RESEARCH REPORTS
Clinical

M. Kebschull[1,7], P. Guarnieri[2,3], R.T. Demmer[4], A.L. Boulesteix[5], P. Pavlidis[6], and P.N. Papapanou[1]*

[1]Division of Periodontics, Section of Oral and Diagnostic Sciences, Columbia University College of Dental Medicine, New York, NY, USA; [2]Biomedical Informatics Shared Resources, Bioinformatics Division, Herbert Irving Comprehensive Cancer Center; [3]Columbia Initiative in Systems Biology, Columbia University Medical Center, New York, NY, USA; [4]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA; [5]Computational Molecular Medicine Research Group, Department of Medical Informatics, Biometry and Epidemiology, University of Munich, Munich, Germany; [6]UBC Centre for High-Throughput Biology, Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada; and [7]Present address, Department of Periodontology, Operative and Preventive Dentistry, University of Bonn, Bonn, Germany; *corresponding author, pp192@columbia.edu

J Dent Res 92(12):1081-1088, 2013

Molecular Differences between Chronic and Aggressive Periodontitis

ABSTRACT

The 2 major forms of periodontitis, chronic (CP) and aggressive (AgP), do not display sufficiently distinct histopathological characteristics or microbiological/immunological features. We used molecular profiling to explore biological differences between CP and AgP and subsequently carried out supervised classification using machine-learning algorithms including an internal validation. We used whole-genome gene expression profiles from 310 'healthy' or 'diseased' gingival tissue biopsies from 120 systemically healthy non-smokers, 65 with CP and 55 with AgP, each contributing with ≥ 2 'diseased' gingival papillae (n = 241; with bleeding-on-probing, probing depth ≥ 4 mm, and clinical attachment loss ≥ 3 mm), and, when available, a 'healthy' papilla (n = 69; no bleeding-on-probing, probing depth ≤ 4 mm, and clinical attachment loss ≤ 4 mm). Our analyses revealed limited differences between the gingival tissue transcriptional profiles of AgP and CP, with genes related to immune responses, apoptosis, and signal transduction overexpressed in AgP, and genes related to epithelial integrity and metabolism overexpressed in CP. Different classifying algorithms discriminated CP from AgP with an area under the curve ranging from 0.63 to 0.99. The small differences in gene expression and the highly variable classifier performance suggest limited dissimilarities between established AgP and CP lesions. Future analyses may facilitate the development of a novel, 'intrinsic' classification of periodontitis based on molecular profiling.

KEY WORDS: pathogenesis, gene expression, transcriptome, microarray analysis, classification, machine learning.

DOI: 10.1177/0022034513506011

Received June 25, 2013; Last revision August 27, 2013; Accepted August 27, 2013

A supplemental appendix to this article is published electronically only at http://jdr.sagepub.com/supplemental.

INTRODUCTION

Two principal forms of periodontitis are currently recognized, chronic (CP) and aggressive periodontitis (AgP). Despite differences in their clinical phenotypes (Armitage and Cullinan, 2010), no unequivocal pathophysiological foundation that differentiates between CP and AgP has been established. Chronic and aggressive periodontitis lesions cannot be distinguished on the basis of histopathologic features (Smith *et al.*, 2010) or microbial colonization profiles (Armitage, 2010), although there is evidence of immunological differences, including the presence of neutrophil abnormalities in AgP (Ryder, 2010).

In other fields, most notably in oncology, tumors of similar histology but disparate clinical behavior are distinguishable based on genome-wide molecular analyses. Specifically, the study of gene expression signatures in tumors with different prognosis has led to an improved classification of breast cancer and has translated into improved therapeutic management (Prat *et al*., 2012).

We hypothesized that the 2 currently accepted major forms of periodontitis, CP and AgP, display characteristic gene expression signatures that allow for a molecular distinction between the 2 entities and reveal differences in underlying pathophysiology.

To address this hypothesis, we utilized whole-genome gene expression profiles of gingival tissue samples from patients with CP or AgP. Based on gingival transcriptomes, we then performed a supervised classification of CP and AgP using machine-learning algorithms.

MATERIALS & METHODS

We used 310 Affymetrix HG-U133Plus2.0 microarray samples (GEO accession number GSE16134) of 'healthy' [n = 69, no bleeding on probing (BoP), probing depth (PD) ≤ 4 mm, and clinical attachment loss (CAL) ≤ 4 mm)] or 'diseased' gingival tissue samples (n = 241; with BoP, PD ≥ 4 mm, and CAL ≥ 3 mm), obtained from 120 non-smoking, systemically healthy individuals with moderate/severe periodontitis (65 with CP and 55 with AgP), as previously described (Demmer *et al.*, 2008; Kebschull and Papapanou, 2010). Excellent correlation of array data and confirmatory qPCR was shown previously (Papapanou *et al.*, 2009).

The diagnosis of either CP or AgP was assigned according to established criteria (Armitage, 1999) after review of patient history and clinical/radiographic records. The characteristics of the participants are described in Appendix Table 1, and those of the sampled sites in Appendix Table 2.

Pre-processed data were analyzed for differential expression by R/Bioconductor (Gentleman *et al.*, 2004) and limma (Smyth, 2004). Patients were modeled as random effects to account for the within-subject correlation of the tissue samples. To account for a potential influence of aging on gingival tissue gene expression – since AgP patients were younger than CP patients – we replicated the aforementioned analyses correcting for participant age.

The resulting ranked gene lists were analyzed for enrichment of functional groups by GSEA (Subramanian *et al.*, 2005) and visualized using a Cytoscape/Enrichment Map (Shannon *et al.*, 2003; Merico *et al.*, 2011).

To test the supervised classification performance, we compared several class prediction algorithms with CP/AgP as the dependent variable using the CMA package (Slawski *et al.*, 2008).

First, the available patients (n = 120) were randomly assigned to either a training set (consisting of 2/3 of all patients) or an evaluation set (the remaining patients). The sets were then populated with all samples belonging to the assigned patients, ensuring that evaluation of the classifiers was performed on samples not previously used for training. The partitioning procedure into training/evaluation sets was repeated 1,000 times.

For each partition, the selection of features (*i.e.*, genes) that could best distinguish between the 2 entities was performed based on the training set using limma's moderated *t* statistics. Different classifier algorithms were then applied to the training set with the top-ranking genes as predictors.

Several candidate classifiers were considered, including: (i) diagonal linear discriminant analysis (DLDA); (ii) partial least-squares (PLS) analysis combined with linear discriminant analysis (PLS-LDA), with the number of PLS components as the tuning parameter; (iii) shrunken centroids discriminant analysis, with the shrinkage parameter as the tuning parameter (scDA, without preliminary feature selection); or (iv) a support vector machine with a linear kernel (SVM), with the cost as the tuning parameter. Where applicable, parameter tuning was performed by 3-fold cross-validation as implemented in the CMA package.

Evaluation of different classifiers and variable numbers of features was then performed in the evaluation set based on sensitivity/specificity of AgP detection, and ROC area under the curve (AUC).

A detailed description is provided in the online Appendix.

RESULTS

AgP and CP Lesions Display Limited Differences

To identify potential differences between AgP and CP at the transcriptome level, we first compared gene expression in 'diseased' tissue samples. This analysis identified a total of 248 differentially regulated probes at an absolute fold change (FC) of ≥ 1.19 ($\log_2$FC 0.25) and a false discovery rate (FDR) of < 0.05 (Tables 1 and 2; Supplemental Material 1, Results from Differential Expression Analyses, panel 1a). The magnitude of the differences was generally modest, with 30 overexpressed and only one underexpressed probe by an absolute change of >1.5 fold in AgP *vs*. CP lesions. A probe coding for Fc receptor-like 5 (FCRL5) was the most overexpressed (by 1.56-fold) and the yet-uncharacterized chromosome 4 open reading frame 26 (C4orf26) was the most repressed probe (by 1.65-fold) in AgP *vs*. CP lesions.

To assess the functional relevance of the identified genes, we performed gene set enrichment analysis. Twenty-five functional groups were significantly enriched in AgP lesions and 14 in CP lesions (Fig. 1; Supplemental Material 1, panels 1b, 1c). Notably, gene sets linked to apoptosis, immune response, and signal transduction were enriched in AgP lesions, while genes sets signifying epithelial integrity and cellular metabolism were enriched in CP lesions.

We then compared clinically 'healthy' gingival tissues from the two disease entities and observed that only 11 probes were differentially expressed at an absolute FC of ≥ 1.19 and an FDR of < .05 (Supplemental Material 1, panel 2), 5 of which were also differentially expressed between 'diseased' tissues.

Next, we identified differentially expressed genes between 'diseased' and 'healthy' gingival tissues in the entire database, regardless of periodontal diagnosis, as well as separately in each diagnostic category (AgP and CP). When 'diseased' and 'healthy' gingival tissues from all 120 patients were compared, regardless of diagnosis, 9,258 probes were differentially expressed at the above-described thresholds (Supplemental Material 1, panel 5). The same comparison yielded 9,833 differentially expressed probes in the subset of the 55 AgP patients, and 9,161 probes in the subset of 65 CP patients (Supplemental Material 1, panel 3). Of these, a majority (8,602 probes) were differentially regulated in a similar fashion (3,520 commonly up-regulated, 5,082 down-regulated) between 'diseased' and 'healthy' gingival tissues in both CP and AgP.

Supplemental Material 1 (panels 6a/6b and 7a/7b) lists all probes that were exclusively differentially regulated between healthy and diseased gingival tissues in one of the two diagnostic categories. The above analyses were repeated after adjustment for age, yielding essentially identical results (data not shown).

Supervised Classification of CP and AgP

To examine whether molecular differences could reliably discriminate CP from AgP lesions, we proceeded with a supervised classification of the 2 disease entities.

Discrimination of the 2 forms of disease ranged from fair to excellent, depending on algorithm and number of features used (Fig. 2, Appendix Figs. 1-3; Supplemental Material 2, Results from Machine Learner Analyses). It was found that scDA and DLDA performed substantially worse than did SVM or PLS-LDA.

DISCUSSION

This study represents the first systematic evaluation of molecular differences between the 2 principal forms of periodontitis, chronic and aggressive. We used high-throughput technology and assessed simultaneously the expression level of > 38,000

Table 1. Probe Sets Overexpressed in 'Diseased' Gingival Tissues of AgP *vs.* CP Patients (FDR < .05, absolute fold change ≥ 1.25)

Name	Fold Change	*p* value	BH Adjusted *p* value
Fc receptor-like 5	1.57	2.09E-05	1.37E-03
Period circadian clock 3	1.46	6.73E-10	2.04E-06
Nuclear receptor subfamily 1, group D, member 2	1.46	4.45E-08	3.25E-05
Tyrosinase-related protein 1	1.46	5.43E-03	4.42E-02
Kelch-like family member 6	1.46	9.54E-07	1.92E-04
Ankyrin repeat domain 36B pseudogene 2	1.45	2.39E-04	6.21E-03
Churchill domain containing 1	1.44	6.81E-11	4.65E-07
Flavin containing monooxygenase 1	1.43	6.46E-11	4.65E-07
Fc receptor-like 5	1.42	9.19E-07	1.88E-04
Nuclear receptor subfamily 1, group D, member 2	1.41	2.67E-13	1.39E-08
Immunoglobulin kappa constant	1.41	4.27E-04	9.05E-03
Fc receptor-like A	1.41	1.00E-04	3.58E-03
Regulator of G-protein signaling 1	1.40	2.20E-04	5.92E-03
Membrane-spanning 4-domains, subfamily A, member 1	1.40	1.14E-04	3.87E-03
Immunoglobulin kappa constant	1.38	1.03E-03	1.55E-02
Fc receptor-like A	1.38	2.26E-05	1.42E-03
Immunoglobulin heavy constant μ	1.38	9.22E-04	1.45E-02
Ankyrin repeat domain 36B pseudogene 2	1.37	1.08E-05	9.47E-04
Lymphocyte transmembrane adaptor 1	1.36	2.56E-05	1.54E-03
KIAA0125	1.35	7.45E-04	1.26E-02
Aldehyde dehydrogenase 1 family, member L2	1.35	7.78E-07	1.65E-04
Membrane metallo-endopeptidase	1.35	1.45E-06	2.53E-04
Fc receptor-like 5	1.34	1.68E-04	4.99E-03
CD38 molecule	1.33	1.74E-05	1.22E-03
CD79a molecule, immunoglobulin-associated alpha	1.33	1.77E-04	5.14E-03
ras homolog family member H	1.33	1.99E-04	5.58E-03
Membrane-spanning 4-domains, subfamily A, member 1	1.33	1.95E-05	1.31E-03
Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	1.32	3.63E-04	8.10E-03
Membrane-spanning 4-domains, subfamily A, member 1	1.32	1.97E-05	1.32E-03
Protease, serine, 35	1.32	5.30E-05	2.41E-03
Platelet/endothelial cell adhesion molecule 1	1.31	1.57E-07	6.19E-05
CD79a molecule, immunoglobulin-associated alpha	1.31	3.77E-04	8.31E-03
Cytidine monophospho-N-acetylneuraminic acid hydroxylase, pseudogene	1.31	1.79E-08	1.74E-05
Fc receptor-like 5	1.30	1.76E-03	2.16E-02
Secreted frizzled-related protein 4	1.30	6.34E-05	2.71E-03
BMS1 homolog, ribosome assembly protein (yeast) pseudogene	1.30	4.65E-03	3.97E-02
Tumor necrosis factor receptor superfamily, member 17	1.30	5.23E-03	4.30E-02
Vimentin	1.29	3.65E-06	4.64E-04
Glucocorticoid induced transcript 1	1.29	4.52E-05	2.17E-03
Sperm associated antigen 4	1.29	1.76E-03	2.16E-02
Family with sequence similarity 46, member C	1.29	2.28E-03	2.54E-02
Period circadian clock 3	1.28	2.90E-08	2.27E-05
FBJ murine osteosarcoma viral oncogene homolog B	1.28	2.94E-03	2.98E-02
Regulator of G-protein signaling 1	1.28	8.58E-04	1.39E-02
sel-1 suppressor of lin-12-like (*C. elegans*)	1.28	1.43E-04	4.53E-03
Glycine dehydrogenase (decarboxylating)	1.28	5.66E-08	3.73E-05
SLAM family member 7	1.28	1.63E-03	2.05E-02
Marginal zone B and B1 cell-specific protein	1.28	1.45E-03	1.92E-02
Zinc finger protein 215	1.28	6.87E-06	6.95E-04
sel-1 suppressor of lin-12-like 3 (*C. elegans*)	1.28	1.03E-03	1.55E-02
Polycystic kidney and hepatic disease 1 (autosomal recessive)-like 1	1.28	9.20E-06	8.49E-04
POU class 2 associating factor 1	1.28	2.65E-03	2.78E-02
Cytidine monophospho-N-acetylneuraminic acid hydroxylase, pseudogene	1.27	1.32E-07	5.70E-05

(continued)

Table 1. (continued)

Name	Fold Change	*p* value	BH Adjusted *p* value
Transmembrane protein 156	1.27	4.55E-09	7.56E-06
Programmed cell death 6 pseudogene	1.27	7.60E-05	3.03E-03
Prostate androgen-regulated mucin-like protein 1	1.27	1.27E-04	4.17E-03
DnaJ (Hsp40) homolog, subfamily C, member 3	1.27	1.48E-04	4.62E-03
Contactin 3 (plasmacytoma-associated)	1.27	1.69E-07	6.31E-05
Signal transducing adaptor family member 1	1.27	5.56E-05	2.49E-03
DnaJ (Hsp40) homolog, subfamily B, member 9	1.27	8.33E-05	3.19E-03
Activating transcription factor 3	1.26	4.39E-03	3.83E-02
RAB30, member RAS oncogene family	1.26	2.44E-06	3.51E-04
Immunoglobulin lambda variable 1-44	1.26	1.60E-03	2.03E-02
CD27 molecule	1.26	9.10E-04	1.44E-02
Chromosome 1 open reading frame 51	1.26	2.00E-06	3.10E-04
GTP binding protein overexpressed in skeletal muscle	1.26	5.00E-08	3.49E-05
Carbohydrate (N-acetylglucosamine-6-O) sulfotransferase 2	1.26	1.44E-04	4.57E-03
Cytohesin 1 interacting protein	1.26	5.67E-04	1.08E-02

Eighty-six probe sets were significant at FDR [Benjamini-Hochberg (BH) adjusted *p* value] < .05, and 14 probe sets met the Bonferroni genome-wide significance threshold (unadjusted *p* value < 9.15E-07). Note that several probes may map to the same gene (for detailed information on the individual probes, see Supplemental Material 1).

genes in clinically 'healthy' or 'diseased' gingival tissue biopsies from a well-phenotyped sample of 120 patients with either CP or AgP who were non-smokers and systemically healthy. Importantly, the analyzed tissue samples were harvested from patients who did not undergo initial non-surgical treatment, and thus reflect the gingival tissue transcriptomes of untreated periodontitis. After rigorous quality control, we analyzed differences in gene expression between gingival lesions from CP and AgP, and carried out analyses to validate the diagnostic classification at the molecular level.

We used several different approaches to identify molecular patterns that distinguish between the 2 forms. The primary analysis assessed expression differences, as well as enrichment of functional groups, between gingival lesions from AgP *vs*. CP patients. Next, we investigated differences in transcriptomes from healthy gingival tissues obtained from AgP or GP patients. In a third step, we examined whether the differences in gene expression between healthy and diseased tissues were similar in the two diagnostic categories.

The primary analysis showed relatively few significantly differentially regulated probes between AgP and CP lesions, all with limited fold changes. To put this observation into perspective, in an earlier comparison between 'diseased' and 'healthy' gingival tissue samples, we observed >10,000 differentially regulated probe sets after Bonferroni correction, with up to five-fold change in expression (Demmer *et al.*, 2008). Nevertheless, these findings cannot preclude the presence of biological differences between the 2 entities, since we compared *established* CP and AgP lesions with similar clinical characteristics. Thus, the obtained transcriptomes may fail to reflect the processes that led to various progression rates at different ages, and may partly represent homeostatic mechanisms rather than disease activity.

We found several immunoglobulin-related overexpressed genes in AgP *vs.* CP lesions, while genes supporting epithelial integrity and focal adhesion (desmocollin 1, laminin γ2, keratin 2, LCE2B) and metabolism-related pathways were underexpressed. Next, we observed that healthy gingiva from CP or AgP patients displayed largely similar expression patterns and pathways. Finally, only a limited number of probes were differentially regulated between healthy and diseased gingival tissues exclusively in only one of the two disease entities—for example, the epithelial and bone integrity-related genes BMP7 and keratin-17 that were found to be repressed only in 'diseased' (as compared with 'healthy') tissues of AgP patients, but not in those of CP patients.

When compared with CP gingival lesions, AgP lesions were largely characterized by the expression of 3 major clusters of functional groups, related to immune response, signal transduction, and programmed cell death. One of the immune response pathways found to be enriched in AgP lesions was the B-cell receptor signaling. B cells are considered critical players in periodontal pathogenesis (Berglundh *et al.*, 2007) and have been shown to be elevated in both localized and generalized AgP *vs*. CP lesions (Sigusch *et al.*, 2006). Analysis of our data indicating a 20% increased expression in the B-cell surface marker CD19 in AgP corroborated these findings. Further, several of the top genes identified in the AgP-CP comparison belonged to the immunoglobulin families, encoding for structural elements of antibodies produced by B cells. Several genes with increased expression in AgP are important for antibody production by B cells, including MZB1 (marginal zone B and B1 cell-specific protein), CD79a, FC receptor-like A, and POU2AF1. Likewise, the induction of FK506 binding protein 11 in AgP suggests a plasmacytic differentiation of B cells. A potential explanation for these differences is that AgP lesions may have been

Table 2. Probe Sets Underexpressed in 'Diseased' Gingival Tissues of AgP *vs.* CP patients (FDR < .05, absolute fold change ≥ 1.25)

Name	Fold Change	*p* value	BH Adjusted *p* value
Chromosome 4 open reading frame 26	–1.65	5.75E-09	8.98E-06
Cytochrome P450, family 3, subfamily A, polypeptide 5	–1.47	2.53E-08	2.16E-05
Keratin 2	–1.47	3.05E-03	3.06E-02
Hemoglobin, beta	–1.44	1.22E-06	2.29E-04
Hemoglobin, beta	–1.42	9.31E-07	1.89E-04
Endoplasmic reticulum aminopeptidase 2	–1.39	4.88E-04	9.82E-03
Histidine ammonia-lyase	–1.38	2.09E-04	5.75E-03
Odontogenic, ameloblast-asssociated	–1.38	4.40E-05	2.14E-03
Microsomal glutathione S-transferase 1	–1.37	2.16E-04	5.87E-03
Hemoglobin, beta	–1.36	1.96E-06	3.07E-04
Filaggrin	–1.36	1.27E-03	1.76E-02
Filaggrin family member 2	–1.36	3.12E-03	3.09E-02
Uncharacterized LOC654433	–1.35	2.43E-07	8.05E-05
Matrix metallopeptidase 28	–1.34	2.69E-09	6.12E-06
Desmocollin 1	–1.33	6.48E-03	4.91E-02
Neurofilament, light polypeptide	–1.32	2.40E-04	6.21E-03
Solute carrier family 13 (sodium-dependent citrate transporter), member 5	–1.32	2.05E-06	3.16E-04
Regulator of G-protein signaling 4	–1.31	1.85E-05	1.27E-03
Uncharacterized LOC654433	–1.30	3.72E-06	4.67E-04
Uroplakin 1A	–1.30	7.38E-04	1.26E-02
Endoplasmic reticulum aminopeptidase 2	–1.30	2.82E-03	2.90E-02
Aryl hydrocarbon receptor nuclear translocator-like	–1.29	5.10E-08	3.49E-05
Pancreatic lipase-related protein 3	–1.28	1.49E-05	1.12E-03
MIR205 host gene (non-protein coding)	–1.28	1.47E-10	7.29E-07
Late cornified envelope 2B	–1.27	4.71E-04	9.59E-03
Glutathione S-transferase theta 2	–1.27	8.59E-04	1.39E-02
Microsomal glutathione S-transferase 1	–1.27	1.94E-04	5.47E-03
Microsomal glutathione S-transferase 1	–1.26	2.47E-04	6.33E-03
Chromosome 6 open reading frame 132	–1.26	9.32E-05	3.42E-03
Defensin, beta 1	–1.26	1.39E-06	2.48E-04
Regulator of G-protein signaling 4	–1.26	3.49E-04	7.91E-03
Uncharacterized LOC100130476	–1.26	2.44E-04	6.29E-03
MIR205 host gene (non-protein coding)	–1.25	2.11E-06	3.22E-04
Hemoglobin, delta	–1.25	2.41E-07	8.04E-05
Family with sequence similarity 213, member A	–1.25	1.95E-06	3.06E-04
DEAD (Asp-Glu-Ala-Asp) box helicase 17	–1.24	5.54E-03	4.47E-02

Thirty-six probe sets were significant at FDR [Benjamini-Hochberg (BH) adjusted *p* value] < .05, and 7 probe sets met the Bonferroni genome-wide significance threshold (unadjusted *p* value < 9.15E-07). Note that several probes may map to the same gene (for detailed information on the individual probes, see Supplemental Material 1).

harvested at a state of 'disease activity' more frequently than CP lesions. Alternatively, these differences may represent responses to different bacterial challenges, as shown in our earlier work (Papapanou *et al.*, 2009).

Moreover, analysis of our data indicates pronounced induction of NK-cell-mediated cytotoxicity in AgP lesions. Interestingly, activation of NK cells was recently associated with increased alveolar bone loss (Chaushu *et al.*, 2012). Supporting this notion, we recently reported that a prominent activating receptor for NK cells (SLAMF7) was significantly induced in AgP lesions (Kramer *et al.*, 2013) and identified invariant NK T-cells to be differentially activated in AgP *vs.* CP (Nowak *et al.*, 2013). In agreement with this observation, the NK-cell-related gene NKTR, as well as the receptor for interferon gamma (the principal cytokine secreted by NK cells), was down-regulated in diseased *vs.* healthy gingival tissues exclusively in CP, but not in AgP.

Signal transduction pathways including Jun and Wnt signaling appeared to be differentially regulated in AgP. Thus, Jun and Jun-related genes (JUND, TNIK) were found to be up-regulated in diseased *vs.* healthy gingival tissues exclusively in AgP but not in CP patients, while Wnt-signaling-related genes such as SFRP4 were overexpressed in AgP *vs.* CP gingival lesions.

The strong enrichment of apoptosis-related pathways in AgP gingival lesions could conceivably relate to the severe loss of

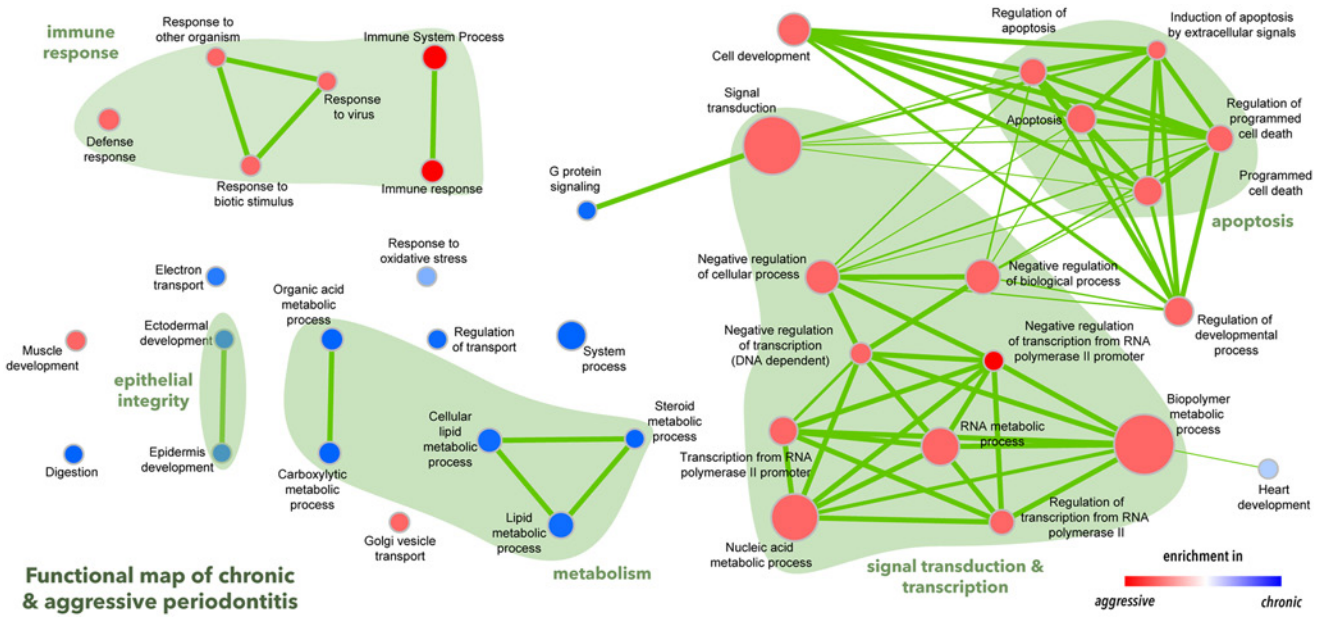

Figure 1. Enrichment map of differences between chronic and aggressive periodontitis lesions. Visualization of gene sets (from the Broad Institute's Molecular Signature Database, collection C2 containing 4,722 curated gene sets) significantly ($p < .05$) enriched in diseased gingival tissues from patients with chronic or aggressive periodontitis. Gene sets are depicted as nodes in a network. Color describes the disease entity (red for AgP and blue for CP), and the color intensity represents the degree of enrichment. The size of the node represents the size of the enriched gene set, and the thickness of the connectors stands for the degree of overlap between the nodes (Merico *et al.*, 2011). The only gene groups that fulfilled the Bonferroni threshold of $p < 1.05\text{x}10\text{E-}05$ were immune response in AgP, and 7 metabolism and epithelium-related groups in CP (see Supplemental Material 1).

attachment that commonly occurs in this disease entity. This finding is in line with prior evidence from *in vitro* studies demonstrating pronounced activation of apoptotic pathways in several cell types after infection with *Aggregatibacter actinomycetemcomitans* that is intimately associated with AgP (Kebschull and Papapanou, 2011).

Nevertheless, it must be emphasized that the differential expression of several genes between the 2 forms of periodontitis alone does not necessarily demonstrate the presence of a biologically distinct 'intrinsic' disease classification. To formally investigate whether transcriptional profiles of CP and AgP gingival lesions are sufficiently distinct to form individual entities, we adopted a classification approach using machine learning classifier algorithms followed by internal validation procedures (*i.e.*, repeated splitting into training and evaluation sets). Gene-expression-based classifiers have been frequently explored in recent years for several pathologic conditions, and resulted in FDA-approved diagnostics for early breast cancer (Gluck *et al.*, 2012). However, there is still no universally accepted methodology for this task, and it is impossible to determine *a priori* the optimal classifier for a particular dataset. Therefore, we used 4 different algorithms to distinguish between AgP and CP lesions and present the diagnostic performance of each, as suggested by good practice guidelines (Dupuy and Simon, 2007), since exclusive focus on the performance of the best classifier results in substantial optimistic bias (Boulesteix and Strobl, 2009). Thus, we acknowledge that application of the algorithms to a new, independent validation dataset would likely lead to worse discriminatory capacity than the one indicated by the largest AUC observed in this study. Interestingly, the ability to distinguish between AgP and CP varied considerably depending on the type of machine learner and the number of features involved. Excellent discrimination was achieved with computationally intensive, tuned algorithms, such as SVM, that incorporate large feature sets, while other classifiers yielded substantially higher misclassification rates. The variable discriminatory performance of the tested algorithms suggests that the molecular differences between AgP and CP were not trivially 'learnable'. This may suggest that the 2 forms represent heterogeneous entities that include multiple subclasses with individual molecular signatures that are diluted when aggregated under the current classification scheme.

We acknowledge that inferences based on linear associations between and among mRNA, protein expression, and phenotypes are crude and simplistic. For example, translation is influenced by small regulatory RNAs, and protein signatures are subject to post-translational modifications. Thus, although expression profiling is a powerful means of analyzing potential molecular mechanisms, it may still fail to account for substantial biological variability. Our group has recently embarked upon the analysis of miRNA expression and function in gingival tissues (Stoecklin-Wasmer *et al.*, 2012). Ultimately, transcriptomic data of gingival tissues, along with data on miRNA expression and target genes, epigenetic regulation, and protein expression, may form the basis for a systems biology approach to the study of the complex pathobiology of periodontitis. Future research will focus on the

biological significance of the identified differentially expressed genes between CP and AgP. Alternatively, identification of *de novo* patient clusters sharing common molecular patterns that translate into distinct clinical phenotypes may lead to a novel, 'intrinsic' classification of periodontitis.

ACKNOWLEDGMENTS

The authors thank Dr. Werner Adler, Erlangen, Germany, for support with the classifier algorithms. This research was supported by the National Institutes of Health (NIH grants DE-015649, DE-021 820, UL1-TR000040) and Colgate-Palmolive, NJ, USA. MK was also supported by the German Research Foundation (KFO208, TP6 & TP9), the German Society for Periodontology (DGParo), the German Society for Oral and Maxillofacial Sciences (DGZMK), and the Neue Gruppe; RTD by NIH K99 DE-018739; and PP by the Michael Smith Foundation for Health Research.

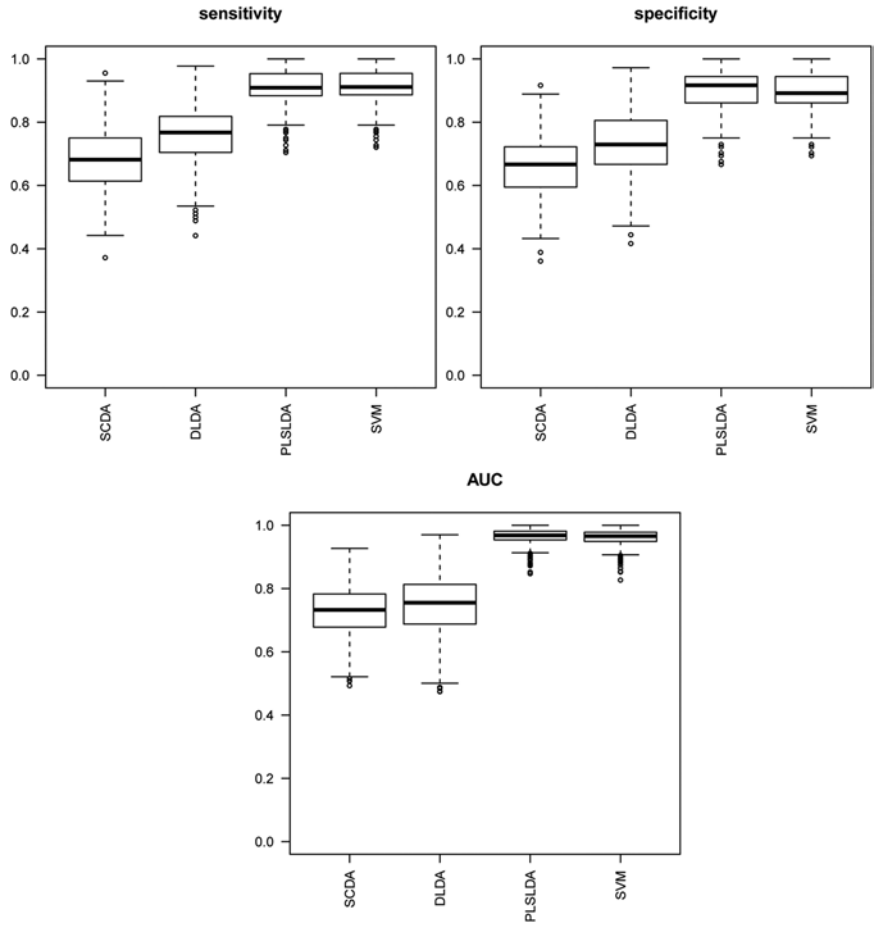

Figure 2. Microarray-based classification of AgP and CP gingival lesions. Four different microarray classifier algorithms were trained to distinguish gingival lesions from AgP or CP patients based on their whole-transcriptome expression profiles. For each of the 1,000 splittings into training/evaluation sets that accounted for multiple tissue samples *per* participant, variable selection was performed based on the training set using a mixed-effects linear model. Subsequently, 4 different classifier algorithms [diagonal linear discriminant analysis (DLDA), partial least-squares analysis combined with linear discriminant analysis (PLS-LDA), shrunken centroids discriminant analysis (scDA), or support vector machines (SVM)] were trained on the training set to distinguish between AgP and CP gingival lesions based on 250 genes (DLDA, PLS-LDA, and SVM). Performance of the algorithms in the classification of the corresponding evaluation sets was then assessed by the sensitivity and specificity of AgP detection, as well as by (ROC) area under the curve (AUC).

REFERENCES

Armitage GC (1999). Development of a classification system for periodontal diseases and conditions. *Ann Periodontol* 4:1-6.

Armitage GC (2010). Comparison of the microbiological features of chronic and aggressive periodontitis. *Periodontol* 2000 53:70-88.

Armitage GC, Cullinan MP (2010). Comparison of the clinical features of chronic and aggressive periodontitis. *Periodontol* 2000 53:12-27.

Berglundh T, Donati M, Zitzmann N (2007). B cells in periodontitis: friends or enemies? *Periodontol* 2000 45:51-66.

Boulesteix AL, Strobl C (2009). Optimal classifier selection and negative bias in error rate estimation: an empirical study on high-dimensional prediction. *BMC Med Res Methodol* 9:85.

Chaushu S, Wilensky A, Gur C, Shapira L, Elboim M, Halftek G, *et al*. (2012). Direct recognition of *Fusobacterium nucleatum* by the NK cell natural cytotoxicity receptor NKp46 aggravates periodontal disease. *PLoS Pathog* 8:e1002601.

Demmer RT, Behle JH, Wolf DL, Handfield M, Kebschull M, Celenti R, *et al*. (2008). Transcriptomes in healthy and diseased gingival tissues. *J Periodontol* 79:2112-2124.

Dupuy A, Simon RM (2007). Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 99:147-157.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, *et al*. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.

Gluck S, Yip AY, Ng EL (2012). Can we replace the microscope with microarrays for diagnosis, prognosis and treatment of early breast cancer? *Expert Opin Ther Targets* 16(Suppl 1):17-22.

Kebschull M, Papapanou PN (2010). The use of gene arrays in deciphering the pathobiology of periodontal diseases. *Methods Mol Biol* 666:385-393.

Kebschull M, Papapanou PN (2011). Periodontal microbial complexes associated with specific cell and tissue responses. *J Clin Periodontol* 38(Suppl 11):17-27.

Kramer B, Kebschull M, Nowak M, Demmer RT, Haupt M, Korner C, *et al*. (2013). Role of the NK cell-activating receptor CRACC in periodontitis. *Infect Immun* 81:690-696.

Merico D, Isserlin R, Bader GD (2011). Visualizing gene-set enrichment results using the Cytoscape plug-in enrichment map. *Methods Mol Biol* 781:257-277.

Nowak M, Kramer B, Haupt M, Papapanou PN, Kebschull J, Hoffmann P, *et al*. (2013). Activation of invariant NK T cells in periodontitis lesions. *J Immunol* 190:2282-2291.

Papapanou PN, Behle JH, Kebschull M, Celenti R, Wolf DL, Handfield M, *et al*. (2009). Subgingival bacterial colonization profiles correlate with gingival tissue gene expression. *BMC Microbiol* 9:221.

Prat A, Ellis MJ, Perou CM (2012). Practical implications of gene-expression-based assays for breast oncologists. *Nat Rev Clin Oncol* 9:48-57.

Ryder MI (2010). Comparison of neutrophil functions in aggressive and chronic periodontitis. *Periodontol* 2000 53:124-137.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al*. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498-2504.

Sigusch BW, Wutzler A, Nietzsch T, Glockmann E (2006). Evidence for a specific crevicular lymphocyte profile in aggressive periodontitis. *J Periodontal Res* 41:391-396.

Slawski M, Daumer M, Boulesteix AL (2008). CMA: a comprehensive Bioconductor package for supervised classification with high dimensional data. *BMC Bioinformatics* 9:439.

Smith M, Seymour GJ, Cullinan MP (2010). Histopathological features of chronic and aggressive periodontitis. *Periodontol* 2000 53:45-54.

Smyth GK (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:Article3. [Epub 2004 Feb 12, 2004].

Stoecklin-Wasmer C, Guarnieri P, Celenti R, Demmer RT, Kebschull M, Papapanou PN (2012). MicroRNAs and their target genes in gingival tissues. *J Dent Res* 91:934-940.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al*. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102:15545-15550.