A genome-wide association study identifies *GLT6D1* as a susceptibility locus for periodontitis

Arne S. Schaefer[1],[*],[†], Gesa M. Richter[1],[†], Michael Nothnagel[2], Thomas Manke[3], Henrik Dommisch[4], Gunnar Jacobs[1], Alexander Arlt[5], Philip Rosenstiel[1], Barbara Noack[6], Birte Groessner-Schreiber[7], Søren Jepsen[4], Bruno G. Loos[8] and Stefan Schreiber[1]

[1]Institute for Clinical Molecular Biology, Christian-Albrechts-University Kiel, Schittenhelmstraße 12, Kiel 24105, Germany, [2]Institute of Medical Informatics and Statistics, University Medical Center Schleswig-Holstein, Campus Kiel, Brunswiker Str. 10, Kiel 24105, Germany, [3]Max Planck Institute for Molecular Genetics, Computational Molecular Biology, Ihnestraße 73, Berlin 14195, Germany, [4]Department of Periodontology, Operative and Preventive Dentistry, University of Bonn, Welschnonnenstraße 17, Bonn 53111, Germany, [5]Clinical for General Medicine I, University Medical Center Schleswig-Holstein, Campus Kiel, House 6, Arnold-Heller- Straße 3, Kiel 24105, Germany, [6]Department of Conservative Dentistry, University Medical Center Carl Gustav Carus, Dresden University of Technology, Fetscherstraße 74, Dresden 01307, Germany, [7]Department of Operative Dentistry and Periodontology, University Medical Center Schleswig-Holstein, Campus Kiel, House 26, Arnold-Heller- Straße 3, Kiel 24105, Germany and [8]Department of Periodontology, Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam and VU University Amsterdam, Louwesweg 1, Amsterdam 1066 EA, The Netherlands

Periodontitis is a widespread, complex inflammatory disease of the mouth, which results in a loss of gingival tissue and alveolar bone, with aggressive periodontitis (AgP) as its most severe form. To identify genetic risk factors for periodontitis, we conducted a genome-wide association study in German AgP patients. We found AgP to be strongly associated with the intronic SNP rs1537415, which is located in the glycosyltransferase gene *GLT6D1*. We replicated the association in a panel of Dutch generalized and localized AgP patients. In the combined analysis including 1758 subjects, rs1537415 reached a genome-wide significance level of $P = 5.51 \times 10^{-9}$, OR = 1.59 (95% CI 1.36–1.86). The associated rare G allele of rs1537415 showed an enrichment of 10% in periodontitis cases (48.4% in comparison with 38.8% in controls). Fine-mapping and a haplotype analysis indicated that rs1537415 showed the strongest association signal. Sequencing identified no further associated variant. Tissue-specific expression analysis of *GLT6D1* indicated high transcript levels in the leukocytes, the gingiva and testis. Analysis of potential transcription factor binding sites at this locus predicted a significant reduction of GATA-3 binding affinity, and an electrophoretic mobility assay indicated a T cell specific reduction of protein binding for the G allele. Overexpression of GATA-3 in HEK293 cells resulted in allele-specific binding of GATA-3, indicating the identity of GATA-3 as the binding protein. The identified association of *GLT6D1* with AgP implicates this locus as an important susceptibility factor, and GATA-3 as a potential signaling component in the pathophysiology of periodontitis.

INTRODUCTION

Periodontitis is a complex chronic inflammatory disease that results in the loss of connective tissue and alveolar bone support of the teeth. Periodontal diseases are highly prevalent (1–2), with chronic periodontitis (CP) being the most common form of disease showing prevalence rates of more than 90% for adults above 60 years. Furthermore, it is the major cause of tooth loss in adults above 40 years, and for the more

*To whom correspondence should be addressed. Tel: +49 4315973882; Fax: +49 4315973730; Email: a.schaefer@ikmb.uni-kiel.de
†The first two authors contributed equally to the study.

severe forms, according to the WHO, human populations worldwide are affected at prevalence rates of up to 20% (3). A pathogenic microflora is the causative agent for the development of periodontitis, which is often promoted by a poor oral hygiene and smoking (4), which both contribute strongly to the disease risk. The microflora of the mouth contains myriads of bacteria which grow in biofilms on tooth surfaces and on the oral mucosa as complex colonies (5). Thus, the oral mucosa is a critical protective interface between external and internal environments, and it can be considered to some extent as a gateway for pathogens to other parts of the body, which eventually induce the development of further inflammatory diseases (6–7). Likewise, epidemiological studies described comorbidity between periodontitis and coronary heart disease (8–9), which is dependent on the severity of periodontal disease (10–11), type 1 and type 2 diabetes (12), rheumatoid arthritis (13), pulmonary diseases (14–15) and intestinal bowel disease (16). Moreover, the association of periodontitis with an increase in premature death in adults was reported (17).

Apart from the environmental factors, formal genetic studies have demonstrated the genetic basis of periodontitis and indicated that about half of the population variance in CP can be attributed to genetic factors, with a concordance rate of 0.23–0.38 for monozygotic twins (18) and a heritability of 50% for dizygotic twins (19). But hitherto, almost all association studies with periodontitis have led to controversial results, leaving the inherited components largely unexplained (20). However, a recently published study could demonstrate that periodontitis is genetically related to coronary heart disease by at least one susceptibility locus (21).

In order to identify genetic risk factors that predispose to periodontitis, we conducted a two-stage genome-wide association study (GWAS) in individuals of aggressive periodontitis (AgP), the most extreme form of periodontal diseases, for which it is assumed that genetic factors play the most prominent role in disease susceptibility (20).

RESULTS

GWAS for AgP

In a first stage, we conducted a GWAS in 141 German cases affected with the sub-phenotype generalized AgP and 500 controls (GWAS1; see Table 1 for phenotype description). Next, we performed a separate GWAS comprising 142 German cases of the sub-phenotype localized AgP and further 472 controls (GWAS2), and analyzed it separately without the knowledge of the GWAS1 results. Several genomic locations were identified as potentially associated with periodontitis by GWAS in stages 1 and 2 (Supplementary Material, Tables S1 and S2). In the analysis of the generalized AgP panel of GWAS1, 197 quality-controlled SNPs passed the pre-assigned threshold for an allelic χ^2 association test, while, using the same test, 244 quality-controlled SNPs passed it in GWAS2.

Next, the significant SNPs were compared for conformity within both GWAS. Out of these SNPs, only rs1537415 remained significant in both stages and showed the same direction of effect in both GWAS periodontitis populations (Supplementary Material, Tables S1 and S2). This SNP is located

within intron 2 of *GLT6D1* (glycosyltransferase 6 domain containing 1; Gene ID 360203) on chromosome 9q34.3. In the generalized AgP of GWAS1, the rare G allele of rs1537415 showed a significant association of $P = 1.8 \times 10^{-4}$ and an odds ratio (OR) of 1.67 [95% confidence interval (CI) 1.27–2.18; Table 2]. In the localized AgP of GWAS2, the rare G allele was similarly associated with $P = 3.1 \times 10^{-4}$ and had an OR of 1.65 (95% CI 1.26–2.17; Table 2).

Validation of the GWAS lead-SNP association

We validated the association of rs1537415 in the Dutch AgP panel of 164 Dutch AgP cases and 368 Dutch controls, which included both generalized and localized AgP patients, with $P = 5.7 \times 10^{-3}$ and an OR of 1.47 (95% CI 1.12–1.93; Table 2).

Combined analysis of the GWAS and validation panels

We additionally performed a Cochran–Mantel–Haenszel (CMH) test in the two GWAS panels and the validation panel for rs1537415, assuming a homogeneous genetic effect of this SNP in all the three samples. In this combined analysis of 438 cases and 1320 controls, rs1537415 showed a P-value of 5.51×10^{-9} and an OR of 1.59 (95% CI 1.36–1.86).

Sequencing of *GLT6D1*

After successful validation of the association of the lead SNP rs1537415, we sequenced the exons (including 200 bp distal to the exon/intron borders), the 3′-UTRs, and 250 bp of the promoter in 47 samples, which consisted of 34 homozygote carriers of the associated minor G allele, two heterozygotes and 11 homozygote carriers of the major C allele (primer sequences are available online in Supplementary Material, Table S3). We found no variant in any of the exons [five very low frequent coding SNPs were annotated in dbSNP (NCBI build36)]. Variants that were already annotated in dbSNP were identified in intron 3 (six), intron 4 (one), distal to the 5′ end (two) and the 3′ end (two) (Supplementary Material, Table S4). We detected two new variants within the 3′-UTR, one (position 137 655 470) was found heterozygous in only one individual. The other SNP (position 137 656 011) was subsequently genotyped and did not show a significant association with periodontitis in the pooled population. The detected 3′-UTR variants rs12336965, rs17040344 and the intronic SNP rs17593794 had already been selected for the subsequent fine-mapping, where they were not associated.

Fine-mapping of the association signals of *GLT6D1*

To narrow down the association signal, we fine-mapped *GLT6D1* in a pooled population of AgP patients ($n = 447$) and controls ($n = 1236$ from the preceded experiments, 236 further population representative German popgen controls). We covered the genomic sequence of *GLT6D1* including ~6 kb upstream and ~2 kb downstream of the transcribed sequences by 23 haplotype tagging SNPs (htSNPs). One SNP (rs1930745) showed significant deviations from Hardy–Weinberg equilibrium in cases and controls ($P < 0.001$) and

Table 1. Characteristics of the study populations

	GWAS1 Cases	Controls	GWAS2 Cases	Controls	Validation (Dutch panel) Cases	Controls	Fine-mapping population Cases	Controls
Subject characteristic, *n* (%)								
Individuals	141	500	142	479	155	341	416	1383
Men, *n*	50	259	48	253	40	187	132	684
Men (%)	35.46	51.80	33.80	53.82	25.81	54.84	31.73	49.46
Women, *n*	91	241	94	226	115	151	284	696
Women (%)	64.54	48.20	66.20	47.18	74.19	44.28	68.27	50.32
Unknown gender, *n* (%)	0	0	0	0	0	3 (0.88)	0	3 (0.22)
Mean age at first diagnosis	30	—	29	—	30	—	30	—
Age at first diagnosis, SD	4.32	—	4.89	—	4.19	—	4.46	—
Mean age at blood donation	32	62	31	40	34	30, 12 n.s.	32	51, 12 n.s.
Age at blood donation, SD	5.63	7.35	6.37	11.30	4.83	5.91	6.08	16.70
Study information								
Recruitment period	2003–2008	2005–2006	2003–2008	2006–2007	2003–2006	2005–2007	2003–2008	2005–2008
Phenotypes, *n* (%)								
Affected teeth, *n*	13.52	—	4.50	—	5.75	—	7.78	—
Affected teeth (%)	5.86	—	2.09	—	3.96	—	5.83	—
Ever smoked, *n*	73	264	55	277	121	137, 9 n.s.	249	691, 9 n.s.
Ever smoked (%)	55.73	52.8	42.31	57.8	78.06	41.27	59.86	50.29
Diabetes mellitus, *n*	3, 29 n.s.	32	0, 19 n.s.	2	0	341 n.s.	3, 53 n.s.	45, 342 n.s.
Diabetes mellitus (%)	2.68	6.40	0	0.42	0	0	0.83	4.32

Values are given as mean [standard deviation (SD)] when appropriate. Affection status was based on reported history (except for the number of affected teeth which was analyzed on the basis of radiographs by the clinicians, smoking was estimated by self-report). Numbers represent successfully genotyped individuals. GWAS, genome-wide association study; n.s., not specified.

was excluded from further analyses. In the pooled population of the fine-mapping, the lead SNP rs1537415 showed the strongest association with $P = 8.02 \times 10^{-7}$, OR = 1.49 (95% CI 1.27–1.75) (Supplementary Material, Table S5, genotypes and frequencies of the htSNPs are available online in Supplementary Material, Table S6). After correction for multiple testing with a conservative Bonferroni threshold that guaranteed an experiment-wise significance level of $\alpha = 0.05$, four htSNPs remained significant, all within the genomic sequence of *GLT6D1*, with rs1537415 being the most significant. The associations remained significant after adjustment for the periodontitis specific covariates gender, smoking and diabetes in a logistic regression analysis and after Bonferroni correction for multiple testing (Fig. 1, Supplementary Material, Table S7). Using Akaike's Information Criterion for choosing the genetic model which best explained the underlying association, a multiplicative mode of inheritance seemed most likely (Supplementary Material, Table S8). Under this genetic model, SNP rs1537415 showed an association with $P = 5.4 \times 10^{-6}$ [OR = 1.51 (95% CI 1.26–1.81)] upon covariate adjustment. The three other significant SNPs spanned the genomic sequence of *GLT6D1* (Fig. 1), with rs11103111 in intron 2 (r^2 to rs1537415 = 0.38), rs1333239 in intron 3 (r^2 to rs1537415 = 0.05) and rs7466817 in intron 4 (r^2 to rs1537415 = 0.17). The latter was located 12 bp downstream of the splice site of exon 5 [$P_{\text{adjusted}} = 9.9 \times 10^{-6}$, OR = 0.65 (95% CI 0.53–0.79)].

Haplotype analysis

Eight haplotypes with a frequency $\geq$ 2% could be inferred from the four significant SNPs (Supplementary Material,

Table S9). For two haplotypes, the frequency distribution showed a highly significant difference between cases and controls. In particular, the GCTC haplotype composed of the rs1537415-G, rs11103111-C, rs1333239-T, and rs7466817-C alleles was 8.6% more frequent in the cases ($P = 6.0 \times 10^{-6}$), and the reciprocal CTCG haplotype was 4.7% more frequent in the controls ($P = 7.0 \times 10^{-4}$). After 100,000 permutations, the lead SNP rs1537415 was marginally stronger associated in the single SNP analysis ($P = 1.0 \times 10^{-5}$) than the GCTC haplotype ($P = 8.00 \times 10^{-5}$). The association observed for rs11103111, rs7466817 and rs1333239 were weaker by one and two orders of magnitude compared to rs1537415 (Supplementary Material, Table S9), indicating that the GWAS lead SNP showed the strongest association signal with periodontitis in this region.

Tissue-specific expression of *glt6d1*

GLT6D1 is a hitherto uncharacterized gene, and little was known on the tissue-specific expression patterns of the transcript. To characterize the expression pattern of this gene in the gingiva (Fig. 2A) as well as in the major human tissues (Fig. 2B), we performed an expression analysis using quantitative RT–PCR. We detected expression of *GLT6D1* in cDNA isolated from both healthy and inflamed gingival tissue samples. In addition, strong gene expression of *GLT6D1* was found in leukocytes and testis, whereas faint expression was documented for specific cDNA of pancreatic tissue samples. In comparison to the gingival epithelium, the gene expression of *GLT6D1* was significantly stronger in the gingival connective tissue ($P < 0.001$; Fig. 3). No significant difference was observed between healthy or inflamed gingival tissue samples.

Table 2. SNP associations of the GWAS lead SNP rs1537415

	Rank	GWAS1	GWAS2	Validation
	87	163	—	
P-value		1.8×10^{-4}	3.1×10^{-4}	5.7×10^{-3}
OR (95% CI)		1.67 (1.27–2.18)	1.65 (1.26–2.17)	1.47 (1.12–1.93)
Controls n 11 (%)		187 (38.3)	185 (38.6)	128 (37.5)
n 12 (%)		236 (48.4)	233 (48.6)	156 (45.7)
n 22 (%)		65 (13.3)	61 (12.7)	57 (16.7)
MAF		37.5	37.1	39.6
Cases n 11 (%)		36 (26.3)	33 (24.6)	38 (25.0)
n 12 (%)		65 (47.4)	70 (52.2)	79 (52.0)
n 22 (%)		36 (26.3)	31 (23.1)	35 (23.0)
MAF		50.0	49.3	49.0

Association statistics are given for the panels of the three stage design generalized AgP (GWAS1), localized AgP (GWAS2) and a Dutch population of both generalized and localized AgP (Validation). Shown are the OR, the 95% confidence intervals (95% CI) and the P-values which were obtained from a χ^2 test, and the genotype counts and frequencies. GWAS SNPs were ranked according to their P-value. GWAS, genome-wide association study, MAF, minor allele frequency. GWAS1 indicates the major allele, GWAS2 indicates the minor allele, AgP, aggressive periodontitis; n, number of carriers.

Reduction in binding affinity of GATA-3 for the disease-associated G allele

The lead SNP rs1537415 is located in intron 2 of *GLT6D1*. To identify a possible causative effect, we conducted a computational analysis of potential transcription factor (TF) binding sites at this genetic locus, utilizing a large compendium of 554 binding models of vertebrate TFs obtained from the TRANSFAC database (22). The rare variant of rs1537415 led to a significant reduction in the predicted binding affinity of GATA-3 (Supplementary Material, Table S10). Next, we experimentally confirmed the predicted change in binding affinity upon rs1537415 C → G exchange by electrophoretic mobility shift analysis. Double-stranded radioactively labeled oligonucleotides which spanned SNP rs1537415 by >20 bp to both sides and were specific for either the C or the G allele (see Materials and Methods section for oligonucleotide sequences), were incubated with nuclear proteins isolated from human epithelial cell line of cervical carcinoma (HeLaS3), embryonic kidney cell line (HEK293), human T cell leukemia Jurkat cells and human acute monocytic cell line (THP-1). The presence of the major C allele showed strong binding of nuclear protein in lymphoblastic Jurkat cells. This binding was found to be decreased in the presence of the disease-associated G allele in this cell type, with an inverse correlation of the amount of the free probe and the band's intensity (Fig. 4). For the other cell types, only very faint protein binding was observed, and this observation was independent of the respective genetic background (Fig. 4). Next, to validate the identity of the nuclear binding protein, we overexpressed GATA-3 in HEK293 cells, where it is normally not expressed (Fig. 5). Whereas no electrophoretic mobility shift could be detected in HEK293 cells, in HEK293 cells overexpressing GATA-3, a shifted band was detected. The size of the band corresponds to the band shift seen with Jurkat nuclear cell extracts. This finding further supports that GATA-3 could bind to the region of rs1537415.

DISCUSSION

In this study, we provide evidence for disease association of the glycosyltransferase *GLT6D1* on chromosome 9q34.3 with AgP. The rare G allele of the lead SNP rs1537415 reached a genome-wide significance level of $P = 5.51 \times 10^{-9}$, OR = 1.59 (95% CI 1.36–1.86) in the combined analysis of GWAS and validation panels. In a subsequent fine-mapping experiment, this SNP was found to exhibit the strongest association with AgP, showing a P-value of 5.4×10^{-6}, OR = 1.51 (95% CI 1.26–1.81) after adjustment for the risk factors smoking, diabetes and gender. The robustness of our finding was also reflected in the highly similar minor allele frequencies (MAF) of rs1537415 in all three case populations, which contributed to the observed combined significant association signal in a similar fashion. The rare G allele was enriched in the German cases with generalized (50.0% MAF) and localized AgP (49.3%) by 12%, when compared with the German controls (37.5 and 37.1% in GWAS1 and GWAS2, respectively). A similar degree of enrichment (9.4%) was observed when we compared the Dutch cases (49.0%) with Dutch controls (39.6%). The allele frequencies in the control populations are in concordance with the European HapMap CEU reference population (39.2%). The observed association implies an involvement of the hitherto uncharacterized gene *GLT6D1* in the pathophysiology of periodontitis. Enzymes of the glycosyltransferase 6 family, a group of single-pass transmembrane proteins, contribute to the synthesis of histo-blood related antigens in the golgi (23). Glycosyltransferases have been shown to play crucial roles in developmental signaling (24–25). It is postulated that these molecules are part of specific cellular mechanisms that regulate the spatial and temporal activity of developmental genes, and control their sensitivity and specificity to regulatory signals (26). As an example, glycosylation of the receptor NOTCH1, which controls cell fate in the various aspects of tissue differentiation by mediating local cell–cell interactions (27), has been shown to be crucial for a number of NOTCH's developmental functions (26).

Our experiments indicated strong *GLT6D1* expression in testis, leukocytes and gingiva, with a significantly stronger gene expression in gingival connective tissue than in the gingival epithelium. No difference in expression was observed between the inflamed and the healthy state of both gingival connective tissue and epithelium. The observed expression of *GLT6D1* pattern might not reflect the overall tissue-specific expression, because the primer design allowed only detection of a region spanning exons 3–5, and potential alternatively spliced products specific to other human tissues might have been missed. But it is likely that it reflects a role of this gene in these specific tissues. Here, the variant rs1537415, located within the second intron of *GLT6D1*, could have a *cis*-acting effect on *GLT6D1* expression. Intronic enhancers (IE) have been previously described for the second intron of other genes, e.g. for *IL-4* (28), where the IE confers GATA-3-dependent enhancement of *IL-4* promoter activity.

Our analysis of potential TF binding affinities indicated the impairment of a potential GATA-3 TF binding site at the position of rs1537415. This prediction was experimentally supported by an electrophoretic mobility assay, which showed
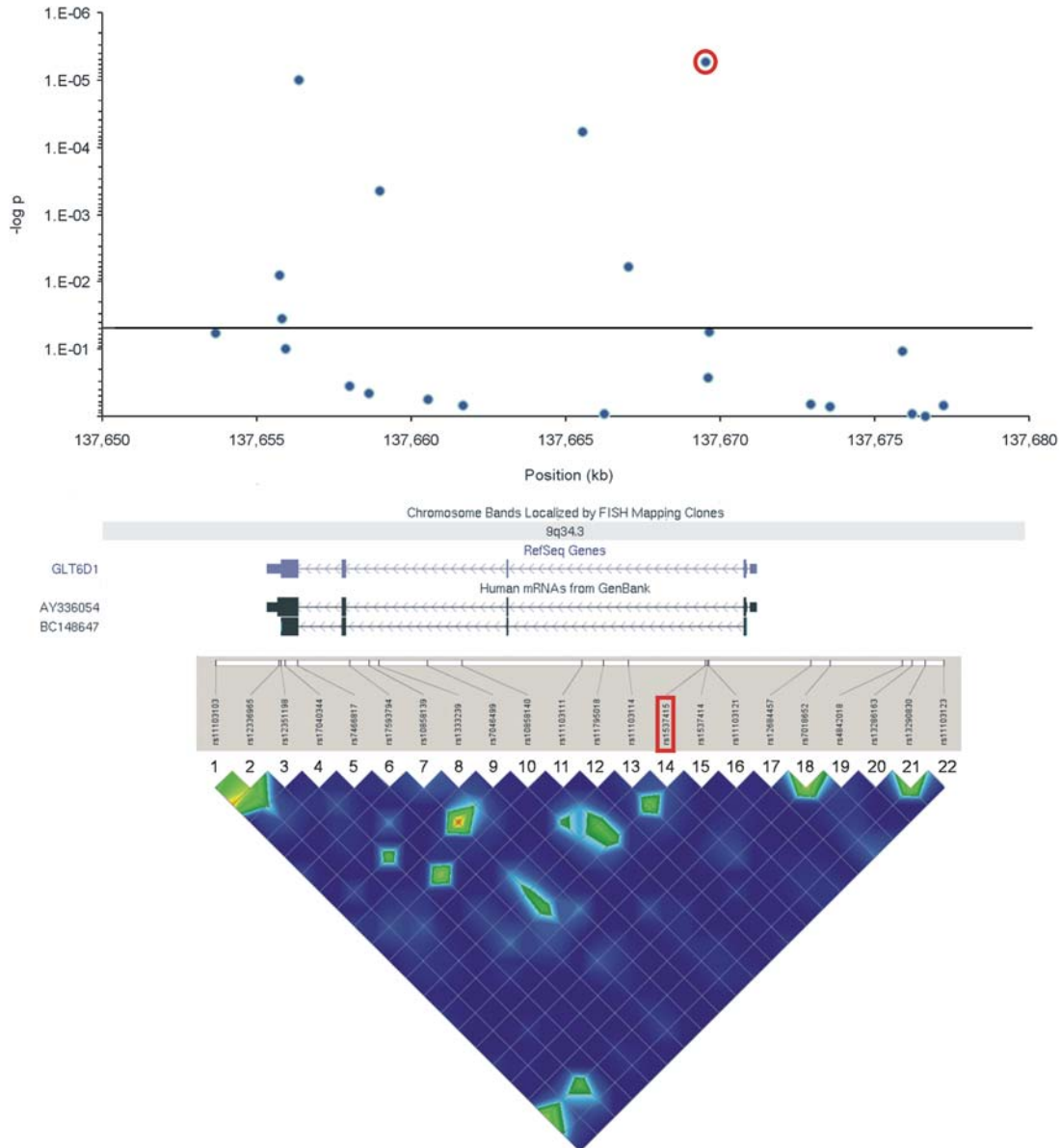
Figure 1. Genetic Region of *GLT6D1*. The top diagram shows the nominal $-\log_{10} p$ under the assumption of a recessive genetic model, plotted as a function of the genomic htSNP position (NCBI build 36). The bottom panel shows pairwise LD in the controls using metric r^2. The htSNPs of the fine-mapping analysis are indicated by numbers (1–22 and the refseq identifiers), the position of the associated SNP rs1537415 on the LD map and the physical gene map is framed by the red square.

GATA-3 binding at this locus in T cells. A recent study identified more than 80 target genes for GATA-3 in Jurkat cells, suggesting a comprehensive range of action for the initiation and maintenance of T helper cell type-2 (Th2) differentiation (29). GATA-3 is considered as the master TF in transcriptional control of Th2 differentiation (30). Th2 cells produce the cytokines interleukin-4 (IL-4), IL-5 and IL-13, which are involved in immunity against extracellular parasites (31) and regulated by GATA-3 binding (32–33). IL-4 is known to inhibit periodontal bone absorption as it serves as an antagonist for Th1 specific cytokines like IL-6 and IFN-γ that in turn trigger the recruitment of macrophages and osteoclast formation (34). Interestingly, the expression profile of CD4[+] cells in inflamed gingiva suggested a continuous

expression of IFN-γ but not for IL-4 (35). The role of GATA-3 in IL-4 regulation has not yet been investigated in this context, but recent studies suggested a potential role of GATA-3 in preventing bone erosion in rheumatoid arthritis (36). Furthermore, IL-4 concentration was observed to be very low in periodontal tissues of advanced AgP patients (37), and the low IL-4 concentrations are associated with periodontal disease progression (38–41). Additionally, an inverse correlation between IL-4 concentration and histopathological markers of periodontitis was recently reported (42). These data become interesting when considering that IL-4 is a potent downregulator of macrophage function and inhibitor of the secretion of pro-inflammatory cytokines such as TNF-α (43).
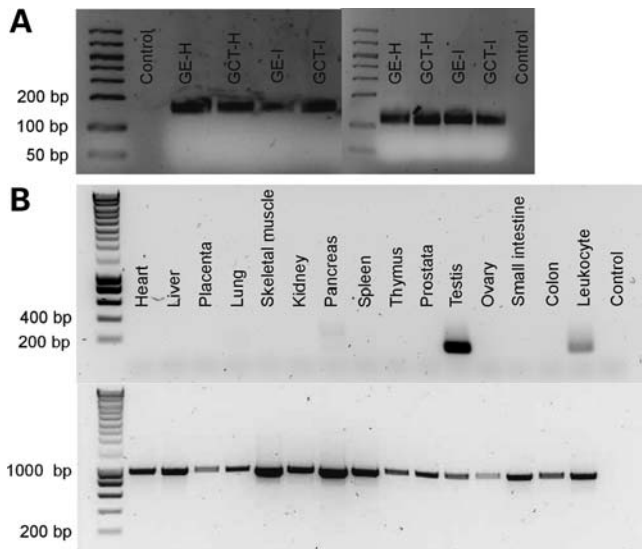
Figure 2. Tissue-specific expression of *GLT6D1*. cDNA was amplified by RT–PCR. The expected product size was 150 bp. (**A**) Tissue-specific expression in healthy and inflamed gingival epithelium (GE) and gingival connective tissue (GCT). H = healthy, I = inflamed, control = water. (**B**) Tissue-specific expression in cDNA samples of the major human tissues (Clontech). control = water.

Figure 3. Quantitative expression of *GLT6D1* in the gingiva. Indicated is the relative gene expression ratio after normalization to the housekeeping gene *GAP-DH* (Ct_{GAP-DH}: H-E = 27.4 ± 0.73, PAR-E = 27.0 ± 1.48; H-C = 27.9 ± 0.53, PAR-C = 28.5 ± 1.43; Ct_{GLT6D1}: H-E = 28.9 ± 0.93, PAR-E = 29.7 ± 0.55, H-C = 28.8 ± 0.51, PAR-C = 29.33 ± 1.06). Ct = cycle threshold, H-E = healthy epithelium, H-C = healthy connective tissue. The cDNA was obtained from human primary gingival epithelial cells and gingival connective tissue (fibroblasts) from healthy individuals and from individuals with periodontitis (each boxplot = 10 samples).
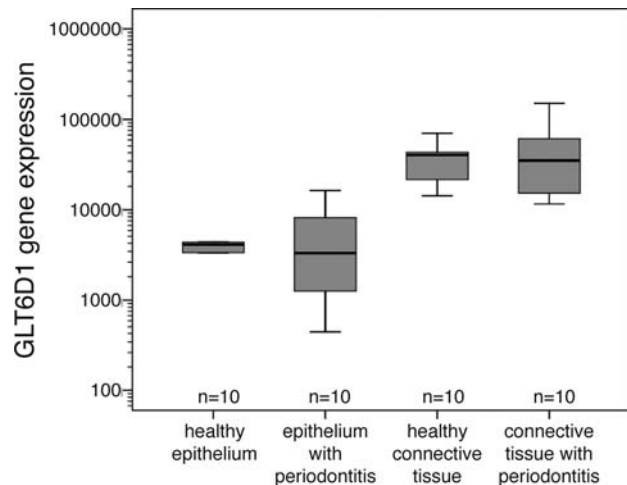
Several recent studies on Th2 differentiation have revealed that GATA-3-dependent transcription of the *IL-4, IL-5* and *IL-13* gene cluster involves complex long-range intrachromosomal interactions that generate loops of DNA and physically juxtapose the promoters of these cytokine genes (44–45). Though highly speculative, the locus at *GLT6D1* might also be part of a potential *trans*-acting regulatory mechanism involving chromatin-remodeling activity, leading to differential cytokine expression pattern in affected individuals. Interestingly, as a prerequisite for the arrangement of equally orientated promoters by chromatin loops, almost all genes distal to *GLT6D1* up to more than 1.0 Mb are transcribed in reverse orientation, whereas almost all genes proximal to *GLT6D1* up to 1.7 Mb are transcribed in forward orientation.

In conclusion, we identified the first genetic risk locus for periodontitis with genome-wide significance. The identified association with the hitherto uncharacterized glycosyltransferase gene *GLT6D1* implicates this locus as an important genetic susceptibility factor, and GATA-3 as a potential signaling component in the pathophysiology of this complex disease. The identification of these primary determinants of periodontitis may not only enhance our understanding of the underlying pathophysiology and ease diagnostics, but may also provide insight into the complex interplay of impaired genetic regulation and disease development.

MATERIALS AND METHODS

Ethics statement

Informed consent was obtained from all participants and the studies were approved by the local Ethics Committees.

Study population

The case population for the GWAS was recruited from throughout Germany between 2003 and 2008. Patients with periodontitis were ≤35 years of age, and the standardized examination criteria have been described before (21). The control population for the GWAS was recruited from the popgen biobank and has been described in detail elsewhere (46). In brief, they were randomly identified on the basis of the population registry of Schleswig-Holstein, Germany (*n* = 500) for stage 1 and by the Blood Service of the University Hospital Schleswig-Holstein (*n* = 472) for stage 2. All controls underwent a physical examination at the popgen facilities to obtain information on the general health status. Information on the oral health status and physical risk factors (gender, smoking, diabetes) was obtained from a questionnaire that was completed during medical consultation. Additionally, a clinical check-up was subsequently performed. The replication cohort was recruited from across the Netherlands between 2003 and 2006 and comprised 164 case individuals of ≤35 years of age and 368 ethnically and age-matched healthy controls. A description of the GWAS and replication populations is given in Table 1.

Genotyping and statistical analysis

Genotyping for the GWAS was performed with the Affymetrix Gene Chip Human Mapping 500K Array Set for patients and controls of GWAS1 and patients of GWAS2. The controls of GWAS2 were genotyped with the Affymetrix Gene Chip 5.0. The genotyping methodology has been described elsewhere (47). GWAS genotype data were automatically called by the BRLMM algorithm (Affymetrix, High Wycombe, UK) and were analyzed using PLINK v1.06 (48). Significance
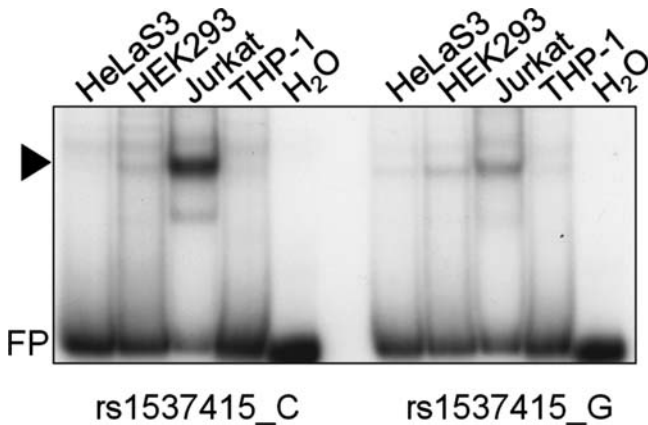
Figure 4. The rs1537415-G allele results in decreased binding of nuclear proteins. The presence of the rs1537415 rare G allele results in decreased binding of nuclear proteins of lymphoblastic Jurkat cells. Band-shift assays were performed with oligonucleotides representing the C or G allele of rs1537415, respectively. FP = free probe, arrowhead = bandshift, H_2O = water control. The nature and origin of the different cell types is described in the text.

of association with single-locus genotypes was assessed using χ^2 tests for allelic 2×2 contingency tables. In brief, SNPs with a genotype call rate less than 90% or a MAF < 5% were excluded. Markers were tested for deviations from Hardy–Weinberg equilibrium in controls before inclusion into the analyses ($\alpha = 0.05$), as previously described (49). Out of a total of 500 568 SNPs, 345 646 (69.05%) SNPs passed general quality-control criteria in GWAS1 and were selected for the subsequent association analyses. In GWAS2, out of 440 794 SNPs that were genotyped on both the Affymetrix 500K Array and the 5.0 GeneChip, 322 825 SNPs (73.24%) passed quality criteria and were subsequently carried forward to association analyses. Before the analysis, we defined a significance threshold of $P \leq 5 \times 10^{-4}$ to follow up only a modest number of SNPs, and manually inspected bad calling of the BRLMM algorithm for the remaining SNPs [$n = 197$ in GWAS stage 1 (Supplementary Material, Table S1) and $n = 244$ in GWAS stage 2 (Supplementary Material, Table S2)]. Only individuals with a call rate of more than 90% were included in the analysis. The genomic inflation factors were $\lambda = 1.04$ in stage 1 and 1.12 in stage 2. For the validation and fine-mapping, SNPlex genotyping (Applied Biosystems, Foster City, CA, USA) was performed, and TaqMan genotyping (Applied Biosystems) was performed for some of the SNPs identified in the resequencing of *GLT6D1* exons and UTRs. Both methodologies have previously been described (21). We assessed the significance of associations with or between single-locus genotypes using χ^2 and Fisher's exact tests for 2×2 and 2×3 contingency tables where applicable, and, for the combined analysis of GWAS and validation genotype data for rs1537415, we performed a CMH test as implemented in PLINK. We used genotype, dominant, multiplicative and recessive genetic models to assess the genetic effect of the associated alleles. Significance was assessed by likelihood-ratio tests, and P-values ≤ 0.05 were considered nominally significant. Logistic regression analysis was performed in the R statistical environment, version 2.7.2 (50). Correction for multiple testing was
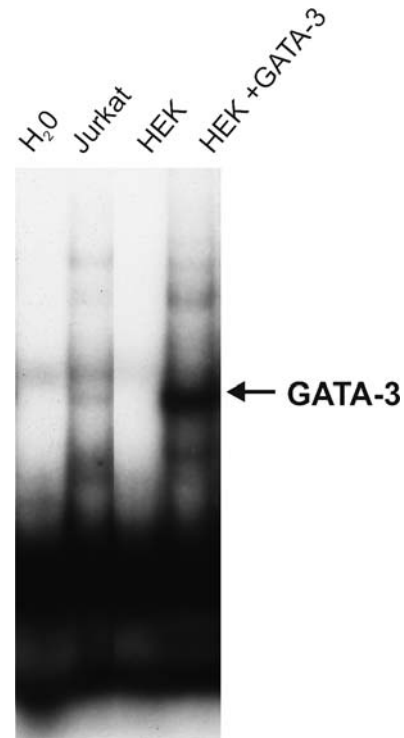
Figure 5. *In vitro* binding of GATA-3 at the predicted binding site at rs1537415. GATA-3 overexpression in HEK293-cells identifies GATA-3 as the nuclear protein that binds to the predicted TF binding site at rs1537415. Given is the electrophoretic band pattern of nuclear protein bound to the DNA sequence flanking rs1537415, in the presence of the common C allele. H_2O = water control, +GATA-3 = GATA-3 overexpressing HEK293 cells. The nature and origin of the different cell types is described in the text.

performed using Bonferroni's method to guarantee an experiment-wise significance level of $\alpha = 0.05$.

Sequencing

Sequencing of genomic DNA was performed using Applied Biosystems BigDye chemistry according to the supplier's recommendations (for primer sequences, see Supplementary Material, Table S3). Traces were inspected for the presence of SNPs and InDels using novoSNP (51).

Fine-mapping

htSNPs were selected from the CEU data of the International HapMap project (http://www.hapmap.org, NCBI build 36, 2008-03) using the Tagger feature (52) of Haploview 4.0 (53). htSNPs were defined by pairwise haplotype tagging ($r^2 > 0.8$). LD measures were calculated with FamHap (54) and plotted with the GOLD program (55).

Haplotype analysis

Haplotype estimation was performed using an EM algorithm (56). Haplotype significance was assessed via 100 000 permutations as implemented in Haploview 4.0 (53).

Isolation of gingival tissues

Gingival samples were obtained from individuals with healthy gingival conditions and from patients with advanced periodontitis. Healthy gingival conditions were defined as follows: no redness, no swelling, no bleeding on probing and with periodontal pockets ≤3 mm. Gingiva of advanced CP showed all typical clinical signs of periodontal inflammation (redness, swelling, bleeding on probing) and a periodontal pocket probing depth ≥6 mm. Before total RNA extraction, the gingival epithelial layer was enzymatically dissected from connective tissue using 3 ml dispase (Cascade Biologics, Paisley, UK) for 6 h on ice. For analysis and quantification of the gene expression in gingival tissues samples, primary human gingival epithelial cells (GE, $n = 3$) and human gingival fibroblasts (gingival connective tissue, GCT, $n = 3$) were collected from healthy individuals who underwent third-molar extraction or periodontal surgery. Primary human epithelial cells were cultured in keratinocyte growth medium using supplements provided with the medium (PAA, Cölbe, Germany) and 1% antibiotics (penicillin, streptomycin, amphotericin), whereas gingival fibroblasts (GCTs) were cultured using Dulbecco's modified Eagle's medium (PAA, Cölbe, Germany) supplemented with 1% antibiotics (penicillin, streptomycin, amphotericin) and 10% fetal bovine serum. GEs and GCTs were incubated at $37°C$ in a humidified atmosphere (5% CO_2). After total-RNA extraction and first strand synthesis, cDNA from both GEs and GCTs served as external standard for the gingival epithelium and the gingival connective tissue, respectively.

Expression analysis

Tissue-specific expression of *GLT6D1* was examined by RT–PCR in healthy and inflamed gingival tissue samples and in the commercial 'human multiple tissue cDNA panel' (Clontech, Mountain View, CA, USA), comprising a variety of healthy tissues of the major human organs. To obtain gingival cDNA, total RNA was extracted from both epithelium and connective tissue, as well as from primary GE and GCT cell culture using the RNAeasy Protection Mini Kit (Qiagen, Hilden, Germany). The reverse transcription reaction was performed following standard protocols (iScript®, Bio-Rad, Munich, Germany). The same primers were used for all cDNA samples (forward primer: 5′-CCT AGA AAA CGC CCT GATG-3′; reverse: 5′-TGC AAA CCT GCC AGT AGC-3′). The expected product size was 150 bp and spanned exons 3–5. All PCR reactions were carried out in a total volume of 20 μl, including 1 μl of cDNA and 250 nM primers under the following conditions: denaturation for 5 min at $95°C$; 38 cycles of 30 s at $95°C$, 20 s at $60°C$, 45 s at $72°C$; final extension for 10 min at $72°C$.

For GEs and GCTs, quantitative RT–PCR was performed. At the end of the quantitative RT–PCR, melting curve analysis was performed to confirm the amplified product was specific. For additional confirmation of the amplified PCR product, gel electrophoresis was performed on 2% agarose gels, and subsequently documented on a BioDoc Analyzer (Biometra, Göttingen, Germany). The quantitative analysis of the cDNA was performed using the iCycler® (Bio-Rad)

and QuantiFast SYBR Green PCR kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. To confirm the use of equal amounts of RNA in each experiment, all samples were analyzed by spectrophotometry using NanoDrop (PeqLab Biotechnology, Erlangen, Germany) and expression analysis of the housekeeping gene glycerylaldehyd-3-dehydrogenase (*GAP-DH*, GeneGlobe primer, Qiagen) was performed. Sample values were normalized to the gene expression of the housekeeping gene *GAP-DH* (GeneGlobe primer, Qiagen), and relative gene expression was calculated using the mathematical model described in Pfaffl (57). Data were statistically analyzed using the Kruskal–Wallis and Mann–Whitney test (SPSS v.17, Germany). The significance level was set at $\alpha = 0.05$.

Computational analysis of TF binding affinities

To determine which TFs were most likely to be affected by SNP rs1537415, we utilized a large compendium of 554 binding models of vertebrate TFs obtained from the TRANSFAC database (22). We extracted a ±50 bp region around rs1537415, and determined the TF binding affinities with the TRAP method as described before (58). In order to compare the effect of sequence variations for different TFs, we utilized a statistical framework as previously defined (59). The *P*-values were estimated from a parameterization of the General Extreme Value distribution. The same procedure was applied to the reference sequence, and the variation yielding 554 pairs of properly normalized binding affinities. For a given TF, any change signaled the effect of the sequence variation on its affinity, and thus allowed the ranking of the effect of the variation in a quantitative manner. The top-ranking factors were those whose affinity was decreased by the SNP (indicating a possible loss of binding site), while the bottom-ranking factors had increased binding affinity (indicating a possible creation of a binding site).

Electrophoretic mobility shift assay

Human epithelial cell line of cervical carcinoma HeLaS3 (ACC161), embryonic kidney cell line HEK293 (ACC305), human T cell leukemia Jurkat cells (AC282) and human acute monocytic cell line THP-1 (ACC16) were purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany). Whole cells were incubated in isotonic lysis buffer (10 mM Tris/HCl pH 7.4, 10 mM NaCl, 1.5 mM $MgCl_2$, 10 mM NaF) for 5 min on ice and cells were disrupted by passing them ×15 through a 26G needle. Nuclei were pelleted by centrifugation (1000g, 10 min) and incubated in hypertonic lysis buffer (20 mM HEPES pH 7.9, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 10 mM NaF, 25% v/v Glycerol) over night at $4°C$ followed by 10 min centrifugation at 20 000g. The supernatant with the nuclear proteins were stored at $−80°C$. Double-stranded radioactively labeled oligonucleotides (14 fmol; G-Fw: 5′-CAT TTT TAA AAA CTA TTA TCT GTC ATG ATA CTA CTG CTG TTC-3′, G-Rev: 5′-GAA CAG CAG TAG TAT CAT GAC AGA TAA TAG TTT TTA AAA ATG-3′, C-Fw: 5′-CAT TTT TAA AAA CTA TTA TCT CTC ATG ATA CTA CTG CTG TTC-3′, C-Rev: 5′-GAA

CAG CAG TAG TAT CAT GAG AGA TAA TAG TTT TTA AAA ATG-3′), spanning SNP rs1537415 (underlined), were incubated with 3 µg nuclear proteins in binding buffer (Promega, Madison, USA) for 30 min on ice, and complexes were separated on a native 6% polyacrylamide/TBE gel for 2 h. Separated complexes were visualized using X-ray film. For overexpression of GATA-3, GATA-3 was amplified from cDNA by using the primers 5′-GAA TTC TAT GGA GGT GAC GGC GG-3′ (forward)/5′-CTC GAG TAA CCC ATG GCG GTG ACC-3′ (reverse) and cloned with *Eco*RI and *Xho*I into a pCEV expression plasmid (kindly provided by T. Miki, National Cancer Institute, Bethesda, MD, USA). The FLAG-tagged GATA-3 expression construct was sequence-verified in an ABI3700 sequencer (Applied Biosystems) before use and transformed into HEK293 cells.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Albandar, J.M. and Rams, T.E. (2002) Global epidemiology of periodontal diseases: an overview. *Periodontol. 2000*, **29**, 7–10.
2. Pihlstrom, B.L., Michalowicz, B.S. and Johnson, N.W. (2005) Periodontal diseases. *Lancet*, **366**, 1809–1820.
3. Petersen, P.E. (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.
4. Bergstrom, J. (2004) Influence of tobacco smoking on periodontal bone height. Long-term observations and a hypothesis. *J. Clin. Periodontol.*, **31**, 260–266.
5. Moore, W.E. and Moore, L.V. (1994) The bacteria of periodontal diseases. *Periodontol. 2000*, **5**, 66–77.
6. Paraskevas, S., Huizinga, J.D. and Loos, B.G. (2008) A systematic review and meta-analyses on C-reactive protein in relation to periodontitis. *J. Clin. Periodontol.*, **35**, 277–290.
7. Wu, T., Trevisan, M., Genco, R.J., Falkner, K.L., Dorn, J.P. and Sempos, C.T. (2000) Examination of the relation between periodontal health status and cardiovascular risk factors: serum total and high density lipoprotein cholesterol, C-reactive protein, and plasma fibrinogen. *Am. J. Epidemiol.*, **151**, 273–282.
8. Bahekar, A.A., Singh, S., Saha, S., Molnar, J. and Arora, R. (2007) The prevalence and incidence of coronary heart disease is significantly increased in periodontitis: a meta-analysis. *Am. Heart J.*, **154**, 830–837.
9. Dietrich, T., Jimenez, M., Krall Kaye, E.A., Vokonas, P.S. and Garcia, R.I. (2008) Age-dependent associations between chronic periodontitis/edentulism and risk of coronary heart disease. *Circulation*, **117**, 1668–1674.
10. Geerts, S.O., Legrand, V., Charpentier, J., Albert, A. and Rompen, E.H. (2004) Further evidence of the association between periodontal conditions and coronary artery disease. *J. Periodontol.*, **75**, 1274–1280.
11. Seinost, G., Wimmer, G., Skerget, M., Thaller, E., Brodmann, M., Gasser, R., Bratschko, R.O. and Pilger, E. (2005) Periodontal treatment improves endothelial dysfunction in patients with severe periodontitis. *Am. Heart J.*, **149**, 1050–1054.
12. Mealey, B.L. (2006) Periodontal disease and diabetes. A two-way street. *J. Am. Dent. Assoc.*, **137** (Suppl. 26), S–31S.
13. Mercado, F.B., Marshall, R.I. and Bartold, P.M. (2003) Inter-relationships between rheumatoid arthritis and periodontal disease. A review. *J. Clin. Periodontol.*, **30**, 761–772.
14. Paju, S. and Scannapieco, F.A. (2007) Oral biofilms, periodontitis, and pulmonary infections. *Oral. Dis.*, **13**, 508–512.
15. Scannapieco, F.A. and Genco, R.J. (1999) Association of periodontal infections with atherosclerotic and pulmonary diseases. *J. Periodontal. Res.*, **34**, 340–345.
16. Brito, F., de Barros, F.C., Zaltman, C., Carvalho, A.T., Carneiro, A.J., Fischer, R.G., Gustafsson, A. and Figueredo, C.M. (2008) Prevalence of periodontitis and DMFT index in patients with Crohn's disease and ulcerative colitis. *J. Clin. Periodontol.*, **35**, 555–560.
17. Söder, B., Jin, L.J., Klinge, B. and Soder, P.O. (2007) Periodontitis and premature death: a 16-year longitudinal study in a Swedish urban population. *J. Periodontal. Res.*, **42**, 361–366.
18. Corey, L.A., Nance, W.E., Hofstede, P. and Schenkein, H.A. (1993) Self-reported periodontal disease in a Virginia twin population. *J. Periodontol.*, **64**, 1205–1208.
19. 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.
20. Loos, B.G., John, R.P. and Laine, M.L. (2005) Identification of genetic risk factors for periodontitis and possible mechanisms of action. *J. Clin. Periodontol.*, **32** (Suppl. 6), 159–179.
21. Schaefer, A.S., Richter, G.M., Groessner-Schreiber, B., Noack, B., Nothnagel, M., El Mokhtari, N.E., Loos, B.G., Jepsen, S. and Schreiber, S. (2009) Identification of a shared genetic susceptibility locus for coronary heart disease and periodontitis. *PLoS Genet.*, **5**, e1000378.
22. Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V. *et al.* (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.*, **31**, 374–378.

23. Turcot-Dubois, A.L., Le Moullac-Vaidye, B., Despiau, S., Roubinet, F., Bovin, N., Le Pendu, J. and Blancher, A. (2007) Long-term evolution of the CAZY glycosyltransferase 6 (ABO) gene family from fishes to mammals—a birth-and-death evolution model. *Glycobiology*, **17**, 516–528.
24. Selleck, S.B. (2000) Proteoglycans and pattern formation: sugar biochemistry meets developmental genetics. *Trends Genet.*, **16**, 206–212.
25. Perrimon, N. and Bernfield, M. (2000) Specificities of heparan sulphate proteoglycans in developmental processes. *Nature*, **404**, 725–728.
26. Grabher, C., von Boehmer, H. and Look, A.T. (2006) Notch 1 activation in the molecular pathogenesis of T-cell acute lymphoblastic leukaemia. *Nat. Rev. Cancer*, **6**, 347–359.
27. Weinmaster, G. and Kintner, C. (2003) Modulation of notch signaling during somitogenesis. *Annu. Rev. Cell. Dev. Biol.*, **19**, 367–395.
28. Henkel, G., Weiss, D.L., McCoy, R., Deloughery, T., Tara, D. and Brown, M.A. (1992) A DNase I-hypersensitive site in the second intron of the murine IL-4 gene defines a mast cell-specific enhancer. *J. Immunol.*, **149**, 3239–3246.
29. Okazaki, M., Maeda, G., Chiba, T., Doi, T. and Imai, K. (2009) Identification of GATA3 binding sites in Jurkat cells. *Gene*, **445**, 17–25.
30. Zheng, W. and Flavell, R.A. (1997) The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. *Cell*, **89**, 587–596.
31. Zhu, J., Min, B., Hu-Li, J., Watson, C.J., Grinberg, A., Wang, Q., Killeen, N., Urban, J.F. Jr, Guo, L. and Paul, W.E. (2004) Conditional deletion of Gata3 shows its essential function in T(H)1-T(H)2 responses. *Nat. Immunol.*, **5**, 1157–1165.
32. Kishikawa, H., Sun, J., Choi, A., Miaw, S.C. and Ho, I.C. (2001) The cell type-specific expression of the murine IL-13 gene is regulated by GATA-3. *J. Immunol.*, **167**, 4414–4420.
33. Lavenu-Bombled, C., Trainor, C.D., Makeh, I., Romeo, P.H. and Max-Audit, I. (2002) Interleukin-13 gene expression is regulated by GATA-3 in T cells: role of a critical association of a GATA and two GATG motifs. *J. Biol. Chem.*, **277**, 18313–18321.
34. Baker, P.J. (2000) The role of immune responses in bone loss during periodontal disease. *Microbes Infect.*, **2**, 1181–1192.
35. Yamamoto, M., Fujihashi, K., Hiroi, T., McGhee, J.R., Van Dyke, T.E. and Kiyono, H. (1997) Molecular and cellular mechanisms for periodontal diseases: role of Th1 and Th2 type cytokines in induction of mucosal inflammation. *J. Periodontal Res.*, **32**, 115–119.
36. van Hamburg, J.P., Mus, A.M., de Bruijn, M.J., de Vogel, L., Boon, L., Cornelissen, F., Asmawidjaja, P., Hendriks, R.W. and Lubberts, E. (2009) GATA-3 protects against severe joint inflammation and bone erosion and reduces differentiation of Th17 cells during experimental arthritis. *Arthritis Rheum.*, **60**, 750–759.
37. Giannopoulou, C., Kamma, J.J. and Mombelli, A. (2003) Effect of inflammation, smoking and stress on gingival crevicular fluid cytokine level. *J. Clin. Periodontol.*, **30**, 145–153.
38. Fujihashi, K., Beagley, K.W., Kono, Y., Aicher, W.K., Yamamoto, M., DiFabio, S., Xu-Amano, J., McGhee, J.R. and Kiyono, H. (1993) Gingival mononuclear cells from chronic inflammatory periodontal tissues produce interleukin (IL)-5 and IL-6 but not IL-2 and IL-4. *Am. J. Pathol.*, **142**, 1239–1250.
39. Kabashima, H., Nagata, K., Hashiguchi, I., Toriya, Y., Iijima, T., Maki, K. and Maeda, K. (1996) Interleukin-1 receptor antagonist and interleukin-4 in gingival crevicular fluid of patients with inflammatory periodontal disease. *J. Oral. Pathol. Med.*, **25**, 449–455.
40. Shapira, L., van Dyke, T.E. and Hart, T.C. (1992) A localized absence of interleukin-4 triggers periodontal disease activity: a novel hypothesis. *Med. Hypotheses*, **39**, 319–322.
41. Tsai, C.C., Ku, C.H., Ho, Y.P., Ho, K.Y., Wu, Y.M. and Hung, C.C. (2007) Changes in gingival crevicular fluid interleukin-4 and interferon-gamma in patients with chronic periodontitis before and after periodontal initial therapy. *Kaohsiung J. Med. Sci.*, **23**, 1–7.
42. Bastos, M.F., Lima, J.A., Vieira, P.M., Mestnik, M.J., Faveri, M. and Duarte, P.M. (2009) TNF-alpha and IL-4 levels in generalized aggressive periodontitis subjects. *Oral Dis.*, **15**, 82–87.
43. te Velde, A.A., Huijbens, R.J., Heije, K., de Vries, J.E. and Figdor, C.G. (1990) Interleukin-4 (IL-4) inhibits secretion of IL-1 beta, tumor necrosis factor alpha, and IL-6 by human monocytes. *Blood*, **76**, 1392–1397.
44. Spilianakis, C.G. and Flavell, R.A. (2004) Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat. Immunol.*, **5**, 1017–1027.
45. Cai, S., Lee, C.C. and Kohwi-Shigematsu, T. (2006) SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.*, **38**, 1278–1288.
46. Krawczak, M., Nikolaus, S., von Eberstein, H., Croucher, P.J., El Mokhtari, N.E. and Schreiber, S. (2006) PopGen: population-based recruitment of patients and controls for the analysis of complex genotype–phenotype relationships. *Community Genet.*, **9**, 55–61.
47. Franke, A., Balschun, T., Karlsen, T.H., Sventoraityte, J., Nikolaus, S., Mayr, G., Domingues, F.S., Albrecht, M., Nothnagel, M., Ellinghaus, D. et al. (2008) Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nat. Genet.*, **40**, 1319–1323.
48. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
49. Wigginton, J.E., Cutler, D.J. and Abecasis, G.R. (2005) A note on exact tests of Hardy–Weinberg equilibrium. *Am. J. Hum. Genet.*, **76**, 887–893.
50. The R Development Core Team, http://www.r-project.org.
51. Weckx, S., Del-Favero, J., Rademakers, R., Claes, L., Cruts, M., De Jonghe, P., Van Broeckhoven, C. and De Rijk, P. (2005) novoSNP, a novel computational tool for sequence variation discovery. *Genome Res.*, **15**, 436–442.
52. de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. and Altshuler, D. (2005) Efficiency and power in genetic association studies. *Nat. Genet.*, **37**, 1217–1223.
53. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
54. Becker, T. and Knapp, M. (2004) A powerful strategy to account for multiple testing in the context of haplotype analysis. *Am. J. Hum. Genet.*, **75**, 561–570.
55. Abecasis, G.R. and Cookson, W.O. (2000) GOLD—graphical overview of linkage disequilibrium. *Bioinformatics*, **16**, 182–183.
56. Qin, Z.S., Niu, T. and Liu, J.S. (2002) Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am. J. Hum. Genet.*, **71**, 1242–1247.
57. Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, **29**, 45.
58. Roider, H.G., Kanhere, A., Manke, T. and Vingron, M. (2007) Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics*, **23**, 134–141.
59. Manke, T., Roider, H.G. and Vingron, M. (2008) Statistical modeling of transcription factor binding affinities predicts regulatory interactions. *PLoS Comput. Biol.*, **4**, e1000039.