Human Molecular Genetics, 2017, Vol. 26, No. 13 2577–2588

doi: 10.1093/hmg/ddx151
Advance Access Publication Date: 25 April 2017
Association Studies Article

ASSOCIATION STUDIES ARTICLE

A genome-wide association study identifies nucleotide variants at *SIGLEC5* and *DEFA1A3* as risk loci for periodontitis

Matthias Munz[1,2], Christina Willenborg[2], Gesa M. Richter[1], Yvonne Jockel-Schneider[3], Christian Graetz[4], Ingmar Staufenbiel[5], Jürgen Wellmann[6], Klaus Berger[6], Bastian Krone[7], Per Hoffmann[8,9], Nathalie van der Velde[10,11], André G. Uitterlinden[10], Lisette C.P.G.M. de Groot[12], Amr H. Sawalha[13,14], Haner Direskeneli[15], Güher Saruhan-Direskeneli[16], Esra Guzeldemir-Akcakanat[17], Huseyin Gencay Keceli[18], Matthias Laudes[19], Barbara Noack[20], Alexander Teumer[21], Birte Holtfreter[22], Thomas Kocher[22], Peter Eickholz[23], Jörg Meyle[24], Christof Doerfer[4], Corinna Bruckmann[25], Wolfgang Lieb[26], Andre Franke[27], Stefan Schreiber[19,27], Rahime M. Nohutcu[17], Jeanette Erdmann[2], Bruno G. Loos[28], Soeren Jepsen[29], Henrik Dommisch[1] and Arne S. Schaefer[1,*]

[1]Department of Periodontology and Synoptic Dentistry, Institute of Dental, Oral and Maxillary Medicine, Charité – University Medicine Berlin, Germany, [2]Institute for Integrative and Experimental Genomics, University Medical Center Schleswig-Holstein - Campus Lübeck, Germany, [3]Department of Periodontology, Clinic of Preventive Dentistry and Periodontology, University Medical Center of the Julius-Maximilians-University, Würzburg, Germany, [4]Department of Operative Dentistry and Periodontology, University Medical Center Schleswig-Holstein, Campus Kiel, Germany, [5]Department of Conservative Dentistry, Periodontology and Preventive Dentistry, Hannover Medical School, Hannover, Germany, [6]Institute of Epidemiology and Social Medicine, University Münster, Germany, [7]Institute of Medical Informatics, Biometry and Epidemiology, University Clinic Essen, Germany, [8]Institute of Human Genetics, University of Bonn, Germany, [9]Human Genomics Research Group, Department of Biomedicine, University Hospital of Basel, Switzerland, [10]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands, [11]Department of Internal Medicine Section of Geriatrics, Amsterdam Medical Center, Amsterdam, The Netherlands, [12]Department of Epidemiology and the EMGO Institute of Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands, [13]Division of Rheumatology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA, [14]Center for Computational Medicine and Bioinformatics, University of Michigan Medical School, USA, [15]Division of Rheumatology, Marmara University, School of

Received: December 14, 2016. **Revised:** March 31, 2017. **Accepted:** April 12, 2017

Medicine, Istanbul, Turkey, [16]Department of Physiology, Istanbul Medical Faculty, Istanbul University, Istanbul, Turkey, [17]Department of Periodontology, Faculty of Dentistry, Kocaeli University, Turkey, [18]Department of Periodontology, Faculty of Dentistry, Hacettepe University, Sihhiye, Ankara, Turkey, [19]Clinic of Internal Medicine, University Clinic Schleswig-Holstein, Kiel, Germany, [20]Clinic of Conservational Dentistry, Center of Dental, Oral and Maxillary Medicine, University Medical Center Carl-Gustav-Carus, Technical University Dresden, Germany, [21]Institute for Community Medicine, University Medicine Greifswald, Germany, [22]Unit of Periodontology, Department of Restorative Dentistry, Periodontology, Endodontology, Preventive Dentistry and Pedodontics, Dental School, University Medicine Greifswald, Germany, [23]Department of Periodontology, Johann Wolfgang Goethe-University, Frankfurt am Main, Germany, [24]Department of Periodontology, University Medical Center Giessen and Marburg, Germany, [25]Department of Conservative Dentistry and Periodontology, Medical University Vienna, School of Dentistry, Vienna, Austria, [26]Institute of Epidemiology, Biobank PopGen, Christian-Albrechts-University, Kiel, Germany, [27]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany, [28]Department of Periodontology and Oral Biochemistry, Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam and VU University Amsterdam, The Netherlands and [29]Department of Periodontology, Operative and Preventive Dentistry, University of Bonn, Bonn, Germany

*To whom correspondence should be addressed at: Charité Centrum Zahn-, Mund- und Kieferheilkunde CC 3, Parodontologie und Synoptische Zahnmedizin, Aßmannshauser Straße 4-6, 14197 Berlin, Germany. Tel: +49 30 450562343; Fax: +49 30 4507562343; Email: arne.schaefer@charite.de

Abstract

Periodontitis is one of the most common inflammatory diseases, with a prevalence of 11% worldwide for the severe forms and an estimated heritability of 50%. The disease is characterized by destruction of the alveolar bone due to an aberrant host inflammatory response to a dysbiotic oral microbiome. Previous genome-wide association studies (GWAS) have reported several suggestive susceptibility loci. Here, we conducted a GWAS using a German and Dutch case-control sample of aggressive periodontitis (AgP, 896 cases, 7,104 controls), a rare but highly severe and early-onset form of periodontitis, validated the associations in a German sample of severe forms of the more moderate phenotype chronic periodontitis (CP) (993 cases, 1,419 controls). Positive findings were replicated in a Turkish sample of AgP (223 cases, 564 controls). A locus at *SIGLEC5* (sialic acid binding Ig-like lectin 5) and a chromosomal region downstream of the *DEFA1A3* locus (defensin alpha 1-3) showed association with both disease phenotypes and were associated with periodontitis at a genome-wide significance level in the pooled samples, with $P = 1.09E-08$ (rs4284742,-G; OR = 1.34, 95% CI = 1.21–1.48) and $P = 5.48E-10$ (rs2738058,-T; OR = 1.28, 95% CI = 1.18–1.38), respectively. SIGLEC5 is expressed in various myeloid immune cells and classified as an inhibitory receptor with the potential to mediate tyrosine phosphatases SHP-1/-2 dependent signaling. Alpha defensins are antimicrobial peptides with expression in neutrophils and mucosal surfaces and a role in phagocyte-mediated host defense. This study identifies the first shared genetic risk loci of AgP and CP with genome-wide significance and highlights the role of innate and adaptive immunity in the etiology of periodontitis.

Introduction

Periodontitis (PD) is among the most common inflammatory diseases, affecting human populations at worldwide prevalence rates of 11% for the severe forms (1). It is the major cause of tooth loss in adults above 40 years (2). A microbial shift of the oral microbiota is considered to trigger the inflammatory reaction and long-term smoking and diabetes add to the disease risk. The inflammation leads to gingival bleeding, gingival pocket formation and to the resorption of alveolar bone, which withdraws from the site of inflammation, a process that eventually causes tooth loss. These steps are unique in the pathogenesis of complex diseases, because the oral cavity is the only part of the human organism where, in an intricate environment of microbes, soft tissues like the gingival mucosa dynamically interact with hard tissues like teeth and skeletal bone. However, the precise molecular mechanisms that drive the individual steps in the pathogenesis of PD are currently unknown. The different phenotypes of PD can be considered as different parts of a large range of similar conditions, which can be attributed to the effects of different combinations of genetic risk loci, which form the individual genetic constitution. This genetic constitution determines the immune response, which is challenged by internal and external factors like the biofilm, smoking, stress, age, excess dietary fat and diabetes. In this view, different disease manifestations are not confined entities but share risk alleles and covariates. Nevertheless, the current study uses samples classified as aggressive and chronic periodontitis, in accordance with the current classification for PD. Aggressive periodontitis (AgP) is a rare disease phenotype with a prevalence of $\sim 0.1\%$ in European Caucasians (3). The diagnosis is based on severe periodontal attachment loss and severe destruction of the alveolar bone in adolescents and young adults (< 35 years of age). Because of the usually very young age of disease onset, the absence of risk factors like long-term smoking or diabetes, often paralleled by the absence of plaque, it is considered that largely unknown genetic risk variants play a strong role in disease onset and progression of AgP. Unraveling the genetic architecture of AgP may help to understand the molecular mechanisms that underlie the disease risk of PD.

We earlier performed a genome-wide association study (GWAS) of AgP with 438 cases and 1,320 controls on the Affymetrix 500K Genotyping System and identified a single

nucleotide polymorphism (SNP) within the gene *GLT6D1* to be associated with AgP at a genome-wide significance level (4), which was later replicated in a sub-Saharan population from Sudan (5). GWAS that focused on chronic periodontitis (CP) suggested various risk variants but have hitherto failed to give clear statistical evidence for association (6–12). This is partly due to the high heterogeneity of CP and the important contributions of confounding risk factors such as smoking and diabetes, but mostly relates to the limited statistical power of these GWAS. We now expand the GWAS in a sample of 896 AgP cases and 7,104 controls from Germany (GER) and The Netherlands (NL), validated the most significant associations in a GWAS of 2,211 German CP cases and 1,817 controls (6), and replicated the variants that were shared between AgP and CP in an independent AgP sample from Turkey (220 cases, 550 controls). Our data give novel insights into the molecular mechanisms of both forms of PD.

Results

Discovery stage

We performed two GWAS in a sample of 717 AgP cases and 4,213 controls from GER, and a sample of 179 AgP cases and 2,891 controls from NL. After the QC steps, we retained 502,332 intersecting autosomal SNPs in 680 cases and 3,973 controls in GER, and 554,855 intersecting autosomal SNPs in 171 cases and 2,607 controls in NL. To combine the dissenting SNP sets, to fill-in missing data and to detect genotyping errors, we imputed genotypes based on reference haplotypes from 1000 Genomes Phase 3. Both GWAS studies were combined in a meta-analysis with association data for 5,870,152 genotyped or imputed autosomal SNPs. Multidimensional scaling showed minimal evidence for population stratification in GER and NL after removal of outliers in the QC procedures (Supplementary Material, Figs S1 and S2). According to the quantile-quantile plots, the distribution of observed *P* values deviated from expected *P* values only in the extreme tail (Supplementary Material, Fig. S3). Genomic inflation factors (λ) for SNPs were 1.05 and 1.02 for GER and NL, respectively, and 0.97 for the two studies combined, indicating population stratification effects were negligible. In the discovery stage, 16 distinct loci surpassed our preassigned selection criteria (Table 1, Supplementary Material, Figs S1–S4).

Adjustment for (i) no covariates, (ii) sex, (iii) smoking status, (iv) MDS components, and (v) all covariates showed only small (maximal up to power of ten) variation of the *P* values in the discovery stage for all top SNPs. This indicates that the covariates only have a minor influence on the level of association (Table 1, Supplementary Material, Table S1).

Validation with CP

16 lead SNPs of the discovery stage were taken forward to validation in 993 German CP cases and 1,419 controls, using imputed genotype data of a GWAS of CP (6). Three SNPs at two loci showed association with CP (Table 1, Fig. 1). At the *DEFA1A3* locus (13), AgP GWAS leads SNP rs2978951, and additionally rs2738058, which was previously found to be associated with CP (6) was included in the validation stage due to its low LD with rs2978951 ($r^2 = 0.28$). rs2738058 showed the statistically most robust signal of association with CP with $P = 6.7 \times 10^{-06}$, OR = 1.33 (95% CI = 1.18–1.51) compared to rs2978951 ($P = 1.5 \times 10^{-02}$, OR = 1.18 [95% CI = 1.03–1.34]). AgP GWAS lead SNP rs4284742 at *SIGLEC5* (sialic acid binding Ig-like lectin 5) showed association with CP with $P = 0.0029$, odds ratio (OR) = 1.34 (95% confidence interval [95% CI] = 1.11-1.63).

The top SNPs at both loci were in low LD with other highly associated variants, suggesting multiple causal variants. However, in the validation step we were only able to validate rs2738058 at *DEFA1A3*, because the SNPs were either not included in CP or the OR strongly deviated from AgP-GER-NL resulting in larger *P* values (Fig. 1). To fully assess whether rs2978951 and rs2738058 are independent variants, we performed a conditional analysis with AgP-GER and AgP-NL using an additive (add), genotypic (gen), recessive (rec), dominant (dom) and heterozygous (het) conditioning model (Supplementary Material, Table S5). According to the *P* values of AgP-GER, independence of the two SNPs was only observed for the het model. SNP rs2978951 had a level of significance of $P = 1.79 \times 10^{-06}$ before and $P = 1.61 \times 10^{-06}$ after conditioning on SNP rs2738058, while the OR and 95% CI remained constant. The *P* value for NL slightly decreased from $P = 2.37 \times 10^{-02}$ to $P = 2.80 \times 10^{-02}$. Conditioning with models add, dom, rec and gen decreased the *P* value up to $P = 4.46 \times 10^{-03}$ (add) in GER and increased it in NL ($P = 6.45 \times 10^{-02}$ [rec]).

Meta-analysis of AgP and CP. The subsequent meta-analysis, which included the AgP discovery and the CP replication sample comprised 1,844 cases and 8,255 controls. It showed genome-wide significant evidence of association for the GWAS lead SNPs at *DEFA1A3* and *SIGLEC5* (Supplementary Material, Table S1). The significance of association for the SNPs rs2978951 and rs2738058 at *DEFA1A3* was $P = 4.59 \times 10^{-08}$ with a genetic effect of OR = 1.26 (95% CI = 1.16–1.36), and $P = 1.29 \times 10^{-09}$ and an OR = 1.28 (95% CI = 1.18–1.39), respectively. For SNP rs4284742 at *SIGLEC5*, it was $P = 4.71 \times 10^{-08}$ with an OR = 1.33 (95% CI = 1.20–1.48). For the other suggested AgP associations, the non-findings in the meta-analysis may be caused by a lack of statistical power of the CP sample, resulting in a false negative finding. Accordingly, no conclusion on a potential role for the etiology of CP can be drawn for these SNPs.

Replication in a Turkish AgP sample

We replicated the associations of the 15 SNPs, which were validated with CP, in an additional AgP case-control sample of Turkish descent (Supplementary Material, Table S1). In this sample, SNP rs1122900 at *SLC1A3* showed nominal significant (i.e. $P < 0.05$) association with $P = 0.0137$ with the same effect direction (OR = 1.32, 95% CI = 1.06–1.65). After pooling with the AgP discovery sample, the association level of rs1122900 was $P = 8 \times 10^{-7}$ (OR = 1.27, 95% CI = 1.16–1.4, Table 2), i.e. stronger compared to its association with the discovery sample alone ($P = 1.77 \times 10^{-05}$) and also compared to the association of the GWAS lead SNP rs6887423 in the discovery sample ($P = 1.42 \times 10^{-06}$).

The other 16 loci showed no individual association with AgP in the small Turkish sample. However, the ORs of multiple SNPs was consistent with ORs in the discovery and resulted in stronger associations after pooling with the AgP discovery sample (Tables 1 and 2).

Meta-analysis of all samples

The final meta-analysis included all AgP samples and the CP validation samples. Here, the *DEFA1A3* SNPs rs2978951 and rs2738058 reached genome-wide significance with $P = 2.06 \times 10^{-08}$ (OR = 1.25, 95% CI = 1.16–1.35; 2,067 cases 8,533 controls) and $P = 6.78 \times 10^{-10}$ (OR = 1.28, 95% CI = 1.18–1.38; 2,067 cases and 8,819 controls), respectively. Likewise, the GWAS lead SNP rs4284742 at *SIGLEC5* was genome-wide significant with $P = 1.34$

Table 1. Three SNPs reached genome-wide significance in the pooled analysis

	SNP	r^2	Locus	Nearest Gene(s)	EA	NEA	EAF	Stage	Adjustment	OR (95% CI)	P	P(Q)	I^2
1	rs2978951		8p23.1	DEFA1A3	A	G	0.41	AgP-Ger	–	1.32 [1.18,1.48]	2.12E-06		
								AgP-Ger	Sex	1.32 [1.18,1.48]	2.38E-06		
								AgP-Ger	Smoking	1.32 [1.18,1.48]	1.84E-06		
								AgP-Ger	MDS1-6	1.32 [1.18,1.48]	1.78E-06		
								AgP-Ger	**Full**	**1.32 [1.18,1.49]**	**1.88E-06**		
								AgP-NL	–	1.25 [1.00,1.56]	4.56E-02		
								AgP-NL	Sex	1.25 [1.00,1.56]	3.73E-02		
								AgP-NL	Smoking	1.25 [1.00,1.56]	3.49E-02		
								AgP-NL	**Full**	**1.25 [1.00,1.57]**	**2.41E-02**		
								Discovery (AgP-GER+AgP-NL)	–	1.31 [1.18,1.45]	3.27E-07		
								Discovery (AgP-GER+AgP-NL)	Sex	1.31 [1.18,1.45]	3.44E-07		
								Discovery (AgP-GER+AgP-NL)	Smoking	1.31 [1.18,1.45]	3.45E-07		
								Discovery (AgP-GER+AgP-NL)	**Full**	**1.31 [1.18,1.45]**	**4.49E-07**	**0.66**	**0**
								Validation (CP-GER)		**1.18 [1.03,1.34]**	**1.50E-02**		
								Replication (AgP-TUR)		**1.19 [0.92,1.55]**	**1.93E-01**		
								Pooled (Disc.+Valid.)		1.26 [1.16,1.36]	4.59E-08	0.43	0
								Pooled (Disc.+Repl.)		1.29 [1.17,1.42]	2.33E-07	0.74	0
								Pooled (Disc.+Valid.+Repl.)		**1.25 [1.16,1.35]**	**2.06E-08**	**0.61**	**0**
1.1	rs2738058	0.28	8p23.1	DEFA1A3	T	C	0.43	AgP-Ger	–	1.29 [1.15,1.45]	1.30E-05		
								AgP-Ger	Sex	1.29 [1.15,1.45]	1.65E-05		
								AgP-Ger	Smoking	1.29 [1.15,1.45]	1.21E-05		
								AgP-Ger	MDS1-6	1.29 [1.15,1.45]	1.70E-06		
								AgP-Ger	**Full**	**1.29 [1.14,1.45]**	**2.23E-06**		
								AgP-NL	–	1.11 [0.89,1.38]	3.52E-01		
								AgP-NL	Sex	1.11 [0.89,1.38]	3.60E-01		
								AgP-NL	Smoking	1.11 [0.89,1.38]	3.15E-01		
								AgP-NL	**Full**	**1.11 [0.89,1.39]**	**3.49E-01**		
								Discovery (AgP-GER+AgP-NL)	–	1.25 [1.12,1.38]	2.63E-05	0.24	0.28
								Discovery (AgP-GER+AgP-NL)	Sex	1.25 [1.13,1.38]	2.62E-05	0.24	0.27
								Discovery (AgP-GER+AgP-NL)	Smoking	1.25 [1.12,1.38]	2.71E-05	0.24	0.27
								Discovery (AgP-GER+AgP-NL)	**Full**	**1.25 [1.12,1.38]**	**3.46E-05**	**0.25**	**0.25**
								Validation (CP-GER)		**1.33 [1.18,1.51]**	**6.65E-06**		
								Replication (AgP-TUR)		**1.25 [0.85,1.83]**	**2.59E-01**		
								Pooled (Disc.+Valid.)		1.28 [1.18,1.39]	1.29E-09	0.38	0
								Pooled (Disc.+Repl.)		1.25 [1.13,1.38]	1.77E-05	0.51	0
								Pooled (Disc.+Valid.+Repl.)		**1.28 [1.18,1.38]**	**6.78E-10**	**0.58**	**0**
2	rs4284742*		19q13.41	SIGLEC5, AC018755.18	G	A	0.76	AgP-Ger	–	1.37 [1.18,1.59]	2.56E-05		
								AgP-Ger	Sex	1.37 [1.18,1.59]	2.45E-05		
								AgP-Ger	Smoking	1.37 [1.18,1.59]	2.63E-05		
								AgP-Ger	MDS1-6	1.37 [1.18,1.59]	1.34E-05		
								AgP-Ger	**Full**	**1.37 [1.19,1.57]**	**1.46E-05**		
								AgP-NL	–	1.19 [0.91,1.56]	1.84E-01		
								AgP-NL	Sex	1.19 [0.91,1.56]	2.39E-01		
								AgP-NL	Smoking	1.19 [0.91,1.56]	1.83E-01		
								AgP-NL	**Full**	**1.19 [0.91,1.56]**	**2.86E-01**		
								Discovery (AgP-GER+AgP-NL)	–	1.33 [1.18,1.50]	3.61E-06	0.36	0.00
								Discovery (AgP-GER+AgP-NL)	Sex	1.33 [1.18,1.50]	3.65E-06	0.37	0.00
								Discovery (AgP-GER+AgP-NL)	Smoking	1.33 [1.18,1.50]	3.72E-06	0.36	0.00
								Discovery (AgP-GER+AgP-NL)	**Full**	**1.33 [1.18,1.51]**	**4.66E-06**	**0.37**	**0.00**
								Validation (CP-GER)		**1.34 [1.11,1.63]**	**2.91E-03**		
								Replication (AgP-TUR)		**1.41 [0.92,2.17]**	**1.14E-01**		
								Pooled (Disc.+Valid.)		**1.33 [1.20,1.48]**	**4.71E-08**	**0.67**	**0**
								Pooled (Disc.+Repl.)		**1.34 [1.19,1.50]**	**1.31E-06**	**0.65**	**0**
								Pooled (Disc.+Valid.+Repl.)		**1.34 [1.21,1.48]**	**1.34E-08**	**0.83**	**0**

*(EA = Effect Allele, NEA = Non-Effect Allele, EAF = Effect Allele Frequency).

*For this SNP the Turkish GWAS controls had a poor genotype quality and had to be excluded from the analysis, which reduced the number of controls for this SNP from 564 to 75.

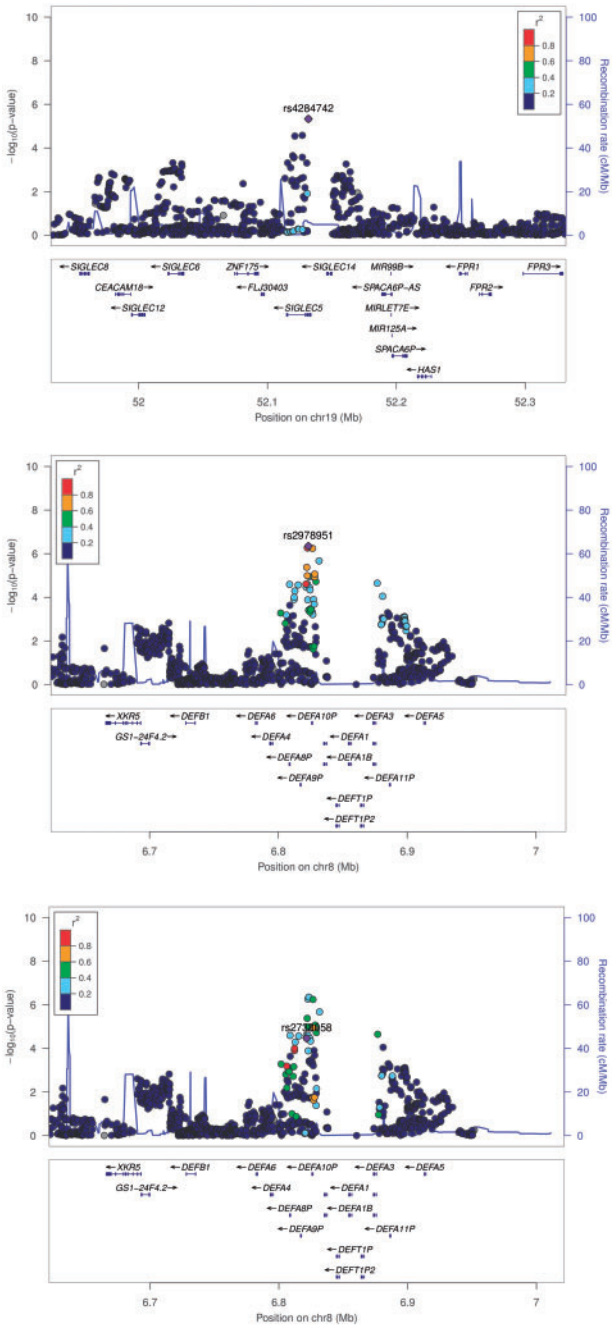

Figure 1. Regional association plots of rs4284742, rs2978951 and rs2738058 in the discovery stage.

$\times 10^{-08}$ (OR = 1.34, 95% CI = 1.21–1.48; 2,027 cases and 8,330 controls).

Expression quantitative trait locus (eQTL) mapping

Examination of the 16 distinct haplotype blocks that suggested association with AgP for expression quantitative trait loci (QTLs) indicated cis- and trans-regulatory effects for various SNPs (Table 3, Supplementary Material, Table S6).

Of the SNPs that showed genome-wide significant association in the meta-analysis, SNP rs2738058 at *DEFA1A3* and rs2977793, which tags GWAS lead SNP rs2978951 at *DEFA1A3* ($r^2 = 0.8$) both show eQTL effects on *DEFA4*, with $P = 1.7 \times 10^{-03}$ and 3.9×10^{-05}, respectively. However, these SNPs map between the pseudogenes *DEFA9P* and *DEFA10P*, 5 kb downstream of *DEFA4*, which are expressed in blood at low levels (GTEx data). Additionally, rs4284742 at *SIGLEC5* showed a cis-effect on the transcription level of *SIGLEC5* with $P = 7.7 \times 10^{-14}$ in whole blood, indicating *SIGLEC5* to be the affected gene of this association.

Experimental data showed that SIGLEC5 efficiently inhibits FCER1G mediated calcium fluxing (14). *FCER1G* was one of the 17 risk loci, which were tagged by SNPs that surpassed our preassigned selection criteria. In the AgP discovery sample, rs2070901, which is located 30 base pairs upstream of the first exon of *FCER1G*, showed an association with $P = 9.45 \times 10^{-06}$, OR = 1.31 (95% CI = 1.16–1.47). eQTL data of this SNP suggested a cis-effect on the expression of *FCER1G* ($P = 9.16 \times 10^{-16}$) in whole blood. However, the most significant effect of rs2070901 was reported to act in *trans* on the expression of *B4GALT3* ($P = 8.43 \times 10^{-26}$). *B4GALT3* expression alters glycosylation and expression of the N-glycan of β1 integrin (ITGB1) (15). SNP rs2144815, another of the 16 GWAS lead SNPs that surpassed our preassigned selection criteria with $P = 1.35 \times 10^{-06}$, OR = 1.39 (95% CI = 1.22–1.59) is located within exon 3 of the long noncoding RNA (lncRNA) RP11-725G5.2. According to GTEx, rs2144815 has a cis-eQTL effect on RP11-725G5.2 (9.86×10^{-11}) in testis and according to GRASP v2 a trans-eQTL effect on *ITGB1* expression in peripheral blood monocytes ($P = 1.22 \times 10^{-06}$). No additional eQTL effect of rs2144815 was reported to date.

For SNP rs1122900 at *SLC1A3*, a cis-eQTL effect for the antisense lncRNA CTD-2353F22.1 ($P = 2.9 \times 10^{-07}$) in lymphocytes was reported in GTEx, which overlaps with the 3' end of *SLC1A3*. Cis genes *DEFA1A3*, *SIGLEC5*, *FCER1G* and CTD-2353F22.1 belong to the same topologically associating domains (TAD, i.e genomic regions defined by interactome boundaries that are mainly cell-type independent), as their linked SNP (Supplementary Material, Table S7) (16).

Relationship with other traits

Among the 16 lead SNPs and their nine neighboring SNPs, which were examined in our analyses, only SNP rs2738058 at the *DEFA1A3* locus has a genome-wide significant association with another trait (IgA nephropathy, $P = 2 \times 10^{-27}$, OR = 1.27, 95% CI = 1.21–1.32, Supplementary Material, Table S8). Moreover, the same SNP was earlier reported as a suggestive risk variant of CP, in the same cohort we used in our validation, before (6). Further associations were related to plasminogen levels (rs4284742), cholesterol (rs17340482, rs4284742), triglycerides (rs17340482), blood pressure (top SNPs rs17340482, rs2978951), arthritis (rs2978951, rs4284742), obesity (rs2070901, rs3830904, rs11382193, rs11633566, rs3830904), cardiac diseases (rs17340482, rs744280), bone (rs1002204), and psychologic disorders (rs3830904, rs6137428, rs1002204, rs11382193, rs3830904). None of the traits was directly related to smoking. A genetic correlation analysis with AgP-GER-NL and 219 other traits showed positive correlations especially to metabolic traits (Supplementary Material, Table S9).

Discussion

The current study identified PD risk loci at SIGLEC5 and DEFA1A3 with genome-wide significance. The various disease phenotypes of PD are part of a large range of similar conditions that are attributed to the effects of different combinations of

Table 2. Lead SNPs at 14 suggestive risk loci

								Discovery		Replication		Pooled			
								(AgP-GER + AgP-NL)		(AgP-TUR)					
	SNP	r^2	Locus	Nearest Gene(s)	EA	NEA	EAF	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	P(Q)	I^2
1	rs17340482		1p36.11	NUDC	G	C	0,90	1.56 [1.31,1.87]	9.99E-07	NA	NA	NA	NA	NA	NA
1.1	rs4970469*	0.97	1p36.11	OSTCP2	G	A	0,90	1.55 [1.30,1.85]	1.29E-06	1.26 [0.69,2.31]	4.55E-01	1.52 [1.29, 1.81]	1.20E-06	6.50E-01	0.00
2	rs3830904		14q32.13	RP11-725G5.2 - SYNE3	T	TC	0,17	1.40 [1.22,1.60]	1.16E-06	NA	NA	NA	NA	NA	NA
2.1	rs2144815	1.00	14q32.13	RP11-725G5.2	T	C	0,17	1.39 [1.22,1.59]	1.53E-06	0.88 [0.67,1.16]	3.71E-01	1.20 [0.91, 1.59]	2.00E-01	1.50E-02	0.76
3	rs6887423		5p13.2	CTD-2353F22.1	T	C	0,42	1.29 [1.17,1.44]	1.42E-06	NA	NA	NA	NA	NA	NA
3.1	rs1122900*	0.65	5p13.2	CTD-2353F22.1	A	C	0,40	1.26 [1.13,1.40]	1.77E-05	1.32 [1.06,1.65]	1.37E-02	1.27 [1.16,1.40]	8.00E-07	7.76E-01	0.00
4	rs747804		9p13.3	PGAM1P2 - CCDS6596.2	G	A	0,64	1.30 [1.17,1.45]	1.65E-06	NA	NA	NA	NA	NA	NA
4.1	rs13283964	1.00	9p13.3	LINC00961 - PGAM1P2	T	C	0,64	1.29 [1.16,1.44]	2.50E-06	1.00 [0.76,1.33]	9.75E-01	1.25 [1.13,1.38]	1.01E-05	2.60E-01	0.26
5	rs6137428		20p13	RP11-128M1.1 - TGM3	C	T	0,90	1.52 [1.27,1.81]	3.92E-06	NA	NA	NA	NA	NA	NA
5.1	rs6047560	0.73	20p13	RP11-128M1.1 - TGM3	A	G	0,87	1.40 [1.20,1.64]	2.20E-05	1.09 [0.68,1.75]	7.17E-01	1.37 [1.18,1.59]	3.43E-05	5.68E-01	0.00
6	rs1380780		3q26.1	LINC01192 - RNU7-82P201	T	A	0,72	1.31 [1.17,1.47]	4.27E-06	NA	NA	NA	NA	NA	NA
6.1	rs1380781	1.00	3q26.1	LINC01192 - RNU7-82P201	C	T	0,72	1.31 [1.17,1.47]	4.29E-06	0.93 [0.69,1.25]	6.26E-01	1.22 [0.99.1.51]	6.00E-02	6.29E-02	0.64
7	rs11382193		11p15.3	MICAL2	A	AT	0,63	1.29 [1.15,1.43]	5.76E-06	NA	NA	NA	NA	NA	NA
8	rs1002204		7q21.12	ABCB1	C	A	0,47	1.27 [1.14,1.40]	7.92E-06	0.99 [0.78,1.27]	9.58E-01	1.22 [1.11,1.34]	4.28E-05	2.00E-01	0.38
9	rs744280		15q12	ATP10A	C	T	0,73	1.31 [1.16,1.47]	7.97E-06	1.07 [0.82,1.40]	6.06E-01	1.26 [1.14,1.41]	1.75E-05	2.82E-01	0.21
10	rs11633566		15q21.1	RP11-325E5.1	G	A	0,75	1.31 [1.17,1.48]	7.99E-06	NA	NA	NA	NA	NA	NA
10.1	rs3088333	0.74	15q21.1	COPS2	C	T	0,79	1.32 [1.16,1.49]	1.83E-05	0.82 [0.62,1.07]	1.39E-01	1.15 [0.86,1.55]	3.50E-01	5.91E-03	0.81
11	rs4731202		7q21.11	RNU6-849P - RPL10P11	T	G	0,66	1.28 [1.15,1.43]	8.54E-06	NA	NA	NA	NA	NA	NA
11.1	rs10238400	0.90	7q21.11	RPL10P11 - GNAI1	A	G	0,65	1.25 [1.13,1.40]	4.09E-05	0.89 [0.61,1.29]	5.32E-01	1.22 [1.10,1.36]	1.62E-04	1.46E-01	0.48
12	rs2070901*		1q23.3	FCER1G	T	G	0,24	1.31 [1.16,1.47]	1.12E-05	1.22 [0.93,1.60]	1.56E-01	1.29 [1.16,1.44]	4.36E-06	8.97E-01	0.00
13	rs7595654		2q37.1	AC006037.2 - RP11-309M7.1	A	G	0,67	1.28 [1.15,1.43]	1.14E-05	NA	NA	NA	NA	NA	NA
14	rs62312993		4q22.1	CCSER1	C	T	0,80	1.34 [1.18,1.53]	1.24E-05	NA	NA	NA	NA	NA	NA
14.1	rs11727861	0.99	4q22.1	CCSER1	C	T	0,80	1.34 [1.17,1.53]	1.48E-05	1.02 [0.73,1.41]	9.16E-01	1.29 [1.14.1.46]	4.94E-05	2.34E-01	0.31

(EA = Effect Allele, NEA = Non-Effect Allele, EAF = Effect Allele Frequency).

*P-value got smaller in pooled analysis using AgP cohorts only.

Result of the unadjusted association tests are shown in Supplementary Material, Table S1.

Table 3. Public domain eQTLs for three SNPs reached genome-wide significance

Locus	SNP	LD SNP	r^2	Tissue	Gene	P
8p23.1	rs2978951	rs2977793	0.80	Blood	DEFA4*	3.90E-05
8p23.1	rs2978951	rs2978951	1.00	Endometrial Tumors	PIGC, MPP3, C20orf173, LOC100289848	6.80E-10
8p23.1	rs2738058	rs2738058	1.00	Blood	DEFA4*	1.70E-03
19q13.41	rs4284742	rs4284742	1.00	Whole Blood	SIGLEC5*	7.72E-14

*Gene resides within topological associated domain (TAD) boundaries of the corresponding LD SNP

genetic variants and environmental factors. In this view, AgP, the most severe and early-onset form of PD, would be affected to a higher degree by genetic risk factors, whereas we consider that CP, which is characterized by a late age of disease-onset, would lack variants with high penetrance, but would be affected stronger by the accumulating negative effects of environmental factors and ageing. Yet, both disease phenotypes, which essentially differ in the speed of progression but not in the clinical picture of alveolar bone loss, share genetic and environmental risk factors. We aimed to elucidate the underlying genetic risk factors of AgP to enhance the current etiological concept.

The risk variant rs2738058, located at the intergenic region that separates the antimicrobial peptides *DEFA1* and *DEFA4* was suggested as a risk allele for CP in a GWAS before (6). These genes belong to the family of alpha defensins that cluster on chromosome 8 and likely play a role in phagocyte-mediated host defense against bacteria, fungi and viruses. The genes *DEFA1* and *DEFA3* are highly copy-variable and differ only by a single base substitution in the coding sequence. They seem to be interchangeable occupants of a 19 kb long copy-variable repeat unit, with both *DEFA1* and *DEFA3* gene number showing variation (17). For this reason, the composite designation *DEFA1A3* was suggested (13). Additionally, some pseudogenes were annotated at this region, but their biological function has yet remained unknown. It is believed that the variability in defensin gene copy numbers could contribute to differences in individual resistance to infections (18). *DEFA1A3* and *DEFA4* are strongly expressed in neutrophils and macrophages, as well as in the epithelia of mucosal surfaces. eQTL analysis suggested a regulatory effect of the rs2738058 on *DEFA4*. However, further studies are needed to characterize the proposed regulatory role of rs2738058 on *DEFA4* and *DEFA1A3* expression and also the context of the observed association with copy number variation.

SIGLEC5 (CD170) is a member of the human CD33-related siglecs and is broadly expressed in various myeloid cells of the innate immune system and in B cells (19). Siglec 5 is classified as an inhibitory receptor with the potential to mediate tyrosine phosphatases SHP-1 and -2 dependent signaling (14). It was suggested that the interaction of Siglec 5 and SHP-1 and -2 could be important in maintaining leukocytes in the quiescent state until activation is triggered via appropriate receptors. Accordingly, SIGLEC5 seems to modulate the activation of myeloid cells to prevent inappropriate reactivity against self-tissues, as recently reviewed (20). The ability to distinguish foreign pathogens from self and to make an appropriate response is essential to avoid bystander damage to host cells. In this context, we point out that an important feature in the pathogenesis of AgP are repeating periods of severe inflammation, which eventually lead to extensive resorption of the alveolar bone. The inflammation often is not accompanied by obvious signs of a particular pathogenic burden such as increased oral plaque, which could reflect an aberrant reactivity of the immune system to yet unknown triggers or intolerance to resident oral bacteria.

The homeostasis of the skeletal bone depends on the balance between bone formation by osteoblasts and bone resorption by osteoclasts. Interestingly, SIGLEC5 was shown to efficiently inhibit FCεRI-mediated activation (14). *FCER1G* (Fc Fragment of IgE Receptor Ig) encodes the gamma subunit of FCεRI and was among the 17 loci, which showed association with AgP with $P < 10^{-06}$ in our GWAS. FCεRI is the high-affinity receptor for the Fc region of immunoglobulin E (IgE), found on epidermal Langerhans cells, mast cells, eosinophils and basophils, and is part of signaling cascades that induce pro-inflammatory cytokine expression (21). It was shown *in vitro* that the osteoclast-associated receptor OSCAR associates with FCER1G, and in the mouse model it was experimentally demonstrated that mice lacking FCER1G exhibit severe osteopetrosis owing to impaired osteoclast differentiation (22). From these studies, it was concluded that FCER1G is an essential constituent for the maintenance of bone homeostasis and it is essential for differentiation of osteoclasts. Another recent study demonstrated that OSCAR is involved in activation of human dendritic cells (23) and enhances the proinflammatory responses of human monocytes and neutrophils (24). In this context, it is tempting to speculate that FCER1G may couple bone resorption to signals of the immune system.

We also observed SNP rs6887423, located downstream to the glutamate transporter gene *SLC1A3* (*GLAST-1*), to be strongly associated with AgP in the German-Dutch discovery sample and to be associated with borderline significance with the small Turkish AgP sample. This SNP showed no association with the CP sample. Larger replication studies are needed to give clear statistical evidence for this association. However, *SLC1A3* is an interesting candidate. Glutamate signaling is implicated in the bone in transmission and response to mechanical stimuli of the environment (25). *SLC1A3* is constitutively expressed in bone forming osteoblasts and in osteocytes and it is inhibited by mechanical loading of these cells (26,27). It is further genome-wide associated with vertebral osteoporosis phenotypes (28).

It is likely that the limitations of the statistical power in our Turkish AgP sample and the CP sample resulted in inability to replicate all relevant signals and that we missed true positive signals. Larger replication samples are needed to fully validate all the signals that were suggested by our explorative study.

In conclusion, of the 17 loci, which showed associations with AgP at a significance level of $P < 10^{-05}$ in the explorative study, *DEFA1A3* and *SIGLEC5* are associated at a genome-wide significance level in the pooled AgP and CP samples, providing statistical evidence for the relevance of these loci in the etiology of PD. We further suggest the *FCER1G* and *SLC1A3* as promising candidates for future association studies in larger case-control samples.

Materials and Methods

Study population

We give a detailed description of the study populations that were included in the discovery stage in Supplementary Material, Tables S10 and S11.

AgP cases and controls. In the discovery stage, we used a German and a Dutch case-control sample of AgP. The German sample consisted of 717 AgP cases (302 males and 415 females) who were recruited across Germany by the biobank Popgen (29), University-Hospital Schleswig-Holstein, Germany, and previously described (30). The German control sample consisted of 4,213 individuals (2,018 males, 2,191 females, four unknown) from North- and West-Germany. They were recruited from the Competence Network "FoCus - Food Chain Plus" (31), the Dortmunder Gesundheitsstudie - DOGS (32) and the Heinz Nixdorf Recall Studies 1-3 (HNR1-3) (33). The Dutch sample consisted of 179 AgP cases who were drawn from across The Netherlands and previously described (30). The Dutch control sample consisted of 2,891 (1,453 males 1,438 females), being individuals from the B-Proof Study (34). The Turkish AgP sample consisted of 220 cases, which we recruited across Turkey. This sample, along with 75 controls, which we recruited at the same locations, was previously described (30); the controls were described in (35).

CP cases and controls. The German CP cases were described in detail in (6). In brief, subjects within the first vs. the third tercile of the study population based on the number of proximal sites with AL $\geq$4 mm were opposed. Individuals aged >60 years were manually excluded from the analysis to increase the statistical power by enriching the case sample with early-onset phenotypes, which are generally considered to be caused to a larger extent by genetic risk factors. Based on these criteria, the sample consisted of 993 CP cases and 1,419 controls, which were genotyped either with the Affymetrix Genome-Wide Human SNP Array 6.0 or the Illumina Human Omni 2.5 array.

Statistical power calculation. The statistical power was calculated with the PS Power and Sample Size Program (36). With the given sample sizes, the probability of exposure among controls is 20% and if the true OR for disease in exposed subjects relative to unexposed subjects is 1.3, we were able to reject the null hypothesis that the OR = 1 with a probability (power) = 0.93 for AgP and 0.75 of for CP.

Ethics statement

The AgP studies have been approved by the ethics commission of the medical faculty of the Christian-Albrechts-Universität zu Kiel, Germany (Vote for cases: A 156/03, Vote for controls: B 231/98). The protocol of the CP study has been approved by the medical ethics committee of the University of Greifswald. Oral and written informed consents were obtained. All clinical investigation has been conducted according to the principles expressed in the Declaration of Helsinki.

Genotyping

Genotyping of the GER and NL AgP cases and controls was performed on an IScan system at the Institute of Clinical Molecular Biology (ICMB), Christian-Albrechts-University (CAU) Kiel, Germany, according to manufacturer's instructions. The AgP cases and controls were genotyped on Omni Bead Chips (Illumina, USA). The specific type of bead chip, which was used for each sample is summarized in Supplementary Material, Table S10.

Genotyping of the Turkish AgP cases and of 75 controls was carried out with iPLEX assays (Sequenom, San Diego, CA, USA), which were designed by the AssayDesigner v.3.1 software (Sequenom). Assay details are available from the authors upon request. Genotyping was performed at the ICMB, CAU Kiel, using Sequenom MassARRAY iPLEX GOLD chemistry according to standard protocols. SpectroCHIPs were analyzed in automated mode by a MassArray MALDI-TOF Compact System 2 with a solid-phase laser mass spectrometer (Bruker Daltonics, Bremen, Germany), called with the real-time SpectroCaller algorithm, and analyzed using the Sequenom SpectroTyper analyzer software v.4.0.20.65 (Sequenom, San Diego, CA, USA). SNPs rs4970469 (nearest genes *NUDC* and *TRNP1*) and rs10238400 (nearest genes *RPL10P11* and *GNAI1*, were genotyped on a 7900HT Fast Real-Time PCR System at the ICMB, CAU Kiel, on 384-well plates using TaqMan assays hCV3230478 and hCV11216204, respectively (Thermo Fisher Scientific, USA), as recently described (37).

The Turkish GWAS controls (N = 489) were genotyped on HumanOmni1-Quad BeadChip human array (Illumina, USA) as described before (35).

Quality control and filtering

Prior to the GWAS analysis, we applied several quality control (QC) steps on the GER and NL AgP cases and controls using the software PLINK v1.07 (38), which reduced the number of SNPs and individuals as shown in the detailed summary of the QC and filtering steps in Supplementary Material, Table S4. Firstly, we imputed sex for individuals with missing sex information and performed a sex check on the remaining individuals. Due to unclear imputation and sex check results eleven, three, and one individuals were excluded from DOGS, HNR-2 and NL-Cont, respectively. We excluded SNPs from the sex chromosomes. The remaining SNPs were lifted over to the human reference genome version GRCh37/hg19. SNP identifiers were updated to the current reference SNP ID (rsID) using the Ensembl Variation Database v84 (39)sub-cohort separately, we excluded SNPs with minor allele frequency (MAF) < 1%, deviations (P value $< 10^{-4}$) from Hardy-Weinberg equilibrium (HWE) or genotype call rates < 98%, in the respective order. Next, we used 1000G Phase 3 data of Northern Europeans from Utah (1000GP3-CEU) to flip all genotypes on the forward strand (40). SNPs with unclear strand direction were removed. A threshold of > 98% for the genotyping call rate per individual resulted in a loss of eleven cases and one control in the German AgP cases, and two cases and 15 controls in Dutch sample. A threshold of > 98% for the genotyping call rate per individual resulted in a loss of 11 cases and one control in the German AgP cases, and two cases and 15 controls in Dutch sample.

Lastly, we investigated population stratification by calculating pairwise identity-by-state (IBS) and identity-by-descent (IBD) estimates on a subset of SNPs (100% genotype call rate and MAF > 5%). 145 (22 cases and 43 controls in GER, four cases and 75 controls in NL duplicates and individuals with familial relationships were removed using the constraint "PI_HAT $\geq$ 0.125 or DST > 0.8" (according to PLINK, PI_HAT is defined as Proportion IBD, i.e. P[IBD = 2]+0.5*P[IBD = 1], DST is defined as IBS distance, i.e. [IBS2 + 0.5×IBS1]/[IBS0 + IBS1 × IBS2]) such that the loss of individuals was minimized. Subsequently, we applied a multidimensional scaling (MDS) and 107 outliers (4 cases and 6 controls

in GER, 4 cases and 93 controls in NL) were manually removed by visualizing the first eight dimensions in a scatterplot matrix (Supplementary Material, Figs S1 and S2). In addition, we evaluated the effects of population stratification for GER and NL separately, by using the additive genetic model and the genomic inflation factor (λ) function of the R package GenABEL (41). This resulted in $\lambda = 1.08$ for GER and $\lambda = 1.01$ for NL, suggesting the population stratification was only negligible for NL. Therefore, we adjusted for MDS components of the filtered cohort and preformed association tests starting with MDS components 1, then 1 and 2; 1, 2 and 3; etc., as covariates. Examining the first ten MDS components, we observed that components 1 to 6 resulted in the lowest genomic inflation factor ($\lambda = 1{,}04$). This combination was assigned as additional covariates for GER.

Genotype imputation

Post QC genotypes were converted from Plink BED/BIM/FAM format to Oxford Gen/Sample format with GTOOL v0.7.5 in order to estimate the haplotypes with the software ShapeIt v2 (42). Then the imputation of ungenotyped SNPs was carried out for the resulting haplotypes using the 1000G Phase 3 data and Impute v2 (43).

Association testing and meta-analysis

Associations of SNPs with AgP in both GER and NL were tested with SNPTEST v2.5.2 using an additive model (with 1 degree of freedom) with the binary variables sex and smoking status (never smoked = 0, ever smoked = 1) as covariates (44). For GER we additionally adjusted for the first 6 MDS components. SNPs in the output were excluded by imputation quality (column frequentist_add_info) < 0.9, average maximal posterior probability (column average_maximum_posterior_call) < 0.9, MAF < 5% in cases or controls and deviation from HWE (P value $< 10^{-4}$) in controls. A total of 6,568,093 and 6,475,582 (6,374,298 intersecting) genotyped and imputed autosomal SNPs from GER and NL were then combined in a discovery meta-analysis. Level of statistical heterogeneity was measured using Cochran's Q and the I-square (I^2) statistic (45). For the discovery meta-analysis as well as subsequent meta-analyses using validation and replication samples, SNPs having P(Q) < 0.05 and $I^2 > 0.5$, i.e. a high amount of heterogeneity, were meta-analyzed with the random effects model, for the remaining SNPs the fixed effects model was used. The potential impact of population stratification as well as the overall significance of the GWAS results were evaluated using quantile-quantile plots and genomic inflation factors (λ).

Selection of SNPs for validation and replication

We selected SNPs to validate their association with CP, which complied with the following criteria in the combined AgP case-control sample: (1) combined P value (GER + NL; $P_{meta} < 10^{-5}$ unadjusted for MDS components (we preassigned this significance threshold as being indicative of suggestive evidence of association with disease (46); (2) $P_{meta} < P_{ger}$ and $P_{meta} < P_{nl}$; (3) SNPs within strong linkage disequilibrium (LD; $r^2 > 0.8$ in 1000GP3-EUR) show association with AgP ($P_{meta} < 10^{-3}$) in the combined sample. (4) SNPs < 500 kb distance to each other among the remaining SNPs were clumped into linked loci and the SNP with the lowest P value was kept.

Imputed genotype information was lacking in the validation cohort of CP for seven SNPs, and we attempted to replace these seven with the best surrogates based on linkage disequilibrium (LD) using genotypes of the European reference population (EUR) of 1000 genomes. Lead SNP rs3830904 was replaced by rs2144815, which showed complete LD ($r^2 = 1$). Lead SNP rs62312993 was replaced by the neighboring SNP rs11727861, which showed almost complete LD ($r^2 = 0.99$). Lead SNPs rs11633566 and rs6137428 were replaced by rs308833 and rs6047560, respectively, which showed strong but not complete LD ($r^2 = 0.74$ and $r^2 = 0.73$, respectively). For two SNPs, rs11382193 and rs7595654, no tagging SNPs in complete or strong LD could be selected from our GWAS data. Accordingly, these SNPs were not tested further.

In the replication cohort of AgP-TUR 489 controls were taken from a previous GWAS dataset, which was not imputed and therefore missed five SNPs at the selected chromosomal regions. To compensate this, we replaced four missing lead SNPs that were tested in AgP-GER-NL and CP by neighboring SNPs with strong or complete LD, based on genotypes of the EUR reference population: SNPs rs1380780, rs17340482, rs4791202, rs747804 were replaced by rs1380781 [$r^2 = 1$], rs4970469 [$r^2 = 0.97$], rs10238400 [$r^2 = 0.9$] and rs13283964 [$r^2 = 1$], respectively. For SNP rs6887423 at *SLC1A3*, a tagging SNP in complete linkage was not available in the Turkish GWAS control sample, and therefore it was replaced by the closest available linked SNP, rs1122900, which showed moderate LD ($r^2 = 0.65$).

eQTL mapping

For the examination of putative *cis*- and *trans*-regulatory effects of the 17 distinct haplotype blocks that suggested association with AgP, the expression quantitative trait locus (eQTL) mapping databases Genotype-Tissue Expression project (GTEx) (47), Haploreg v4.1 (48), GRASP v2 (49) and Blood eQTL Browser (50) were used.

Relationship with other traits analyses

For each of the 16 loci in our top hit list we extracted all traits within position of the top SNP +/- 200kb from the NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog, January 2017) (51). In a LD score regression analysis, the genetic correlation between adjusted summary statistics of AgP-GER-NL and 219 other traits was computed using LD Hub (52).

Supplementary Material

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

Funding

The German Research Foundation DFG (Deutsche Forschungsgemeinschaft; GZ: SCHA 1582/3-1) supported this study. Collection of the AgP cases was additionally supported by the German Ministry of Education and Research through the POPGEN biobank project (01GR0468). The Dortmund Health Study (DHS) is 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

(collection of sociodemographic and clinical data). Blood collection in the Dortmund Health Study was done through funds from the Institute of Epidemiology and Social Medicine University of Muenster and genotyping supported by the German Ministry of Research and Education (BMBF, grant no. 01ER0816). FOCUS was supported by the Federal Ministry of Education and Research BMBF (FKZ 0315540A). The HNR study is supported by the Heinz Nixdorf Foundation (Germany). Additionally, the HNR study was funded by the German Ministry of Education and Science and the German Research Council (DFG; Project SI 236/8-1, SI236/9-1, ER 155/6-1). The genotyping of the Illumina HumanOmni-1 Quad BeadChips of the HNR subjects was financed by the German Centre for Neurodegenerative Diseases (DZNE), Bonn. SHIP is part of the Community Medicine Research net (CMR, http://www.community-medicine.de) of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). This project was granted by BMBF-01-ZZ-9603/0 and BH was supported by GABA, Switzerland. Generation of genome-wide SNP data has been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg, West Pomerania. The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH. This research was financially supported by the Turkish Scientific and Research Council (TUBITAK) (Grant no: 114S543).

References

1. Marcenes, W., Kassebaum, N.J., Bernabe, E., Flaxman, A., Naghavi, M., Lopez, A. and Murray, C.J. (2013) Global burden of oral conditions in 1990-2010: a systematic analysis. *J. Dental Res.*, **92**, 592–597.
2. 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.
3. Susin, C., Haas, A.N. and Albandar, J.M. (2014) Epidemiology and demographics of aggressive periodontitis. *Periodontology 2000*, **65**, 27–45.
4. Schaefer, A.S., Richter, G.M., 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**, 553–562.
5. Hashim, N.T., Linden, G.J., Ibrahim, M.E., Gismalla, B.G., Lundy, F.T., Hughes, F.J. and El Karim, I.A. (2015) Replication of the association of GLT6D1 with aggressive periodontitis in a Sudanese population. *J. Clin. Periodontol.*, **42**, 319–324.
6. Teumer, A., Holtfreter, B., Volker, U., Petersmann, A., Nauck, M., Biffar, R., Volzke, H., Kroemer, H.K., Meisel, P., Homuth, G. *et al.* (2013) Genome-wide association study of chronic periodontitis in a general German population. *J. Clin. Periodontol.*, **40**, 977–985.
7. 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**, 555–561.
8. Hong, K.W., Shin, M.S., Ahn, Y.B., Lee, H.J. and Kim, H.D. (2015) Genomewide association study on chronic periodontitis in Korean population: results from the Yangpyeong health cohort. *J. Clin. Periodontol.*, doi: 10.1111/jcpe.12437.
9. 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.
10. Divaris, K., Monda, K.L., North, K.E., Olshan, A.F., Reynolds, L.M., Hsueh, W.C., Lange, E.M., Moss, K., Barros, S.P., Weyant, R.J. *et al.* (2013) Exploring the genetic basis of chronic periodontitis: a genome-wide association study. *Hum Mol Genet*, **22**, 2312–2324.
11. 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**, 21S–28S.
12. Kasbohm, E., Holtfreter, B., Volker, U., Petersmann, A., Samietz, S., Biffar, R., Volzke, H., Meisel, P., Kacprowski, T., Homuth, G. *et al.* (2016) Exome variant analysis of chronic periodontitis in 2 large cohort studies. *J. Dent. Res.*, **96**, 73–80.
13. Aldred, P.M., Hollox, E.J. and Armour, J.A. (2005) Copy number polymorphism and expression level variation of the human alpha-defensin genes DEFA1 and DEFA3. *Hum. Mol. Genet.*, **14**, 2045–2052.
14. Avril, T., Freeman, S.D., Attrill, H., Clarke, R.G. and Crocker, P.R. (2005) Siglec-5 (CD170) can mediate inhibitory signaling in the absence of immunoreceptor tyrosine-based inhibitory motif phosphorylation. *J. Biol. Chem.*, **280**, 19843–19851.
15. Chen, C.H., Wang, S.H., Liu, C.H., Wu, Y.L., Wang, W.J., Huang, J., Hung, J.S., Lai, I.R., Liang, J.T. and Huang, M.C. (2014) beta-1,4-Galactosyltransferase III suppresses beta1 integrin-mediated invasive phenotypes and negatively correlates with metastasis in colorectal cancer. *Carcinogenesis*, **35**, 1258–1266.
16. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
17. Khan, F.F., Carpenter, D., Mitchell, L., Mansouri, O., Black, H.A., Tyson, J. and Armour, J.A. (2013) Accurate measurement of gene copy number for human alpha-defensin DEFA1A3. *BMC Genomics*, **14**, 719.
18. Linzmeier, R.M. and Ganz, T. (2005) Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. *Genomics*, **86**, 423–430.
19. Crocker, P.R. and Varki, A. (2001) Siglecs, sialic acids and innate immunity. *Trends Immunol.*, **22**, 337–342.
20. Crocker, P.R., Paulson, J.C. and Varki, A. (2007) Siglecs and their roles in the immune system. *Nat. Rev. Immunol.*, **7**, 255–266.
21. von Bubnoff, D., Novak, N., Kraft, S. and Bieber, T. (2003) The central role of FcepsilonRI in allergy. *Clin. Exp. Dermatol.*, **28**, 184–187.
22. Koga, T., Inui, M., Inoue, K., Kim, S., Suematsu, A., Kobayashi, E., Iwata, T., Ohnishi, H., Matozaki, T., Kodama, T. *et al.* (2004) Costimulatory signals mediated by the ITAM motif cooperate with RANKL for bone homeostasis. *Nature*, **428**, 758–763.
23. Merck, E., Gaillard, C., Gorman, D.M., Montero-Julian, F., Durand, I., Zurawski, S.M., Menetrier-Caux, C., Carra, G., Lebecque, S., Trinchieri, G. *et al.* (2004) OSCAR is an

FcRgamma-associated receptor that is expressed by myeloid cells and is involved in antigen presentation and activation of human dendritic cells. *Blood*, **104**, 1386–1395.
24. Merck, E., Gaillard, C., Scuiller, M., Scapini, P., Cassatella, M.A., Trinchieri, G. and Bates, E.E. (2006) Ligation of the FcR gamma chain-associated human osteoclast-associated receptor enhances the proinflammatory responses of human monocytes and neutrophils. *J. Immunol.*, **176**, 3149–3156.
25. Mason, D.J. (2004) Glutamate signalling and its potential application to tissue engineering of bone. *Eur. Cells Mater.*, **7**, 12–25. discussion 25-16.
26. Mason, D.J., Suva, L.J., Genever, P.G., Patton, A.J., Steuckle, S., Hillam, R.A. and Skerry, T.M. (1997) Mechanically regulated expression of a neural glutamate transporter in bone: a role for excitatory amino acids as osteotropic agents? *Bone*, **20**, 199–205.
27. Mason, D.J. and Huggett, J.F. (2002) Glutamate transporters in bone. *J. Musculoskelet.Neuronal Interact.*, **2**, 406–414.
28. Nielson, C.M., Liu, C.T., Smith, A.V., Ackert-Bicknell, C.L., Reppe, S., Johanna, J., Wassel, C., Register, T.C., Oei, L., Alonso Lopez, N. *et al.* (2016) Novel genetic variants are associated with increased vertebral volumetric BMD, reduced vertebral fracture risk, and increased expression of SCL1A3 and EPHB2. *J. Bone Miner. Res.*, **31**, 2085–2097.
29. 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.
30. 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.
31. 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.
32. Berger, K. (2012) [DHS: The Dortmund health study]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, **55**, 816–821.
33. 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.
34. van Wijngaarden, J.P., Dhonukshe-Rutten, R.A., van Schoor, N.M., van der Velde, N., Swart, K.M., Enneman, A.W., van Dijk, S.C., Brouwer-Brolsma, E.M., Zillikens, M.C., van Meurs, J.B. *et al.* (2011) Rationale and design of the B-PROOF study, a randomized controlled trial on the effect of supplemental intake of vitamin B12 and folic acid on fracture incidence. *BMC Geriatrics*, **11**, 80.
35. Renauer, P.A., Saruhan-Direskeneli, G., Coit, P., Adler, A., Aksu, K., Keser, G., Alibaz-Oner, F., Aydin, S.Z., Kamali, S., Inanc, M. *et al.* (2015) Identification of Susceptibility Loci in IL6, RPS9/LILRB3, and an Intergenic Locus on Chromosome 21q22 in Takayasu Arteritis in a Genome-Wide Association Study. *Arthritis Rheumatol.*, **67**, 1361–1368.
36. Dupont, W.D. and Plummer, W.D. Jr. (1998) Power and sample size calculations for studies involving linear regression. *Control Clin. Trials*, **19**, 589–601.
37. 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.
38. 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.
39. Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S. *et al.* (2015) Ensembl 2015. *Nucleic Acids Res.*, **43**, D662–D669.
40. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A. *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
41. Aulchenko, Y.S., Ripke, S., Isaacs, A. and van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, **23**, 1294–1296.
42. Delaneau, O., Marchini, J. and Zagury, J.F. (2012) A linear complexity phasing method for thousands of genomes. *Nat. Methods*, **9**, 179–181.
43. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
44. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
45. Higgins, J., Thompson, S., Deeks, J. and Altman, D. (2002) Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J. Health Serv. Res. Policy*, **7**, 51–61.
46. Erdmann, J., Grosshennig, A., Braund, P.S., Konig, I.R., Hengstenberg, C., Hall, A.S., Linsel-Nitschke, P., Kathiresan, S., Wright, B., Tregouet, D.A. *et al.* (2009) New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat. Genet.*, **41**, 280–282.
47. Consortium, G.T. (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648–660.
48. Ward, L.D. and 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**, D930–D934.
49. Eicher, J.D., Landowski, C., Stackhouse, B., Sloan, A., Chen, W., Jensen, N., Lien, J.P., Leslie, R. and Johnson, A.D. (2015) GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res.*, **43**, D799–D804.
50. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E. *et al.* (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.*, **45**, 1238–1243.

51. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
52. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Early, G. *et al.* (2017) LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, **33**, 272–279.