Identification of single nucleotide pleomorphisms associated with periodontal disease in head and neck cancer irradiation patients by exome sequencing

Valentina Talevi, MS,[a,b] Jia Wen, PhD,[b] Rajesh V. Lalla, DDS,[c] Michael T. Brennan, DDS,[a] Farah B. Mougeot, PhD,[a] and Jean-Luc C. Mougeot, PhD[a,b]

Objective. Periodontal disease (PD) is a common oral complication in patients with head and neck cancer (HNC) undergoing radiation therapy (RT). Our objective was to identify candidate single nucleotide polymorphisms (SNPs) associated with PD in radiation-treated patients with HNC.

Study Design. DNA was extracted from the saliva of patients with HNC (n = 69) before RT. Clinical attachment loss (CAL) increment greater than 0.2 mm over 24 months after RT was used to define PD progression. After exome sequencing, SNPs associated with post-RT PD progression were identified by using logistic regression and homozygosity analyses. The web tools STRING, the Database for Annotation, Visualization and Integrated Discovery (DAVID), GeneCodis, and Ensembl Variant Effect Predictor were used for functional analysis.

Results. Of the 48 patients with HNC with post-RT PD progression, 24 had no tooth with 5 mm or greater pocket depth before RT, whereas of the 21 patients with HNC without progression, 11 had PD initially. A total of 330 SNPs (249 genes) with over-represented homozygous genotype (98.5% variant allele) were found to be associated with post-RT PD. Sixty of these corresponded to PD-related pathways, including previously identified genes. In patients with HNC with post-RT PD progression, SNPs were found in genes (n = 10) in contrast to those without progression (n = 7).

Conclusions. The SNPs of collagen genes were identified, potentially defining susceptibility to PD in patients with HNC, and this could be further investigated to characterize PD drug targets. (Oral Surg Oral Med Oral Pathol Oral Radiol 2020;130:32−42)

In 2019, 53,000 new cases of head and neck cancer (HNC) of the oral cavity and pharynx, including 12,410 laryngeal HNC, were diagnosed in the United States.[1] The 5-year survival rate is 65.3% for cancers of the oral cavity and pharynx and 60.3% for laryngeal cancer.[1] Approximately 95% of HNCs are squamous cell carcinomas (SCCs).[2] Other (non-SCC) HNCs include those that occur in the salivary glands and thyroid gland.[2-4]

Patients with HNC usually receive radiation therapy (RT) as standard treatment, with or without induction or concurrent chemotherapy. Because of the damage to oral tissues, almost all RT-treated patients with HNC develop at least 1 acute or chronic oral complication.[5-7] Acute oral complications, such as mucositis, taste loss, and dry mouth, may occur during RT and up to 3 to 4 weeks after RT.[8] Chronic oral complications include caries, osteoradionecrosis, trismus, and progression of periodontal disease (PD).[8]

The mechanisms by which RT-associated PD develops in some patients with HNC at different levels of severity are not well understood, although polymorphisms in specific gene sets have been implicated in PD susceptibility (e.g., genes encoding interleukins, matrix metalloproteinases [MMPs], and fragment crystallizable gamma receptors [FcγR]).[7,9-11] The term *PD* refers to inflammation of both gingival tissue and bone supporting the tooth, corresponding to a broad spectrum of conditions that may be defined as gingivitis or periodontitis.[12] As PD progresses, jaws can be destroyed and teeth can be lost.[13,14] The more severe form of PD, periodontitis, can be classified into 4 categories based on the severity of disease, considering several factors, such as clinical attachment loss (CAL), probing depth, tooth mobility, and tooth and bone loss.[15]

CAL, defined as the measurement of the position of the soft tissue in relation to the cementoenamel junction, was shown to increase, on average, by 0.2 mm or greater per year in patients without HNC but with established periodontitis.[16,17] Ammajan et al. reported that 61.5% of RT-treated patients with HNC with mandibular teeth affected and 34.4% patients with maxillary teeth

Statement of Clinical Relevance

In patients with head and neck cancer undergoing radiation therapy, susceptibility to progression of post-radiation periodontal disease may be governed by different sets of gene polymorphisms, depending on the periodontal disease reaching a certain stage before cancer therapy.

[a]Department of Oral Medicine, Carolinas Medical Center, Atrium Health, Charlotte, NC, USA.
[b]College of Computing and Informatics, Department of Bioinformatics and Genomics, UNC-Charlotte, Charlotte, NC, USA.
[c]Section of Oral Medicine, University of Connecticut Health, Farmington, CT, USA.
Received for publication Dec 6, 2019; returned for revision Feb 5, 2020; accepted for publication Feb 16, 2020.

2212-4403/$-see front matter
https://doi.org/10.1016/j.oooo.2020.02.013

affected showed attachment loss greater than 0.2 mm.[18] This study showed an average CAL increase of 0.27 mm ± 0.31 for mandibular teeth and 0.15 mm ± 0.22 for maxillary teeth in 6 months' interval time.[18] RT can exacerbate PD progression and increase the risk of osteoradionecrosis as a result of CAL increase.[19] Several genetic studies, including genome-wide association studies (GWAS), have sought to better understand why certain individuals are more susceptible to PD.[20-22]

Our clinical team has previously investigated oral health complications, including PD, in a large cohort of patients with HNC who underwent intensity-modulated radiation therapy (IMRT) (OraRad study No. U01 DE022939). This study showed that 47% of patients had at least 1 tooth with a probing depth 5 mm or greater before RT.[23] In the present study, we identified the candidate single nucleotide polymorphisms (SNPs) associated with post-RT PD progression in a subset of the HNC patient population in our OraRad study by using exome sequencing of DNA extracted from saliva.

MATERIALS AND METHODS

Patient recruitment

The study cohort included patients with HNC (N = 69; males = 56; females = 13; mean age [standard deviation {SD}] = 56.9 [11.3]; range = 26−87), diagnosed with SCC (n = 56) or non-SCC (n = 13) oral cancers (coincidental gender numbers matching). The patients were recruited at the Carolinas Medical Center (CMC)−Atrium Health (Charlotte, NC) and the University of Connecticut Health (Farmington, CT). Ethnicity included 87% white, 10.2% black, 1.4% American Indian, and 1.4% unknown. All patients with HNC received curative RT to the head and neck region with or without concurrent chemotherapy (Table I). The total dose consisted of 6500 to 7000 cGy distributed as 180 to 200 cGy daily doses, 4 days a week over 6 to 7 weeks.

This study was approved by the institutional review boards at both centers, and all patients signed informed consent forms (No. 11-13-04 A).

PD assessment

The progression of PD was evaluated by using the CAL index. The differences among the CAL measurements before RT (T0), at the visit at 12 months (T12), and at the visit at 24 months (T24) were calculated. Pre-RT CAL measurement was performed between day −42 and day 0, with day 0 corresponding to the day of the first administered RT dose. The cohort (N = 69) was divided into 2 groups: patients without PD progression (no-PDP group) and those with PD progression (PDP group) after RT.

The no-PDP group included all RT-treated patients with HNC not showing an increase in CAL greater than 0.2 mm over a period of 24 months (n = 21), that is, T0−T12, T12−T24, or T0−T24 (see Table I). The PDP group included all patients showing an increment of CAL greater than 0.2 mm (n = 48) in this same period (see Table I). The cutoff value of greater than 0.2 mm CAL (whole mouth average) was chosen to

Table I. Demographic and clinical data of patients with head and neck cancer undergoing RT

Criteria	*No-PDP group**	*PDP group†*	*P value‡*
Patient count (males/females)	21 (18/3)	48 (38/10)	
Age (years)			
Mean (SD)	55.5 (11.3)	57.5 (11.4)	
Median	54	57	
Range	32−72	26−87	
Ethnicity§			
Males: C/B/AI/U	14/2/1/1	37/1/0/0	
Females: C/B/AI/U	2/1/0/0	7/3/0/0	
Pre-RT probing depth (pd) measurements			
Range of number of teeth with pd ≥ 5 mm	0−12	0−14	
Number of patients with 1 or more teeth with pd ≥ 5 mm	11	24	
Percentage of patients with 1 or more teeth with pd ≥ 5 mm	52.4%	50%	
Pre-RT CAL			
Mean (SD)	1.910 (0.741)	1.631 (0.709)	
Median	1.720	1.557	.1556
Range	1.113−3.438	0.527−3.933	
CAL change			
T0 to T12			
Mean (SD)	−0.051 (0.179)	0.154 (0.593)	
Median	0.005	0.284	.00906
Range	−0.424 to 0.166	−1.970 to 1.282	
T0 to T24			
Mean (SD)	−0.119 (0.299)	0.442 (0.613)	
Median	0.007	0.384	.00014
Range	−0.865to 0.187	−1.298 to 1.961	
T12 to T24			
Mean (SD)	0.011 (0.189)	0.357 (0.433)	
Median	0.082	0.333	.0035
Range	−0.441 to 0.180	−0.846 to 1.250	

CAL, clinical attachment loss; *HNC,* head and neck cancer; *No-PDP,* no periodontal disease progression; *PDP,* periodontal disease progression; *RT,* radiation therapy; *SD,* standard deviation.

*No-PDP group includes patients with HNC showing a change of CAL less than 0.2 mm (whole mouth average) at 12 (T12) and/or 24 (T24) months after RT.

†PDP group includes patients showing a change of CAL greater than 0.2 mm after RT.

‡Mann-Whitney U-test

§Ethnicity: C (white), B (black), AI (American Indian), U (unknown). Patients with HNC consisted of 56 patients with squamous cell carcinoma (SCC) and 13 patients without SCC (not stratified in table).

define the occurrence of PD progression in patients with HNC, characterized or not characterized by chronic PD progression at baseline (T0). This choice was based on the study by Machtei et al. mentioned earlier, showing an average increase of 0.2 mm or greater per year in patients without HNC but with established periodontitis.[17] The rationale was to increase the likelihood of capturing PD progression that may be impacted by RT in our patient cohort, consisting of fast and slow progressors.

The presence of PD before RT was defined as a patient having 1 or more teeth with probing depth 5 mm or greater, which was the criterion used in the OraRad study, because a probing depth 5 mm or greater may indicate the presence of moderate periodontitis.[23,24] In the no-PDP group, 11 of 21 patients with HNC (52.4%) had PD before RT, whereas in the PDP group 24 of 48 patients (50%) had PD before RT (see Table I). The details of each patient with HNC are given in Supplemental Table S1.

Saliva sample collection, DNA isolation, exome sequencing, and SNP determination

Saliva samples from patients with HNC were collected before RT into saliva collection kits (OGR-500; DNA GenoTek, Ontario, Canada), processed according to the manufacturer's instructions, and stored at −80°C until further analysis. DNA was isolated by using the prepIT-L2P reagent procedure, according to the manufacturer's instructions (DNA GenoTek Inc.). DNA was quantified and assessed for quality control by using Nanodrop for determination of the A260 nm/A280 nm absorbance ratio and the Qubit assay (Thermo Fisher, Waltham, MA).

The Ion Ampliseq Exome Kit (Life Technologies, Carlsbad, CA) was used for the preparation of sequencing libraries, according to the manufacturer's protocol, and the libraries were clonally polymerase chain reaction amplified on ion sphere particles by using the Ion PI Template OT2 200 Kit v2 (Life Technologies/Thermo Fisher Scientific, Waltham, MA). Ion PI Sequencing 200 Kit v2 on an Ion Proton Sequencer and Ion PI Chips (Life Technologies/Thermo Fisher Scientific) were used to enrich the template-positive ion sphere particles to maximize the number of sequencing reads.

For SNP genotyping, Ion Torrent Suite Software (Thermo Fisher) was used. After processing, Variant Call Format (VCF) files were generated by using the Torrent Mapping Alignment Program (integrating BWA-short,[25] BWA-long,[26] and SSAHA[27] algorithms), the Super-Maximal Exact Matching program,[28] and the Torrent Variant Caller plugin. For base-space SNP prediction, the Bayesian method was used for low-medium coverage ($\times$ 10−50), and the frequentist approach was applied for higher coverage ($> \times$ 60).

Variants quality control and GWAS

Initial VCF files were filtered for quality control by using the VCF tools v0.1.15 program package.[29] SNPs with a high genotyping failure rate were removed, and the SNPs present in at least 80% of the patients were kept for further analysis. SNPs with minor allele frequency less than 5% were filtered out. Multiallelic variants were removed, and biallelic variants were retained for further analysis.[30] From this process, 11,934 autosomal informative SNPs remained for further analysis.

Statistical analyses

Logistic regression was implemented in PLINK[31] v1.9 to identify significant SNPs associated with PD (95% confidence level). Covariates (e.g., gender, age, and race) were included in the logistic regression model, but not HNC cancer type because of the lack of degrees of freedom. The "adjust" *P* values feature was used to correct for multiple testing (significance level α= 0.05). To determine overrepresentation of the homozygote genotype for the alternative and reference alleles in the no-PDP and PDP groups, a 2-tailed z-test was performed on candidate SNPs (significance level α = 0.05).[32]

Gene network, gene ontology, and variant interpretation

Query Kaviar[33] and SNPnexus[34] online programs were used to identify the SNPs and the genes overlapping candidate SNP locations. STRING[35] v11.0 was used to generate a gene interaction network for candidate genes with predominant homozygote SNPs, identified via z-tests. The highest confidence score (90%) and several active interaction sources, such as text mining, experiments, databases, coexpression, and co-occurrence, were used.[35] Gene ontology (biological processes) was determined for the genes with at least 1 connection within the molecular network by using Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8[36,37] and GeneCodis[38-40] online tools. Genes with one or more SNPs overrepresented in the no-PDP group were considered as genes containing "potentially protective" SNPs. The genes with SNPs overrepresented in the PDP group were considered "potentially detrimental," based on the assumption that such SNPs could confer susceptibility to PD.

Genes included in the main biological processes relevant to PD were integrated in an extended network containing genes identified by previous PD-related studies. A first set of genes was extracted from 4 PD genetic association studies,[11,41-43] and a second set was extracted from 3 PD functional genomics studies.[44-46] Gene ontology analysis was repeated for these additional sets of genes to identify those also identified by our analysis, as described above. The potential impact of SNPs located in the genes identified within the PD-related biological

processes were interpreted by using the Ensembl Variant Effect Predictor (VEP) online tool.[47]

Comparison tests

By using the classification of "no-PDP" and "PDP" groups and the respective subcategories (patients with PD before RT and those without), a 2-tailed z-test and a χ^2 test were performed in 4 comparisons. These comparisons were made to identify SNPs distinguishing the no-PDP group from the PDP group and the subcategories within the groups (see Supplemental Figure S1 for comparisons A−D). Additionally, we compared the SNP genotypes of a no-PDP patient who had the highest number of teeth with probing depths 5 mm or greater before RT with those of this patient's counterpart in the PDP group.

In these comparisons, the genes identified through the regression analysis were considered to have a genotype potentially detrimental (overrepresented in the PDP group) or protective (overrepresented in the no-PDP group) when the SNPs were found to be significant via both the z-test and the χ^2 test (P values $< .05$).

RESULTS

The flow chart in Figure 1 summarizes the main steps and results of the SNP analysis and the systems biology analyses.

SNP candidates associated with PD, gene ontology, and molecular network analysis

The logistic regression analysis showed that 489 genetic variants were associated with PD progression (adjusted P value $< .05$), whereas age, gender, and ethnicity were not (adjusted P value $\geq .05$). From the 2-tailed z-test for the alternative variant allele, we identified 325 SNPs/244 genes that were predominant in either the no-PDP or the PDP group. We identified 5 SNPs/3 genes for the reference allele in the PDP group only.

STRING[35] analysis generated 17 tight interaction subnetworks containing 64 genes (90% confidence level) (Supplemental Table S2a). Of these, 36 genes were overrepresented in the PDP group. Gene ontology analysis performed with DAVID[36,37] and GeneCodis[38-40] tools identified 6 major biological processes relevant to PD and corresponding to 26 unique genes of the 64 genes (Table II). The biological processes represented in the network were angiogenesis, cell adhesion, DNA repair, extracellular matrix organization, immune response, and innate immune response (Figure 2). Of the 26 genes, 14 were identified with a potentially protective SNP genotype (Table III). VEP annotation[46] showed that 13 SNPs (9 genes) had a moderate impact and 18 SNPs (9 genes) had a low impact (Table IV) and that 44 SNPs (26 genes) had a modifier impact (data not shown). Because of the relevance of collagen genes in PD, VEP[43] was also used to extend the analysis to those identified in our study: *COL1 A2, COL6 A5, COL14 A1*, and *COL27 A1*. Most potential effects reported by VEP were modifiers (Table V).

In addition, by repeating the gene ontology analysis that included previously identified PD-related genes, the 6 biological processes mentioned earlier were enriched by 49 genes (Supplemental Table S2b), including 38 SNPs from previous genetic association studies.[11,41-45] The gene ontology *immune response* and *innate immune response* biological processes were enriched by 35 genes, and 14 genes enriched *cell adhesion* and *extracellular matrix organization* biological processes. Finally, 4 of the 49 genes were also identified by our SNP analysis: *COL14 A1*,[45] *HLA-DQB1*,[46] *HLA-B*,[43] and *MMP2*.[11]

Comparison of the no-PDP and PDP groups with or without PD before RT

The A to D comparisons (see Supplemental Figure S1) were performed, excluding genes less relevant to PD, that is, those belonging to the *angiogenesis* and *DNA repair* biological processes among the 6 identified. Only 1 gene was found significant in comparison A (*VCAN*) and one in comparison B (*ITGA1*). Patients with pre-RT PD were compared with patients without pre-RT PD within the PDP (comparison A) and the no-PDP (comparison B) groups, as determined by the z-test and the χ^2 test (P value $< .05$) (Supplemental Table S3). Both genes may carry potentially detrimental SNPs and are involved in cell adhesion and extracellular matrix organization processes. Comparison C, between the no-PDP and PDP patients not having pre-RT PD, showed 8 out of 21 significant genes (see Supplemental Table S3). Five of 8 genes may carry potentially detrimental SNPs, and each of the 4 biological processes cell adhesion, extracellular matrix organization, immune response, and innate immune response contains at least 1 of these 8 genes. Comparison D, between the no-PDP and PDP patients with pre-RT PD showed 11 significant genes out of 21. Six of the 11 genes may carry the potentially detrimental SNPs (comparison D; see Supplemental Table S3). Among the 18 genes found significant, 8 genes in comparison C and 11 genes in comparison D, *COL6 A5* was found in both comparisons.

The genotype characterization for the 11 genes identified in comparison D (see Supplemental Table S3) for the 1 patient in the PDP group (patient 65; see Supplemental Table S2) with the highest number of teeth with probing depths 5 mm of greater (n = 14) showed the presence of homozygous SNPs (alternate allele) in 4 of 6 genes with potentially detrimental genotypes (66.7%; *COL27 A1, COL6 A5, RPS6 KA1*, and *VCAN*). This patient had only 1 of the 5 genes with potentially protective genotypes (20%; *TNFRSF10 A*). The same genotype characterization done for the no-PDP patient counterpart (patient 12;

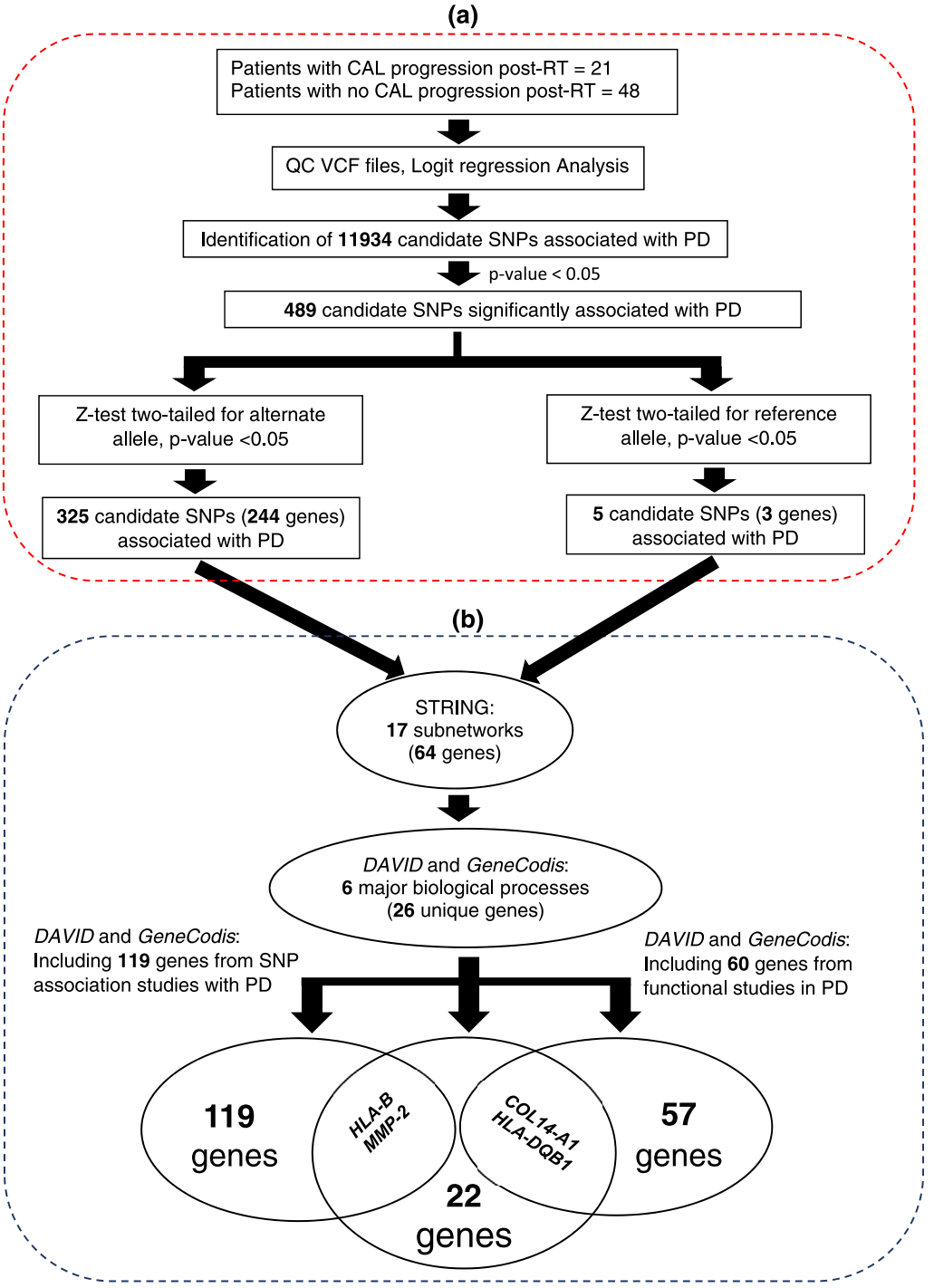

Fig. 1. Analytical strategy of single nucleotide polymorphism (SNP) analysis of radiation-treated patients with head and neck cancer (HNC) at risk for developing periodontal disease. **(A)** Filtering and statistical analyses: The combined Variant Call Format (VCF) file, obtained from n = 69 patients, was filtered for minor allele frequency > 5%; genotype rate > 80%; and biallelic sites. Logit regression (adjusted $P < .05$) and z-test ($P < .05$) identified 247 genes (244 alternative variant alleles and 3 reference alleles) with candidate SNPs potentially associated with periodontal disease. **(B)** Gene network and gene ontology analyses: STRING tool determined 17 gene networks (64 genes) based on the 247 genes input. The 64 genes were subjected to gene ontology analysis using GeneCodis (https://genecodis.genyo.es/) and DAVID (Database for Annotation, Visualization and Integrated Discovery; https://david.ncifcrf.gov/). As a result, 6 biological processes highly associated with periodontal disease, including 26 unique genes, were identified. Gene ontology analysis was repeated, including genes identified in previous studies as being associated with periodontal disease. This study confirmed that the *HLA-B, MMP2, COL14 A1*, and *HLA-DQB1* genes, previously shown to contain SNPs, were associated with periodontal disease. *CAL,* clinical attachment loss; *RT,* radiation therapy.

Table II. Periodontal disease (PD) relevant biological processes identified by gene network and gene ontology analyses including 26 unique genes

*GO pathway**	*Biological process*†	*Genes*‡	*Protective (Yes/No)*§
GO:0007155	Cell adhesion	CDH23	Y
		COL14 A1	N
		COL1 A2	Y
		COL27 A1	N
		COL6 A5	N
		HSPG2	Y
		ITGA1	N
		ITGAL	Y
		VCAN	N
GO:0030198	Extracellular matrix organization	ANXA2	N
		COL14 A1	N
		COL1 A2	Y
		COL27 A1	N
		COL6 A5	N
		HSPG2	Y
		ITGA1	N
		ITGAL	Y
		MMP2	N
		VCAN	N
GO:0001525	Angiogenesis	HSPG2	Y
		ANXA2	N
		MMP2	N
GO:0006281	DNA repair	BARD1	Y
		GEN1	Y
		GTF2 H1	Y
		SMC6	Y
		XRCC1	Y
GO:0006955	Immune response	CNR2	Y
		IKBKE	N
		HLA-DQB1	Y
		HLA-C	Y
		HLA-B	Y
		TNFRSF10 A	Y
GO:0045087	Innate immune response	IKBKE	N
		RPS6 KA2	N
		RPS6 KA1	N
		MAP3 K1	N
		TRIM5	N

*Gene ontology (GO) pathway identities for biological processes.
†Genes included in the 6 major biological processes.
‡Biological processes categorized on the assumption that the effects of single nucleotide polymorphisms (SNPs) in genes might potentially be protective.
§Genes protective (Y) if they were overrepresented in patients with no post-radiation therapy (RT) progression of PD (i.e., clinical attachment loss [CAL] increase less than 0.2 mm over 24 months; no-PDP group).

see Supplemental Table S2) with the highest number of teeth with probing depths 5 mm or greater (n = 12) showed SNPs in 5 of all 5 genes with potentially protective genotypes (100%; *COL1 A2, CNR2, HLA-DQB1, ITGAL*, and *TNFRSF10 A*) and only 1 of the 6 genes with potentially detrimental genotypes (16.7%; *RPS6 KA1*).

DISCUSSION

This is the first study reporting the use of exome-sequencing in the identification of SNPs associated with PD in RT-treated patients with HNC.

Significance of SNPs in biological processes, cell adhesion, and extracellular matrix organization

As anticipated for an SNP association study, we identified significantly more SNPs/genes (325/244) for the alternative (variant) allele associated with PD progression compared with the reference allele (5/3). Among the 247 genes, gene ontology identified six biological processes relevant to PD,[47] including processes required for cell−cell or cell−matrix organization[48] and for structural organization of the extracellular matrix.[49] Eight genes were in common among these processes, including the collagen genes *COL1 A2, COL6 A5, COL14 A1*, and *COL27 A1*. Collagen type I is among these 4 collagen genes. *COL1 A2* encodes the protein for type I collagen, which represents the major component of the periodontium lost during PD progression.[50,51]

In our study, we found that the patients in the no-PDP group carrying the intronic rs389328 variant homozygote polymorphism (potentially protective) in *COL1 A2* might benefit from such a variant. However, according to the VEP analysis, the rs389328 variant genotype has a "modifier impact," which is difficult to predict without experimental evidence. SNPs in intronic regions might influence splicing during RNA transcription, leading to the production of different protein isoforms possibly more beneficial than the original form. Conversely, if the intron mutation is in a seed region, complementary to a miRNA sequence, this variation could affect regulation of the gene. If miRNA−mRNA binding is weak, the degradation may be slower or totally inhibited. *COL6 A5, COL14 A1,* and *COL27 A1* are, instead, overrepresented in PDP group, suggesting a role in PD susceptibility in patients with HNC.

Mutations in the nonsense-mediated decay (NMD) target region of the *COL6 A5* gene could potentially generate truncated COL6 A5 proteins that will not be functional in cell adhesion and extracellular matrix organization. Indeed, NMD is a quality-control mechanism, and its main function is eliminating mRNA transcripts that encode premature stop codons to prevent the translation of nonfunctional proteins.[52] The rs993823 mutation in *COL14 A1* and the rs2567711 mutation in *COL27 A* are intron variants in the protein-

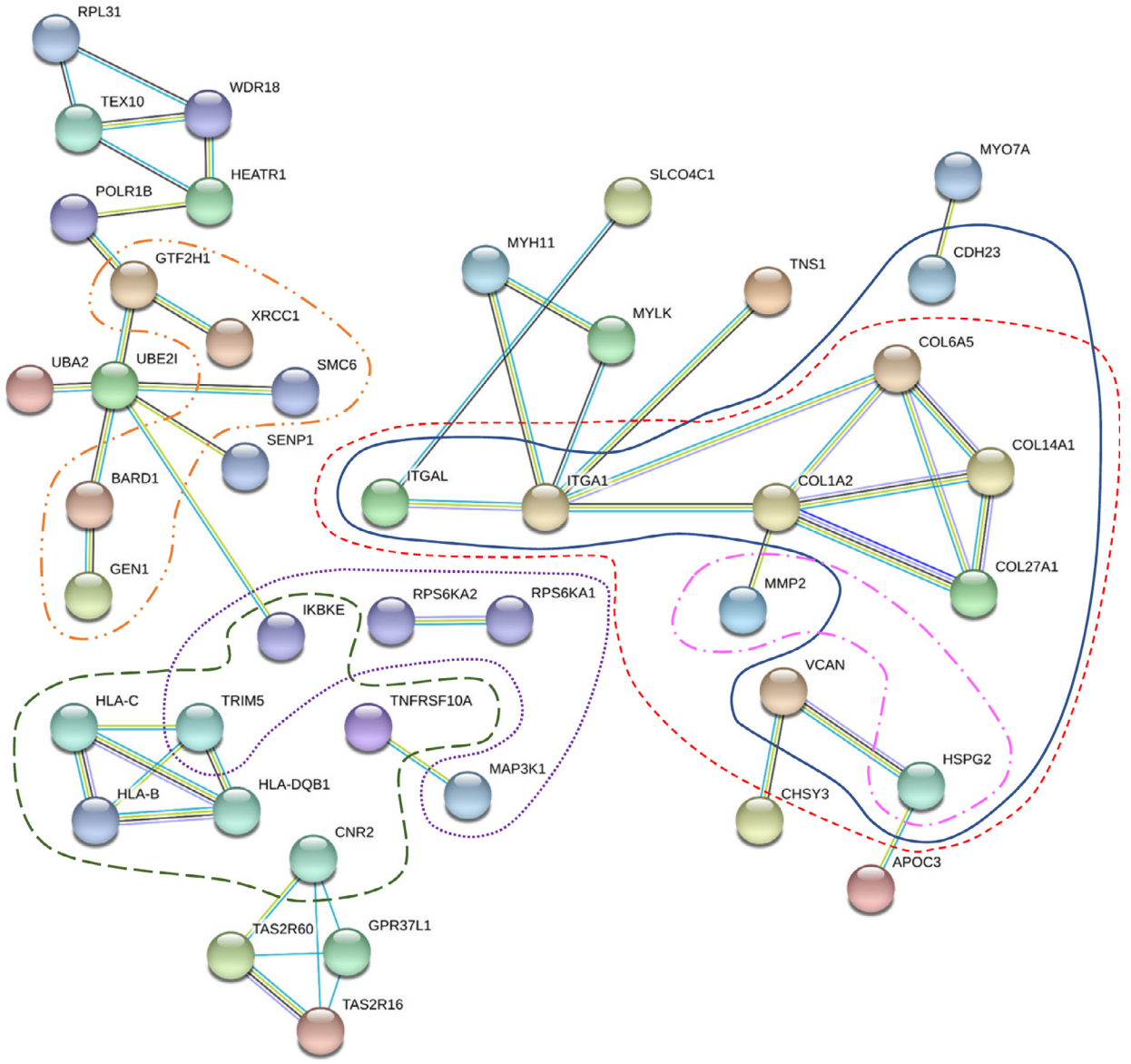

Fig. 2. Molecular network representing 6 major biological processes with relevance to periodontal disease identified in this study. A molecular interaction network of 64 among 247 candidate genes associated with periodontal disease was identified by using the STRING tool. The network includes 17 subnetworks. Of these 64 genes, 26 are part of six major biological processes identified by using the DAVID and GeneCodis tools. Blue line *(solid)* circled group of 9 genes represents "cell adhesion" biological process, and red line *(short dash)* represents "extracellular matrix organization" (10 genes). Both biological processes include 4 collagen genes. The pink line *(alternating dot and dash)* represents the biological process "angiogenesis" (2 genes). Orange line *(alternating double-dot and dash)* identifies the "DNA repair" process (5 genes). Dark green line *(long dash)* represents "immune response" (7 genes), and purple line *(dotted)* represents "innate immune response" (5 genes).

coding region, which could affect mRNA splicing and produce different protein isoforms. For the *COL14 A1* and *COL27 A* genes, we hypothesize that isoforms are weaker or nonfunctional because the 2 mutations are mainly identified in the PDP group of patients with HNC. Nonfunctional collagen proteins could affect the structure of the periodontium because collagen is one of the primary components of periodontal tissue.[53]

Furthermore, identification of 3 mutations in 3 collagen genes overrepresented in the PDP group suggests possible weakening of the collagen structure. RT could further weaken this structure as a result of its direct effects on the periodontium,[54,55] thereby contributing to PD progression in patients with HNC. As shown in the cell adhesion and extracellular biological processes (see Figure 2), collagen genes interact with other genes, including *MMP2*. The SNPs in *MMP2* identified in this

Table III. Annotation of the single nucleotide polymorphisms (SNPs) localized in the 26 genes associated with the 6 biological processes relevant to periodontal disease (PD) identified in this study

Genes	*Chr**	*Position**	*Ref*†	*Alt*†	*Rs ID*‡	*Protective*§
HSPG2	1	22216574	C	A	2254358	Y
	1	22216604	C	G	2254357	Y
CNR2	1	24200903	G	C	2229586	Y
	1	24200969	T	C	2229584	Y
	1	24200983	C	T	2229583	Y
	1	24201094	G	C	2229581	Y
	1	24201109	C	T	2229580	Y
	1	24201262	A	G	2502993	Y
	1	24201357	A	G	4649124	Y
	1	24201448	C	T	3003336	Y
	1	24201643	G	A	2501431	Y
IKBKE	1	206648378	G	A	1059704	N
RPS6 KA1	1	26900735	C	G	282179	N
BARD1	2	215645464	C	G	2229571	Y
GEN1/ SMC6	2	17954647	T	G	300178	Y
	2	17962518	C	T	300169	Y
	2	17942775	T	A	1812152	Y
COL6 A5	3	130162395	A	G	9883988	N
ITGA1	5	52157374	T	C	2447867	N
VCAN	5	82833145	G	A	2548541	N
MAP3 K1	5	56177443	G	A	702689	N
HLA-DQB1	6	32629802	A	G	1049092	Y
	6	32629935	C	G	1063322	Y
HLA-C	6	31237014	G	A	3998381	Y
	6	31237048	A	G	9264594	Y
	6	31237048	T	C	1130838	Y
	6	31237162	C	G	35708511	Y
	6	31237230	A	G	9264596	Y
HLA-B/HLA-C	6	31238851	T	C	2308604	Y
	6	31237323	A	G	68094471	Y
	6	31237333	T	A	66772001	Y
	6	31237405	C	T	56010430	Y
	6	31238135	G	A	1050317	Y
	6	31238138	C	T	1050320	Y
	6	31238230	G	C	1050716	Y
	6	31238234	G	A	1050344	Y
RPS6 KA2	6	167271716	T	C	9347162	N
COL1 A2	7	94040133	T	A	389328	Y
COL14 A1	8	121267617	C	T	993823	N
TNFRSF10 A	8	23049292	C	T	2230229	Y
COL27 A1	9	116958216	C	T	2567711	N
CDH23	10	73515103	G	T	3747858	Y
GTF2 H1	11	18379629	C	A	4150661	Y
TRIM5	11	5906048	G	A	4758168	N
	11	5906192	A	C	1453435	N
	11	5906073	G	T	4757986	N
ANXA2	15	60648214	T	C	12898604	N
ITGAL	16	30495412	T	C	2285459	Y
MMP2	16	55523705	T	C	243849	N
XRCC1	19	44079687	G	A	3213245	Y

The *HSPG2, CNR2, GEN1, SMC6, HLA-DQB1, HLA-C, HLA-B*, and *TRIM5* genes include 2 or more SNPs. *GEN1* and *SMC6* overlap the same 3 mutations, and the *HLA-B* and *HLA-C* genes overlap the same 8 mutations.

**Chr* and *Position* refer to the location of the SNP.

†*Ref* and *Alt* indicate the reference and the alternative allele of the SNP.

‡*Rs ID* refers to the accession number for each individual SNP.

§*Protective* specifies if the SNP is a "potentially protective" SNP against PD (Y = yes) or an SNP that could confer susceptibility to periodontal disease (PD) (N = no, "potentially detrimental").

study are localized in the protein-coding region as synonymous variants or intron variants. In case of the synonymous variant, the impact on protein coding will be low. In case of the intron variant, however, we could expect overexpression of the MMP2 gelatinase. For this reason, in patients carrying polymorphism rs243849, overrepresented in the PDP group, it could confer susceptibility to PD. In addition, Makela et al.[56] found that the MMP9 and MMP1 are highly present in the crevicular gingival fluid of patients with PD, thereby potentially contributing to PD when overactivity is not properly countered by tissue inhibitors of metalloproteinase (TIMP).[57]

Comparison of no-PDP and PDP groups with or without pre-RT PD

COL6 A5 appeared to be the only gene with a significant *P* value, overrepresented in patients with CAL progression greater than 0.2 mm, regardless of the presence of pre-RT PD. *COL6 A5* is also the only gene in the collagen network, with its SNP having a moderate impact compared with the other collagen genes containing modifier SNPs (see Table V). The moderate impact refers to a nondisruptive variant that might change the effectiveness of the protein.[47] As shown in Supplemental Table S3, except for *COL6 A5*, genes with significant *P* values in comparison C were not significant in comparison D, and vice versa. Indeed, in the PDP group of patients with pre-RT PD, mutations in *COL14 A1, COL6 A5, ITGA1, MMP2*, and *RPS6 KA2* could potentially have contributed to CAL increase (comparison C), whereas in patients without pre-RT PD, mutations located in *COL27 A1, COL6 A5, VCAN, IKBKE, RPS6 KA1*, and *TRIM5* (comparison D) would be responsible for CAL increase (see Supplemental Table S3).

In addition, we found that in the no-PDP group, the patient with HNC and the highest pre-RT PD status had the variant allele homozygous genotype potentially conferring protection in all the significant genes identified. However, the patient in the PDP group with highest pre-RT PD status presented the variant allele mostly in genes with detrimental SNPs. Furthermore, overall, patients in both no-PDP and PDP groups had similar representation of primary cancer sites (see Supplemental Table S1). Fregnani et al. have previously shown that IMRT affected the lower and upper teeth with an exposure ranging from 19.6 to 43.6 Gy (average [SD] = 32.82 [7.30]), suggesting that the progression of PD is not related to major differences between

Table IV. Variant effect of single nucleotide polymorphisms (SNPs) located in protein coding regions with moderate or low impact in the 26 unique genes of 6 biological processes with relevance to periodontal disease (PD) identified in this study

Feature type/biotype	*Impact*	*SO term**	*Genes*	*Rs ID†*
Transcript/protein coding	Moderate	Missense variant	BARD1	2229571
			COL6 A5	9883988
			GEN1	1812152
				300169
			HLA-B	1050716
			HLA-C	35708511
				1130838
				1050716
			HLA-DQB1	1049092
				1063322
			MAP3 K1	702689
			RPS6 KA2	9347162
			TNFRSF10 A	2230229
		Splice region variant	HLA-C	35708511
	Low	Synonymous variant	CNR2	2229581
				2229580
				2502993
				4649124
				3003336
				2501431
			HLA-B/HLA-C	1050317
				1050320
				1050344
			HLA-C	2308604
			HLA-DQB1	1049092
			HSPG2	2254358
				2254357
			IKBKE	1059704
			ITGA1	2447867
			MMP2	243849
			VCAN	2548541
		Splice region variant	HLA-B/HLA-C	68094471
				2308604
		3 prime UTR variant	HLA-B/HLA-C	1050344
				2308604

The table shows only the SNPs with the highest severity in terms of impact, among the 50 reported in Table III. Because the study did not identify any high impact, only moderate and low impacts are reported. The moderate effects result from missense polymorphisms and splice region variants. The low impacts are caused by synonymous polymorphisms, splice region, and 3' UTR variants.
UTR, untranslated region.
**SO* (Sequence Ontology) refers to a specific consequence that is assigned to each combination of a polymorphism and a transcript.
†*Rs ID* refers to the accession number for each individual SNP.

local teeth exposure overall.[58] Indeed, both patient 12 (no-PDP group, 12 teeth with ≥ 5 mm probing depth before RT) and patient 65 (PDP group, 14 teeth with ≥ 5 mm probing depth before RT) in this study were diagnosed with SCC of the submandibular gland (see Supplemental Table S1).

As exemplified by many SNP studies using different technologic platforms (e.g., full genome sequencing, SNP arrays), reproducibility is often achieved at the pathway level, not the gene level, especially with smaller sample sizes. Thus, with our patient cohort (N = 69), with the use of exome sequencing (testing for 11,934 filtered informative SNPs of 253,230 initially detected), we found SNPs in the previously identified genes *HLA-DQB1*,[43] *HLA-B*,[46] *COL14 A1*,[11] and *MMP2*[45] (see Table III). Moreover, the role of *MMP9* in PD has been described by Tarannum et al.,[46] and the roles of *COL11 A1* and *COL4 A6* in PD have been described by Garzon et al.[45] Although we did not identify SNPs in these genes, we identified *MMP2, COL1 A2, COL6 A5, COL14 A1*, and *COL27 A1* representing similar/connected pathways (see Figure 2).

Limitations

Although our results provide plausible insights into PD progression of RT-treated patients with HNC, the small sample size in this study is a limiting factor. In addition, although we had information regarding the total radiation dose delivered to the primary site, we did not have precise topological information about the radiation field associated with single tooth follow-up observations; thus, we recognize the limited granularity in our data regarding the direct local impact of RT on CAL.[59] Collection of extensive topologic radiation data on a large cohort of patients with HNC undergoing RT will be valuable to better investigate susceptibility to PD. Validation of the effects of SNPs at the functional level and how these effects alter the integrity of the periodontium will be critical to identifying drug targets and to serve as predictor of PD progression in RT-treated patients with HNC.[60]

CONCLUSIONS

Our data suggest that certain SNPs may confer protection, whereas others confer susceptibility to PD in RT-treated patients with HNC. Also, different sets of genes may be involved, depending on whether or not these patients had pre-RT PD. Thus, different preventive strategies for PD treatment may be considered. Mutations in the collagen genes could be important factors conferring susceptibility to PD in RT-treated patients with HNC, thereby representing potential targets for therapy.

Table V. Annotation of the significant single nucleotide polymorphisms (SNPs) identified in 4 collagen genes

Gene	*Rs ID**	*Feature type/biotype*	*Impact*	*Sequence Ontology*†	*Protective*‡
COL1 A2	389328	Transcript/protein coding	Modifier	Intron variant	Y
		Transcript/upstream gene variant	Modifier	Retained intron	Y
COL6 A5	9883988	Transcript/protein coding	Moderate§	Missense variant	N
		Transcript/nonsense mediated decay	Modifier‖	NMD transcript variant¶	N
COL14 A1	993823	Transcript/protein coding	Modifier	Intron variant	N
COL27 A1	2567711	Transcript/protein coding	Modifier	Intron variant	N
		Transcript/nonsense mediated decay	Modifier	NMD transcript variant	N

The table shows the potential effects of the 4 variants in the 4 collagen genes based on the regions where they are localized. rs9883988 polymorphism localized in *COL6 A5* is the only variant that could have a moderate impact in the transcriptome by causing a missense mutation. All the other variants have *modifier* impacts. rs389328 in *COL1 A2* is potentially a protective polymorphism.
**Rs ID* refers to the accession number for each individual SNP.
†*Sequence Ontology* refers to a specific consequence that is assigned to each combination of a polymorphism and a transcript.
‡*Protective* indicates if the SNP could potentially be beneficial ("protective") for the carriers (Y = yes) or could confer susceptibility to periodontal disease (N = no).
§*Moderate* refers to a nondisruptive variant that might change protein effectiveness.
‖*Modifier* indicates noncoding mutations or mutations affecting noncoding genes, where predictions for this type of variants are difficult or there is no evidence of impact.
¶*NMD* refers to nonsense-mediated decay.

ACKNOWLEDGMENTS

We are thankful to all those who contributed to this study: clinical staff at Carolinas Center for Oral Health at Atrium Health and University of Connecticut Health in Farmington, for performing CT for patient recruitment and saliva sample collection; Dr. James Hodges and other members of the Biostatistics Division at the University of Minnesota, for collection and preliminary analysis of clinical and demographic data; oral medicine research and molecular service core laboratories at Atrium Health, for saliva samples processing, exome sequencing, and providing the VCF files; and Dr. Xinghua Shi for her helpful discussion at the initial planning of the SNP analysis.

FUNDING

This study was supported by Atrium Health Foundation research fund.

REFERENCES

1. National Cancer Institute. Cancer stat facts: oral cavity and pharynx cancer. Available at: https://seer.cancer.gov/statfacts/html/oralcav.html. Accessed August 25, 2019.
2. Reis A, da Cruz AD. The impact of human papillomavirus on cancer risk in penile cancer. In: Vanden Broeck D, ed. *Human Papillomavirus and Related Diseases: From Bench to Bedside - A Clinical Perspective*, Rijeka, Croatia: InTech; 2012.
3. Memorial Sloan Kettering Cancer Center. Head & neck cancers. Available at: https://www.mskcc.org/cancer-care/types/head-neck. Accessed August 25, 2019.
4. National Cancer Institute. Head and neck cancers. Available at: https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet. Accessed August 25, 2019.
5. Bensadoun RJ, Patton LL, Lalla RV, Epstein JB. Oropharyngeal candidiasis in head and neck cancer patients treated with radiation: update 2011. *Support Care Cancer*. 2011;19:737-744.
6. Brailo V, Boras VV, Juras DV, Rogulj AA, Brzak BL, Alajbeg I. Oral side effects of head and neck irradiation. In: Akarslan Z, ed. *Head and Neck Cancer*, London, UK: IntechOpen; 2017.
7. Hong CH, Napenas JJ, Hodgson BD, et al. A systematic review of dental disease in patients undergoing cancer therapy. Oral Care Study Group, Multi-national Association of Supportive Care in Cancer (MASCC)/International Society of Oral Oncology (ISOO). *Support Care Cancer*. 2010;18:1007-1021.
8. Sroussi HY, Epstein JB, Bensadoun RJ, et al. Common oral complications of head and neck cancer radiation therapy: mucositis, infections, saliva change, fibrosis, sensory dysfunctions, dental caries, periodontal disease, and osteoradionecrosis. *Cancer Med*. 2017;6:2918-2931.
9. Kassim N, Sirajuddin S, Biswas S, Rafiuddin S, Apine A. Iatrogenic damage to the periodontium caused by radiation and radiotherapy. *Open Dent J*. 2015;9:182-186.
10. Sadetzki S, Chetrit A, Sgan-Cohen HD, et al. Long-term effects of exposure to ionizing irradiation on periodontal health status—the tinea capitis cohort study. *Front Public Health*. 2015;3:226.
11. da Silva MK, de Carvalho ACG, Alves EHP, da Silva FRP, Pessoa LDS, Vasconcelos DFP. Genetic factors and the risk of periodontitis development: findings from a systematic review composed of 13 studies of meta-analysis with 71,531 participants. *Int J Dent*. 2017;2017:1914073.
12. Armitage GC. Development of a classification system for periodontal diseases and conditions. *Ann Periodontol*. 1999;4:1-6.
13. European Federation of Periodontology. What is periodontitis? Available at: https://www.efp.org/patients/what-is-periodontitis.html. Accessed August 25, 2019.
14. Novak JM. Classification of diseases and conditions affecting the periodontium. In: Newman MG, Takei HH, Klokkevold PR, Carranza FA, eds. *Carranza's Clinical Periodontology*, 10th ed., St. Louis, MO: Saunders; 2006:100-109.
15. Caton JG, Armitage G, Berglundh T, et al. A new classification scheme for periodontal and peri-implant diseases and conditions—introduction and key changes from the 1999 classification. *J Periodontol*. 2018;89(suppl 1):S1-S8.
16. Fritz PC. Clinical attachment level—how to calculate and interpret this important measurement. *Oral Health*. 2013. October 1. https://www.oralhealthgroup.com/features/clinical-attachment-level-how-to-calculate-and-interpret-this-important-measurement/.
17. Machtei EE, Norderyd J, Koch G, Dunford R, Grossi S, Genco RJ. The rate of periodontal attachment loss in subjects with established periodontitis. *J Periodontol*. 1993;64:713-718.
18. Ammajan RR, Joseph R, Rajeev R, Choudhary K, Vidhyadharan K. Assessment of periodontal changes in patients undergoing radiotherapy for head and neck malignancy: a hospital-based study. *J Cancer Res Ther*. 2013;9:630-637.
19. Irie MS, Mendes EM, Borges JS, Osuna LG, Rabelo GD, Soares PB. Periodontal therapy for patients before and after radiotherapy: a review of the literature and topics of interest for clinicians. *Med Oral Patol Oral Cir Bucal*. 2018;23:e524-e530.

20. Trott JR, Cross HG. An analysis of the principle reasons for tooth extractions in 1813 patients in Manitoba. *Dent Pract Dent Rec*. 1966;17:20-27.
21. Divaris K, Monda KL, North KE, et al. Exploring the genetic basis of chronic periodontitis: a genome-wide association study. *Hum Mol Genet*. 2013;22:2312-2324.
22. Rhodin K, Divaris K, North KE, et al. Chronic periodontitis genome-wide association studies: gene-centric and gene set enrichment analyses. *J Dent Res*. 2014;93:882-890.
23. Lalla RV, Long-Simpson L, Hodges JS, et al. Clinical registry of dental outcomes in head and neck cancer patients (OraRad): rationale, methods, and recruitment considerations. *BMC Oral Health*. 2017;17:59.
24. American Academy of Periodontology Task Force Report on the Update to the 1999 Classification of Periodontal disease and conditions. *J. Periodontol*. 2015;86:835-838.
25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754-1760.
26. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589-595.
27. Ning Z, Cox AJ, Mullikin JC. SSAHA: a fast search method for large DNA databases. *Genome Res*. 2001;11:1725-1729.
28. Li H. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics*. 2012;28:1838-1844.
29. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCF tools. *Bioinformatics*. 2011;27:2156-2158.
30. Marees AT, de Kluiver H, Stringer S, et al. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int J Methods Psychiatr Res*. 2018;27:e1608.
31. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-575.
32. Social Science Statistics. Z Score calculator for 2 population proportions. Available at: https://www.socscistatistics.com/tests/ztest/Default2.aspx. Accessed August 25, 2019.
33. Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. Kaviar: an accessible system for testing SNV novelty. *Bioinformatics*. 2011;27:3216-3217.
34. Dayem Ullah AZ, Oscanoa J, Wang J, Nagano A, Lemoine NR, Chelala C. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Res*. 2018;46:W109-W113.
35. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607-D613.
36. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44-57.
37. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1-13.
38. Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A. GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acids Res*. 2012;40:W478-W483.
39. Nogales-Cadenas R, Carmona-Saez P, Vazquez M, et al. GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. *Nucleic Acids Res*. 2009;37:W317-W322.
40. Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biol*. 2007;8:R3.
41. Goncalves PF, Harris TH, Elmariah T, Aukhil I, Wallace MR, Shaddox LM. Genetic polymorphisms and periodontal disease in populations of African descent: a review. *J Periodontal Res*. 2018;53:164-173.
42. Nibali L, Di Iorio A, Tu YK, Vieira AR. Host genetics role in the pathogenesis of periodontal disease and caries. *J Clin Periodontol*. 2017;44:S52-S78.
43. Gandhi M, Kothiwale S. Association of periodontal diseases with genetic polymorphisms. *Int J Genet Eng*. 2012;2:19-27.
44. Song L, Yao J, He Z, Xu B. Genes related to inflammation and bone loss process in periodontitis suggested by bioinformatics methods. *BMC Oral Health*. 2015;15:105.
45. Garzón I, Roa A, Moreu G, et al. Development of a diagnostic algorithm in periodontal disease and identification of genetic expression patterns: a preliminary report. *J Dent Sci*. 2012;7:48-56.
46. Tarannum F, Faizuddin M. Effect of gene polymorphisms on periodontal diseases. *Indian J Hum Genet*. 2012;18:9-19.
47. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17:122.
48. Lorencini M, Silva JA, Almeida CA, Bruni-Cardoso A, Carvalho HF, Stach-Machado DR. A new paradigm in the periodontal disease progression: gingival connective tissue remodeling with simultaneous collagen degradation and fibers thickening. *Tissue Cell*. 2009;41:43-50.
49. Smith CL, Blake JA, Kadin JA, Richardson JE, Bult CJ, Mouse Genome Database G. Mouse Genome Database (MGD)—2018: knowledge base for the laboratory mouse. *Nucleic Acids Res*. 2018;46:D836-D842.
50. Honibald EN, Mathew S, Padmanaban J, Sundaram E, Ramamoorthy RD. Perioceutics: matrix metalloproteinase inhibitors as an adjunctive therapy for inflammatory periodontal disease. *J Pharm Bioallied Sci*. 2012;4:S417-S421.
51. GeneCards. *COL1 A2* gene (protein coding). Available at:https://www.genecards.org/cgi-bin/carddisp.pl?gene=COL1A2. Accessed August 25, 2019.
52. Baker KE, Parker R. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Curr Opin Cell Biol*. 2004;16:293-299.
53. Romanos GE, Bernimoulin JP. Collagen as a basic element of the periodontium: Immunohistochemical aspects in the human and animal. 1. Gingiva and alveolar bone. *Parodontol*. 1990;1:363-375. [in German].
54. Yusof ZW, Bakri MM. Severe progressive periodontal destruction due to radiation tissue injury. *J Periodontol*. 1993;64:1253-1258.
55. Epstein JB, Lunn R, Le N, Stevenson-Moore P. Periodontal attachment loss in patients after head and neck therapy. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 1998;86:676-677.
56. Makela M, Salo T, Uitto VJ, Larjava H. Matrix metalloproteinases (MMP-2 and MMP-9) of the oral cavity: cellular origin and relationship to periodontal status. *J Dent Res*. 1994;73:1397-1406.
57. Sapna G, Gokul S, Bagri-Manjrekar K. Matrix metalloproteinases and periodontal diseases. *Oral Dis*. 2014;20:538-550.
58. Fregnani ER, Parahyba CJ, Morais-Faria K, et al. IMRT delivers lower radiation doses to dental structures than 3 DRT in head and neck cancer patients. *Radiat Oncol*. 2016;11:116.
59. Marques MA, Dib LL. Periodontal changes in patients undergoing radiotherapy. *J Periodontol*. 2004;75:1178-1187.
60. Yoshie H, Kobayashi T, Tai H, Galicia JC. The role of genetic polymorphisms in periodontitis. *Periodontol 2000*. 2007;43:102-132.

Reprint requests:
Jean-Luc Charles Mougeot
Farah Bahrani Mougeot
Department of Oral Medicine
Carolinas Medical Center-Atrium Health
Charlotte
NC 28232-2861
USA.
jean-luc.mougeot@atriumhealth.org

Table S1. CAL change at different points, primary site of cancer, and total dose of radiation therapy for each HNC patient (N=69)

Patient IDs	*CAL change [a]T12 minus [b]T0*	*CAL change [c]T24 minus T0*	*CAL change T24 minus T12*	*[d]PDP group*	*Number of teeth with pd ≥5mm, T0*	*Primary cancer sites*	*Total RT dose to primary cancer site*
Pt 1	-0.001	0.115	0.116	No	2	Base of Tongue	7000
Pt 2	0.328	-0.095	-0.423	Yes	0	Tonsil	7000
Pt 3	0.113	0.518	0.405	Yes	0	Nasopharynx	7000
Pt 4	0.051	0.187	0.136	No	0	Neck	6000
Pt 5	0.428	0.637	0.209	Yes	0	Tonsil	7000
Pt 6	0.154	0.29	0.136	Yes	0	Base of Tongue	7000
Pt 7	0.313	0.16	-0.153	Yes	1	Tongue	5400
Pt 8	/	0.007	/	No	1	Tonsil	7000
Pt 9	/	-0.213	/	No	10	Tonsil	7000
Pt 10	-0.053	/	/	No	0	Tonsil	6600
Pt 11	-0.216	-0.259	-0.043	No	6	Tonsil	6600
Pt 12	-0.424	-0.865	-0.441	No	12	Submandibular Gland	6600
Pt 13	/	0.043	/	No	0	Base of Tongue	6996
Pt 14	-0.166	0.464	0.63	Yes	0	Mandible	6000
Pt 15	-0.239	0.083	0.322	Yes	3	Nasopharynx	7020
Pt 16	0.154	/	/	No	5	Nasopharynx	7000
Pt 17	0.095	0.137	0.042	No	0	Base of Tongue	7000
Pt 18	0.012	0.141	0.129	No	7	Parotid	5525
Pt 19	-0.15	0.03	0.18	No	0	Mandible	6600
Pt 20	-0.721	-0.388	0.333	Yes	3	Larynx	7000
Pt 21	-0.445	-0.112	0.333	Yes	4	Tongue	6000
Pt 22	0.365	/	/	Yes	0	Pharynx	7000
Pt 23	-0.775	0.425	1.2	Yes	1	Tonsil	7000
Pt 24	0.256	/	/	Yes	0	Buccal/Labial Mucosa	6600
Pt 25	/	-0.47	/	No	0	Tonsil	7000
Pt 26	0.146	0.688	0.542	Yes	0	Base of Tongue	6000
Pt 27	/	/	0.071	No	0	Tonsil	7000
Pt 28	0.577	-0.269	-0.846	Yes	0	Tonsil	7000
Pt 29	0.101	0.28	0.179	Yes	0	Parotid	6600
Pt 30	0.612	1.056	0.444	Yes	1	Buccal/Labial Mucosa	6000
Pt 31	-0.347	-0.253	0.094	No	0	Nasopharynx	6996
Pt 32	0.711	1.961	1.25	Yes	0	Tonsil	6996
Pt 33	0.321	0.208	-0.113	Yes	0	Buccal Mucosa	5000
Pt 34	0.748	1.869	1.121	Yes	1	Unknown	7000
Pt 35	0.333	/	/	Yes	0	Oral Tongue	5400
Pt 36	-0.411	0.351	0.762	Yes	1	Oral Tongue	6000
Pt 37	0.572	0.656	0.084	Yes	1	Tonsil	7000
Pt 38	0.03	-0.143	-0.173	Nes	0	Tonsil	7000
Pt 39	0.93	0.944	0.014	Yes	0	Base of Tongue	6996
Pt 40	0.432	/	/	Yes	1	Nasopharynx	7000
Pt 41	1.245	/	/	Yes	2	Submandibular Gland	6000
Pt 42	0.696	/	/	Yes	0	Parotid	6000
Pt 43	-0.255	/	/	No	1	Base of Tongue	7000
Pt 44	-0.055	/	/	No	1	Base of Tongue	6000
Pt 45	0.166	/	/	No	1	Oral Cavity	7000
Pt 46	0.037	/	/	No	2	Neck	6996
Pt 47	-0.574	0.173	0.747	Yes	4	Unknown	7000
Pt 48	-1.97	-1.298	0.672	Yes	4	Unknown	6600
Pt 49	-0.284	-0.007	0.277	Yes	2	Base of Tongue	7000
Pt 50	-1.215	-0.804	0.411	Yes	2	Tonsil	6600
Pt 51	-0.089	0.185	0.274	Yes	0	Not specified/Neck	5940
Pt 52	0.089	0.327	0.238	Yes	0	Tonsil	6600
Pt 53	0.833	0.893	0.06	Yes	0	Base of Tongue	7000
Pt 54	0.165	0.543	0.378	Yes	3	Base of Tongue	7000
Pt 55	0.317	0.893	0.576	Yes	/	Oropharynx	7000
Pt 56	0.136	/	/	No	0	Nasopharynx	7000
Pt 57	0.03	0.22	0.19	Yes	0	Oral Cavity	7000

(*continued*)

Table S1. Continued

Patient IDs	*CAL change [a]T12 minus [b]T0*	*CAL change [c]T24 minus T0*	*CAL change T24 minus T12*	*[d]PDP group*	*Number of teeth with pd ≥5mm, T0*	*Primary cancer sites*	*Total RT dose to primary cancer site*
Pt 58	-0.288	0.581	0.869	Yes	0	Oral Cavity	6000
Pt 59	0.284	0.342	0.058	Yes	4	Tonsil	7000
Pt 60	0.375	0.417	0.042	Yes	4	Major Salivary Gland	6600
Pt 61	0.284	0.884	0.6	Yes	2	Unknown	7000
Pt 62	0.764	1.486	0.722	Yes	0	Base of Tongue	7000
Pt 63	-0.618	0.215	0.833	Yes	2	Tonsil	6000
Pt 64	0.178	0.982	0.804	Yes	0	Oropharynx	7000
Pt 65	0.438	0.77	0.332	Yes	14	Submandibular Gland	6300
Pt 66	1.282	0.917	-0.365	Yes	11	Hypopharynx/Larynx/ Supraglottis	7000
Pt 67	0.1	0.234	0.134	Yes	4	Tonsil	7000
Pt 68	0.316	/	/	Yes	0	Base of Tongue	7000
Pt 69	0.369	/	/	Yes	0	Submandibular Gland	6600

aT12 refers to visit at 12 months after radiation therapy.
bT0 refers to visit prior to radiation therapy.
cT24 refers to visit at 24 months after radiation therapy.
dPDP (periodontal disease progression) group includes patients showing an annual change of CAL >0.2mm following radiation therapy (RT). pd is pocket depth.

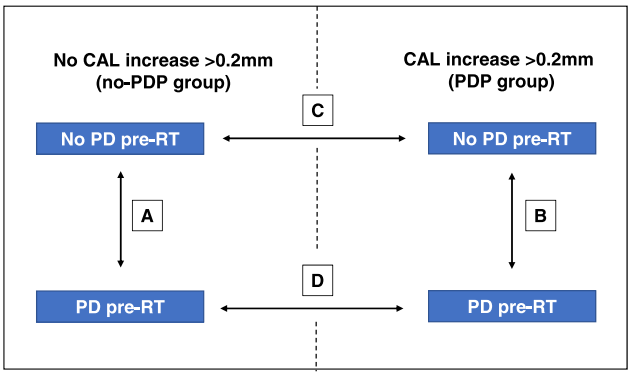

Figure S1. Comparisons (A-D) of SNP prevalence in RT-treated HNC patients who exhibited periodontal disease progression (PDP) *vs.* those who did not (no-PDP), whether or not they had PD pre-RT.

Comparison-A was performed between the patients who did not have PD pre-RT and those who had PD pre-RT within the no-PDP group. **Comparison-B** compared the patients who did not have PD pre-RT to the patients who did within the PDP group. **Comparison-C** was performed between no-PDP and PDP patients who did not have PD pre-RT. **Comparison-D** was done between no-PDP and PDP patients who had PD pre-RT.

CAL= Clinical Attachment Loss
RT= Radiation Therapy
PD= Periodontal Disease

Table S2a. List of 64 gene symbols detected by STRING analysis and the corresponding gene names

Gene Symbol	*Gene name*
ANXA2	Annexin A2
APOC3	Apolipoprotein C3
BARD1	BRCA1 Associated RING Domain 1
CC2D2A	Coiled-Coil and C2 Domain Containing 2A
CDH23	Cadherin Related 23
CHSY3	Chondroitin Sulfate Synthase 3
CNR2	Cannabinoid Receptor 2
COL14A1	Collagen Type XIV Alpha 1 Chain
COL1A2	Collagen Type I Alpha 2 Chain
COL27A1	Collagen Type XXVII Alpha 1 Chain
COL6A5	Collagen Type VI Alpha 5 Chain
GEN1	GEN1 Holliday Junction 5′ Flap Endonuclease
GPR37L1	G Protein-Coupled Receptor 37 Like 1
GTF2H1	General Transcription Factor IIH Subunit 1
HBE1	Hemoglobin Subunit Epsilon 1
HBG2	Hemoglobin Subunit Gamma 2
HEATR1	HEAT Repeat Containing 1
HECW2	HECT, C2 And WW Domain Containing E3 Ubiquitin Protein Ligase 2
HLA-B	Major Histocompatibility Complex, Class I, B
HLA-C	Major Histocompatibility Complex, Class I, C
HLA-DQB1	Major Histocompatibility Complex, Class II, DQ Beta 1
HSPG2	Heparan Sulfate Proteoglycan 2
IKBKE	Inhibitor of Nuclear Factor Kappa B Kinase Subunit Epsilon
INPP1	Inositol Polyphosphate-1-Phosphatase
ITGA1	Integrin Subunit Alpha 1
ITGAL	Integrin Subunit Alpha L
KATNAL1	Katanin Catalytic Subunit A1 Like 1
KATNB1	Katanin Regulatory Subunit B1
KLHL5	Kelch Like Family Member 5
KRT4	Keratin 4
KRT9	Keratin 9
MAP3K1	Mitogen-Activated Protein Kinase Kinase Kinase 1
MARS	Methionyl-TRNA Synthetase
MMP2	Matrix Metallopeptidase 2
MYH11	Myosin Heavy Chain 11
MYLK	Myosin Light Chain Kinase
MYO7A	Myosin VIIA
NDE1	NudE Neurodevelopment Protein 1
NOTCH3	Notch Receptor 3
POGLUT1	Protein O-Glucosyltransferase 1
POLR1B	RNA Polymerase I Subunit B
RPL31	Ribosomal Protein L31
RPS6KA1	Ribosomal Protein S6 Kinase A1
RPS6KA2	Ribosomal Protein S6 Kinase A2
SARS	Seryl-TRNA Synthetase
SENP1	SUMO Specific Peptidase 1
SLC18A2	Solute Carrier Family 18 Member A2
SLCO4C1	Solute Carrier Organic Anion Transporter Family Member 4C1
SMC6	Structural Maintenance of Chromosomes 6
SYN3	Synapsin III
SYNJ2	Synaptojanin 2
TAS2R16	Taste 2 Receptor Member 16
TAS2R60	Taste 2 Receptor Member 60
TEX10	Testis Expressed 10
TNFRSF10A	TNF Receptor Superfamily Member 10a
TNS1	Tensin 1
TRIM5	Tripartite Motif Containing 5
TUBGCP6	Tubulin Gamma Complex Associated Protein 6
UBA2	Ubiquitin Like Modifier Activating Enzyme 2
UBE2I	Ubiquitin Conjugating Enzyme E2 I
VARS2	Valyl-TRNA Synthetase 2, Mitochondrial
VCAN	Versican
WDR18	WD Repeat Domain 18
XRCC1	X-Ray Repair Cross Complementing 1

(*continued*)

Table S2a. Continued

Table S2b. List of 49 gene symbols associated with periodontal disease from literature and corresponding gene names

Gene Symbol	*Gene name*
BPI	Bactericidal Permeability Increasing Protein
CCL5	C-C Motif Chemokine Ligand 5
CCR10	C-C Motif Chemokine Receptor 10
CCR5	C-C Motif Chemokine Receptor 5 (Gene/Pseudogene)
CD14	CD14 Molecule
CDH16	Cadherin 16
CDH2	Cadherin 2
CDH7	Cadherin 7
COL4A6	Collagen Type IV Alpha 6 Chain
CTLA4	Cytotoxic T-Lymphocyte Associated Protein 4
CTSS	Cathepsin S
CYBA	Cytochrome B-245 Alpha Chain
DEFB1	Defensin Beta 1
EMR1	Adhesion G Protein-Coupled Receptor E1
FCAR	Fc Fragment of IgA Receptor
FCGR2A	Fc Fragment of IgG Receptor IIa
FCGR2B	Fc Fragment of IgG Receptor IIb
FCGR3A	Fc Fragment of IgG Receptor IIIa
FCGR3B	Fc Fragment of IgG Receptor IIIb
FGB	Fibrinogen Beta Chain
HLA-B	Major Histocompatibility Complex, Class I, B
HLA-C	Major Histocompatibility Complex, Class I, C
ICAM1	Intercellular Adhesion Molecule 1
IL10	Interleukin 10
IL18	Interleukin 18
IL1A	Interleukin 1 Alpha
IL1B	Interleukin 1 Beta
IL1RN	Interleukin 1 Receptor Antagonist
IL2	Interleukin 2
IL3	Interleukin 3
IL4	Interleukin 4
IL6	Interleukin 6
IL8	C-X-C Motif Chemokine Ligand 8
LTA	Lymphotoxin Alpha
LTF	Lactotransferrin
MICA	MHC Class I Polypeptide-Related Sequence A
MMP9	Matrix Metallopeptidase 9
NCR2	Natural Cytotoxicity Triggering Receptor 2
NOD2	Nucleotide Binding Oligomerization Domain Containing 2

(*continued*)

Table S2b. Continued

Gene Symbol	*Gene name*
NPY	Neuropeptide Y
S100A8	S100 Calcium Binding Protein A8
TENM2	Teneurin Transmembrane Protein 2
TLR4	Toll Like Receptor 4
TNF	Tumor Necrosis Factor
TNFRSF11B	TNF Receptor Superfamily Member 11b
TNFSF14	TNF Superfamily Member 14
TRIM5	Tripartite Motif Containing 5
VAV1	Vav Guanine Nucleotide Exchange Factor 1
WNT5A	Wnt Family Member 5A

Table S3. SNP over-representation of PD-related genes among radiation treated HNC patient subgroups

Gene ([e]Pathway[f] Protective)	*[a]Comparison-A (Z-test / Chi-squared p values)*	*[b]Comparison-B (Z-test/ Chi-square p values)*	*[c]Comparison-C (Z-test/Chi-square p values)*	*[d]Comparison-D (Z-test/ Chi-square p values)*
COL14A1 (E,C / N)			[g]Yes (0.0131/0.0130)	
COL27A1 (E,C / N)				Yes (0.0108/0.0108)
COL6A5 (E,C / N)			Yes (0.0444/0.0447)	Yes (0.0114/0.0115)
ITGA1 (E,C / N)		Yes (0.0203/0.0201)	Yes (0.0071/0.0071)	
VCAN (E,C / N)	Yes (0.0198/0.0197)			Yes (0.0039/0.0038)
MMP2 (E / N)			Yes (0.0188/0.0188)	
ANXA2 (E / N)				
IKBKE (IM, IN / N)				Yes (0.0198/0.0200)
RPS6KA1 (IN / N)				Yes (0.0039/0.0038)
RPS6KA2 (IN / N)			Yes (0.0238/0.0239)	
MAP3K1 (IN / N)				
TRIM5 (IN / N)				Yes (0.0039/0.0038)
COL1A2 (E,C / Y)				Yes (0.0293/0.0.0292)
HSPG2 (E,C / Y)				
ITGAL (E,C / Y)				Yes (0.0019/0.0019)
CDH23 (C / Y)			Yes (0.0278/0.0275)	
CNR2 (IM / Y)				Yes (0.0278/0.0281)
HLA-DQB1 (IM / Y)				Yes (0.0039/0.0038)
HLA-B (IM / Y)			Yes (0.0466/0.0464)	
HLA-C (IM / Y)			Yes (0.0466/0.0464)	
TNFRSF10A (IM / Y)				Yes (0.0067/0.0067)

The table reports genes in four different biological processes (cell adhesion, extracellular matrix organization, immune response, and innate immune response) with p values< 0.05, obtained in each type of comparison using two-tailed z-test and chi-squared test. Only one significant gene out of 21 genes (4.76%) is found for **Comparison-A** and **-B**. Eight out of 21 genes (38.09%) are significant in **Comparison-C** and 11 out of 21 genes are significant in **Comparison-D** (52%).

a**Comparison-A** refers to the comparison between patients without PD at baseline and patients with PD at baseline, overall not showing a progression in CAL over 24 months.

b**Comparison-B** refers to the comparison between patients without PD at baseline and patients with PD at baseline, overall showing an increment of CAL >0.2 mm over 24 months.

c**Comparison-C** refers to the comparison between patients without progression of CAL >0.2 mm and patients with CAL progression, overall not having PD at baseline.

d**Comparison-D** refers to the comparison between patients without progression of CAL>0.2 mm and patients with CAL progression, overall having PD at baseline.

e[d] indicates the four different pathways (E stands for extracellular matrix organization, C for cell adhesion, IM for immune response, IN for innate immune response).

fProtective indicates if a gene carries a potentially protective SNP genotype (Y=yes) or not (N=no).PD refers to periodontal disease status defined as a patient having at least one tooth with $\geq$5mm per previously published clinical study (OraRad study U01DE022939) and PD study by the American Academy of Periodontology Task Force [23, 24].