

OPEN

# Translation of mouse model to human gives insights into periodontitis etiology

Aysar Nashef<sup>1,2,16,17</sup>, Munz Matthias<sup>3,4,17</sup>, Ervin Weiss<sup>5</sup>, Bruno G. Loos<sup>6</sup>, Søren Jepsen<sup>7</sup>, Nathalie van der Velde<sup>8,9</sup>, André G. Uitterlinden<sup>8</sup>, Jürgen Wellmann<sup>10</sup>, Klaus Berger<sup>10</sup>, Per Hoffmann<sup>11,12</sup>, Matthias Laudes<sup>13</sup>, Wolfgang Lieb<sup>14</sup>, Andre Franke<sup>15</sup>, Henrik Dommisch<sup>2</sup>, Arne Schäfer<sup>3,4\*</sup>, Yael Houri-Haddad<sup>1\*</sup> & Fuad A. Iraqi<sup>16\*</sup>

To suggest candidate genes involved in periodontitis, we combined gene expression data of periodontal biopsies from Collaborative Cross (CC) mouse lines, with previous reported quantitative trait loci (QTL) in mouse and with human genome-wide association studies (GWAS) associated with periodontitis. Periodontal samples from two susceptible, two resistant and two lines that showed bone formation after periodontal infection were collected during infection and naïve status. Differential expressed genes (DEGs) were analyzed in a case-control and case-only design. After infection, eleven protein-coding genes were significantly stronger expressed in resistant CC lines compared to susceptible ones. Of these, the most upregulated genes were *MMP20* ( $P = 0.001$ ), *RSPO4* ( $P = 0.032$ ), *CALB1* ( $P = 1.06 \times 10^{-4}$ ), and *AMTN* ( $P = 0.05$ ). In addition, human orthologous of candidate genes were tested for their association in a case-controls samples of aggressive (AgP) and chronic (CP) periodontitis (5,095 cases, 9,908 controls). In this analysis, variants at two loci, *TTL11/PTGS1* (rs9695213,  $P = 5.77 \times 10^{-5}$ ) and *RNASE2* (rs2771342,  $P = 2.84 \times 10^{-5}$ ) suggested association with both AgP and CP. In the association analysis with AgP only, the most significant associations were located at the HLA loci *HLA-DQH1* (rs9271850,  $P = 2.52 \times 10^{-14}$ ) and *HLA-DPA1* (rs17214512,  $P = 5.14 \times 10^{-5}$ ). This study demonstrates the utility of the CC RIL populations as a suitable model to investigate the mechanism of periodontal disease.

Periodontitis (PD) is one of the most common complex inflammatory diseases in human. The disease is believed to be a multifactorial trait which initiated by compositional shift of the oral micro biome because of different stimuli resulting in destruction of tissues surrounding the teeth. The precise mechanisms underlying individual disease susceptibility or that drive the individual steps in the pathogenesis of PD have largely remained unknown. Previous genome-wide association studies (GWAS) of chronic periodontitis (CP) reported several 'suggestive' susceptibility loci but failed to produce genome-wide significant evidence of association<sup>1-6</sup>. GWAS that used the

<sup>1</sup>Department of Prosthodontics, Dental school, The Hebrew University, Hadassah Jerusalem, Israel. <sup>2</sup>Department of Oral and Maxillofacial surgery, Poriya Medical center, Poriya, Israel. <sup>3</sup>Department of Periodontology and Synoptic Medicine, Institute for Dental and Craniofacial Sciences, Charité – University Medicine Berlin, Berlin, Germany. <sup>4</sup>Institute for Cardiogenetics, University of Lübeck, 23562, Lübeck, Germany. <sup>5</sup>School of Dental Medicine, Tel-Aviv University, Tel-Aviv, Israel. <sup>6</sup>Department of Periodontology and Oral Biochemistry, Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam and Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. <sup>7</sup>Department of Periodontology, Operative and Preventive Dentistry, University of Bonn, Bonn, Germany. <sup>8</sup>Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. <sup>9</sup>Department of Internal Medicine section of Geriatrics, Amsterdam Medical Center, Amsterdam, The Netherlands. <sup>10</sup>Institute of Epidemiology and Social Medicine, University Münster, Münster, Germany. <sup>11</sup>Institute of Human Genetics, University of Bonn, Bonn, Germany. <sup>12</sup>Human Genomics Research Group, Department of Biomedicine, University Hospital of Basel, Basel, Switzerland. <sup>13</sup>Department of Medicine, University of Kiel, Kiel, Germany. <sup>14</sup>Institute of Epidemiology, Christian-Albrechts-University, Kiel, Germany. <sup>15</sup>Institute of Clinical Molecular Biology, Christian-Albrechts-University, Berlin, Germany. <sup>16</sup>Department of Clinical Microbiology and Immunology, Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel. <sup>17</sup>These authors contributed equally: Aysar Nashef and Munz Matthias \*email: [arne.schaefer@charite.de](mailto:arne.schaefer@charite.de); [yaelho@ekmd.huji.ac.il](mailto:yaelho@ekmd.huji.ac.il); [fuadi@post.tau.ac.il](mailto:fuadi@post.tau.ac.il)

rare but severe and early-onset form aggressive periodontitis (AgP), which is believed to have a stronger genetic component, or studies that combined AgP and CP analysis samples to increase the statistical power, identified five genome-wide significant loci<sup>7–9</sup>. Compared to other complex diseases, the low number of statistically significant genetic risk loci likely reflects the relatively small sample sizes of the available PD cohorts. Given these size limitations and the general scarcity of high-quality phenotype-genotype cohorts for PD, GWAS alone are unlikely to identify the missing susceptibility loci. Here, the mouse model provides an appropriate and powerful system to identify genes that have a role in the etiology of complex human diseases. Given that expression patterns in mouse models recapitulate those in humans<sup>10</sup>, the examination of differences in gene expression as a response to different environmental exposures in genetically identical mice and between genetically different mice living in the same environmental context, can give direct insight into the mechanism that lead to disease susceptibility. Additionally, the genes underlying these differences can be mapped as quantitative trait loci (QTL)<sup>11</sup>. These QTLs and the differently expressed genes (DEGs) can be translated to the human orthologous genes and serve as candidates for genetic association studies in human case-control samples. Recently, we demonstrated the strength of combining genome-wide expression profiling and QTL-mapping in a F<sub>2</sub> mouse population and identified the human chromosomal region at *PF4/PPBP/CXCL5* to carry genetic susceptibility variants of AgP and CP<sup>12</sup>. In the current study, we used recombinant inbred mouse lines of the Collaborative Cross (CC) that were descended from eight divergent strains of mice: A/J, C57BL/6J, 129S1/SvImJ, NOD/LtJ, NZO/HiLtJ, CAST/Ei, PWK/PhJ, and WSB/EiJ and specifically designed for high-resolution mapping of QTLs<sup>13</sup>. Briefly, this population provide an increased level of genetic and phenotypic diversity compared to other existing mouse genetic reference populations<sup>14</sup>. The full description of the CC population under development in our labs at Tel-Aviv University and its power for dissecting complex traits is described in recent publications<sup>13,15–20</sup>. In previous studies, we identified CC-RILs with increased susceptibility and resistance to PD as a result of mixed infection with the oral bacteria *Porphyromonas gingivalis* (*P.g.*) and *Fusobacterium nucleatum* (*F.n.*)<sup>21</sup> and mapped two QTLs that were associated with alveolar bone loss in the CC mouse population<sup>22</sup>. In the current study, we hypothesized that in response to an identical challenge of bacterial infection, resistant and susceptible CC-RILs have different gene expression profiles. The identification of the significant DEGs could pinpoint putative molecular pathways and mechanisms that contribute to disease resistance and susceptibility. These genes also provide reasonable candidate genes for association studies in human case-control analyses populations of PD. To test our hypothesis, we investigated the expression patterns of oral tissues in six susceptible and resistant CC-RILs before and after mixed infection with *P.g.* and *F.n.*<sup>21</sup>. We identified a set of genes to be significantly upregulated in resistant CC-RILs after bacterial infection. We present new data that support the previously suggested susceptibility genes *TLL11*, *HLA-DOA* and *LBP*, and propose *HLA-DQH1* and *RNASE2* as novel risk factors for AgP and PD.

## Results

**Gene enrichment analysis.** The Principle Component Analysis (PCA) showed that at 42 days after infection, the largest variation within the 24 samples was between the genetically different RILs (e.g. resistant vs. susceptible vs. bone formation group) rather than between the infection/control conditions (Appendix-Fig. 1). All of RNAseq raw data were submitted to GEO website for allowing freely public access. The accession number of the raw data is **GSE145474**, and can be access by website: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145474>.

**Differential gene expression in periodontal tissues in susceptible, resistant, and bone-formation RILs after oral bacterial infection compared to mock-infection.** For each of the three phenotypes, we tested the differential gene expression between infected vs. non-infected individuals (Appendix-Table 1). This analysis showed how each of the three groups specifically responded to bacterial infection on the transcription level, giving a potential mechanistic link between the phenotype and specific DEGs. Eleven DEGs were observed in the phenotype “resistant”, and two within the phenotype “susceptible”. No significant DEG was observed for the phenotype “bone formation”. All DEGs except of *Bpifa6* could be mapped to an orthologous autosomal gene in the human genome (*Xist* is a non-coding RNA gene that is located on the X-chromosome. This chromosome was excluded from the human GWAS data set and the subsequent analyses). *Obp2b* showed homology with the two human genes *OBP2A* and *OBP2B* (Appendix-Table 1).

**Differential gene expression in periodontal tissues between the susceptible, resistant and bone-formation RILs after oral bacterial infection.** We next analyzed the differences in gene expression between the three genetically different phenotype groups after bacterial infection. The gene expression after bacterial infection of the CC-RILs of one phenotype group was compared with the gene expression after bacterial infection of the CC-RILs of the other two-phenotype groups. 1,326 genes were differentially expressed between infected susceptible and resistant RILs, 823 genes were differentially expressed between infected susceptible RILs compared to infected RILs of the bone formation, and 1,572 genes were differentially expressed between infected resistant RILs compared to the bone formation group (Padj < 0.05, log<sub>2</sub>FC > 1). Heatmaps of the expressed values for each comparison are shown in Appendix-Fig. 2. The ten strongest up regulated and the ten strongest down-regulated genes of each comparison are shown in Appendix Table 2. The regulated genes in the susceptible group were significantly related to immune response pathways FDR < 0.05 as shown in Appendix-Fig. 3. Functional network analyses of DEGs within the observed pathways are shown in Appendix Figs. 4, -5 and -6.

To nominate candidate genes that have a role in the etiology of PD, we selected the ten most up- or down-regulated DEGs, the DEGs that mapped to the QTLs *Perio3* and *Perio4* or that were suggested as risk genes by published human GWAS on PD. The QTLs *Perio3* and *Perio4* comprised 81 mouse genes (Appendix Table 3). Six DEGs (*PARP1*, *DUSP23*, *PCDH17*, *PSEN2*, *H3f3aos*, *EPHX1*) located to the QTLs *Perio3* and *Perio4* (Table 1, *H3f3aos* and *Dusp23* have no human orthologous and showed only borderline significance with P = 0.07). 15

Gene Name	Description	Chr.	Padj (Comparison)
Psen2	presenilin_2	1	0.037 (O/R)
Dusp23	dual_specificity_phosphatase_23	1	0.07 (O/R)
Ephx1	epoxide_hydrolase_1,_microsomal	1	0.01 (S/O)
H3f3aos	H3_histone_family_3A,_opposite_strand	1	0.07 (S/R)
Pcdh17	protocadherin_17	14	0.09 (S/R); 0.04(O/R)
Parp1	poly_(ADP-ribose)_polymerase_family,_member_1	1	0.05 (O/R); 0.039 (S/O)

**Table 1.** The table lists six genes and their Padj, which underlie the previous QTL (Perio3 and Perio4) and showed significant differential expression between one or more of three comparisons. (O; bone formation group, S; susceptible group, R; resistant group).

DEGs had been reported in previous human GWAS on periodontitis with  $P < 9 \times 10^{-6}$  (*C1orf87*, *HLA-DOA*, *BIRC5*, *CCDC13*, *GLDC*, *GPR141*, *OTOF*, *IFI16*, *ETNK2*, *TTL11*, *ATP5S*, *GRID1*, *LBP*, *NIN*, and *VAV1*; Appendix Table 4). A potential gene regulatory network between QTLs associated with susceptibility to alveolar bone loss in CC mice with DEGs in susceptible vs. resistant lines and bone formation groups was proposed and shown in Appendix Figure 3; two genes within the significant QTL *Perio3*, *Psen2* and *Parp1*, showed a HUB gene pattern.

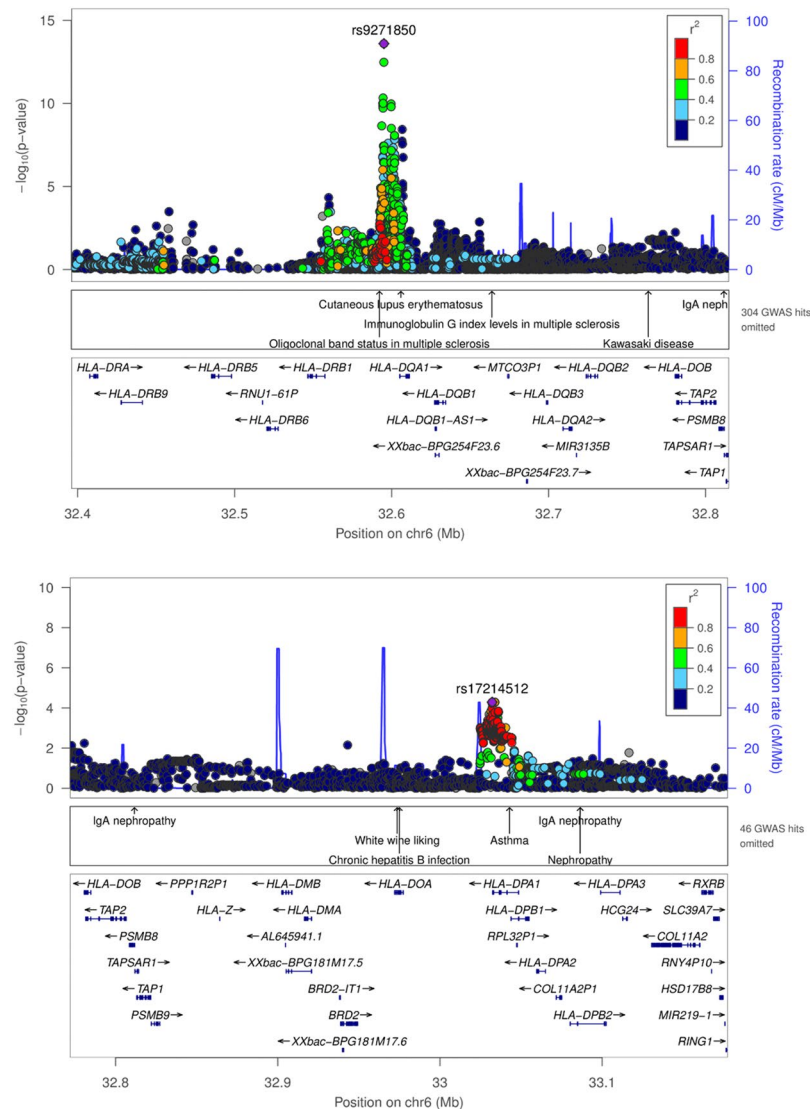
**Candidate gene association study in case-control samples of AgP and CP.** We selected the human orthologous genes of the 13 DEGs (and 200 kb up- and downstream to include putative regulatory sequences) that were identified after comparing the transcriptional response of the CC-RILs to bacterial infection with mock-infection. First, the common SNPs of these regions were examined for associations with the disease phenotype AgP. The smallest p-value was observed for a region upstream *OBP2A* (rs1329505,  $P = 7.85 \times 10^{-4}$ ), located ~73 kb upstream of SNP rs1537415 at *GLT6D1* that was previously reported to be associated with AgP with genome-wide significance<sup>8</sup>. Next, the variants at these loci were tested for association in a meta-analysis of AgP and CP. In this meta-analysis, the strongest association was observed for a large chromosomal region at *RPL29* (chr3:51,874,679–52,220,203; with  $p = 9.3 \times 10^{-5}$ ; rs16943). This region was previously reported to be associated with macrophage inflammatory protein 1b levels at a genome-wide significance level<sup>23</sup>. According to the GWAS Catalog, the tested genetic regions comprise 55 genome-wide significant associations and 61 suggestive associations with various phenotypes, as well as 5,903 eQTLs. These variants were tested for association in AgP and the meta-analysis of AgP and CP, if genotypes where available for both, the AgP and CP samples.

Next, we tested the associations of the common variants of the human orthologous genes (+ 200 kb up- and downstream) that corresponded to the ten most up- or downregulated DEGs, the DEGs that mapped to the QTLs *Perio3* and *Perio4* or that were suggested as risk genes by published human GWAS on PD (Table 1+ Appendix Table 2+ Appendix Table 4). In total, 23,413 SNPs were tested for association with AgP (896 cases, 7,104 controls). In this analysis, 6,423 SNPs indicated an association with  $p < 0.05$ . The most significant associations were located at the HLA loci *HLA-DQH1* (rs9271850,  $P = 2.52 \times 10^{-14}$ ) and *HLA-DPA1* (rs17214512,  $P = 5.14 \times 10^{-5}$ ; Fig. 1, Table 2). In the AgP-CP meta-analysis that combined the AgP and CP samples (5,095 cases, 9,908 controls), the HLA loci lost significance. Instead, a genetic region at *TTL11* (tubulin tyrosine ligase-like family, member 11), upstream *PTGS1* (prostaglandin G/H synthase 1), (chr9, rs9695213,  $P = 5.77 \times 10^{-5}$  [in AgP only the association was  $P = 8.1 \times 10^{-3}$ ]) and at *RNASE2* (Ribonuclease A Family Member 2) (chr14, rs2771342,  $P = 2.84 \times 10^{-5}$  [in AgP only the association was not significant]) showed the most significant associations (Fig. 2, Table 2). The chromosomal region at *TTL11* was previously suggested to be associated with increased quantities of *Pg* colonization of the human oral cavity before<sup>1</sup>. In addition to these loci, among the DEGs that mapped to the QTLs *Perio3* and *Perio4* and had additionally been suggested in previous human GWAS on PD as putative risk loci of PD, variants at *LBP* (chr.20) suggested association in the AgP-CP Meta-analysis (rs1780616,  $P = 1.67 \times 10^{-3}$  [AgP; rs1780616,  $P = 7.78 \times 10^{-3}$ ]).

## Discussion

This is the first study to investigate the transcriptome of periodontal tissues from susceptible, resistant and “bone-formation” RILs after infection with the periodontal bacteria *Pg* and *Fn* and at a naive status.

The comparison of gene expression with and without bacterial infection in mice of the resistant RILs, the susceptible RILs and the bone forming RILs, showed that the difference in gene expression in response to bacterial infection was strongest in the resistant RILs. Here, we identified eleven significant DEGs. Among the most upregulated genes were *AMTN* and *CALB1*, which are involved in calcium-binding, a relevant process in bone formation. Likewise, the genes *PAPLN* and *FBN2* are glycoproteins of the extracellular matrix, which is involved in regrowth and healing of tissue. Notably, after bacterial or mock-infection we did not identify significant DEGs in the susceptible and bone formation RILs. This may relate to the small sample size, which putatively caused false negative findings. Otherwise, the functional context of the significant DEGs we identified in the resistant RILs may point to the relevance of processes of tissue regeneration to maintain oral health under bacterial challenge. This may indicate that under particular strong bacterial infection as in our experimental setting, the resistance of bone loss was caused by a particular capacity in the regeneration of tissue integrity that was certainly harmed by the invading pathogens. Correspondingly, in our experimental setting disease susceptibility or resistance did not result from special features of the immune response.



**Figure 1.** Regional association plots of the most significant associations with AgP case-control cohort that located at the HLA loci, **(A)** HLA-DQH1 and **(B)** HLA-DPA1.

The screening for genes that were differentially expressed between the three RILs after bacterial infection identified a very large number of several hundreds to thousands of DEGs. This could relate to the extensive genetic variability between these RILs, which may have induced “transcriptional noise” that precluded the identification of DEGs relating to a different response to bacterial infection. However, the number of DEGs between the mock-infected RILs of the different phenotype groups was considerably lower compared to the number of genes that differed in expression between the RILs after bacterial infection, indicating that the response to the infection caused most of the observed differential expression. The comprehensive set of genes that were differentially regulated after oral bacterial infection can provide a useful reference for human studies, which relate to bacterial induced oral diseases like PD. More specifically, we note that the chromosomal region at *TLL11* showed the smallest p-value of association in the CP-AgP meta-analysis together with a significant association in the AgP-only analysis. The chromosomal region at *TLL11* was previously described to be associated with increased quantities of *Pg*.<sup>1</sup> However, the causative variants and their putative target genes are unknown and we note that the association also locates upstream to the cyclooxygenase *PTGS1*. However, the independent rediscovery of this locus despite the lack of bacterial data together with the suggestive association with AgP and in the meta-analysis of AgP and CP, argues for a true positive association finding. The missing statistical evidence for the suggested risk loci may generally relate to the relatively small sample size of the available PD analysis samples, compared to other complex common diseases. Given the size limitations of the available PD cohorts, true positive associations of genetic variants with moderate effects or with a lower MAF are unlikely to generate p-values, which pass a correction threshold necessary for thousands of independent tests. Thus, the sample size was a major limitation of the human association study. But similarly, this limitation explicitly argues for a complimentary approach to GWAS, in a manner described in the current work.

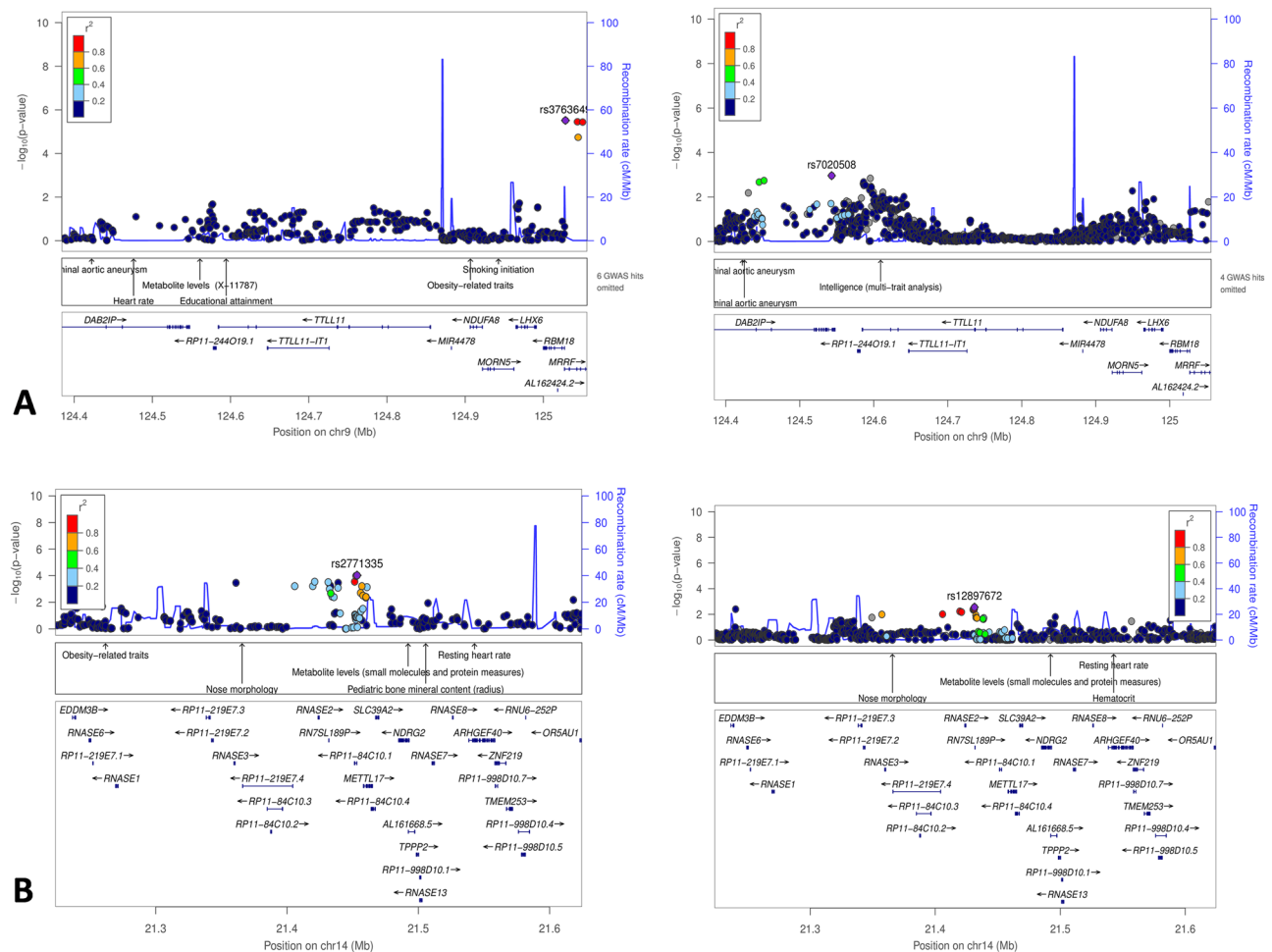
	Locus	Nearest Gene(s)	A1	A2	Stage	OR	L95	U95	P(Q)	I2	Mod	P
1	rs2771342											
	14q11.2	<i>METTL17</i> (downstream to <i>RNASE2</i> )	C	T	AgP-Ger	1.12	0.99	1.26				0.08
					AgP-NL	0.92	0.73	1.16				0.49
					CP-EA-mod	1.15	1.04	1.28				0.00
					CP-EA-sev	1.13	0.98	1.30				0.08
					Pooled	1.12	1.05	1.19	0.37	0.04	FE	7.41E-04
					CP-Ger	1.19	1.05	1.36				0.01
					Pooled (all)	1.13	1.07	1.20	0.42	0.00	FE	2.84E-05
2	rs9695213											
	9p24.1	<i>MRRF</i> (upstream to <i>TLL11</i> , downstream <i>PTGS1</i> )	G	A	AgP-Ger	1.19	1.00	1.42				0.05
					AgP-NL	1.38	0.99	1.93				0.06
					CP-EA-mod	1.27	1.09	1.48				0.00
					CP-EA-sev	1.22	0.99	1.50				0.07
					Pooled	1.24	1.13	1.37	0.86	0.00	FE	7.91E-06
					CP-Ger	1.01	0.84	1.22				0.88
					Pooled (all)	1.19	1.09	1.30	0.34	0.12	FE	5.77E-05
3	rs1780616											
	20q11.23	AL391095.1 - <i>LBP</i>	T	C	AgP-Ger	1.14	1.01	1.29				0.04
					AgP-NL	1.25	0.98	1.58				0.07
					CP-EA-mod	1.07	0.97	1.19				0.18
					CP-EA-sev	1.17	1.01	1.35				0.04
					Pooled	1.13	1.05	1.20	0.62	0.00	FE	5.44E-04
					CP-Ger	1.01	0.88	1.15				0.92
					Pooled (all)	1.10	1.04	1.17	0.42	0.00	FE	1.67E-03
4	rs9271850											
		<i>HLA-DQH1</i>										
			A	G	AgP-Ger	1.53	1.34	1.74				2.54E-10
					AgP-NL	1.66	1.32	2.09				1.75E-05
					Meta-AgP	1.56	1.39	1.75	0.54	0.00	FE	2.52E-14
5	rs17214512											
		<i>HLA-DPA1</i>										
			A	G	AgP-Ger	1.31	1.12	1.54				9.53E-04
					AgP-NL	1.44	1.07	1.93				1.64E-02
					Meta-AgP	1.34	1.16	1.54			FE	5.14E-05

**Table 2.** Five SNPs showed significant association with one of the tested three cohorts either separately or/ and as a pooled samples. (AgP-Ger; The German AgP samples, AgP-NL; Dutch AgP samples, Mod-CP; European American moderate CP cases, Sev-CP; European American severe CP cases, CP-Ger; German CP samples, A1 = Effect allele; A2 = Non-effect allele; EAF = Effect allele frequency; Cas = Cases; Con = Controls; OR = Odds ratio; CI = Confidence interval; P(Q) = Cochranes Q P-value; I2 = Heterogeneity index; Mod = Model; RE = Random effects; FE = Fixed effects; P = P-value).

Likewise, we identified a suggestive association with the gene *LBP* (Lipopolysaccharide binding protein) that was earlier reported to be associated with CP<sup>4</sup>. The soluble acute-phase protein LBP binds to bacterial lipopolysaccharide (LPS) to elicit immune responses. Although the association at this locus was weak in our study, the independent rediscovery argues for a true positive finding.

In the association study that excluded CP patients, we observed a strong association with the *HLA-DOA/-DPA1* locus, which was reported in a GWAS on CP before. However, this association was only significant in our AgP case-control sample and not in the AgP-CP meta-analysis. This can indicate a false negative finding due to different classification of CP in our study and the previous GWAS. However, a true negative finding cannot be ruled out at this stage and this association needs clarification. Likewise the strong association with *HLA-DQH1* was only observed in our AgP case-control sample but not in the CP sample.

In summary, this study generated valuable information by giving independent support for the genetic regions at *DAB2IP/TLL11*, *HLA-DOA* and *LBP* to be involved in the genetic etiology of PD and propose *HLA-DQH1* and *RNASE2* as novel risk factors for AgP and CP. This demonstrates the utility of RILs of the CC as a controlled model system of bacteria induced PD for the discovery of susceptibility genes of human PD.



**Figure 2.** Regional association plots of the loci *DAB2IP/TTL11* and *RNASE2*, which suggested association in the meta-analysis of AgP and CP with  $p < 10^{-4}$ . **(A)** *DAB2IP/TTL11*. Left panel: The association upstream *TTL11*, tagged by rs3763649, was observed in the discovery meta-analysis using AgP and CP cases and controls. Right panel: After exclusion of the CP case-control samples, the associated region was identical to the genetic locus associated with increased quantities of *P.G.* in humans (Red rectangle = region associated with increased quantities of *P.G.* in humans). **(B)** *RNASE2*. Left panel: The association of the suggested novel susceptibility locus of PD was strongest in the meta-analysis using AgP and CP cases and controls. Right panel: After exclusion of the CP case-control samples, the associated region was less significant, possibly reflecting the reduced statistical power.

## Materials and Methods

**CC-RIL selection and assessment.** Experiments were performed at the Department of Prosthodontics, Faculty of Dental Medicine, and Hadassah Medical Center, Israel and at the Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Israel. All mouse experimental protocols used in this project approved by the Institutional Animal Care and Use Committee (IACUC) at Tel Aviv University in Israel, with numbers M-08-044, which adheres to the Israeli guidelines, which follows the National Institute of Health (NIH) of United States of America (USA) animal care and use protocols. Mice were monitored daily by the student for their overall health status. Mice, which were observed to suffer (less movement and activity) from the infection, and based on the consultation with the Veterinarian at the small animal unit, were terminated. Furthermore, based on the approved protocol, mice which lost 10% of their bodyweight between two measured points, or 20% overall, of their initial bodyweight, were also terminated. Termination procedure of mice during assessment based animal suffering or at endpoint of the experiment was based on injection of Ketamine and Xylazine with dose of 100 mg/kg (body weight) and 10 mg/kg (bodyweight), respectively for initial euthanasia, and subsequently by CO<sub>2</sub> chamber.

Experimental CC-RILs mice at age 8 to 12 wk of the International Livestock (IL) cohort, were provided by the Small Animal Facility, Faculty of Medicine, Tel Aviv University, Israel, and were assessed for their alveolar bone changes in response to a mixed oral infection with a well-known periodontal pathogen. In the current study we used the mixed infection model using the *P.g.* bacteria as a known key bacteria in periodontitis<sup>24,25</sup> mixed with *F.n.* as another known pathogenic bacteria from the orange complex<sup>26</sup>. As we showed in our previous studies, the

Cases-Ctrls sample	n of CC_RILs (names)	phenotype	n_cases	n_ctrls	total
#1	2 (TAU-IL188, TAU-IL111)	resistant	4 (2 per line)	4 (2 per line)	8
#2	2 (TAU-IL785, TAU-IL551)	susceptible	4 (2 per line)	4 (2 per line)	8
#3	2 (TAU-IL2124, TAU-IL2126)	bone formation	4 (2 per line)	4 (2 per line)	8
Total	6 lines		12	12	24

**Table 3.** Number and characteristics of RILs used for differential expression analysis. Ctrl; control, n; number, RILs; recombinant inbred lines.

mixed bacteria mimic the periodontal infection in human and induce a prominent bone loss in mice<sup>27</sup>. Briefly, each CC line was divided in an infected and a control group. The mice were treated with Sulfamethoxazole (0.8 mg/ml) in drinking water for a continuous period of ten days, followed by an antibiotic-free period of three days, before oral application of mixed culture of *P.g. (strain 381)* and *F.n. (strain PK1594)* (control groups were treated with PBS and 2% Carboxymethylcellulose)<sup>27,28</sup>. 42 days after the last infection, the mice were sacrificed and the alveolar bone volume at the second molar of the left hemi-maxilla was evaluated by micro-computed tomography ( $\mu$ CT)<sup>29</sup>. The CC lines showed significant variation in their response to the mixed infection. While susceptible CC lines showed significant decrease in bone volume ( $P < 0.05$ ) after infection, the resistant CC lines did not show significant change in bone volume after infection. In addition, a group of CC-RILs showed increase in bone volume and was named “bone formation group”<sup>22</sup>. For each of the phenotypes “resistant”, “susceptible” and “bone formation”, two strains were selected for differential gene expression analysis (Appendix Figure 8). Because sex effect on changes in the alveolar bone volume was not observed, both sexes were treated equally. Tissue samples from gingiva, alveolar bone, periodontal ligament and teeth were collected as mixed tissues from two resistant CC lines (TAU-IL188, TAU-IL111), two susceptible CC lines (TAU-IL785, TAU-IL551) and two CC lines that showed bone formation after 42 days of oral bacterial infection (TAU-IL2124, TAU-IL2126). Of each strain, total RNA of two biological replicates were sequenced (RNAseq) after bacterial infection and after mock-infection. In the current study, these are designated as cases and controls, respectively. Table 3 summarizes the experimental setting.

**RNA isolation, RNA-sequencing and Differential gene expression analysis in CC-RILs.** Total high molecular weight genomic RNA was purified using the RNeasy kit (Qiagen, Hilden, Germany). RNA quality was assessed using a BioAnalyzer (Agilent Technologies, Palo Alto, CA, USA). Samples with the RNA integrity number (RIN)  $\geq 7$  was used to prepare a library of template molecules. The mRNA was sequenced with a single-end technology (read length of 75) on NextSeq 500 platform (Illumina, San Diego, USA). Alignment was performed to mouse genome version GRCm38. Quality control assays and differential expression results, were calculated and visualized in R (version 3.3.1)<sup>30</sup> using R package DEseq 2<sup>23</sup>. The significance threshold was taken as  $\text{Padj} < 0.05$ . First, we analyzed the differences in gene expression in response to the different environmental exposure (bacterial infection vs. mock-infection) within each phenotype group (susceptible, resistant or bone formation). For each phenotype group, two independent CC-RILs with the same phenotype were analyzed. From each of these CC-RILs, two biological replicates were included. These two replicates were combined with the other two replicates of the second CC-RIL with the same phenotype, to exclude random effects of a single line, resulting in four samples (see Table 3). Second, between the CC-RILs that developed different phenotypes under the same environmental exposure, we analyzed the differences in gene expression. In a case-only design, the gene expression levels of infected CC-RILs of one phenotype group were compared with the gene expression levels of infected CC-RILs of the other two phenotype groups (e.g. susceptible vs. resistant and susceptible vs. bone formation).

**Biological processes associated with the regulated genes.** We tested the general differences between each type of response after bacterial infection (resistant, bone formation and susceptible) and considered the basal gene expression differences of the tested CC lines. Biological processes at  $\text{Padj} < 0.05$  associated with the generated differentially expressed genes and potential regulatory networks between the generated DEGs and the previous periodontitis QTL in CC mice were defined using the IPA software (QIAGEN Inc. <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>). Subsequently, we performed functional network analyses of each gene underlying the most significant pathway.

**Selection criteria of candidate DEGs for association mapping in the human case-control samples.**

1. DEGs derived from the comparison infected vs. non-infected for each of the three phenotypes ( $N = 13$ ).
2. The ten most up regulated and the ten most down-regulated DEGs of the comparisons of the different phenotypes after bacterial infection (resistant vs. susceptible vs. bone formation group at infection status) ( $N = 60$ ).
3. DEGs that were previously reported to be associated with PD in GWAS ( $N = 15$ ).
4. DEGs that located to the novel QTLs *Perio3* and *Perio4* ( $N = 6$ ).

**Human gene homology mapping and annotation.** Identified candidate genes in mice were mapped to human genes by using the GeneHopper database<sup>31</sup>. GeneHopper contains 27,443 orthologous gene mappings between mouse and human (Ensembl)<sup>32</sup>. All regions of the human orthologous genes  $\pm 200$  kb were annotated

with GWAS associations with  $P < 10^{-5}$  from the NHGRI-EBI GWAS Catalog<sup>33,34</sup>. Currently, the GWAS Catalog contains 150 associations ( $P < 5 \times 10^{-6}$ ) with a PD phenotype (Appendix table 5). Moreover, the human orthologues were annotated for expression quantitative trait loci (eQTLs) of the public domain by using Qtlizer ([www.genehopper.de/qtlizer](http://www.genehopper.de/qtlizer)), an annotation method of the Genehopper web application, which is integrating data from multiple eQTL databases.

**Participating human studies.** The meta-analysis samples consisted of case-control GWAS of German and Dutch AgP<sup>9</sup> and of European American<sup>2</sup> and German CP<sup>4</sup> patients. The German AgP sample (AgP-Ger) included 680 cases and 3,973 controls. Cases were recruited across Germany by the biobank Popgen<sup>35</sup>, University-Hospital Schleswig-Holstein, Germany. Controls originated from North- and West-Germany and were recruited from the Competence Network “FoCus - Food Chain Plus”<sup>36</sup>, the Dortmund Gesundheitsstudie – DOGS<sup>37</sup> and the Heinz Nixdorf Recall Studies 1–3<sup>38</sup>. The Dutch AgP sample (AgP-NL) consisted of 171 cases and 2,607 controls. The Dutch cases were recruited from the ACTA (Academisch Centrum Tandheelkunde Amsterdam) and the Dutch controls were recruited from Rotterdam and Wageningen by the B-Proof Study<sup>39</sup>. Inclusion criteria for AgP were  $\geq 2$  affected teeth with  $\geq 30\%$  bone loss in patients  $< 36$  years of age; disease phenotype was diagnosed by full mouth dental radiographs. Genotype imputation was performed using 1000 Genomes Phase 3 reference panel<sup>40</sup>. A detailed quality control and imputation workflow is described in<sup>9</sup>. 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 and were described before<sup>41</sup>. In brief, the patients were classified by the Centers for Disease Control/American Academy of Periodontology (CDC/AAP) consensus three-level classification system<sup>42</sup>. The CDC/AAP taxonomy uses clinical attachment loss (CAL) and PD criteria to define three CP categories as healthy-mild, moderate, and severe CP cases, the first being the control. 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. The German CP (CP-Ger) sample, which consisted of 993 cases and 1,419 controls from a meta-analysis of SHIP and SHIP-TREND cohorts<sup>43–45</sup> was used for validation of the results of the explorative meta-analysis and also included in the pooled analysis that comprised all samples. In brief, subjects within the first and the third tertile of proportion of proximal sites with attachment loss (AL)  $\geq 4$  mm were contrasted after stratification by sex and 10-year age groups. Age-specific tertiles were defined to include severely diseased cases within each age stratum. Thus, also young and severely diseased subjects were captured and included in the third tertile. Otherwise, those in the third tertile would have been the older ones and those in the first tertile would have included the younger ones only. Individuals aged  $> 60$  years were excluded. The identification of these subjects is important as we assume that genetic predispositions might manifest especially in younger age, while in older subjects these effects are overlaid with environmental risk factor associated disease progression. This case-control sample was previously described in detail<sup>4</sup>. 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.

**Filtering.** In the post imputation QC processing, variants were excluded, which passed the following cut-off criteria: AgP: Hardy-Weinberg-Equilibrium P-value ( $P_{\text{HWE}}$ )  $< 10^{-4}$ , imputation quality (INFO) of  $< 0.8^9$ ; CP (Germany):  $P_{\text{HWE}} \leq 0.001$ , imputation quality ( $r^2_{\text{HAT}}$ )  $\leq 0.3^4$ ; CP (European-US):  $P_{\text{HWE}} < 10^{-5}$  (ARIC) and  $P_{\text{HWE}} < 10^{-6}$  (Health ABC), imputation quality  $< 0.8^2$ . Moreover, we filtered out variants with a minor allele frequency (MAF)  $< 0.05$  because our study lacked statistical power to analyze rare variants. The variant sets did not completely overlap between the different genotype data sets that was mainly due to the different human genome references that were used to impute the different data sets. Only variants with genotype data available in each study were analyzed in the meta-analysis.

**Meta-analysis and mapping to DEGs.** We meta-analysed the genome-wide association scans of AgP-Ger, AgP-NL, CP-EA-sev, CP-EA-mod and CP-Ger based on the additive model using logistic regression. By default, we applied a fixed effects model. However, for variants with a high degree of heterogeneity, i.e. a P-value of Cochran’s Q  $P(Q) < 0.05$  and a heterogeneity index  $I^2 > 0.5$ , we applied a random effects model instead. Subsequently, we extracted all associations in the range of the candidate genes  $\pm 200$  kilobases and loci with an association of  $P_{\text{meta}} < 10^{-3}$  were selected.

### Data availability

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>), respectively. Summary statistics for CP-EU can also be downloaded (see<sup>2</sup>).

Received: 3 June 2019; Accepted: 3 March 2020;

Published online: 17 March 2020

### References

1. Divaris, K. *et al.* Genome-wide Association Study of Periodontal Pathogen Colonization. *J. Dent. Res.* **91**, S21–S28 (2012).
2. Divaris, K. *et al.* Exploring the genetic basis of chronic periodontitis: A genome-wide association study. *Hum. Mol. Genet.* **22**, 2312–2324 (2013).
3. Shimizu, S. *et al.* A Genome-wide Association Study of Periodontitis in a Japanese Population. *J. Dent. Res.* **94**, 555–561 (2015).



4. Teumer, A. *et al.* Genome-wide association study of chronic periodontitis in a general German population. *J. Clin. Periodontol.* **40**, 977–985 (2013).
5. Hong, K.-W., Shin, M.-S., Ahn, Y.-B., Lee, H.-J. & Kim, H.-D. Genomewide association study on chronic periodontitis in Korean population: Results from the Yangpyeong health cohort. *J. Clin. Periodontol.* **42** (2015).
6. Sanders, A. E. *et al.* Chronic Periodontitis Genome-wide Association Study in the Hispanic Community Health Study / Study of Latinos. *J. Dent. Res.* **96**, 64–72 (2017).
7. Munz, M. *et al.* Meta-analysis of genome-wide association studies of aggressive and chronic periodontitis identifies two novel risk loci. *Eur. J. Hum. Genet.* 102–113 <https://doi.org/10.1038/s41431-018-0265-5> (2018).
8. Schaefer, A. S. *et al.* A genome-wide association study identifies GLT6D1 as a susceptibility locus for periodontitis. *Hum. Mol. Genet.* **19**, 553–562 (2010).
9. Munz, M. *et al.* A genome-wide association study identifies nucleotide variants at SIGLEC5 and DEFA1A3 as risk loci for periodontitis. *Hum. Mol. Genet.* **26**, 2577–2588 (2017).
10. Takao, K. & Miyakawa, T. Genomic responses in mouse models greatly mimic human inflammatory diseases. *Proc. Natl. Acad. Sci. USA* **112**, 1167–72 (2015).
11. Iraqi, F. Fine mapping of quantitative trait loci using advanced intercross lines of mice and positional cloning of the corresponding genes. *Exp. Lung Res.* **26**, 641–649 (2000).
12. Shusterman, A. *et al.* The PF4/PPBP/CXCL5 Gene Cluster Is Associated with Periodontitis. *J. Dent. Res.* 002203451770631 <https://doi.org/10.1177/0022034517706311> (2017).
13. Iraqi, F. A., Churchill, G. & Mott, R. The Collaborative Cross, developing a resource for mammalian systems genetics: a status report of the Wellcome Trust cohort. *Mamm. Genome* **19**, 379–81 (2008).
14. The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics* **190**, 389–401 (2012).
15. Iraqi, Fa *et al.* Heritability and coefficient of genetic variation analyses of phenotypic traits provide strong basis for high-resolution QTL mapping in the Collaborative Cross mouse genetic reference population. *Mamm. Genome* **25**, 109–119 (2014).
16. Soller, M. & Iraqi, Fa The Collaborative Cross - A next generation mouse genetic resource population for high resolution genomic analysis of complex traits. *Livest. Sci.* **166**, 19–25 (2014).
17. Vered, K., Durrant, C., Mott, R. & Iraqi, Fa Susceptibility to klebsiella pneumoniae infection in collaborative cross mice is a complex trait controlled by at least three loci acting at different time points. *BMC Genomics* **15**, 865 (2014).
18. Durrant, C. *et al.* Collaborative Cross mice and their power to map host susceptibility to *Aspergillus fumigatus* infection. *Genome Res.* **21**, 1239–48 (2011).
19. Churchill, Ga *et al.* The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* **36**, 1133–1137 (2004).
20. Aylor, D. L. *et al.* Genetic analysis of complex traits in the emerging Collaborative Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Res.* **21**, 1213–1222 (2011).
21. Shusterman, A. *et al.* factor of host susceptibility to periodontitis in the CollaboratiGenotype is an important determinantv Cross and inbred mouse populations. *BMC Genet.* **14**, 68 (2013).
22. Nashef, A. *et al.* Integration of Murine and Human Studies for Mapping Periodontitis Susceptibility. *J. Dent. Res.* 002203451774418 <https://doi.org/10.1177/0022034517744189> (2018).
23. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome Biol.* **15**, 1–21 (2014).
24. Hajishengallis, G., Darveau, R. P. & Curtis, M. A. The keystone-pathogen hypothesis. *Nat. Rev. Microbiol.* **10**, 717–25 (2012).
25. Darveau, R. P., Hajishengallis, G. & Curtis, M. A. Porphyromonas gingivalis as a potential community activist for disease. *J. Dent. Res.* **91**, 816–820 (2012).
26. Socran.sky, S. S., Haffajee, A. D., Cugini, M. A. & Smith, C. K. J. R. Microbial complexes in subgingival plaque. *J Clin Periodontol* **25**, 134–144 (1998).
27. Polak, D. model of experimental periodontitis induced by P. gingivalis/Fusobacterium nucleatum infection: bone loss and host response. *et al.* Mouse model of experimental periodontitis induced by Porphyromonas gingivalis/Fusobacterium nucleatum infection: bone loss and host response. *J. Clin. Periodontol.* **36**, 406–10 (2009).
28. Baker, P. J., Evans, R. T. & Roopenian, D. C. Oral infection with Porphyromonas gingivalis and induced alveolar bone loss in immunocompetent and severe combined immunodeficient mice. *Arch. Oral Biol.* **39**, 1035–40 (1994).
29. Wilensky, A., Gabet, Y., Yumoto, H., Hourri-Haddad, Y. & Shapira, L. Three-dimensional quantification of alveolar bone loss in Porphyromonas gingivalis-infected mice using micro-computed tomography. *J. Periodontol.* **76**, 1282–6 (2005).
30. Chalovich, J. M. & Eisenberg, E. NIH Public Access. *Biophys. Chem.* **257**, 2432–2437 (2005).
31. Munz, M., Tönnies, S., Balke, W. T. & Simon, E. Multidimensional gene search with GeneHopper. *Nucleic Acids Res.* **43**, W98–W103 (2015).
32. Herrero, J. *et al.* Ensembl comparative genomics resources. *Database* **2016**, 2016 (2016).
33. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
34. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–6 (2014).
35. 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).
36. 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).
37. Berger, K. Dogs. *Bundesgesundheitsblatt - Gesundheitsforsch. - Gesundheitsschutz* **55**, 816–821 (2012).
38. Schermund, 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).
39. 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).
40. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
41. Hill, C. *et al.* The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am. J. Epidemiol.* **129**, 687–702 (1989).
42. Page, R. C. & Eke, P. I. Case Definitions for Use in Population-Based Surveillance of Periodontitis. *J. Periodontol.* **78**, 1387–1399 (2007).
43. John, U. *et al.* Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Soz. Praventivmed.* **46**, 186–94 (2001).
44. 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).
45. Volzke, H. *et al.* Cohort Profile: The Study of Health in Pomerania. *Int. J. Epidemiol.* **40**, 294–307 (2011).

## Acknowledgements

We thank Dr. Yuval Nevo, Dr. Sharona Algavish and Dr. Hadar Benimen for the bioinformatics support. There are no competing financial interests in relation to the work described. This work was supported by the German Research Foundation DFG (Deutsche Forschungsgemeinschaft; GZ: SCHA 1582/3-1 and SCHA 1582/4-1), the Israel Science Foundation [grant number 429/09], the US-American National Institute of Health (NIH): RO1-DE022527, UL1-TR001111, and RO1-DE021418, The Federal Ministry of Education and Research (01GR0468), German Migraine & Headache Society (DMKG) and 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, The Federal Ministry of Research and Education (BMBF, grant no. 01ER0816), the Federal Ministry of Education and Research BMBF (FKZ 0315540 A), the Heinz Nixdorf Foundation (Germany), the Federal Ministry of Education and Science and the German Research Council (DFG; Project SI 236/8-1, SI 236/9-1, ER 155/6-1). The German Centre financed Genotyping of the Illumina HumanOmni-1 Quad BeadChips of the HNR subjects for Neurodegenerative Diseases (DZNE), Bonn.

## Author contributions

Aysar Nashef, performed the mouse experiments and translation to human, RNAseq analysis, wrote the draft of the MS and final approval; Munz Matthias, performed the human bioinformatics analysis and made MS and final approval; Ervin Weiss, was involved in designing the mouse experiments and data analysis; Bruno G. Loos, Søren Jepsen, Nathalie van der Velde, André G. Uitterlinden, Jürgen Wellmann, Klaus Berger, Per Hoffmann, Matthias Laudes, Wolfgang Lieb, Andre Franke, Henrik Dommisch, were involved with collecting and genotyping analysis of human samples and all approved final version of the MS; Arne Schäfer, was involved with human bioinformatics analysis and writing the first and final version of the MS; Yael Hourri-Haddad was involved with designing the mouse experiments, data analysis and drafting the MS and final approval of the MS; Fuad A. Iraqi was involved with designing the mouse experiments and translation to human, data analysis and drafting the MS and final approval of the MS.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-61819-0>.

**Correspondence** and requests for materials should be addressed to A.S., Y.H.-H. or F.A.I.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**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) 2020