

# SCIENTIFIC REPORTS

OPEN

## Genome-wide association meta-analysis of coronary artery disease and periodontitis reveals a novel shared risk locus

Matthias Munz<sup>1,2</sup>, Gesa M. Richter<sup>1</sup>, Bruno G. Loos<sup>4</sup>, Søren Jepsen<sup>5</sup>, Kimon Divaris<sup>6,7</sup>, Steven Offenbacher<sup>8</sup>, Alexander Teumer<sup>9</sup>, Birte Holtfreter<sup>10</sup>, Thomas Kocher<sup>10</sup>, Corinna Bruckmann<sup>11</sup>, Yvonne Jockel-Schneider<sup>12</sup>, Christian Graetz<sup>13</sup>, Loreto Munoz<sup>2,3,14</sup>, Anita Bhandari<sup>2,3,14</sup>, Stephanie Tennstedt<sup>2,3,14</sup>, Ingmar Staufienbiel<sup>15</sup>, Nathalie van der Velde<sup>16,17</sup>, André G. Uitterlinden<sup>16</sup>, Lisette C. P. G. M. de Groot<sup>18</sup>, Jürgen Wellmann<sup>19</sup>, Klaus Berger<sup>19</sup>, Bastian Krone<sup>20</sup>, Per Hoffmann<sup>21,22</sup>, Matthias Laudes<sup>23</sup>, Wolfgang Lieb<sup>24</sup>, Andre Franke<sup>24</sup>, Henrik Dommisch<sup>1</sup>, Jeanette Erdmann<sup>2,3,14</sup> & Arne S. Schaefer<sup>1</sup>

Evidence for a shared genetic basis of association between coronary artery disease (CAD) and periodontitis (PD) exists. To explore the joint genetic basis, we performed a GWAS meta-analysis. In the discovery stage, we used a German aggressive periodontitis sample (AgP-Ger; 680 cases vs 3,973 controls) and the CARDIoGRAMplusC4D CAD meta-analysis dataset (60,801 cases vs 123,504 controls). Two SNPs at the known CAD risk loci *ADAMTS7* (rs11634042) and *VAMP8* (rs1561198) passed the pre-assigned selection criteria ( $P_{\text{AgP-Ger}} < 0.05$ ;  $P_{\text{CAD}} < 5 \times 10^{-8}$ ; concordant effect direction) and were replicated in an independent GWAS meta-analysis dataset of PD (4,415 cases vs 5,935 controls). SNP rs1561198 showed significant association (PD[Replication]:  $P = 0.008$  OR = 1.09, 95%

<sup>1</sup>Charité – University Medicine Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute for Dental and Craniofacial Sciences, Department of Periodontology and Synoptic Dentistry, Berlin, Germany. <sup>2</sup>Institute for Cardiogenetics, University of Lübeck, 23562, Lübeck, Germany. <sup>3</sup>DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, Lübeck, 23562, Germany. <sup>4</sup>Department of Periodontology and Oral Biochemistry, Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam and Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. <sup>5</sup>Department of Periodontology, Operative and Preventive Dentistry, University of Bonn, Bonn, Germany. <sup>6</sup>Department of Pediatric Dentistry, School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill, USA. <sup>7</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, USA. <sup>8</sup>Department of Periodontology, School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill, USA. <sup>9</sup>Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. <sup>10</sup>Unit of Periodontology, Department of Restorative Dentistry, Periodontology, Endodontology, Preventive Dentistry and Pedodontics, Dental School, University Medicine Greifswald, Greifswald, Germany. <sup>11</sup>Department of Conservative Dentistry and Periodontology, Medical University Vienna, School of Dentistry, Vienna, Austria. <sup>12</sup>Department of Periodontology, Clinic of Preventive Dentistry and Periodontology, University Medical Center of the Julius-Maximilians-University, Würzburg, Germany. <sup>13</sup>Department of Conservative Dentistry, Unit of Periodontology, University Medical Center Schleswig-Holstein, Campus Kiel, Germany. <sup>14</sup>University Heart Center Luebeck, 23562, Lübeck, Germany. <sup>15</sup>Department of Conservative Dentistry, Periodontology and Preventive Dentistry, Hannover Medical School, Hannover, Germany. <sup>16</sup>Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands. <sup>17</sup>Department of Internal Medicine section of Geriatrics, Amsterdam Medical Center, Amsterdam, The Netherlands. <sup>18</sup>Wageningen University, Division of Human Nutrition, Wageningen, The Netherlands. <sup>19</sup>Institute of Epidemiology and Social Medicine, University Münster, Münster, Germany. <sup>20</sup>Institute of Medical Informatics, Biometry and Epidemiology, University Clinic Essen, Essen, Germany. <sup>21</sup>Institute of Human Genetics, University of Bonn, Bonn, Germany. <sup>22</sup>Human Genomics Research Group, Department of Biomedicine, University Hospital of Basel, Basel, Switzerland. <sup>23</sup>Department of Medicine 1, University of Kiel, Kiel, Germany. <sup>24</sup>Institute of Epidemiology, Christian-Albrechts-University, Kiel, Germany. Jeanette Erdmann and Arne S. Schaefer contributed equally. Correspondence and requests for materials should be addressed to A.S.S. (email: [arne.schaefer@charite.de](mailto:arne.schaefer@charite.de))

Received: 15 December 2017

Accepted: 31 August 2018

Published online: 12 September 2018

CI = [1.02–1.16]; PD [Discovery + Replication]:  $P = 0.0002$ , OR = 1.11, 95% CI = [1.05–1.17]). For the associated haplotype block, allele specific *cis*-effects on *VAMP8* expression were reported. Our data adds to the shared genetic basis of CAD and PD and indicate that the observed association of the two disease conditions cannot be solely explained by shared environmental risk factors. We conclude that the molecular pathway shared by CAD and PD involves *VAMP8* function, which has a role in membrane vesicular trafficking, and is manipulated by pathogens to corrupt host immune defense.

Coronary artery disease (CAD) has a well-established genetic basis and a strong inflammatory component (reviewed in<sup>1</sup>). An association between CAD and the common oral inflammatory disease condition, periodontitis (PD), was reported in several clinical and observational studies (previously reviewed<sup>2</sup>). Because of the high prevalence of CAD and PD this association is potentially of public health importance. Recent evidence indicates that the observed association between CAD and PD is independent of smoking<sup>2</sup> and obesity<sup>3</sup>. However, it could be explained in parts by other shared risk factors like diabetes, and age. In this context, the knowledge of shared genetic risk variants could substantially contribute to the understanding of the mechanisms that underlie the epidemiological associations.

Recently, we demonstrated shared associations of two CAD risk loci, *ANRIL*<sup>4–7</sup> and *PLG*<sup>8,9</sup>, with aggressive periodontitis (AgP), a severe early-onset form of PD. In addition, we observed suggestive evidence for shared association of a rare genetic variant at *VAMP3*<sup>10</sup>, located at a chromosomal region that was earlier described to be associated with increased colonization of oral periodontal pathogens<sup>11</sup>. AgP, which has a global prevalence of 0.1%<sup>12</sup>, is characterized by a particular early age of disease onset (<35 years of age), which is why patients with AgP generally do not suffer from late-onset diseases like CAD or diabetes. By contrast, chronic periodontitis (CP), a widespread form of PD, has a prevalence of >9% in adults aged  $\geq 30$  years<sup>13</sup> and is mainly observed in the elderly. AgP and CP have a similar histopathology and can be considered as parts of the same disease spectrum, which attribute to the effects of different combinations of genetic risk loci that determine the individual immune response. Because AgP cases have a high heritability as reflected by the early-age of disease onset and are not confounded by risk factors such as diabetes and age, AgP is particularly suitable to explore the shared genetic basis of CAD and PD<sup>14</sup>. The identified variants can be subsequently validated for the relevance in the more moderate form of CP.

In the current study, we performed a meta-analysis using data from the “Coronary Artery Disease Genome-wide Replication and Meta-analysis plus The Coronary Artery Disease” (CARDIoGRAMplusC4D) consortium and based on genome-wide association studies (GWAS) datasets for AgP and CP. We provide evidence for the shared association of variant rs1561198 with CAD and PD, which has reported *cis*-effects on the expression of the adjacent gene *VAMP8*.

## Materials and Methods

**Participating studies.** The discovery sample consisted of a GWAS of German AgP cases and controls, which was previously described<sup>15</sup>, and of a case-control GWAS meta-analysis of CAD from the CARDIoGRAMplusC4D consortium<sup>16</sup>. The replication was carried out with GWA studies of case-control samples of Dutch AgP<sup>15</sup> and CP case-control samples of German<sup>17</sup> and of European American descent<sup>18</sup>. Summary statistics of all studies were based on the additive model.

The German AgP sample (AgP-Ger) included 680 cases and 3,973 controls. Cases were recruited across Germany by the biobank Popgen<sup>19</sup>, University-Hospital Schleswig-Holstein, Germany. Controls originate from North- and West-Germany and were recruited from the Competence Network “FoCUS - Food Chain Plus”<sup>20</sup>, the Dortmunder Gesundheitsstudie – DOGS<sup>21</sup> and the Heinz Nixdorf Recall Studies 1–3<sup>22</sup>. Genotyping of the cases was performed on Illumina Omni Bead Chips and the imputation was based on the 1000 Genomes Phase 3 reference panel<sup>23</sup>.

The dataset of the CARDIoGRAMplusC4D consortium included an assembly of 60,801 cases and 123,504 controls from 48 studies, mainly of European, South Asian East Asian descent<sup>16</sup>. Genotyping was carried out using Affymetrix or Illumina platforms. Subsequent genotype imputation was based on the 1000 Genomes phase 1 version 3 reference. The meta-analysis was performed by using either the fixed-effects model or the random-effects model, depending on the level of statistical heterogeneity.

The Dutch AgP sample (AgP-NL) consisted of 171 cases of Dutch descent that were recruited from the ACTA (Academisch Centrum Tandheelkunde Amsterdam) and of 2,607 population representative controls from the B-Proof Study<sup>24</sup>, which were collected at Rotterdam and Wageningen. Genotyping of the cases and imputation was performed together with AgP-Ger.

The German CP (CP-Ger) sample consisted of 993 cases and 1,419 controls from two independent cross-sectional studies SHIP and SHIP-TREND, recruited at the University Medicine Greifswald<sup>25–27</sup>. Cases and controls were defined by contrasting subjects within the first vs the third tertile of proportion of proximal sites with attachment loss (AL)  $\geq 4$  mm. To increase the statistical power by enriching the case sample with early-onset phenotypes, subjects aged >60 years were excluded. Cases and controls were genotyped either with the Affymetrix Genome-Wide Human SNP Array 6.0 or the Illumina Human Omni 2.5 array and imputed on the 1000 Genomes Phase 1 reference.

The European American CP (CP-EA) sample included 958 severe (sev) CP cases, 2,293 moderate (mod) CP cases and 1,909 controls from the Atherosclerosis Risk in Communities (ARIC) Study<sup>28</sup>. Genotyping was carried out using the Affymetrix Genome-Wide Human SNP Array 6.0 and the subsequent genotype imputation was performed on the HapMap Phase II reference with individuals of Northern and Western European (CEU) ancestry.

**Discovery.** In the discovery stage, we compared variants in GWAS of AgP-Ger and CAD regarding their effect direction and significance level. At first, variants with a minor allele frequency (MAF)  $< 0.05$  were filtered out and only variants with genotypes available for both study samples were kept.

Then we selected variants showing the same effect direction in AgP-Ger and CAD and surpassing the pre-assigned restrictions  $P_{\text{AgP-GER}} < 0.05$  and  $P_{\text{CAD}} < 5 \times 10^{-8}$ . We chose these P-value thresholds to account for the huge difference in sample size (CAD sample  $\sim 35\times$  larger than AgP-Ger sample). Subsequently, we pruned the remaining variants using intermediate linkage disequilibrium. We chose a pruning threshold of  $r^2 \geq 0.2$  to account for association tails of variants arising from their low correlation with the truly associated haplotype block.

**Replication meta-analysis.** Variants that passed the discovery step were taken forward to replication stage and meta-analysed in additional PD GWAS samples AgP-NL, CP-Ger, CP-EA-mod and CP-EA-sev. By default, we applied the fixed effects model. However, for variants having a high amount of heterogeneity, i.e. where the P-value of Cochran's Q  $P(Q) < 0.05$  and heterogeneity index  $I^2 > 0.5$ , we applied the random effects model instead. Correction for multiple testing was performed using the method of Bonferroni. Additionally, we used the adjustment method for shared controls of Zaykin and Kozbur to check the P-value inflation when combining CP-EA-mod and CP-EA-sev<sup>29</sup>.

**Statistical power calculation for the discovery stage.** The statistical power of the AgP-Ger sample at a significance level of 0.05 was calculated with Genetic Association Study (GAS) Power Calculator<sup>30</sup>. When assuming an additive model and a prevalence rate of 0.1% for AgP in the general population<sup>12</sup>, risk variants with minor allele frequencies of  $> 0.4$  can be detected at a probability (power) of 0.8 if the genotype relative risk (GRR) is  $\sim 1.18$ . The GRR is an estimator for the OR.

**Functional annotation.** Variants were annotated using the Genepopper database (DB)<sup>31</sup>. Genepopper DB integrates data of many public sources by applying periodically executed extraction, transformation and loading (ETL) processes. Specifically, we used integrated datasets of linkage disequilibrium (LD), expression quantitative trait loci (eQTL) mappings, topologically associated domain boundaries (TADs) and GWAS studies to annotate our findings.

Identified loci were annotated using LD information (correlation measures  $r^2$  and  $D'$ ) from the European reference population (EUR) of 1000 Genomes Phase 3<sup>23</sup>. eQTL mapping information was gathered from Genotype-Tissue Expression project (GTEx)<sup>32</sup>, Haploreg v4.1<sup>33,34</sup>, GRASP v2.0<sup>35</sup>, GEUVADIS<sup>36</sup>, SCAN<sup>37</sup>, seeQTL<sup>38</sup> and Blood eQTL Browser<sup>39</sup>. Information about TAD boundaries was taken from Dixon *et al.*<sup>40</sup>. TADs are genomic regions defined by mainly cell-type independent interactome boundaries representing a spatial compartment in the genome. Physical interactions of regulatory DNA elements and gene promoters occur more frequently within a TAD. Thus, we refer to *cis* regulation, if a SNP with a putative effect on gene expression (expression SNP; eSNP) resides upstream or downstream from the gene but within the same TAD. In contrast, if a gene is located in a different TAD than the corresponding eSNP, we define this as a putative trans-regulatory effect, being most likely affected indirectly, e.g. via intermediate genes in a pathway. This dataset contained TADs with a length of  $\sim 853$  kilo base pairs (kb) in average (maximum length = 4.44 mega base pairs [mb], minimum [min] length = 0.8 kb). Variant consequence information was taken from Ensembl Variation DB and additionally, we annotated variants using combined annotation dependant depletion (CADD) score<sup>41,42</sup>. The CADD score is calculated by applying the formula  $-10 \times \log_{10}(\text{rank}/\text{total})$  on each variant in a ranking of single nucleotide variants which is created by combining several other variant annotation scores using machine learning techniques. Accordingly, a CADD score of  $\geq 5.22$  indicates that the variant belongs to the 20% most deleterious substitutions in the human genome and a score  $\geq 10$  indicates that the variant belongs to the 10% most deleterious substitutions. To elucidate the relationship to other traits and diseases, we extracted information about phenotype associations from the NHGRI-EBI Catalog<sup>43</sup>.

**Genetic risk score calculation.** In an additional analysis, we assessed the genetic relation of CAD and PD by calculating a genetic risk score (GRS) based on known CAD risk variants and their corresponding effect sizes in CAD and by applying this score to the sample of AgP-Ger and respective controls (Supplementary Table 1). The GRS was calculated per individual using the formula  $GRS = \sum w_x n_x$ , with  $w_x$  being the effect size (Odds ratio [OR]) and  $n_x$  being the number of risk alleles of variant  $x$ . GRS of cases and controls of AgP-Ger were then compared using basic statistics.

## Results

**Discovery meta-analysis.** The variant sets of AgP-Ger and CAD consisted of 6,416,838 and 9,455,779 variants. A total of 5,519,261 genetic variants were existent in both studies. In the discovery stage which included AgP-Ger and CAD, 276 variants in three distinct loci at 9p21.3, 15q25.1 and 2p11.2 surpassed our pre-assigned selection criteria (Supplementary Table 1).

Locus 9p21.3 reached the highest level of significance in PD with  $P = 1.23 \times 10^{-4}$  (OR = 1.26, 95% CI = [1.12–1.41]) and  $P = 1.56 \times 10^{-32}$  (OR = 1.21, 95% CI = [1.17–1.25]) for lead variant rs10116277 in AgP and CAD, respectively (Table 1 and Fig. 1). This single nucleotide polymorphism (SNP) is located in the intronic region of the large non-coding antisense RNA *CDKN2B-AS1* (alternative name *ANRIL*) which has been reported as risk factor for both AgP and CAD before<sup>7,44</sup>.

The lead variant of the second locus at 15q25.1 was SNP rs11634042, intronic of the gene *MORF4L1* (Mortality factor 4 like 1) and 21 kb upstream of *ADAMTS7* (ADAM metalloproteinase with thrombospondin type 1 motif, 7) with  $P = 5.67 \times 10^{-3}$  (OR = 1.18, 95% CI = [1.05–1.32]) and  $P = 2.17 \times 10^{-13}$  (OR = 1.08, 95% CI = [1.06–1.10]) for AgP and CAD.

	Lead Variant	Locus	Nearest Gene(s)	A1	A2	EAF Cas	EAF Con	Stage	OR [95% CI]	P
1	rs10116277	9p21.3	CDKN2B-AS1	T	G	0.52	0.46	AgP-Ger	1.26 [1.12–1.41]	1.23E-04
								CAD	1.21 [1.17–1.25]	1.56E-32
2	rs11634042	15q25.1	MORF4L1	C	T	0.60	0.56	AgP-Ger	1.18 [1.05–1.32]	5.67E-03
								CAD	1.08 [1.06–1.10]	2.17E-13
								AgP-NL	1.00 [0.80–1.25]	1
								CP-EA-mod	1.00 [0.90–1.10]	9.53E-01
								CP-EA-sev	1.04 [0.91–1.20]	5.34E-01
								CP-Ger	0.96 [0.85–1.09]	5.40E-01
								<b>Pooled (Replication)</b>	<b>1.00 [0.94–1.06]</b>	<b>9.50E-01</b>
<b>Pooled (PD)</b>	<b>1.04 [0.98–1.10]</b>	<b>1.99E-01</b>								
3	rs1561198	2p11.2	VAMP8 - VAMP5	T	C	0.48	0.44	AgP-Ger	1.19 [1.05–1.34]	4.57E-03
								CAD	1.06 [1.04–1.08]	6.37E-10
								AgP-NL	0.96 [0.77–1.20]	7.10E-01
								CP-EA-mod	1.09 [1.00–1.20]	5.81E-02
								CP-EA-sev	1.09 [0.95–1.24]	2.08E-01
								CP-Ger	1.13 [1.00–1.28]	5.86E-02
								<b>Pooled (Replication)</b>	<b>1.09 [1.02–1.16]</b>	<b>6.79E-03</b>
								<b>Pooled (PD)</b>	<b>1.11 [1.05–1.17]</b>	<b>2.02E-04</b>

**Table 1.** Three loci were identified in the discovery step (AgP-Ger, CAD): 9p21.3, 15q25.1 and 2p11.2. Since the locus at 9p21.3 is already a known shared risk locus for PD and CAD, we only took the loci at 15q25.1 and 2p11.2 forward to replication in PD (Pooled[Replication] = AgP-NL + CP-EA-mod + CP-EA-sev + CP-Ger). In the replication step, only locus 2p11.2 could be successfully replicated. At this locus, SNP rs1561198 showed the strongest association with PD in the overall PD sample (Pooled[PD] = AgP-Ger + AgP-NL + CP-EA-mod + CP-EA-sev + CP-Ger). A1 = Effect allele; A2 = Non-effect allele; EAF = Effect allele frequency; Cas = Cases; Con = Controls; OR = Odds ratio; CI = Confidence interval; P = P-value; CAD = Coronary artery disease; AgP = Aggressive periodontitis; CP = Chronic periodontitis; Ger = German; NL = Dutch; EA = European American; sev = severe; mod = moderate.

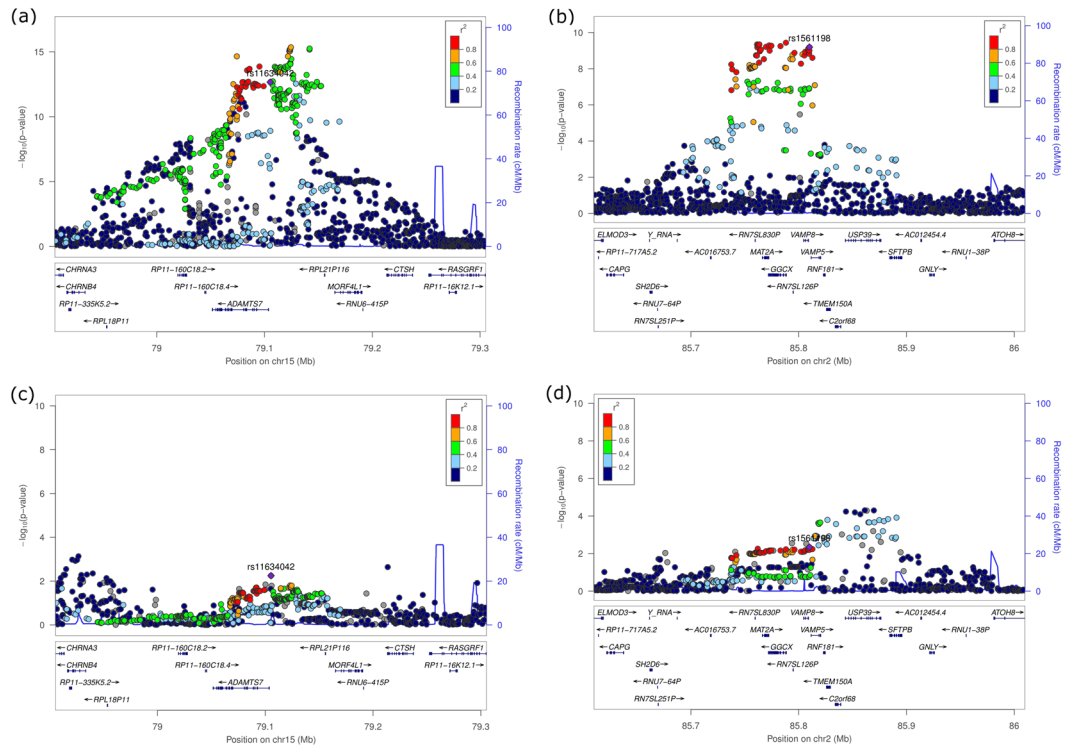
The third associated region at 2p11.2 showed association with  $P = 4.57 \times 10^{-3}$  (OR = 1.19, 95% CI = [1.05–1.34]) and  $P = 6.37 \times 10^{-10}$  (OR = 1.06, 95% CI = [1.04–1.08]) for variant rs1561198 in AgP-Ger and CAD, respectively. Variant rs1561198 is an intergenic SNP located 0.8 kb downstream of *VAMP8* (vesicle-associated membrane protein 8) and 1.5 kb upstream of *VAMP5* (vesicle-associated membrane protein 5). Also the loci at *MORF4L1* and *VAMP8* are already known susceptibility loci for CAD<sup>16,45</sup>, however they have not yet been reported to be associated with PD.

**Replication.** Of the three loci that we identified in the discovery meta-analysis, *ANRIL* had repeatedly been replicated as a genetic risk factor of PD<sup>7,46–48</sup>. Accordingly, we only selected the two novel suggestive risk loci at 15q25.1 and 2p11.2 for replication. To address putative multiple independent association signals in these two loci, we first applied variant pruning on the 75 variants that passed our selection criteria in the discovery meta-analyses, using intermediate linkage disequilibrium of  $r^2 \geq 0.2$  (Supplementary Table 2). Only the two lead variants rs11634042 at 15q25.1 and rs1561198 at 2p11.2 remained after pruning, indicating an association of a single haplotype block at each locus. Therefore, only SNPs rs11634042 and rs1561198 as well as their high LD variants ( $r^2 \geq 0.8$ ; in the following we call these variant sets LD blocks) were selected for replication in PD (Supplementary Table 3). At 15q25.1, SNP rs11634042 (intronic of *MORF4L1*) had the lowest P-value in the replication meta-analysis, with  $P = 0.95$  (OR = 1, 95% CI = [0.94–1.06]), showing no association with PD (Table 1 and Fig. 2).

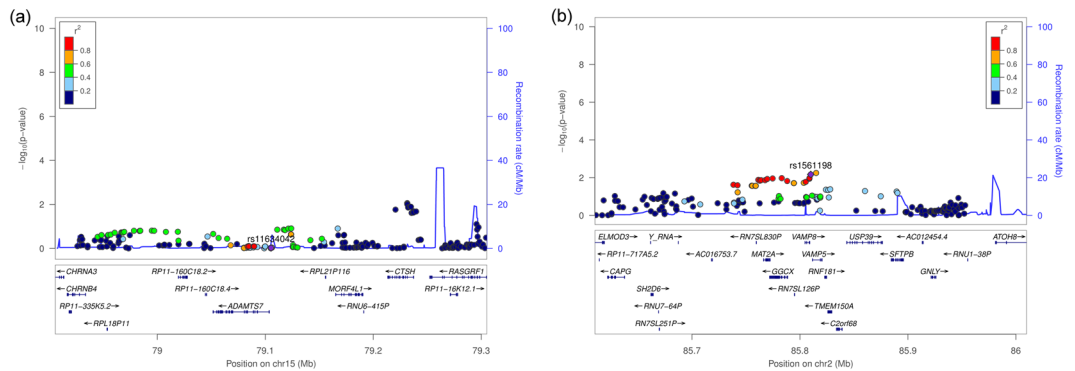
At 2p11.2, SNP rs1561198 showed the strongest association ( $P = 0.007$ ; OR = 1.09; 95% CI = [1.02–1.16]) among 15 out of a total of 38 strongly linked variants of this haplotype block, for which genotype data were available for all PD samples. The association of rs1561198 with PD was significant in the replication after correction for multiple testing, and after pooling all PD samples, SNP rs1561198 had a P-value of  $P = 2.02 \times 10^{-4}$  (OR = 1.11; 95% CI = [1.05–1.17]).

A comparison of pooled P-value for CP-EA-mod and CP-EA-sev with and without adjustment for shared controls indicated a reasonable inflation of less than one potency (Supplementary Table 4).

**In-silico characterization of selected variant effects.** Next, we investigated the associated genetic region at rs1561198 for putative regulatory effects on other genes and associations with other traits. First, we grouped genes in *cis* and *trans* with respect to SNP rs1561198 and its 38 variants in the LD block using information about topologically associated domain boundaries (TADs). The LD block of rs1561198 is located within a TAD spanning 2 mega base pairs (mb) on chromosome 2. This TAD harbours 14 genes of which seven are protein coding (Supplementary Table 5). Examination of the 39 variants in the haplotype block of rs1561198 indicated study-wide significant *cis*- and *trans*-regulatory effects on multiple genes (Supplementary Table 6). Table 2 shows those study-wide significant eSNP effects in *cis* with  $P < 10^{-10}$  for blood and gastrointestinal tissues that, to our knowledge, have been reported to date. For these *cis*-located genes the strongest eQTLs were found in blood for *VAMP8* (Vesicle associated membrane



**Figure 1.** Regional association plots of the two loci at 15q25.1 and 2p11.2 that were identified in the discovery meta-analysis and not yet known to be associated with PD. The plots show a region  $\pm 200$  kb of the discovery lead variants. (a) rs11634042 in CAD (b) rs1561198 in CAD (c) rs11634042 in AgP-Ger and (d) rs1561198 in AgP-Ger.



**Figure 2.** Regional association plots of (a) rs11634042  $\pm 200$  kb at 15q25.1 and (b) rs1561198  $\pm 200$  kb at 2p11.2 based on the pooled replication samples AgP-NL, CP-EA-mod, CP-EA-sev, CP-Ger.

protein 8;  $P = 9.8 \times 10^{-198}$ ), *MAT2A* (Methionine adenosyltransferase 2A;  $P = 1.4 \times 10^{-76}$ ), *USP39* ( $P = 9.8 \times 10^{-38}$ ), *VAMP5* ( $P = 8.2 \times 10^{-32}$ ) and *GGCX* (Gamma-glutamyl carboxylase;  $P = 3.7 \times 10^{-33}$ ).

Additionally, we assessed the relative pathogenicity of the variants in the haplotype block using the combined annotation dependant depletion score (CADD) to explore their potential of being causal for the observed association (Supplementary Table 7). Lead SNP rs1561198 was among 12 variants with  $CADD \geq 5$ , indicating a comparably high probability for having a causal effect (Fig. 3). The intergenic SNP rs2166529 ( $r^2 = 0.82$  with rs1561198), was assigned with the highest CADD score (15.29). Spearman correlation of  $r_s = 0.26$  ( $P = 0.34$ ; Alternative hypothesis: true  $r_s$  is not equal to 0) between CADD scores and P-values of the associations with PD indicated a low non-significant correlation.

Furthermore, we searched the LD block of rs1561198 for the presence of GWAS lead SNPs of other diseases and traits as listed in the in the NHGRI-EBI GWAS Catalog with  $P < 10^{-5}$ . According to the catalogue, the LD block is associated with CAD (rs7568458,  $P = 4 \times 10^{-10}$ ), myocardial infarction (SNP rs10176176,  $P = 3 \times 10^{-10}$ ) and prostate cancer (rs3731827,  $P = 3 \times 10^{-9}$ ) with genome-wide significance (Supplementary Table 8). Moreover, the chromosomal region at rs1561198  $\pm 500$  kb is associated with additional phenotypes including

Lead SNP	Gene	Samples (Best P-value)
rs1561198	GGCX	Whole blood (3.7e-33), Artery - Tibial (6.2e-25), Artery - Aorta (1.2e-21), Esophagus - Mucosa (2e-19), Atherosclerotic aortic root (1.8e-15), Cells - EBV-transformed lymphocytes (1.8e-14), Muscle skeletal (6.3e-13), Heart - Left ventricle (5.5e-11)
rs1561198	MAT2A	Blood (1.4e-76), Artery - Mammary (1.4e-14)
rs1561198	USP39	Peripheral blood (9.8e-38)
rs1561198	VAMP5	Peripheral blood (8.2e-32), Blood (1.5e-11), Muscle skeletal (1.5e-11)
rs1561198	VAMP8	Peripheral blood (9.8e-198), Blood (1.4e-76), Esophagus - Mucosa (2.9e-35), Cells - Macrophages (1.4e-24), Cells - Monocytes (1.6e-19)

**Table 2.** *Cis* eQTLs with  $P < 10^{-10}$  in blood and/or gastrointestinal tissue for the LD block of replicated SNP rs1561198. The smallest P-value per sample is given in brackets.

genome-wide significant associations with blood protein levels and pulse pressure (SNP rs7577293,  $P = 5 \times 10^{-76}$  and rs11689667,  $P = 2 \times 10^{-08}$ , respectively; Supplementary Table 9).

**Genetic risk score.** To date, 99 variants could be robustly identified to predispose for CAD according the current literature, (Supplementary Table 10). For all variants except SNP rs1878406, we had genotypes available in both case and control cohorts in the AgP-Ger sample and no genotypes were missing. Thus, the CAD-GRS, based on 98 CAD risk variants, was calculated in the German cases-control sample for AgP and revealed no differences in mean GRS between AgP-Ger cases ( $\mu = 107.46$ ;  $\sigma^2 = 6.43$ ) and controls ( $\mu = 107.48$ ;  $\sigma^2 = 6.48$ ; Supplementary Fig. 1).

## Discussion

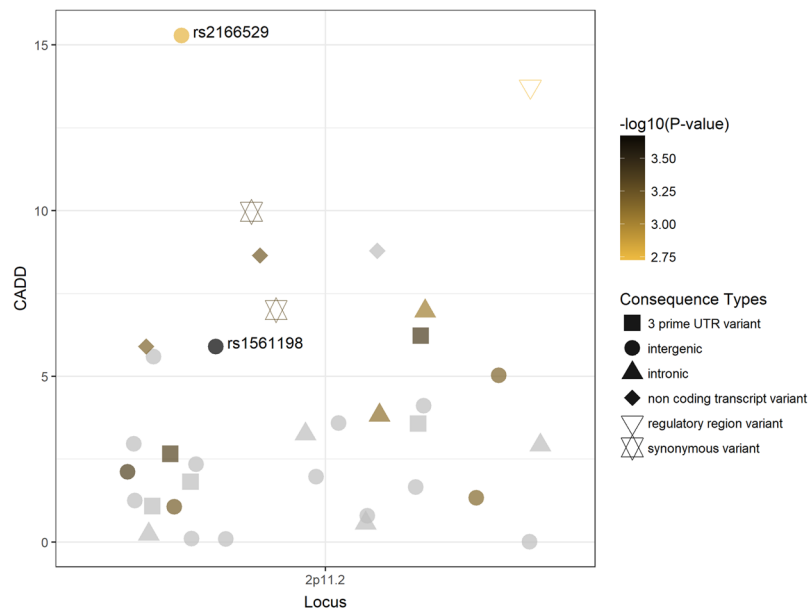
In this study, we aimed to identify novel genetic risk factors that are shared between CAD and PD, in order to improve the current pathogenic understanding of both diseases and highlight possible common biological underpinnings. We provide evidence for an association with PD for SNP rs1561198, located at *VAMP8*, which is a known genetic risk variant of CAD. We were able to replicate our observation from the discovery sample in an independent case-control sample of PD.

A limitation of the study was the relatively small sizes of the available PD case-controls samples compared to the CARDIoGRAMplusC4D sample. The CARDIoGRAMplusC4D sample was ~12 times larger than the pooled PD samples. This is why we consider that the association of SNP rs1561198 reached genome-wide significance in CAD, whereas the same variant with a similar effect size had a significance level in the pooled PD sample several orders of magnitudes lower. However, despite the comparatively lower association P-value of rs1561198 with PD with  $P = 2.02 \times 10^{-4}$  in the pooled PD samples, the association passed the Bonferroni-corrected significance threshold in the replication and we conclude that the association of rs1561198 with PD is a true positive finding. Additional evidence for association is provided by the observation that rs1561198 – together with several highly linked variants – showed consistent association of the same risk allele with PD in all individual PD samples, with the exception of the Dutch AgP sample that included the smallest number of cases ( $n = 179$ ), but a comparably large number of controls ( $n = 2,891$ ). Because of the low number of cases, we argue that this sample is highly susceptible to random errors (i.e. chance effects) and, at the same time, results in effect confidence intervals for which the sizes are biased by the large control sample. For the same reason we think, that the inter-sample heterogeneity would have increased when adding the Dutch AgP case-control sample to the discovery stage and deterred us from using the fixed effects model in this stage. Because of that and to decrease the chances of false positive association findings transferred from the discovery stage to the replication, we decided to exclude the Dutch sample from the discovery stage, but to add it to the replication stage instead.

In the discovery stage, the LD block of rs1561198 had an OR of ~1.19 in AgP-Ger, which is close to the estimations of detectable ORs according to our power calculations. Compared with the CAD (OR = 1.06) and CP samples (OR = 1.09) that were included in our analyses, the higher OR in AgP-Ger might be exclusive to AgP, which is influenced to greater extent by genetic factors compared to the late onset diseases CAD and CP, or reflect the Winner's curse bias, which results in an overestimation of the effect size.

The power of our study is also demonstrated by the re-discovery of the known association at 9p21.3 (*CDKN2B-AS1*), which is well established for both CAD<sup>4-6</sup> and PD<sup>7</sup>. The lead-SNP of the association of this locus, rs10116277, had an OR = 1.26 (MAF = 52%) in AgP-Ger, which was comparable in size and direction to the CAD association (OR = 1.21). This SNP is among the variants with the highest OR of all currently known genetic loci of CAD and AgP (Supplementary Table 9)<sup>49</sup>. The other previously reported shared risk loci at *VAMP3* (rs10864294)<sup>10</sup> and *PLG* (rs4252120)<sup>9</sup> were not re-discovered because they did not pass the pre-defined criterion  $P_{CAD} < 5 \times 10^{-8}$ .

Our finding of shared risk variants at *VAMP8* support and complement results of a previous report addressing the shared molecular mechanisms of both diseases. In this study, a transcriptome-wide shRNA knock-down approach demonstrated that *CDKN2B-AS1* and *VAMP3* expression is correlated on the RNA and protein level, and a rare variant upstream of *VAMP3* (rs17030881) was suggested to be associated with AgP and CAD<sup>10</sup>. *VAMP3* and *VAMP8* are simultaneously expressed in various cell types, e.g. in the secretory granules of platelets<sup>50</sup> and mast cells<sup>51</sup>, where they form complexes with platelet syntaxin 4 during platelet secretion, to release inflammatory cytokines, or in adipocytes, where they promote the fusion of glucose transporter type 4 (*GLUT4*)<sup>52</sup>. *VAMP3* and *VAMP8* mediated membrane trafficking in platelets also plays an important role in thrombosis and wound healing, processes with established relevance for the etiology of CAD and PD. In addition, VAMP mediated platelet secretion has an antimicrobial role in the host response to bacterial infection (previously reviewed<sup>53,54</sup>).



**Figure 3.** rs1561198 and the variants in its LD block ( $r^2 \geq 0.8$ ) at 2p11.2 are plotted against their CADD scores. The colour of each symbol refers the pooled P-value of all PD samples. Within the LD block, rs1561198 showed the lowest association P-value whereas rs2166529 showed the second highest association P-value and the highest CADD score. Since CADD scores are calculated for SNPs only, we four didn't include four insertions and deletions in this plot.

PD is caused by bacterial pathogens and it was suggested that the risk of CAD is putatively increased by bacterial infections<sup>55,56</sup>. Likewise, two clinical trials involving full-mouth mechanical debridement resulted in a transient deterioration of surrogate markers of CAD<sup>57,58</sup>, which supports the notion that bacterial inoculation and other procedure-related inflammation resulting from mechanical debridement has immediate negative effects on risk for CAD. In addition, living oral pathogens were detected in atheromatous plaques from coronary arteries<sup>59,60</sup>. In this context it is interesting that various studies determined that SNARE (soluble N-ethylmaleimide-sensitive factor attachment receptor) proteins like VAMP3 and VAMP8, are manipulated by specific pathogens to corrupt host membrane vesicular trafficking (reviewed in<sup>61</sup>). E.g., *Chlamydia* species mimic VAMP3 and VAMP8 to establish within the host cell<sup>62</sup>, *Shigella* species stimulate endocytosis of VAMP3 and VAMP8 in the internalization process of Shiga toxin<sup>63</sup>, and virulence factors of *Leishmania* parasites cleave VAMP3 and VAMP8 to evade phagocytosis<sup>64</sup>. We also note that a GWAS on periodontal pathogen colonization reported a region within *CAMTA1*, adjacent to *VAMP3*, to be associated with increased quantities of oral pathogenic bacteria<sup>1,65</sup>. Taken together, we note a possible yet speculative link of the observed associations with *VAMP3* and *VAMP8* and the increased risk of PD and CAD with effects of the putative causative variants on functions of coagulation, wound healing and thrombosis, as well as the invasion of some bacteria into host cells. However, the putative causal variant(s) of the *VAMP8* associated haplotype block and their likely effects on gene regulation still need to be identified. Currently, their effects in the susceptibility for CAD and PD are speculative.

We conclude that our finding of a haplotype block at *VAMP8*, which increases the risk for CAD and PD adds to the previously reported shared risk variants and indicates that the observed association of CAD and PD reported in prior epidemiological studies (previously reviewed<sup>2</sup>) cannot be solely explained by shared environmental risk factors. In addition, the current study probably missed possible associations with a shared role in the genetic etiology of CAD and PD. This is because of a lack of statistical power of PD to identify shared risk variant of CAD and PD with ORs <1.1 and an incomplete coverage of common and rare variants in the GWAS data sets. To identify the full spectrum of pleiotropic variants, larger sample sizes are needed for future studies.

### Data Availability

The CARDIoGRAMplusC4D summary statistics dataset is freely available on the consortium website (<http://www.cardiogramplusc4d.org>). Similarly, summary statistics for CP-EU-mod and CP-EU-sev can also be downloaded (see respective publication). Genotype data for aggressive periodontitis samples and the chronic periodontitis sample with German descent are available upon request from the biobanks PopGen (<https://www.epidemiologie.uni-kiel.de/biobanking>) and ShIP (<https://www.fvcm.med.uni-greifswald.de/index.html>).

### References

- Hansson, G. K. Inflammation, Atherosclerosis, and Coronary Artery Disease. *N. Engl. J. Med.* **352**, 1685–1695 (2005).
- Lockhart, P. B. *et al.* Periodontal Disease and Atherosclerotic Vascular Disease: Does the Evidence Support an Independent Association? *Circulation* **125** (2012).
- Shungin, D. *et al.* Using genetics to test the causal relationship of total adiposity and periodontitis: Mendelian randomization analyses in the Gene-Lifestyle Interactions and Dental Endpoints (GLIDE) Consortium. *Int. J. Epidemiol.* **44**, 638–650 (2015).

4. Wellcome Trust Case Control Consortium, T. W. T. C. C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–78 (2007).
5. Helgadóttir, A. *et al.* A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction. *Science* (80–). **316**, 1491–1493 (2007).
6. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–91 (2007).
7. Schaefer, A. S. *et al.* Identification of a shared genetic susceptibility locus for coronary heart disease and periodontitis. *Plos Genet.* **5**, e1000378 (2009).
8. Deloukas, P. *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**, 25–33 (2012).
9. Schaefer, A. S. *et al.* Genetic Evidence for Plasminogen as a Shared Genetic Risk Factor of Coronary Artery Disease and Periodontitis. *Circ. Cardiovasc. Genet.* **8**, 159–167 (2015).
10. Bochenek, G. *et al.* The large non-coding RNA ANRIL, which is associated with atherosclerosis, periodontitis and several forms of cancer, regulates ADIPOR1, VAMP3 and C11ORF10. *Hum. Mol. Genet.* **22**, 4516–4527 (2013).
11. Divaris, K. *et al.* Genome-wide association study of periodontal pathogen colonization. *J. Dent. Res.* **91**, 21S–28S (2012).
12. Susin, C., Haas, A. N. & Albandar, J. M. Epidemiology and demographics of aggressive periodontitis. *Periodontol.* **2000** **65**, 27–45 (2014).
13. Eke, P. I. *et al.* Update on Prevalence of Periodontitis in Adults in the United States: Nhanes 2009 to 2012. *J. Periodontol.* **86**, 611–622 (2015).
14. Gibson, G. Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–45 (2012).
15. Munz, M. *et al.* A genome-wide association study identifies nucleotide variants at SIGLEC5 and DEFA1A3 as risk loci for periodontitis. *Hum. Mol. Genet.* **33**, 272–279 (2017).
16. The CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
17. Teumer, A. *et al.* Genome-wide association study of chronic periodontitis in a general German population. *J. Clin. Periodontol.* **40**, 977–985 (2013).
18. Divaris, K. *et al.* Exploring the genetic basis of chronic periodontitis: a genome-wide association study. *Hum. Mol. Genet.* **22**, 2312–24 (2013).
19. Krawczak, M. *et al.* PopGen: Population-Based Recruitment of Patients and Controls for the Analysis of Complex Genotype-Phenotype Relationships. *Public Health Genomics* **9**, 55–61 (2006).
20. Müller, N. *et al.* IL-6 blockade by monoclonal antibodies inhibits apolipoprotein (a) expression and lipoprotein (a) synthesis in humans. *J. Lipid Res.* **56**, 1034–42 (2015).
21. Berger, K. DOGS. *Bundesgesundheitsblatt - Gesundheitsforsch. - Gesundheitsschutz* **55**, 816–821 (2012).
22. Schmermund, A. *et al.* 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. *Am. Heart J.* **144**, 212–218 (2002).
23. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
24. van Wijngaarden, J. P. *et al.* 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 Geriatr.* **11**, 80 (2011).
25. John, U. *et al.* Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Soz. Präventivmed.* **46**, 186–94 (2001).
26. Hensel, E. *et al.* Study of Health in Pomerania (SHIP): a health survey in an East German region. Objectives and design of the oral health section. *Quintessence Int.* **34**, 370–8 (2003).
27. Volzke, H. *et al.* Cohort Profile: The Study of Health in Pomerania. *Int. J. Epidemiol.* **40**, 294–307 (2011).
28. Hill, C. *et al.* The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am. J. Epidemiol.* **129**, 687–702 (1989).
29. Zaykin, D. V. & Kozbur, D. O. P-value based analysis for shared controls design in genome-wide association studies. *Genet. Epidemiol.* **34**, 725–38 (2010).
30. Skol, A. D., Scott, L. J., Abecasis, G. R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–13 (2006).
31. Munz, M., Tönnies, S., Balke, W.-T. & Simon, E. Multidimensional gene search with GeneHopper. *Nucleic Acids Res.* **43**, 98–103 (2015).
32. GTEx Consortium, Gte. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–60 (2015).
33. Ward, L. D. & Kellis, M. HaploReg v4: Systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44**, D877–D881 (2016).
34. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–4 (2012).
35. Eicher, J. D. *et al.* GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res.* **43**, D799–804 (2015).
36. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–11 (2013).
37. Gamazon, E. R. *et al.* SCAN: SNP and copy number annotation. *Bioinformatics* **26**, 259–62 (2010).
38. Xia, K. *et al.* seeQTL: a searchable database for human eQTLs. *Bioinformatics* **28**, 451–2 (2012).
39. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–43 (2013).
40. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–80 (2012).
41. Chen, Y. *et al.* Ensembl variation resources. *BMC Genomics* **11**, 293 (2010).
42. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
43. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–6 (2014).
44. Samani, N. J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–53 (2007).
45. Coronary Artery Disease (CAD) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.* **43**, 339–44 (2011).
46. Schaefer, A. S. *et al.* CDKN2BAS is associated with periodontitis in different European populations and is activated by bacterial infection. *J. Med. Genet.* **48**, 38–47 (2011).
47. Ernst, F. D. *et al.* Replication of the association of chromosomal region 9p21.3 with generalized aggressive periodontitis (gAgP) using an independent case-control cohort. *BMC Med. Genet.* **11**, 119 (2010).
48. Schaefer, A. S. *et al.* Validation of reported genetic risk factors for periodontitis in a large-scale replication study. *J. Clin. Periodontol.* **40**, 563–72 (2013).
49. Howson, J. M. M. *et al.* Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat. Genet.* **49**, 1113–1119 (2017).
50. Polgár, J., Chung, S.-H. & Reed, G. L. Vesicle-associated membrane protein 3 (VAMP-3) and VAMP-8 are present in human platelets and are required for granule secretion. *Blood* **100**, 1081–3 (2002).



51. Paumet, F. *et al.* Soluble NSF attachment protein receptors (SNAREs) in RBL-2H3 mast cells: functional role of syntaxin 4 in exocytosis and identification of a vesicle-associated membrane protein 8-containing secretory compartment. *J. Immunol.* **164**, 5850–7 (2000).
52. Zhao, P. *et al.* Variations in the requirement for v-SNAREs in GLUT4 trafficking in adipocytes. *J. Cell Sci.* **122**, 3472–3480 (2009).
53. Sun, H. The Interaction Between Pathogens and the Host Coagulation System. *Physiology* **21**, 281–288 (2006).
54. Yeaman, M. R. Platelets: at the nexus of antimicrobial defence. *Nat. Rev. Microbiol.* **12**, 426–437 (2014).
55. Beck, J. D. *et al.* Periodontal Disease and Coronary Heart Disease. *Circulation* **112**, 19 LP–24 (2005).
56. Roivainen, M. *et al.* Infections, inflammation, and the risk of coronary heart disease. *Circulation* **101**, 252–7 (2000).
57. Tonetti, M. S. *et al.* Treatment of Periodontitis and Endothelial Function. *N. Engl. J. Med.* **356**, 911–920 (2007).
58. Minassian, C., D'Aiuto, F., Hingorani, A. D. & Smeeth, L. Invasive Dental Treatment and Risk for Vascular Events. *Ann. Intern. Med.* **153**, 499 (2010).
59. Gaetti-Jardim, E., Marcelino, S. L., Feitosa, A. C. R., Romito, G. A. & Avila-Campos, M. J. Quantitative detection of periodontopathic bacteria in atherosclerotic plaques from coronary arteries. *J. Med. Microbiol.* **58**, 1568–1575 (2009).
60. Cavrini, F. *et al.* Molecular detection of *Treponema denticola* and *Porphyromonas gingivalis* in carotid and aortic atheromatous plaques by FISH: report of two cases. *J. Med. Microbiol.* **54**, 93–96 (2005).
61. Wesolowski, J. & Paumet, F. SNARE motif: a common motif used by pathogens to manipulate membrane fusion. *Virulence* **1**, 319–24 (2010).
62. Delevoye, C. *et al.* SNARE protein mimicry by an intracellular bacterium. *PLoS Pathog.* **4**, e1000022 (2008).
63. Renard, H.-F., Garcia-Castillo, M. D., Chambon, V., Lamaze, C. & Johannes, L. Shiga toxin stimulates clathrin-independent endocytosis of the VAMP2, VAMP3 and VAMP8 SNARE proteins. *J. Cell Sci.* **128**, 2891–2902 (2015).
64. Matte, C. & Descoteaux, A. Exploitation of the Host Cell Membrane Fusion Machinery by *Leishmania* Is Part of the Infection Process. *Plos Pathog.* **12**, e1005962 (2016).
65. Rhodin, K. *et al.* Chronic periodontitis genome-wide association studies: gene-centric and gene set enrichment analyses. *J. Dent. Res.* **93**, 882–90 (2014).

## Acknowledgements

Data on coronary artery disease have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from [www.CARDIOGRAMPLUSC4D.ORG](http://www.CARDIOGRAMPLUSC4D.ORG). This work was supported by a research grant of the German Research Foundation DFG (Deutsche Forschungsgemeinschaft; G.Z.: SCHA 1582/3-1). Funding was provided by the German Federal Ministry of Education and Research (BMBF) in the context of the e:Med program (e:AtheroSysMed and sysINFLAME), the FP7 European Union project CVgenes@target (261123) and a grant from the Fondation Leducq (CADgenomics: Understanding Coronary Artery Disease Genes, 12CVD02). This study was also supported through the Deutsche Forschungsgemeinschaft (DFG) cluster of excellence 'Inflammation at Interfaces'. Collection of the AgP cases was additionally supported by the German Ministry of Education and Research through the POPGEN biobank project (01GR0468 and 01EY1103). 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.

## Author Contributions

Study design: Matthias Munz, Steven Offenbacher, Thomas Kocher, André G. Uitterlinden, Lisette C.P.G.M. de Groot, Jeanette Erdmann, Arne S. Schaefer. Study management: Gesa M. Richter, Bruno G. Loos, Søren Jepsen, Kimon Divaris, Steven Offenbacher, Alexander Teumer, Birte Holtfreter, Thomas Kocher, Corinna Bruckmann, Yvonne Jockel-Schneider, Christian Graetz, Ingmar Staufenberg, Nathalie van der Velde, André G. Uitterlinden, Lisette C.P.G.M. de Groot, Jürgen Wellmann, Klaus Berger, Bastian Krone, Per Hoffmann, Matthias Laudes, Wolfgang Lieb, Andre Franke, Henrik Dommisch, Jeanette Erdmann, Arne S. Schaefer. Subject recruitment: Bruno G. Loos, Steven Offenbacher, Birte Holtfreter, Thomas Kocher, Corinna Bruckmann, Yvonne Jockel-Schneider, Christian Graetz, Ingmar Staufenberg, Nathalie van der Velde, André G. Uitterlinden, Lisette C.P.G.M. de Groot, Jürgen Wellmann, Klaus Berger, Bastian Krone, Per Hoffmann, Henrik Dommisch, Jeanette Erdmann, Arne S. Schaefer. Genotyping: Steven Offenbacher, Alexander Teumer, Christian Graetz, Matthias Laudes, Wolfgang Lieb, Andre Franke, Arne S. Schaefer. Methods and analysis: Matthias Munz, Kimon Divaris, Alexander Teumer, Nathalie van der Velde, Arne S. Schaefer. Interpretation of results: Matthias Munz, Gesa M. Richter, Arne S. Schaefer. Drafting of manuscript: Matthias Munz, Gesa M. Richter, Arne S. Schaefer. Critical review of manuscript: Gesa M. Richter, Kimon Divaris, Alexander Teumer, Birte Holtfreter, Loreto Munoz, Stephanie Tennstedt, Anita Bhandari, Wolfgang Lieb, Jeanette Erdmann, Arne S. Schaefer.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-31980-8>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018