The *PF4/PPBP/CXCL5* Gene Cluster Is Associated with Periodontitis

A. Shusterman[1], M. Munz[2,3], G. Richter[2], S. Jepsen[4], W. Lieb[5], B. Krone[6], P. Hoffman[7,8], M. Laudes[9], J. Wellmann[10], K. Berger[10], T. Kocher[11], S. Offenbacher[12], K. Divaris[13,14], A. Franke[15], S. Schreiber[9,15], H. Dommisch[2], E. Weiss[16], A.S. Schaefer[2*], Y. Houri-Haddad[1*], and F.A. Iraqi[17*]

Abstract

Periodontitis is a common dysbiotic inflammatory disease with an estimated heritability of 50%. Due to the limited sample size of available periodontitis cohorts and the underlying trait heterogeneity, genome-wide association studies (GWAS) of chronic periodontitis (CP) have been unsuccessful in discovering susceptibility factors. A strategy that combines agnostic GWAS with a well-powered candidate-gene approach has the potential to discover novel loci. We combined RNA-seq data from gingival tissues with quantitative trait loci (QTLs) that were identified in a F_2-cross of mice resistant and susceptible to infection with oral bacterial pathogens. Four genes, which were located within the mapped QTLs, showed differential expression. The chromosomal regions across the human orthologous were interrogated for putative periodontitis-associated variants using existing GWAS data from a German case-control sample of aggressive periodontitis (AgP; 651 cases, 4,001 controls), the most severe and early onset form of periodontitis. Two haplotype blocks, one upstream to the coding region of *UGT2A1* (rs146712414, $P = 9.1 \times 10^{-5}$; odds ratio [OR], 1.34; 95% confidence interval [CI], 1.16–1.56) and one downstream of the genes *PF4/PPBP/CXCL5* (rs1595009, $P = 1.3 \times 10^{-4}$; OR, 1.32; 95% CI, 1.15–1.52), were associated with AgP. The association of rs1595009 was validated in an independent cohort of CP of European Americans (1,961 cases and 1,864 controls; $P = 0.03$; OR, 1.45; 95% CI, 1.01–1.29). This association was further replicated in another sample of 399 German CP cases (disease onset <60 y of age) and 1,633 controls ($P = 0.03$; OR, 1.75; 95% CI, 1.06–2.90). The combined estimates of association from all samples were $P = 2.9 \times 10^{-5}$ (OR, 1.2; 95% CI, 1.1–1.3). This study shows the strength of combining QTL mapping and RNA-Seq data from a mouse model with association studies in human case-control samples to identify genetic risk variants of periodontitis.

Keywords: QTL mapping, association, genetic, alveolar bone loss, mice model, GWAS

[1]Department of Prosthodontics, Hadassah Faculty of Dental Medicine, Hebrew University, Jerusalem, Israel
[2]Department of Periodontology and Synoptic Medicine, Institute for Dental and Craniofacial Sciences, Charité–University Medicine Berlin, Berlin, Germany
[3]Institute for Integrative and Experimental Genomics, University Medical Center Schleswig-Holstein–Campus Lübeck, Lübeck, Germany
[4]Department of Periodontology, Operative and Preventive Dentistry, University of Bonn, Bonn, Germany
[5]Institute of Epidemiology, Biobank popgen, Christian-Albrechts-University, Kiel, Germany
[6]Institute of Medical Informatics, Biometry and Epidemiology, University Clinic Essen, Essen, Germany
[7]Institute of Human Genetics, University of Bonn, Bonn, Germany
[8]Germany und Human Genomics Research Group, Department of Biomedicine, University Hospital of Basel, Basel, Switzerland
[9]Clinic of Internal Medicine, University Clinic Schleswig-Holstein, Kiel, Germany
[10]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany
[11]Unit of Periodontology, Department of Restorative Dentistry, Periodontology, Endodontology, Preventive Dentistry and Pedodontics, Dental School, University Medicine Greifswald, Greifswald, Germany
[12]University of North Carolina–Chapel Hill, School of Dentistry, Department of Periodontology, Chapel Hill, NC, USA
[13]University of North Carolina–Chapel Hill, School of Dentistry, Department of Pediatric Dentistry, Chapel Hill, NC, USA
[14]University of North Carolina–Chapel Hill, Gillings School of Global Public Health, Department of Epidemiology, Chapel Hill, NC, USA
[15]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany
[16]Maurice and Gabriella Goldschleger School of Dental Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[17]Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
*Authors contributing equally to this article.

A supplemental appendix to this article is available online.

Corresponding Authors:
Y. Houri-Haddad, Department of Prosthodontics, Faculty of Dental Medicine, Hebrew University, Hadassah, Jerusalem 9112102 Israel.
Email: yaelho@ekmd.huji.ac.il
F. Iraqi, Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel.
Email: fuadi@post.tau.ac.il
A.S. Shaefer, Department of Periodontology and Synoptic Dentistry, Institute of Dental, Oral and Maxillary Medicine, Charité–University Medicine Berlin, Aßmannshauser Str. 4-6, 14197 Berlin, Germany.
Email: arne.schaefer@charite.de

Introduction

Periodontitis is one of the most common inflammatory diseases, with a prevalence rate for the severe forms of approximately 11% worldwide (Marcenes et al. 2013). Chronic periodontitis (CP) is the most common form and characterized by late onset and slow disease progression (Eke et al. 2012). It is the major cause of tooth loss in adults older than 40 y (Petersen 2003; Eke et al. 2012; Marcenes et al. 2013). The disease is characterized by destruction of the alveolar bone due to an aberrant host inflammatory response to a dysbiotic subgingival microbiome. The estimated heritability of 50% indicates the important yet postulated role of genetic susceptibility factors (Michalowicz et al. 1991; Corey et al. 1993; Michalowicz 1994). On the other hand, aggressive periodontitis (AgP) is a rare but severe and early onset form of the disease; it is believed that it has a strong genetic component, but specific implicated molecular disease mechanisms remain largely unknown. These different phenotypes of periodontitis can be considered different parts of a large range of similar conditions, which are attributed to the effects of different combinations of genetic risk loci, which form the personal genetic constitution and determine the individual immune response. In this view, the different disease manifestations are not confined entities but share risk alleles and covariates.

Previous genome-wide association studies (GWAS) of CP have reported several "suggestive" susceptibility loci but have hitherto failed to produce genome-wide significant evidence of association for common variants (Divaris et al. 2012, 2013; Teumer et al. 2013; Hong et al. 2015; Shimizu et al. 2015; Sanders et al. 2016), although a recent publication reported the rare variant rs149133391 (minor allele frequency [MAF] = 0.01%, TSNAX-DISC1) to be associated with CP at genome-wide significance in a Hispanic population (Sanders et al. 2016). These unremarkable results are likely reflections of the relatively small sample size of periodontitis cohorts and the possible underlying trait heterogeneity. To date, a single genome-wide significant locus has been reported for AgP (Schaefer et al. 2010), and a few additional loci/genes below the genome-wide significance threshold were validated in independent studies (Schaefer et al. 2009, 2011; Bochenek et al. 2013; Schaefer et al. 2015).

Given the sample size limitations of the available samples and the general scarcity of high-quality phenotype-genotype cohorts of periodontitis, GWAS alone are unlikely to identify the missing susceptibility loci. In this case, association studies of well-prioritized candidate genes may be more promising, as confined genetic regions require a smaller number of tests and thus are more efficient. An ideal strategy would combine the advantages of the unbiased hypothesis-free GWAS with the increased power of candidate-gene studies. Such a strategy requires a system for unbiased de novo generation of hypotheses on candidate genes. To this end, we developed a genetic approach using a mouse model and human AgP and CP GWAS samples. In an earlier study, we sought to identify and prioritize quantitative trait loci (QTLs) conferring susceptibility to alveolar bone loss upon infection with periodontal pathogens in mice. Specifically, we crossed the mouse line BALB/cJ, which is susceptible to infection with the human periodontal pathogens *Porphyromonas gingivalis* (PG) and *Fusobacterium nucleatum* (FN), with the resistant mouse line A/J into the F_2 generation. In addition, we demonstrated the heritability of the variation in disease susceptibility between these founder lines and identified 2 significant QTLs. These QTLs are located on the murine chromosomes 5 and 3 (at genomic intervals 67.39–90.35 and 43.53–53 Mb), include approximately 200 genes, and were designated *Perio* 1 and 2, respectively (Shusterman et al. 2013).

Given that expression patterns in mouse models recapitulate those among humans (Takao and Miyakawa 2015), the examination of gene expression differences within a mapped QTL between susceptible and resistant mice is a promising strategy to identify a limited number of candidate genes, which subsequently can be tested for disease association in humans. In the current study, we provide evidence in support of murine models to study the role of gene variants and environmental factors on human disease. We give evidence that the combination of RNA-seq and QTL mapping in mice allows the identification of promising candidate genes with relevance to human diseases, as well as report the association of CP and AgP with the *PF4/PPBP/CXCL5* gene cluster.

Materials and Methods

RNA Sequencing

Eight samples (RNA integrity ≥7) from right hemi-maxilla of the F_2 population from 4 resistant and 4 susceptible mice were used to prepare a library of template molecules for subsequent sequencing on a HiSeq platform (Illumina) as described in detail in the online supplementary data (Supplement Material 1). Raw data were normalized and analyzed for differential expression using the software cuffdiff from the CuffLinks package v2.1.1 (Trapnell et al. 2012) and DESeq v1.18.0 (Bioconductor). Differentially expressed genes (DEGs) were required to have a *q* value <0.1, a FPKM (Fragments per Kilobase of transcript per Million mapped reads) expression level >0 in both conditions, and a false discovery rate <0.1.
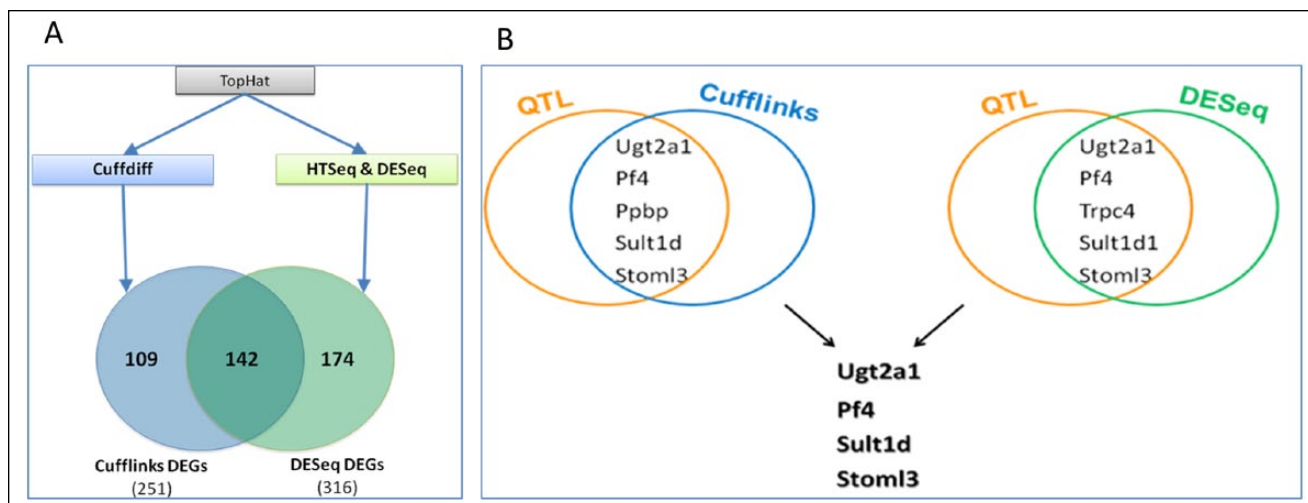
Candidate Gene Selection

Candidate genes were compared between QTLs and CuffLinks results and between QTL and DESeq results. Genes that were shared between these 2 data sets were selected (Shusterman et al. 2013: Appendix Table).

Human Study Populations

A description of the human study population is given in the online supplementary data.

(1) AgP (Germany). The AgP patients were recruited throughout Germany and Vienna, Austria, and were described before

Figure 1. Candidate gene selection by gene expression. (**A**) Schematic workflow of gene expression analyses using CuffLinks and DESeq. The data were analyzed with the software packages CuffLinks and DESeq. A subset of 251 differentially expressed genes (DEGs) was detected by CuffLinks and 316 DEGs were detected by DESeq. In total, 142 DEGs were detected by both analyses. (**B**) Strategy for candidate gene selection. The candidate gene selection was based on the combination of quantitative trait loci (QTL) mapping and the comparison of gene expression by CuffLinks and DESeq between resistant and susceptible F_2 offspring. Genes that were identified by both software tools and were located at a QTL were selected as candidate genes for association studies in the human samples.

(Schaefer et al. 2015). The AgP control sample consisted of German individuals, who were recruited from the Competence Network "FoCus—Food Chain Plus" (Muller et al. 2015), the Dortmund Health Study (Berger 2012), and the Heinz Nixdorf Recall Studies 1 to 3 (Schmermund et al. 2002).

CP (United States). The American CP cohort was used as described in Divaris et al. (2012) and consisted of European American participants of the Dental Atherosclerosis Risk In Communities (ARIC) study. Individuals who were periodontally healthy or had mild periodontitis were used as controls (ARIC Investigators 1989).

CP (Germany). The German CP cases and controls were first described in Schaefer et al. (2011).

Genotyping, Association Tests, and eQTL Search

Genotyping of the AgP sample was performed with OmniExpress arrays on an iScan System (Illumina). Single-nucleotide polymorphism (SNP) rs1595009 was genotyped for the German CP sample on 384-well plates using TaqMan assay hCV8839363 (Applied Biosystems) as previously described (de Jong et al. 2014). Association tests were performed using the additive model in SNPTEST v2.5.2 (Marchini et al. 2007). Linkage disequilibrium (LD) was calculated with Plink v1.9 by using a window size of 1 MB and the 1000 Genomes phase III data with European descent ($n = 680$) as the LD reference. Putative effects of the disease-associated SNPs rs1595009 and rs2472649 were identified by a bioinformatics query of the databases haploreg (v4.1) (Ward and Kellis 2012) and GTEx (v.6) (GTEx Consortium 2015) for expression quantitative trait loci (eQTL) according to the reference SNP cluster ID (rsID) and the hg19 position.

Results

Generation of Candidate Genes by RNA-seq

The analysis of the RNA-seq data using the software tools CuffLinks and DESeq identified 251 and 316 DEGs between the 2 groups of F_2 mice, which were susceptible or resistant to oral infection with PG and FN. The 2 groups showed an overlap of 142 genes (Fig. 1A, Appendix Table). Five DEGs, predicted by CuffLinks, were located within the previously identified 2 QTLs (Shusterman et al. 2013). Within the first QTL located on chromosome 3 (chr3), this was *Stoml3* (stomatin-like 3; chr3:53,488,653-53,508,502; GRCm38). Within the more significant second QTL on chr5, these were *Ugt2a1* (UDP glucuronosyltransferase 2 family, polypeptide A1; chr5:87,459,490-87,490,871), *Sult1d1* (sulfotransferase family 1D member 1, pseudogene; chr5:87,554,645-87,569,027), *Ppbp* (pro-platelet basic protein; chr5:90,768,518-90,770,063), and *Pf4* (platelet factor 4; chr5:90,772,435-90,773,383).

Two DEGs, which were predicted by DESeq, *Stoml3* and *Trpc4* (transient receptor potential cation channel, subfamily C, member 4; chr3:54,156,035-54,318,471, which was not detected by CuffLinks), were located within the QTL on chr3. Within the QTL on chr5, DESeq proposed 3 DEGs, which were also predicted by CuffLinks (*Ugt2a1, Sult1d1, Pf4*). *Ppbp*, which is located upstream of *Pf4*, was not detected by DESeq. Four DEGs were selected as candidate genes for association tests in the human case-control samples, as illustrated in Figure 1B. These 4 genes were *Ugt2a1, Pf4, Sult1d1*, and *Stoml3*.

Association Study in a Case-Control Sample of AgP

Of the 4 selected murine genes, the human genome contains protein coding orthologues for 3: *UGT2A1* (GRCh37/hg19,
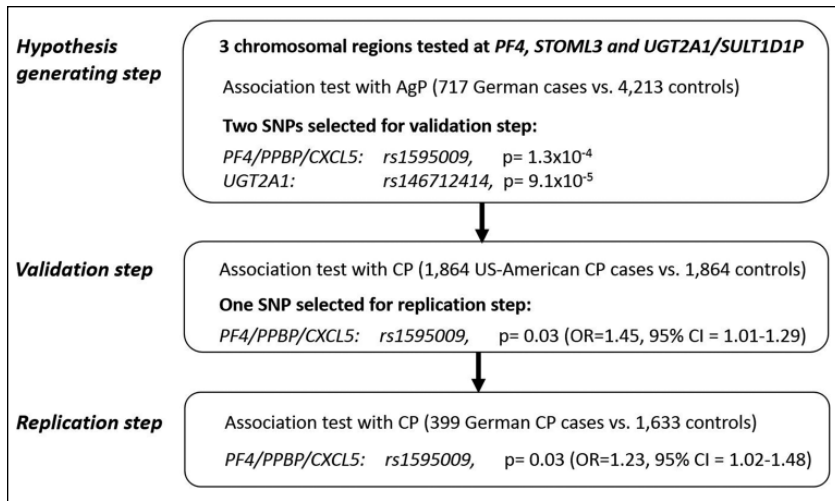
Figure 2. Schematic workflow of the study design of the human association analysis. (1) Hypothesis-generating step: all single-nucleotide polymorphisms (SNPs) within 200 kb up- and downstream of the candidate genes were tested for association with aggressive periodontitis (AgP), the most severe phenotype (*PF4/PPBP/CXCL5*: 1,130 SNPs; *STOML3*: 2,098 SNPs; *UGT2A1-SULT1D1P*: 363 SNPs). SNPs with an association of $P < 10^{-3}$ (additive model) were located at *PF4/PPBP/CXCL5* and *UGT2A1*. (2) The most significant tagging SNPs of each of the 2 AgP-associated haplotype blocks at *PF4/PPBP/CXCL5* and *UGT2A1* were selected for validation in a large sample of the moderate phenotype chronic periodontitis (CP). The association of rs1595009 at *PF4/PPBP/CXCL5* was nominally significant (additive model), but not rs146712414 at *UGT2A1*. (3) rs1595009 was selected for replication in an independent case-control sample of CP and showed nominal significant association at the recessive model. (After increasing the size of the control sample by adding the German AgP controls, the association was also nominally significant at the additive model; see the main text.)

chr4:70,454,135-70,518,967), *PF4* (chr4:74,846,542-74,847, 841), and *STOML3* (chr13:39,540,062-39,564,996). For the murine gene *Sult1d1*, the human orthologous gene was annotated as the pseudogene *SULT1D1P* (chr4:70657590-70679606). To search the chromosomal regions across the human orthologous genes for putative periodontitis-associated SNPs, we leveraged the early onset disease phenotype AgP in the hypothesis-generating step. AgP is considered to be enriched in genetic risk factors compared with late-onset, more moderate forms such as CP, thus offering benefits in terms of decreasing the probability of false-positive findings carried into the replication.

All SNPs within 200 kilobases (kb) up and downstream of the candidate genes were included in the analysis, which correspond with a total of 1,130, 2,098, and 363 SNPs for *PF4/PPBP/CXCL5, STOML3*, and *UGT2A1-SULT1D1P*, respectively, in our imputed genotype data. In the 2 candidate gene loci *PF4/PPBP/CXCL5* and *UGT2A1*, we identified SNPs with $P < 10^{-3}$ (Fig. 2). A haplotype block downstream of *PF4/PPBP/CXCL5* showed an association with AgP (haplotype-tagging index SNP rs1595009; $P = 1.3 \times 10^{-4}$; odds ratio [OR], 1.32; 95% confidence interval [CI], 1.15–1.52; MAF cases = 20%, MAF controls = 16%; Table 1, Fig. 3). Of note, a GWAS lead-SNP of inflammatory bowel disease (IBD), rs2472649, is located upstream of *PF4* and *PPBP* but downstream of *CXCL5* (Jostins et al. 2012). This SNP ($r^2 = 0.71$ with rs1595009, MAF cases = 18%, MAF controls = 15%) was associated with AgP,

albeit with lower significance ($P = 3.0 \times 10^{-4}$; OR, 1.31; 95% CI, 1.13–1.51) and with the same effect direction as for IBD.

Upstream of the coding region of *UGT2A1*, a haplotype block was also associated with AgP (lead SNP: rs146712414, $P = 9.1 \times 10^{-5}$; OR, 1.34; 95% CI, 1.16–1.56; MAF cases = 18%, MAF controls = 14%). The chromosomal regions across *STOML3* and *SULT1D1P* were not associated with AgP.

Validation in a US Case-Control Sample of CP

Next, we selected the most significant tagging SNPs of each of the 2 haplotype blocks at *PF4/PPBP/CXCL5* and *UGT2A1* for validation in a cohort of European Americans (American whites) with moderate periodontitis ($n = 1,961$). Individuals who were periodontally healthy or had only mild periodontitis were used as controls ($n = 1,864$) (Divaris et al. 2013). In this sample, which had a statistical power >0.90, rs1595009 was associated with $P = 0.03$ (OR, 1.45; 95% CI, 1.01–1.29; Table 1). The IBD GWAS lead SNP rs2472649 showed borderline association with $P = 0.05$ (OR, 1.15; 95% CI, 1.0–1.31).

When we used the smaller case sample of US cases with severe CP with the same controls, we also tested the available sample of US severe CP cases ($n = 785$) versus the healthy controls. This sample gave a statistical power of <0.75. With this sample, both SNPs showed no associations. The haplotype upstream of *UGT2A1* showed no association with moderate or severe CP.

Replication in a German Case-Control Sample of CP

Of SNPs that were tested in the validation step, only the haplotype at *PF4/PPBP/CXCL5* indicated association with CP, with rs1595009 showing the strongest association. Thus, rs1595009 was selected for replication in a German CP case-control sample (399 cases; age of disease onset <60 y; 1,633 healthy controls). In this sample, rs1595009 showed no association in the additive model but a borderline significant association in the recessive model, with $P = 0.03$ (OR, 1.75; 95% CI, 1.06–2.90). When the German CP cases were matched to the larger control sample of the AgP panel, rs1595009 was associated with CP in the allelic model ($P = 0.03$; OR, 1.23; 95% CI, 1.02–1.48) as well as the recessive model ($P = 0.0004$; OR, 2.29; 95% CI, 1.43–3.65). The combined estimates of association from the German AgP sample, the European American moderate CP sample, and the German CP sample were $P_{pooled} = 2.9 \times 10^{-5}$ (OR, 1.2; 95% CI, 1.1–1.3).

Table 1. Estimates of Association for the *PF4/CXCL5* Locus.

Locus	SNP	Sample	P Value	OR (95% CI)
PF4/CXCL5	rs1595009	AgP (German)	1.3×10^{-4}	1.32 (1.2–1.5)
		Severe CP (United States)	7.3×10^{-1}	1.03 (0.9–1.2)
		Moderate CP (United States)	2.9×10^{-2}	1.45 (1.01–1.3)
		CP (Germany)	3×10^{-2a}	1.75 (1.1–2.9)
	rs2472649	AgP (German)	7.5×10^{-4}	0.78 (0.7–0.9)
		Severe CP (United States)	3.1×10^{-1}	1.10 (0.9–1.3)
		Moderate CP (United States)	5.1×10^{-2}	1.15 (1.0–1.3)

AgP, aggressive periodontitis; CI, confidence interval; CP, chronic periodontitis; OR, odds ratio; SNP, single-nucleotide polymorphism.
[a]Recessive model.
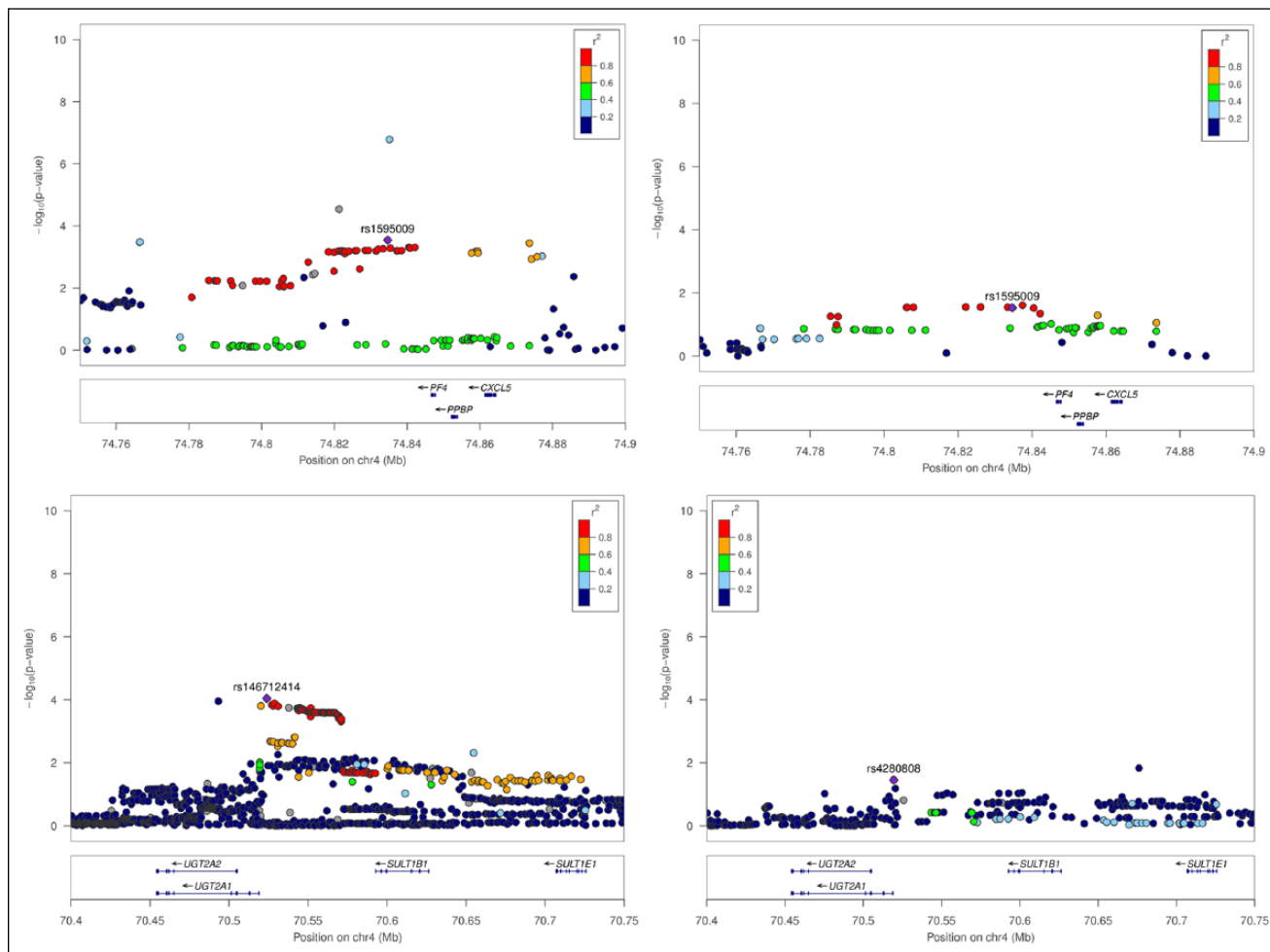
Figure 3. Regional association plots of *PF4, UGT2A2, STOML3,* and *SULT1D1P*. The $-\log_{10} P$ values of the analyzed single-nucleotide polymorphisms (SNPs) were plotted as a function of the genomic SNP position (NCBI build 36, hg19/1000 Genomes 2012, CEU). Annotations were generated by LocusZoom. **Top left:** *PF4/CXCL5* locus—aggressive periodontitis (AgP); **top right:** *PF4/CXCL5* locus—moderate chronic periodontitis (CP); **lower left:** *UGT2A2/SULT1B1* locus—AgP; **lower right:** *UGT2A2/SULT1B1* locus—moderate CP.

In Silico Prediction of Downstream Effects on Trans-*Gene Expression*

Examination of the eQTL mapping databases GTEx (Gene-Tissue-Expression Project; GTEx Consortium 2015) and haploreg (v.4.1; Ward and Kellis 2012) indicated effects on the transcription level for the 2 AgP- and CP-associated SNPs at

PF4/PPBP/CXCL5, rs1595009 and rs2472649. It also indicated regulatory effects of the AgP-associated SNP rs146712 414 on *UGT2A1* (Table 2).

For both *PF4/PPBP/CXCL5* SNPs, the most significant effects were found in blood for the distant gene *PF4V1* (platelet factor 4 variant 1; chr4:74,719,013-74,720,198 [GRCh37/hg19]). *PF4V1* is also located within QTL2 of the F_2 mouse

Table 2. Expression Effects of rs1595009 and rs2472649 (gtexportal.org).

SNP	Target Gene	P Value	Effect	Source
rs1595009	*PF4V1*	3.7×10^{-10}	*trans*	Whole blood
	CXCL5	6.5×10^{-6}	*cis*	
rs2472649	*PF4V1*	5.6×10^{-8}	*trans*	
	CXCL5	9.0×10^{-9}	*cis*	
rs146712414	*1S1NSC*	1.1×10^{-5}	*trans*	Fibroblasts

SNP, single-nucleotide polymorphism.
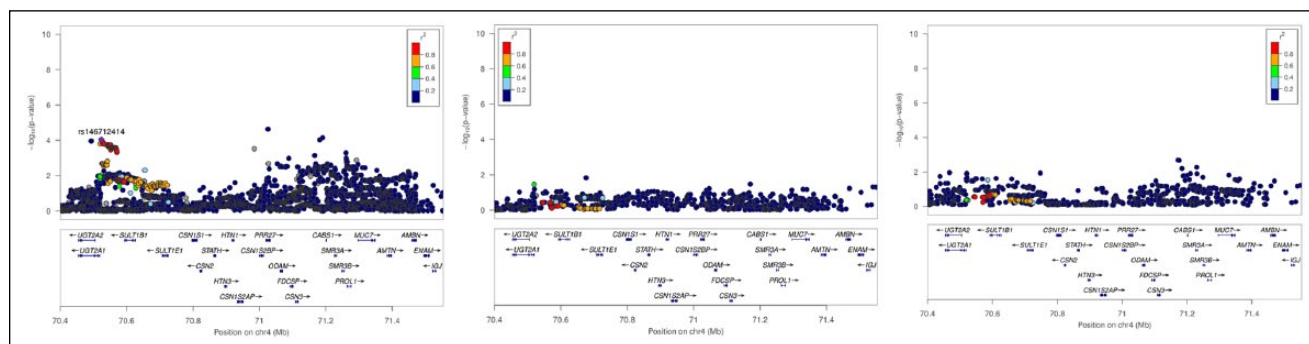
Figure 4. Association plots of the casein gene cluster. The $-\log_{10} P$ values of the analyzed single-nucleotide polymorphisms (SNPs) were plotted as a function of the genomic SNP position (NCBI build 36, hg19/1000 Genomes 2012, CEU). SNP annotations were generated by LocusZoom. The casein genes span the region from *CSN1S1* to *ENAM*. **Left panel:** aggressive periodontitis (AgP); **middle panel:** moderate chronic periodontitis (CP); **right panel:** severe CP.

cross. This effect was more significant for rs1595009 ($P = 3.7 \times 10^{-10}$) compared to rs2472649 ($P = 5.6 \times 10^{-8}$). For both SNPs, significant *cis*-regulation on the nearby gene *CXCL5* (chemokine [C-X-C motif] ligand 5) was reported in blood, with $P = 6.5 \times 10^{-6}$ and $P = 9.0 \times 10^{-9}$, respectively. *CXCL5* was located at QTL2 in the F_2 mouse cross.

The AgP-associated haplotype block-tagging SNP rs146712414 (*UGT2A1*) showed a *trans*-effect on *CSN1S1* (casein alpha s1) expression in fibroblasts, with $P = 1.1 \times 10^{-5}$. Importantly, *CSN1S1* was also located within QTL2 of the F_2 mouse cross.

The Casein Gene Cluster Indicates Association with AgP and CP

Because the eQTLs indicated effects on *PF4V1* and the casein genes of the secretory calcium-binding phosphoprotein (SCPP) gene family (Kawasaki et al. 2011), we tested these genetic regions for putative associations with AgP and CP in a separate experiment. While there were no AgP or CP risk variants in *PF4V1* (Appendix Figure), the chromosomal region spanning the casein gene cluster indicated the presence of AgP- and severe CP-associated variants from *CSN1S2AP* (casein alpha s2-like B, pseudogene) to *AMBN* (ameloblastin) (Fig. 4). Three SNPs (rs1480739, rs9329032, rs2127787) in high LD ($r^2 > 0.85$; HapMap CEU reference population) were associated with both diseases. However, they showed opposite effect directions in the European AgP sample and the white US CP sample, with the tagging SNP rs2127787 showing associations with $P =$

0.007 (OR, 1.17; 95% CI, 1.1–1.3) and $P = 0.03$ (OR, 0.866; 95% CI, 0.76–0.99) with AgP and severe CP, respectively.

Discussion

In the current study, we integrated QTL mapping and differential gene expression in a mouse model with association testing in humans. This approach identified a haplotype block downstream the *PF4/PPBP/CXCL5* gene cluster as a shared genetic risk locus for AgP and CP. Interestingly, the associated SNPs showed an eQTL effect in *cis*- on *CXCL5* but not on *PF4* or *PPBP*. *CXCL5* encodes a member of the CXC subfamily of chemokines, which recruits neutrophils, promotes angiogenesis, and remodels connective tissues (provided by RefSeq). *CXCL5*, similar to *PF4*, is stored in the alpha-granules of platelets. However, the strongest eQTL effects of the associated SNPs were reported to be *trans*, on the chemokine *PF4V1* gene, which is very similar to *PF4* and shares its antiangiogenic and anticoagulant effects (RefSeq). The human eQTL effects of the best-associated SNPs do not directly report the candidate genes chosen based on the murine expression data. However, eQTLs tend to be tissue specific, and likewise, eQTLs that were compared across different human tissues found <50% to up to 80% sharing, as previously reviewed in Nica and Dermitzakis (2013). The reported human eQTL effects were derived from human blood, whereas the expression data were generated from murine gingival tissues. eQTL data from human gingiva do not currently exist. Thus, clear evidence of the true target genes that are affected in the gingiva

by the associated haplotypes is lacking. In addition, eQTL effects could differ between humans and mice, which would further contribute to the observed differences between the selected candidate genes and the target genes of the eQTL effects.

The association at *PF4/PPBP/CXCL5* was stronger for AgP than for CP, which is in line with the concept that the severe early onset phenotype AgP has a stronger genetic underpinning, whereas additive negative effects of external risk factors such as smoking and diabetes, which accumulate during aging, have a stronger influence on the moderate late-onset phenotype CP. Yet, the association of rs1595009 was borderline significant in 2 independent CP samples. However, the association was not significant when using the European American sample of severe CP cases. This case sample was relatively small, and a false-negative association cannot be excluded. Likewise, the association with the small German CP case sample was only significant in the additive genetic model when all German controls were included. If the German AgP controls were excluded, the association was only significant in the recessive genetic model. We argue that testing the same SNP in 2 genetic models is not fully independent. With this in mind, we consider the association at *PF4/PPBP/CXCL5* as validated.

We also observed a haplotype block upstream of *UGT2A1* to be associated with AgP, which was not validated in the CP samples. Of note, the AgP-associated haplotype upstream of *UGT2A1*, tagged by SNP rs146712414, showed a *trans*-regulatory eQTL effect on the expression of the gene *CSN1S1* in fibroblasts, which encodes a Ca-binding phosphoprotein of the casein gene family. All casein genes belong to the SCPP gene family, which arose by gene duplication, and all are clustered on this chromosomal region (Kawasaki et al. 2011). This gene family has an important role in the regulation of the Ca-phosphate concentration of the extracellular environment (Kawasaki and Weiss 2003). In the context of periodontitis, it is of interest that almost 70% of alveolar bone consists of calcium phosphates. Some of these genes, such as *STATH, ODAM, SCPPPQ1*, and *ENAM*, are used in mineralization of the bones and tooth surfaces through regulation of extracellular Ca-phosphate concentrations; in addition, some of their functions are related to barrier integrity and immunity (RefSeq). Subsequent analysis of this locus showed association of the casein gene cluster at a significance level comparable to the haplotype block upstream of *UGT2A1*. The severe CP sample also indicated association of this gene cluster but with an opposite effect-direction. However, the moderate CP case sample did not produce any association at this locus. This could point to a false-positive association finding. Otherwise, it could be a true-positive association, which may be particularly relevant for AgP or for severe forms of disease in general, rather than for moderate ones. The reversed ORs in the European and European American samples could be indicative, given that the association is true, that the causative variants were not identified. We note that the periodontitis-associated variant rs9329032 is located 1.2 kb upstream of the rare variant rs17714883 (MAF <3% in the 1000 Genomes EUR reference populations), which

was reported to be associated with plasma levels of polychlorinated biphenyls at a genome-wide significance level. Both SNPs are located within the same intron of the retrotransposon *AF424542*. To determine the true nature of the observed association of the casein gene cluster, which was weaker compared to the association at *PF4/PPBP/CXCL5*, replication in additional large case-control samples of severe forms of periodontitis is required.

In conclusion, using murine models to study the role of gene variants and environmental factors on human disease, we identified the *PF4* locus as a genetic risk factor of human periodontitis, and we suggest that the casein gene cluster is a putative risk factor for severe forms of periodontitis. The association at *PF4/PPBP/CXCL5* is weaker compared to those commonly identified by GWAS, but it is shared by different disease phenotypes and may contribute to the missing heritability of periodontitis. In this context, we argue that mouse models are able to recapitulate human conditions and provide an efficient approach to identify and prioritize loci that are likely to be missed by traditional GWAS.

Author Contributions

A. Shusterman, M. Munz, G. Richter, S. Jepsen, W. Lieb, B. Krone, P. Hoffman, M. Laudes, J. Wellmann, K. Berger, T. Kocher, S. Offenbacher, K. Divaris, A. Franke, S. Schreiber, H. Dommisch, E. Weiss, A.S. Schaefer, Y. Houri-Haddad, F.A. Iraqi, contributed to conception, design, and data acquisition, drafted and critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of the work.

Authors' Note

The third and fourth affiliations have been switched from their original order in the initial OnlineFirst version. The following authors' affiliations have also been updated: S.J., W.L., B.K., P.H., M.L., J.W., and S.S.

Acknowledgments

References

ARIC Investigators. 1989. The Atherosclerosis Risk In Communities (ARIC) study: design and objectives. The ARIC Investigators. Am J Epidemiol. 129(4):687–702.

Berger K. 2012. DHS: The Dortmund health study [in German]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 55(6–7):816–821.

Bochenek G, Hasler R, El Mokhtari NE, Konig IR, Loos BG, Jepsen S, Rosenstiel P, Schreiber S, Schaefer AS. 2013. The large non-coding RNA ANRIL, which is associated with atherosclerosis, periodontitis and several forms of cancer, regulates ADIPOR1, VAMP3 and C11ORF10. Hum Mol Genet. 22(22):4516–4527.

Corey LA, Nance WE, Hofstede P, Schenkein HA. 1993. Self-reported periodontal disease in a Virginia twin population. J Periodontol. 64(12):1205–1208.

de Jong TM, Jochens A, Jockel-Schneider Y, Harks I, Dommisch H, Graetz C, Flachsbart F, Staufenbiel I, Eberhard J, Folwaczny M, et al. 2014. SLC23A1 polymorphism rs6596473 in the vitamin C transporter SVCT1 is associated with aggressive periodontitis. J Clin Periodontol. 41(6):531–540.

Divaris K, Monda KL, North KE, Olshan AF, Lange EM, Moss K, Barros SP, Beck JD, Offenbacher S. 2012. Genome-wide association study of periodontal pathogen colonization. J Dent Res. 91(Suppl 7):21S–28S.

Divaris K, Monda KL, North KE, Olshan AF, Reynolds LM, Hsueh WC, Lange EM, Moss K, Barros SP, Weyant RJ, et al. 2013. Exploring the genetic basis of chronic periodontitis: a genome-wide association study. Hum Mol Genet. 22(11):2312-2324.

Eke PI, Dye BA, Wei L, Thornton-Evans GO, Genco RJ; CDC Periodontal Disease Surveillance workgroup: James Beck (University of North Carolina, Chapel Hill, USA), Gordon Douglass (Past President, American Academy of Periodontology), Roy Page (University of Washington). 2012. Prevalence of periodontitis in adults in the United States: 2009 and 2010. J Dent Res. 91(10):914–920.

GTEx Consortium. 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 348(6235):648–660.

Hong KW, Shin MS, Ahn YB, Lee HJ, Kim HD. 2015. Genomewide association study on chronic periodontitis in Korean population: results from the Yangpyeong health cohort. J Clin Periodontol. 42(8):703–710.

Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumm LP, Sharma Y, Anderson CA, et al. 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 491(7422):119–124.

Kawasaki K, Lafont AG, Sire JY. 2011. The evolution of milk casein genes from tooth genes before the origin of mammals. Mol Biol Evol. 28(7):2053–2061.

Kawasaki K, Weiss KM. 2003. Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. Proc Natl Acad Sci U S A. 100(7):4060–4065.

Marcenes W, Kassebaum NJ, Bernabe E, Flaxman A, Naghavi M, Lopez A, Murray CJ. 2013. Global burden of oral conditions in 1990–2010: a systematic analysis. J Dent Res. 92(7):592–597.

Marchini J, Howie B, Myers S, McVean G, Donnelly P. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 39(7):906–913.

Michalowicz BS. 1994. Genetic and heritable risk factors in periodontal disease. J Periodontol. 65(Suppl 5):479–488.

Michalowicz BS, Aeppli D, Virag JG, Klump DG, Hinrichs JE, Segal NL, Bouchard TJ Jr, Pihlstrom BL. 1991. Periodontal findings in adult twins. J Periodontol. 62(5):293–299.

Muller N, Schulte DM, Turk K, Freitag-Wolf S, Hampe J, Zeuner R, Schroder JO, Gouni-Berthold I, Berthold HK, 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(5):1034–1042.

Nica AC, Dermitzakis ET. 2013. Expression quantitative trait loci: present and future. Philos Trans R Soc Lond B Biol Sci. 368(1620):20120362.

Petersen PE. 2003. The World Oral Health Report 2003: continuous improvement of oral health in the 21st century–the approach of the WHO Global Oral Health Programme. Community Dent Oral Epidemiol. 31(Suppl 1): 3–23.

Sanders AE, Sofer T, Wong Q, Kerr KF, Agler C, Shaffer JR, Beck JD, Offenbacher S, Salazar CR, North KE, et al. 2016. Chronic periodontitis genome-wide association study in the Hispanic Community Health Study/ Study of Latinos. J Dent Res. 96(1):64–72.

Schaefer AS, 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(1):159–167.

Schaefer AS, Richter GM, Dommisch H, Reinartz M, Nothnagel M, Noack B, Laine ML, Folwaczny M, Groessner-Schreiber B, Loos BG, et al. 2011. CDKN2BAS is associated with periodontitis in different European populations and is activated by bacterial infection. J Med Genet. 48(1):38–47.

Schaefer AS, Richter GM, Groessner-Schreiber B, Noack B, Nothnagel M, El Mokhtari NE, Loos BG, Jepsen S, Schreiber S. 2009. Identification of a shared genetic susceptibility locus for coronary heart disease and periodontitis. PLoS Genet. 5(2):e1000378.

Schaefer AS, Richter GM, Nothnagel M, Manke T, Dommisch H, Jacobs G, Arlt A, Rosenstiel P, Noack B, Groessner-Schreiber B, et al. 2010. A genome-wide association study identifies GLT6D1 as a susceptibility locus for periodontitis. Hum Mol Genet. 19(3):553–562.

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(2):212–218.

Shimizu S, Momozawa Y, Takahashi A, Nagasawa T, Ashikawa K, Terada Y, Izumi Y, Kobayashi H, Tsuji M, Kubo M, et al. 2015. A genome-wide association study of periodontitis in a Japanese population. J Dent Res. 94(4):555–561.

Shusterman A, Durrant C, Mott R, Polak D, Schaefer A, Weiss EI, Iraqi FA, Houri-Haddad Y. 2013. Host susceptibility to periodontitis: mapping murine genomic regions. J Dent Res. 92(5):438–443.

Takao K, Miyakawa T. 2015. Genomic responses in mouse models greatly mimic human inflammatory diseases. Proc Natl Acad Sci U S A. 112(4):1167–1172. Published erratum in: Proc Natl Acad Sci U S A. 2015;112(10):E1163–E1167.

Teumer A, Holtfreter B, Volker U, Petersmann A, Nauck M, Biffar R, Volzke H, Kroemer HK, Meisel P, Homuth G, et al. 2013. Genome-wide association study of chronic periodontitis in a general German population. J Clin Periodontol. 40(11):977–985.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and CuffLinks. Nat Protoc. 7(3):562–578.

Ward LD, Kellis M. 2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 40(Database issue):D930–D934.