PCTS and clinical traits were created using SAS PROC CORR. We carried forward to GWAS only those PCTs that were significantly associated with clinical disease (6 out of 11). The resulting principal components comprised the six PCTs (PCT1–6). Each PCT was characterized by a distinct microbial community structure and the level of IL-1β that were quantified by eigenvalues (loadings).

Genotyping, imputation and quality control

Blood collection and DNA extraction as performed at a central ARIC laboratory in Houston, TX has been described previously (14). Genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 chip which offers 906 600 SNP markers. The quality control procedures and SNP imputation based on HapMap Phase II CEU build 36 is described by Divaris *et al.* (18).

Estimates of relatedness and population stratification in ARIC was performed as described previously (14) to create a subset of 85 947 'high-quality' SNPs that met the following criteria: MAF ≥ 0.1, call rate >99.5%, Hardy–Weinberg equilibrium $P \geq 10^{-3}$. First- and second-degree relatives were identified and excluded, and 10 ancestry principal components were generated using EIGENSTRAT as previously described (15).

Analytical strategy

The association between individual SNPs and the six PCTs was tested using linear regression models assuming additive allelic effects, adjusting for age, sex, examination center and ancestry (10 first ancestral PCs). Beta coefficients, standard errors and *P*-values were estimated for each trait. We applied a multiple-test correction assuming 1 million independent tests, resulting to a genome-wide significance level of $P < 5 \times 10^{-8}$ using the ProbABEL software (60). Q–Q plots and Manhattan plots were generated and lambda coefficients were calculated for each complex trait using custom programs (14). LocusZoom ver.1.1 (61) was used to visualize genomic regions of interest. Coding, non-synonymous SNPs were interrogated using PolyPhen-2 for the prediction of potentially damaging protein structural changes (29).

Additional online resources included tools available at the National Center for Biotechnology Information (NCBI- http://www.ncbi.nlm.nih.gov/) and the UCSF Genome browser v274 at (http://genome.ucsc.edu/). Reporting of genes was based on the 'HUGO Gene Nomenclature' naming convention (http://www.genenames.org). The full names and genomic locations of the reported genes are presented in Supplementary Material, Table S2. Individual loci were annotated using 1000 genomes, when possible.

Gene-centric analyses of PCTs

Genome-wide association results of the six PCTs were interrogated for gene and gene-set enrichment using Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) using methods previously described (15). Gene-centric association results were corrected for gene size including a 110 kb upstream and 40 kb downstream region, number of SNPs and local linkage disequilibrium characteristics and a multiple testing-corrected *P*-value threshold of $0.05/18307 = 2.7 \times 10^{-6}$ was employed.

PCT single-marker association testing with chronic periodontitis in the larger dental ARIC cohort and aggressive periodontitis in an independent Northern European cohort

A true replication attempt of the PCT findings is currently impossible given the lack of other cohorts with 'deep' periodontal, microbial and inflammatory phenotyping and genotype data. For this reason, we examined the association of the identified loci with clinical disease, in the larger Dental ARIC cohort and an independent German sample as a validation attempt. This was done to determine any potential clinical relevance of these candidates emerging from the PCT analyses.

First, significant genes identified with MAGENTA for PCT1-6 were tested for association with periodontal disease status (extent interproximal attachment loss ≥3 mm) using orthogonal analysis. We conducted a GWA analysis as described above, adjusting for the other 10 PCTs (excluding the index trait) using EXTCAL3i+ as the phenotype. Then MAGENTA was used to obtain gene-centric *P*-values for the 80 genes identified as being significantly associated with the six PCTs. A stringent clinical definition of disease as a continuous variable was selected for association testing (EXTCAL3i+), using a Bonferroni corrected *P*-value of $0.05/80 = 0.00062$ as the statistical significance threshold. This allowed for the testing of candidate genes in the context of their index PCT, reflecting distinct microbial community structures and levels of IL-1β. The results of this analytical step should be interpreted with caution as there is the potential for false-positive findings due to collinearity. To examine for possible issues of collinearity (due to the inclusion of PCTs as covariates), we obtained VIFs via the SAS PROC REG '/VIF' option for the models used to produce Table 5. Although there is no consensus with critical VIF values to denote collinearity, VIF >10 is generally considered indicative of serious collinearity (37).

Furthermore, we obtained association results (betas and *P*-values) with aggressive periodontitis of SNPs highlighted in our single-marker and gene-centric analyses from an independent North and West German sample that has been reported by Schaefer and colleagues (32–35). This sample included 717 aggressive periodontitis cases [AgP; 302 males, 415 females, 355 smokers (155 unknown)] (33) and 4213 controls [2018 males, 2191 females (four unknown)], 2239 smokers (161 unknown) from the FOCUS (Food Chain Plus) (33), DOGS (Dortmunder Gesundheitsstudie) (34) and HNR1-3 (Heinz Nixdorf Recall Studie) (35) cohorts. German participants were genotyped on an IScan system with HumanOmni BeadChips (Illumina) as recently described (32). A description of this validation sample is given in Supplementary Material, Tables S6 and S7, and in Supplementary Material, Methods. Evidence of SNP association validation in the German cohort was determined by (i) multiple testing-corrected statistical significance ($P < 0.00135$); (ii) nominal statistical significance; (iii) directional consistency, accounting for testing of 37 SNPs.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

The authors thank the staff and participants of the ARIC study for their important contributions.

Genotyping of the AgP cases was performed at the Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Germany. We specially thank Andre Franke and Stefan Schreiber. The collection of sociodemographic and clinical data in the Dortmund Health Study was supported by the German Migraine & Headache Society (DMKG) and by unrestricted grants of equal share from Almirall, Astra Zeneca, Berlin Chemie, Boehringer, Boots Health Care, Glaxo-Smith-Kline, Janssen Cilag, McNeil Pharma, MSD Sharp & Dohme and Pfizer to the University of Muenster. Blood collection in the Dortmund Health Study was done through funds from the Institute of Epidemiology and Social Medicine University of Muenster. We specially thank Klaus Berger and Juergen Wellmann, Institute of Epidemiology and Social Medicine, University Münster, Germany. We are extremely grateful to all investigators who contributed to the generation of this data set. Specially, we gratefully thank Per Hoffmann, Institute of Human Genetics, University of Bonn, Germany and Human Genomics Research Group, Department of Biomedicine, University Hospital of Basel, Switzerland, and Bastian Krone, Institute of Medical Informatics, Biometry and Epidemiology, University Clinic Essen, Germany.

Conflict of Interest statement. None declared.

Funding

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); National Institutes of Health contract (HHSN268200625226C); National Institute of Environmental Health Sciences grant (P30ES010126); and National Institute of Dental and Craniofacial Research grants (R01DE11551, R01DE021418). Infrastructure was partly supported by grant number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The GWAS of AgP was supported by a research grant of the Deutsche Forschungsgemeinschaft DFG (GZ: SCHA 1582/3-1). The Heinz-Nixdorff-Recall (HNR) study is supported by the Heinz Nixdorf Foundation (Germany). Additionally, the study is funded by the German Ministry of Education and Science and the German Research Foundation (DFG; Project SI 236/8-1, SI236/9-1, ER 155/6-1). The genotyping of the Illumina HumanOmni-1 Quad Bead Chips of the HNR subjects was financed by the German Centre for Neurodegenerative DISEASES (DZNE), Bonn, Germany. Funding to pay the Open Access publication charges for this article was provided by Dr Offenbacher's Kenan Professorship funds (Steven Offenbacher DDS, PhD, MMSc WR Kenan Jr, Distinguished Professor).

References

1. Eke, P.I., Dye, B.A., Wei, L., Thornton-Evans, G.O. and Genco, R.J. (2012) Prevalence of periodontitis in adults in the United States: 2009 and 2010. *J. Dent. Res.*, **91**, 914–920.
2. Eke, P.I., Dye, B.A., Wei, L., Slade, G.D., Thornton-Evans, G.O., Borgnakke, W.S., Taylor, G.W., Page, R.C., Beck, J.D. and Genco, R.J. (2015) Update on prevalence of periodontitis in adults in the United States: NHANES 2009–2012. *J. Periodontol.*, **86**, 1–18.
3. Merchant, A., Pitiphat, W., Douglass, C.W., Crohin, C. and Joshipura, K. (2002) Oral hygiene practices and periodontitis in health care professionals. *J. Periodontol.*, **73**, 531–535.
4. Grossi, S.G., Genco, R.J., Machtet, E.E., Ho, A.W., Koch, G., Dunford, R., Zambón, J.J. and Hausmann, E. (1995) Assessment of risk for periodontal disease. II. Risk indicators for alveolar bone loss. *Periodontology*, **66**, 23–29.
5. Haffajee, A.D. and Socransky, S.S. (2001) Relationship of cigarette smoking to attachment level profiles. *J. Clin. Periodontol.*, **28**, 283–295.
6. Michalowicz, B.S., Aeppli, D., Virag, J.G., Virag, J.G., Klump, D.G., Hinrichs, J.E., Segal, N.L., Bouchard, T.J. and Pihlstroml, B.L. (1991) Periodontal findings in adult twins. *J. Periodontol.*, **62**, 293–299.
7. Michalowicz, B.S., Diehl, S.R., Gunsolley, J.C., Sparks, B.S., Brooks, C.N., Koertge, T.E., Califano, J.V., Burmeister, J.A. and Schenkein, H.A. (2000) Evidence of a substantial genetic basis for risk of adult periodontitis. *J. Periodontol.*, **71**, 1699–1707.
8. Hajishengallis, G. and Lamont, R.J. (2012) Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. *Mol. Oral. Microbiol.*, **27**, 409–419.
9. Teles, R., Teles, F., Frias-Lopez, J., Paster, B. and Haffajee, A. (2013) Lessons learned and unlearned in periodontal microbiology. *Periodontology 2000*, **62**, 95–162.
10. Teles, R.P., Gursky, L.C., Faveri, M., Rosa, E.A., Teles, F.R.F., Feres, M., Socransky, S.S. and Haffajee, A.D. (2010) Relationships between subgingival microbiota and GCF biomarkers in generalized aggressive periodontitis. *J. Clin. Periodontol.*, **37**, 313–323.
11. Lourenço, T.G.B., Heller, D., Silva-Boghossian, C.M., Cotton, S.L., Paster, B.J. and Colombo, A.P.V. (2014) Microbial signature profiles of periodontally healthy and diseased patients. *J. Clin. Periodontol.*, **41**, 1027–1036.
12. Abusleme, L., Dupuy, A.K., Dutzan, N., Silva, N., Burleson, J.A., Strausbaugh, L.D., Gamonal, J. and Diaz, P. (2013) The subgingival microbiome in health and periodontitis and its relationship with community biomass and inflammation. *ISME. J.*, **7**, 1016–1025.
13. Saygun, I., Kubar, A., Sahin, S., Sener, K. and Slots, J. (2008) Quantitative analysis of association between herpes viruses and bacterial pathogens in periodontitis. *J. Periodont. Res.*, **43**, 352–359.
14. Divaris, K., Monda, K.L., North, K.E., Olshan, A.F., Reynolds, L.M., Hsueh, W.C., Lange, E.M., Moss, K., Barros, S., Weyant, R.J. *et al.* (2013) Exploring the genetic basis of chronic periodontitis: a genome-wide association study. *Hum. Mol. Genet.*, **22**, 2312–2324.
15. Rhodin, K., Divaris, K., North, K.E., Barros, S.P., Moss, K., Beck, J.D. and Offenbacher, S. (2014) Chronic Periodontitis genome-wide association studies: gene-centric and gene set enrichment analyses. *J. Dent. Res.*, **93**, 882–890.
16. Shaffer, J.R., Polk, D.E., Wang, X., Feingold, E., Weeks, D.E., Lee, M.K., Cuenco, K.T., Weyant, R.J., Crout, R.J., McNeil, D.W. and Marazita, M.L. (2014) Genome-wide association study of periodontal health measured by probing depth in adults ages 18–49 years. *G3 (Bethesda)*, **4**, 307–314.
17. Teumer, A., Holtfreter, B., Volker, U., Petersmann, A., Nauck, M., Biffar, R., Volzke, H., Kroemer, H.K., Meisel, P., Homuth, G. and Kocher, T. (2013) Genome-wide association study of chronic periodontitis in a general German population. *J. Clin. Periodontol.*, **40**, 977–985.
18. Divaris, K., Monda, K.L., North, K.E., Olshan, A.F., Lange, E.M., Moss, K., Barros, S.P., Beck, J.D. and Offenbacher, S. (2012) Genome-wide association study of periodontal pathogen colonization. *J. Dent. Res.*, **91**, S21–S28.

19. Freitag-Wolf, S., Dommisch, H., Graetz, C., Jockel-Schneider, Y., Harks, I., Staufenbiel, I., Meyle, J., Eickholz, P., Noack, B., Bruckmann, C. *et al.* (2014) Genome-wide exploration identifies sex-specific genetic effects of alleles upstream NPY to increase the risk of severe periodontitis in men. *J. Clin. Periodontol.*, **41**, 1115–1121.
20. Elinav, E., Strowig, T., Kau, A.L., Henao-Mejia, J., Thaiss, C.A., Booth, C.J., Peaper, D.R., Bertin, J., Eisenbarth, S.C., Gordon, J.I. and Flavell, R.A. (2011) NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell*, **145**, 745–757.
21. Suo, C., Toulopoulou, T., Bramon, E., Walshe, M., Picchioni, M., Murray, R. and Ott, J. (2013) Analysis of multiple phenotypes in genome-wide genetic mapping studies. *BMC Bioinform.*, **14**, 151–157.
22. Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M.M., Hysi, P.G., Wollstein, A., Lao, O., Bruijne, M.D., Ikram, M.A. *et al.* (2012) A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet.*, **8**, e1002932.
23. Karasik, D., Cupples, L.A., Hannan, M.T. and Kiel, D.P. (2004) Genome screen for a combined bone phenotype using principal component analysis: the Framingham study. *Bone*, **34**, 547–556.
24. McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P.A. and Hirschhorn, J.N. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev.*, **9**, 356–369.
25. Stranger, B.E., Stahl, E.A. and Raj, T. (2011) Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*, **187**, 367–383.
26. Klein, P. (1994) *An Easy Guide to Factor Analysis*. Taylor Francis Ltd, UK.
27. Büyüköztürk, Ş. (2004) *Social sciences data analysis hand book*, pp. 140–175. Pegem Publishing, Ankara.
28. Socransky, S.S., Haffajee, S.S., Cugini, A.D. and Smith, C.K., Jr (1998) Microbial complexes in subgingival plaque. *J. Clin. Periodontol.*, **25**, 134–144.
29. Jin, T., Perry, A., Jiang, J., Smith, B., Curry, J.A., Unterholzner, L., Jiang, Z., Horvath, G., Rathinam, V.A., Johnstone, R.W. *et al.* (2012) Structures of the HIN domain: DNA complexes reveal ligand binding and activation mechanisms of the AIM2 inflammasome and IFI16 receptor. *Immunity*, **36**, 561–571.
30. Dzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods.*, **7**, 248–249.
31. Segrè, A.V., Groop, L., Mootha, V.K., Daly, M.J. and 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.
32. Schaefer, A.S., Bochenek, G., Jochens, A., Ellinghaus, D., Dommisch, H., Guzeldemir-Akcakanat, E., Graetz, C., Harks, I., Jockel-Schneider, Y., Weinspach, K. *et al.* (2015) Genetic evidence for plasminogen as a shared genetic risk factor of coronary artery disease and periodontitis. *Circ. Cardiovasc. Genet.*, **8**, 159–167.
33. Muller, N., Schulte, D.M., Turk, K., Freitag-Wolf, S., Hampe, J., Zeuner, R., Schroder, J.O., Gouni-Berthold, I., Berthold, H.K., Krone, W. *et al.* (2015) IL-6 blockade by monoclonal antibodies inhibits apolipoprotein (a) expression and lipoprotein (a) synthesis in humans. *J. Lipid Res.*, **56**, 1034–1042.
34. Berger, K. (2012) DHS: The Dortmund health study. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* **55**, 816–821.
35. Schmermund, A., Mohlenkamp, S., Stang, A., Gronemeyer, D., Seibel, R., Hirche, H., Mann, K., Siffert, W., Lauterbach, K., Siegrist, J. *et al.* (2002) Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk factors, evaluation of coronary calcium and lifestyle. *Am. Heart J.*, **144**, 212–218.
36. MacRae, C.A. and Pollak, M.R. (2015) Effect size does matter the long road to mechanistic insight from genome-wide association. *Circulation*, **132**, 1943–1945.
37. Belsley, D.A., Kuh, E. and Welsch, R.E. (2005) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, New York, NY.
38. Offenbacher, S., Barros, S.P., Paquette, D.W., Winston, J.L., Biesbrock, A.R., Thomason, R.G., Gibb, R.D., Fulmer, A.W., Tiesman, J.P., Juhlin, K.D. *et al.* (2009) Gingival transcriptome patterns during induction and resolution of experimental gingivitis in humans. *J. Periodontol.*, **80**, 1963–1982.
39. Dębniak, T., Soczawa, E., Boer, M., Rózewicka-Czabańska, M., Wiśniewska, J., Serrano-Fernandez, P., Mirecka, A., Paszkowska-Szczur, K., Lubinski, J., Krysztoforska, L. *et al.* (2014) Common variants of ZNF750, RPTOR and TRAF3IP2 genes and psoriasis risk. *Arch. Dermatol. Res.*, **306**, 231–238.
40. Ciccacci, C., Biancone, L., Di Fusco, D., Ranieri, M., Condino, G., Giardina, E., Onali, S., Lepre, T., Pallone, F., Novelli, G. *et al.* (2013) TRAF3IP2 gene is associated with cutaneous extraintestinal manifestations in inflammatory bowel disease. *J. Crohns. Colitis*, **7**, 44–52.
41. Fredman, G., Oh, S.F., Ayilavarapu, S., Hasturk, H., Serhan, C.N. and Van Dyke, T.E. (2011) Impaired phagocytosis in localized aggressive periodontitis: rescue by Resolvin E1. *PLoS One*, **6**, e24422.
42. Khocht, A. and Albandar, J.M. (2014) Aggressive forms of periodontitis secondary to systemic disorders. *Periodontology 2000*, **65**, 134–148.
43. Tyagi, S.R., Uhlinger, D.J., Lambeth, J.D., Champagne, C. and Van Dyke, T.E. (1992) Altered diacylglycerol level and metabolism in neutrophils from patients with localized juvenile periodontitis. *Infect. Immun.*, **60**, 2481–2487.
44. Emingil, G., Çinarcik, C., Baylas, H., Çoker, I. and Hüseyinov, A. (2001) Levels of leukotriene B4 in gingival crevicular fluid and gingival tissue in specific periodontal diseases. *J. Periodontol.*, **72**, 1025–1031.
45. Kantarci, A., Oyaizu, K. and Van Dyke, T.E. (2003) Neutrophil-mediated tissue injury in periodontal disease pathogenesis: findings from localized aggressive periodontitis. *J. Periodontol.*, **74**, 66–75.
46. Pradeep, A.R., Manjunath, S.G., Swati, P.P., Shikha, C. and Sujatha, P.B. (2007) Gingival crevicular fluid levels of leukotriene B4 in periodontal health and disease. *J. Periodontol.*, **78**, 2325–2330.
47. el-Ahmady, O., Mansour, M., Zoeir, H. and Mansour, O. (1997) Elevated concentrations of interleukins and leukotriene in response to Mycobacterium tuberculosis infection. *Ann. Clin. Biochem.*, **34**, 160–164.
48. Tobin, D.M., Roca, F.J., Ray, J.P., Ko, D.C. and Ramakrishnan, L. (2013) An enzyme that inactivates the inflammatory mediator leukotriene b4 restricts mycobacterial infection. *PLoS One*, **8**, e67828.
49. Abnave, P., Mottola, G., Gimenez, G., Boucherit, N., Trouplin, V., Torre, C., Conti, F., Ben Amara, A., Lepolard, C., Djian, B. *et al.* (2014) Screening in planarians identifies MORN2 as a key component in LC3-associated phagocytosis

and resistance to bacterial infection. *Cell Host Microbe*, **16**, 338–350.
50. Liu, B., Faller, L.L., Klitgord, N., Mazumdar, V., Ghodsi, M., Sommer, D., Gibbons, T.R., Treangen, T.J., Chang, Y., Li, S. *et al.* (2012) Deep sequencing of the oral microbiome reveals signatures of periodontal disease. *PLoS One*, **7**, e37919.
51. Duran-Pinedo, A.E., Chen, T., Teles, R., Starr, J.R., Wang, X., Krishnan, K. and Frias-Lopez, J. (2014) Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis. *ISME J.*, **8**, 1659–1672.
52. Vartoukian, S.R., Downes, J., Palmer, R.M. and Wade, W.G. (2013) Fretibacterium fastidiosum gen. nov., sp. nov., isolated from the human oral cavity. *Int. J. Syst. Evol. Microbiol.*, **63**, 458–463.
53. Wade, W.G. (2013) Characterisation of the human oral microbiome. *J. Oral Biosci.*, **55**, 143–148.
54. Beck, J.D., Elter, J.R., Heiss, G., Couper, D., Mauriello, S.M. and Offenbacher, S. (2001) Relationship of periodontal disease to carotid artery intima–media wall thickness: the atherosclerosis risk in communities (ARIC) study. *Arter. Thromb. Vasc. Biol.*, **21**, 1816–1822.
55. Offenbacher, S., Barros, S.P., Singer, R.E., Moss, K., Williams, R.C. and Beck, J.D. (2007) Periodontal disease at the biofilm–gingival interface. *J. Periodontol.*, **78**, 1911–1925.
56. Elter, J.R., Champagne, C.M.E., Offenbacher, S. and Beck, J.D. (2004) Relationship of periodontal disease and tooth loss to prevalence of coronary heart disease. *J. Periodontol.*, **75**, 782–790.
57. Page, R.C. and Eke, P.I. (2007) Case definitions for use in population-based surveillance of periodontitis. *J. Periodontol.*, **78**, 1387–1399.
58. Champagne, C.M.E., Buchanan, W., Reddy, M.S., Preisser, J.S., Beck, J.D. and Offenbacher, S. (2003) Potential for gingival crevice fluid measures as predictors of risk for periodontal diseases. *Periodontology 2000*, **31**, 167–180.
59. Offenbacher, S., Lin, D., Strauss, R., McKaig, R., Irving, J., Barros, S.P., Moss, K., Barrow, D., Hefti, A. and Beck, J.D. (2006) Effects of periodontal therapy during pregnancy on periodontal status, biologic parameters, and pregnancy outcomes: a pilot study. *J. Periodontol.*, **77**, 2011–2024.
60. Aulchenko, Y.S., Struchalin, M.V. and van Duijn, C.M. (2010) ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinform.*, **11**, 134–144.
61. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.