Eur J Oral Sci 2004; 112: 216–223
Printed in UK. All rights reserved

European Journal of
Oral Sciences

Gene expression signatures in chronic and aggressive periodontitis: a pilot study

Panos N. Papapanou[1], Armin Abron[1], Miguel Verbitsky[2], Doros Picolos[1], Jun Yang[1], Jie Qin[2], James B. Fine[1], Paul Pavlidis[2]

[1]Division of Periodontics, Section of Oral and Diagnostic Sciences, School of Dental and Oral Surgery, Columbia University, New York, USA; [2]Genome Center, Columbia University, New York, USA

Papapanou PN, Abron A, Verbitsky M, Picolos D, Yang J, Qin J, Fine JB, Pavlidis P. Gene expression signatures in chronic and aggressive periodontitis: a pilot study. Eur J Oral Sci 2004; 112: 216–223. © Eur J Oral Sci, 2004

This pilot study examined gene expression signatures in pathological gingival tissues of subjects with chronic or aggressive periodontitis, and explored whether new subclasses of periodontitis can be identified based on gene expression profiles. A total of 14 patients, seven with chronic and seven with aggressive periodontitis, were examined with respect to clinical periodontal status, composition of subgingival bacterial plaque assessed by checkerboard hybridizations, and levels of serum IgG antibodies to periodontal bacteria assayed by checkerboard immunoblotting. In addition, at least two pathological pockets/patient were biopsied, processed for RNA extraction, amplification and labeling, and used to study gene expression using Affymetrix U-133 A arrays. Based on a total of 35 microarrays, no significantly different gene expression profiles appeared to emerge between chronic and aggressive periodontitis. However, a *de novo* grouping of the 14 subjects into two fairly robust clusters was possible based on similarities in gene expression. These two groups had similar clinical periodontal status and subgingival bacterial profiles, but differed significantly with respect to serum IgG levels against the important periodontal pathogens *Porphyromonas gingivalis*, *Tannerella forsythensis* and *Campylobacter rectus*. These early data point to the usefulness of gene expression profiling techniques in the identification of subclasses of periodontitis with common pathobiology.

Panos N. Papapanou, Division of Periodontics, Section of Oral and Diagnostic Sciences, Columbia University School of Dental and Oral Surgery, 630 West 168th Street, PH-7E-110, New York, NY 10032, USA

Telefax: +1–212–3059313
E-mail: pp192@columbia.edu

Key words: periodontitis; gene expression; microarray; classification

Accepted for publication February 2002

A few years ago, a revised classification of periodontal diseases was introduced by an international workshop (1). The revised system abolished several previously accepted categories such as early onset periodontitis, adult, rapidly progressive and refractory periodontitis, and termed the two major forms of periodontal disease chronic and aggressive periodontitis.

The advantages of the new classification system when compared to its 10-yr-old predecessor were discussed in detail by ARMITAGE (2). However, a persisting important shortcoming of the new system is the fact that it lacks an unequivocal, pathobiology-based foundation. Instead, the definition of the two main forms of periodontitis is still based on combinations of clinical signs and symptoms of questionable discriminatory ability. As underscored in recent publications (2,3), there is a clear need for additional research focusing on the biological basis underlying the distinction between the different periodontal disease entities. Such studies can be greatly facilitated by the use of contemporary microarray technology.

Gene expression profiling is a powerful means of generating comprehensive genome-level data sets on complex diseases and has provided enormous insights in other fields in which it has been applied, mostly cancer research (4). There are two general analytic approaches that have been used in this context: The first is class validation, where a given classification scheme which is based on more 'traditional' diagnostic methods is explored at the level of gene expression. These studies can provide support for, or refute, the biological relevance of a diagnostic classification scheme, and can also identify specific genes which are relevant to the distinction between classes. For example, recent work by VAN'T VEER *et al.* (5) identified genes whose expression is associated with metastatic, as opposed to localized, breast cancer, and can also be used to predict survival (6), resulting in the development of a new diagnostic test for breast cancer (7). Similarly, POMEROY *et al.* (8) found gene expression signatures relevant to the clinical outcome in cases of embryonal tumors of the central nervous system. The second analytical approach is class discovery, in which a set of samples with no previous diagnostic distinctions are examined for evidence of new classifications, which can then be related to relevant clinical variables. For example, ALIZADEH *et al.* (9) showed that diffuse large B-cell lymphomas could be classified in two groups based on gene expression data, and that these groups differed in long-term prognosis. PEROU *et al.* (10) found that otherwise heterogeneous breast cancer samples could be classified into a small number of distinct groups characterized by major differences in gene

expression. Such successful applications of microarrays have not been restricted to cancer research. Recent studies have examined muscular dystrophy (11), Alzheimer's disease (12), arthritis (13), and asthma (14).

To our knowledge, a gene expression-based approach has not yet been applied in the classification of periodontitis. We thus initiated a study to: (i) examine whether distinct gene expression signatures can be identified in pathological gingival tissues obtained from subjects suffering from chronic or aggressive periodontitis; and (ii) explore the feasibility of a new classification scheme based on similarities in gene expression. The aims of the present pilot study were to: (i) determine the within-subject variance in gene expression obtained from different pathological sites; (ii) obtain preliminary data on differences in gene expression between chronic and aggressive periodontitis; and (iii) explore new subclasses of disease based on gene expression profiles.

Material and methods

The design and procedures of the study were approved by the Columbia University Institutional Review Board.

Subjects

A total of 14 subjects, seven with chronic periodontitis and seven with aggressive periodontitis, were recruited among the patients referred for periodontal therapy to the Clinic for Post-doctoral Periodontics, Columbia University School of Dental and Oral Surgery, NY, USA. A periodontist (author P. N. P.) screened the referral records and the accompanying full-mouth series of intraoral radiographs of incoming patients to make a preliminary assessment of eligibility for participation. Eligible patients were (i) at least 13 yr old; (ii) had a minimum of 24 teeth present; (iii) had no past history of periodontal therapy other than prophylaxis provided by the referring general dentist; (iv) had received no systemic antibiotics or anti-inflammatory drugs for at least 6 months; (v) harbored a minimum of four teeth with radiographic bone loss; (vi) did not suffer from diabetes mellitus; (vii) did not suffer from any of the systemic conditions or genetic disorders that entail the diagnosis of periodontitis associated with systemic disease (15); (viii) were not pregnant; and (ix) were not current users of tobacco products or of nicotine replacement medication. Patients interested in participating signed an informed consent form prior to enrolment. The demographic characteristics of the patients are shown in Table 1.

Clinical examination and periodontal diagnosis

All participants underwent a full mouth examination of the periodontal tissues at six sites per tooth, using a manual probe. The examination included assessments of presence/absence of dental plaque and bleeding on probing, and linear measurements of probing depth and clinical attachment level.

Two periodontists (authors P.N.P. and J.B.F.) jointly reviewed the radiographic and clinical records of each patient, and assigned a consensus diagnosis of either chronic or aggressive periodontitis according to the established criteria (2). Specifically, aggressive periodontitis patients were required to: (i) be systemically healthy; (ii) display a severity of disease disproportionate to the amount of local etiologic factors; and (iii) show evidence of rapid attachment loss and bone loss either with respect to their age, or based on comparisons of available sets of clinical or radiographic records obtained at different time points. In addition, a reported history of familial aggregation of periodontal disease was interpreted as strongly suggestive of aggressive periodontitis. In all other cases, a diagnosis of chronic periodontitis was assigned. Following the assignment of diagnosis, patients were further subdivided according to the extent of attachment and bone loss into a localized or a generalized periodontitis subgroup (16, 17).

Gingival tissue donor areas

Following the initial examination and oral hygiene instructions, maxillary segments of the dentition that required periodontal surgery were identified, according to established treatment criteria, i.e. on the basis of deep inflamed pockets with loss of periodontal tissue support. In these areas, interproximal sites with a probing depth ranging between 6 mm and 8 mm and clinical attachment level between 4 mm and 6 mm were selected. A single surgical region encompassing three to four posterior consecutive teeth with sites that fulfilled the criteria above were selected. No subgingival instrumentation was carried out at the surgical region before tissue sample collection, to avoid affecting the local gene expression at the sampled sites.

Blood and subgingival plaque samples

Five milliliters of blood were collected by venepuncture (18). Individual subgingival plaque samples were obtained from the mesio-palatal surface of each tooth present (excluding third molars) in the quadrant where the segment selected for surgery was located, and from each tooth in its diagonally opposite lower quadrant (14 sites per subject), according to the description by Papapanou *et al*. (18). An additional four plaque samples were harvested from the disto-palatal aspects of the four consecutive teeth included in the surgical area. Thus, a total of 18 subgingival plaque samples were obtained from each subject.

Collection of gingival tissue samples

After infiltration with an appropriate local anesthetic, submarginal incisions 1–2 mm apical to the gingival margin were made at the segment of the dentition that harbored the qualifying sites. After reflection of mucoperiosteal flaps, the portion of each interproximal gingival papilla that adhered to the root surface was carefully dissected. This section comprised the ulcerated epithelial lining of the periodontal pocket and a portion of the underlying connective tissue. After dissection, the gingival tissue specimens obtained were thoroughly rinsed with sterile normal saline solution and transferred into Eppendorf tubes containing five volumes of a RNA stabilization reagent (RNA*later*; Ambion, Austin, TX, USA). A total of four specimens were collected from each surgical area, including three specimens from qualifying sites according to the criteria outlined above, and one from an adjacent site displaying a probing depth of ≤ 3 mm, attachment loss of ≤ 2 mm, and no bleeding on probing, obtained from the portion of the periodontal flap reflected for mere access purposes. After collection of the gingival tissue specimens, pocket elimination/reduction

periodontal surgery was completed according to standard procedures. The standard postsurgical protocol was followed.

RNA extraction, reverse transcription, *in vitro* cRNA synthesis

The gingival tissue specimens were stored in RNA*later* overnight at 4°C and thereafter snap-frozen and stored in liquid nitrogen. All further processing occurred simultaneously for gingival biopsies originating from patients with chronic and aggressive periodontitis. Gingival tissue specimens were homogenized in Trizol reagent (Invitrogen Life Technologies, Carlsbad, CA, USA). After incubation with chloroform and centrifugation at 12 000 *g*, RNA collected in the upper aqueous phase was precipitated by mixing with isopropyl alcohol, further centrifugation and washing in 75% ethanol. Further purification of the extracted RNA was achieved by the use of the RNeasy total RNA isolation kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. The purified RNA was quantified spectrophotometrically. Five micrograms of total RNA originating from each of two diseased samples was reverse transcribed using the SuperScript double stranded cDNA synthesis kit (Invitrogen) according to the manufacturer's instructions. The T7-$(dT)_{24}$ oligomer (Proligo; Genset, LaJolla, CA, USA) was used for priming first strand cDNA synthesis in place of the oligo (dT) or random primers provided with the SuperScript kit. Cleanup of the double-stranded cDNA was achieved by phenol–chloroform extraction with the use of Phase Lock Gels (Eppendorf, Hamburg, Germany) followed by ethanol precipitation. Synthesis of botin-labeled cRNA was performed by the use of the Enzo BioArray High Yield RNA Transcript Labeling kit (Affymetrix, Santa Clara, CA, USA). Cleaning up of the *in vitro* transcription (IVT) products was carried out by using RNeasy spin columns (Qiagen) followed by ethanol precipitation. The cRNA yield was determined spectrophotometrically at 260 nm. Finally, the cRNA was fragmented by incubation in fragmentation buffer (200 mM Tris acetate, pH 8.1, 500 mM K acetate, 150 mM Mg acetate) at 94°C for 35 min and stored at −80°C until hybridizations.

Affymetrix GeneChip hybridizations

Affymetrix Human Genome U-133 A arrays were used representing more than 22 000 transcripts derived from approximately 18 000 well-substantiated human genes. Hybridizations, probe array scanning and gene expression analysis were performed using the Gene Chip Core Facility at the Columbia University Genome Center, according to the manufacturer's instructions.

Checkerboard DNA–DNA hybridizations

Analysis of plaque samples with respect to the 12 bacterial species listed in Table 3 was performed according to the 'checkerboard' DNA–DNA hybridization method (19), as described earlier (18), with the following modifications: the chemiluminescent substrate used for detection was CSPD (Roche/Boehringer-Mannheim, Indianapolis, IN, USA). Evaluation of the chemiluminescence signal was performed in a LumiImager F1 Workstation (Roche/Boehringer-Mannheim) by comparing the obtained signals with the ones generated by pooled standard samples containing 10^6 or 10^5 of each of the species. Standard curves were generated for each species by means of the LUMIANALYST software (Roche/Boehringer-Mannheim), and the chemiluminescent signals obtained were ultimately transformed into bacterial counts and exported into EXCEL files.

Table 1

Demographic characteristics of subjects in the four diagnostic categories

Diagnosis	N	Age (yr)/Mean (SD)/Min–max	Gender	Race/ethnicity
Localized chronic	1	33	M	H
Generalized chronic	6	41.3 (6.8) 35–54	3 F, 3 M	5 H, 1 W
Localized aggressive	1	16	F	AA
Generalized aggressive	6	31.7 (3.9) 26–36	4 F, 2 M	5 H, 1 AA
All	14	34.8 (8.7)	8 F, 6 M	11 H, 2 AA, 1 W

F, female; M, male. H, Hispanic; AA, African American; W, white.

Checkerboard immunoblotting

The level of serum IgG antibodies against the same bacterial species used in the checkerboard hybridizations was assessed by checkerboard immunoblotting (20) as described elsewhere (18) with the following modifications: human IgG standards (Sigma-Aldrich, St Louis, MO, USA) were prepared at dilutions of 333, 500, 1000, and 2000 ng ml^{-1}. Chemiluminescent signals were assessed in the same LumiImager F1 Workstation used in the DNA hybridizations, were quantified in comparison to a curve generated by protein A (Sigma-Aldrich) and the four human IgG standards by using the LUMIANALYST software, and were exported into EXCEL files.

Data analysis

The clinical periodontal variables were averaged for each patient. In order to circumvent the problem of raw values equal to zero, the microbiological and serum antibody data were log transformed as follows: log (raw value + 1). First, within each subject, a mean log-transformed bacterial load across the 18 microbiologically sampled sites was calculated for each investigated species. Subsequently, species-specific infection ratios (antibody titer over homologous mean bacterial load) were calculated as follows: each subject's log transformed specific antibody titer was divided by the individual subject's mean bacterial load for the homologous species described above, generating a species-specific 'infection ratio'. In other words, the infection ratios (antibody concentration/bacterial load) described a measure of antibody responsiveness for each infecting bacterial species.

Initial microarray data analysis, consisting of calculating expression values from the raw digital image of the fluorescent microarray, was conducted using the Affymetrix Microarray Suite Version 5.0 software. Data were screened to identify arrays with poor signal/background ratios, or which exhibited low RNA quality, as indicated by the Affymetrix software, using criteria defined by Affymetrix. Such 'outlier' samples were excluded. The 'signal' values for each high-quality array were used as measure of expression for subsequent stages of the analysis.

To address the question of whether gene expression differences exist between chronic and aggressive periodontitis, a supervised analysis was carried out. In this analysis, all microarray data obtained from diseased sites from each patient were combined by calculating the mean of all available values for each gene. In the description that follows, each resulting set of values was treated as if it were a single sample, although it was combined data from several samples. We used a Student's *t*-test to search for genes that were differentially expressed between groups, applying a multiple testing procedure based on the false discovery rate (21). To test whether either type of periodontitis could be 'diagnosed' using microarray data, we used a supervised learning procedure. Supervised learning involves training a learner (in this case, a computer algorithm) to recognize the difference between two or more categories of labeled training examples, which in this case were individual microarray samples from chronic and aggressive periodontitis, labeled as to their membership in each group. Training is based on genes selected from the training set that distinguish the classes (feature selection). Once trained, the learner can be evaluated by testing its ability to correctly classify test examples, the classification of which is known by the experimenter but hidden from the learner. This basic procedure (supervised learning using cross-validation with feature selection) represents standard methodologies which have been previously applied to microarray data in numerous studies (8, 22, 23). The supervised learning algorithm that we selected was the support vector machine (SVM), which has been previously used to classify tumor samples based on microarray data (24–26). We used an existing SVM implementation cowritten by author P. P. (http://microarray.cpmc.columbia.edu/gist).

To explore a potentially new classification scheme based on the gene expression profiles, we employed a unsupervised analysis using clustering, similar to the approach that has been successfully used in other microarray studies to identify groups of samples that are clinically relevant, but which were previously unknown (9). We employed the commonly used hierarchical clustering (27) and evaluated the robustness of the identified clusters by a non-parametric bootstrap (28).

Results

A single sample from healthy gingival tissue (control sample) was analysed from a total of six subjects (three with chronic and three with aggressive periodontitis). A minimum of two diseased gingival tissue samples were analysed from each subject. Thus, the total number of microarrays processed in this report was 36. Based on quality controls (background and noise measurements provided by the Affymetrix software), the data for one sample from a patient with chronic periodontitis was deemed to be of low quality and was excluded from the analysis. All other RNA and microarray hybridization data were of excellent quality. Thus, the final data set includes microarrays for six control samples, 14 diseased tissue samples from chronic periodontitis subjects, and 15 such samples from aggressive periodontitis subjects (35 arrays in total).

We first sought to explore the within-subject variance in gene expression. The findings from this analysis are presented in Fig. 1. Figure 1A illustrates a representative plot of expression values for two samples obtained from a single subject. As illustrated by the majority of the points that are close to the diagonal, the samples showed an overall high similarity. Figure 1B illustrates the full correlation matrix for all 36 microarrays processed, including the one that was discarded from further analysis because of poor quality (sample 7-2). As indicated by the highly compressed scale, correlations below 0.9 are visualized in black, while correlations of 1.0 are

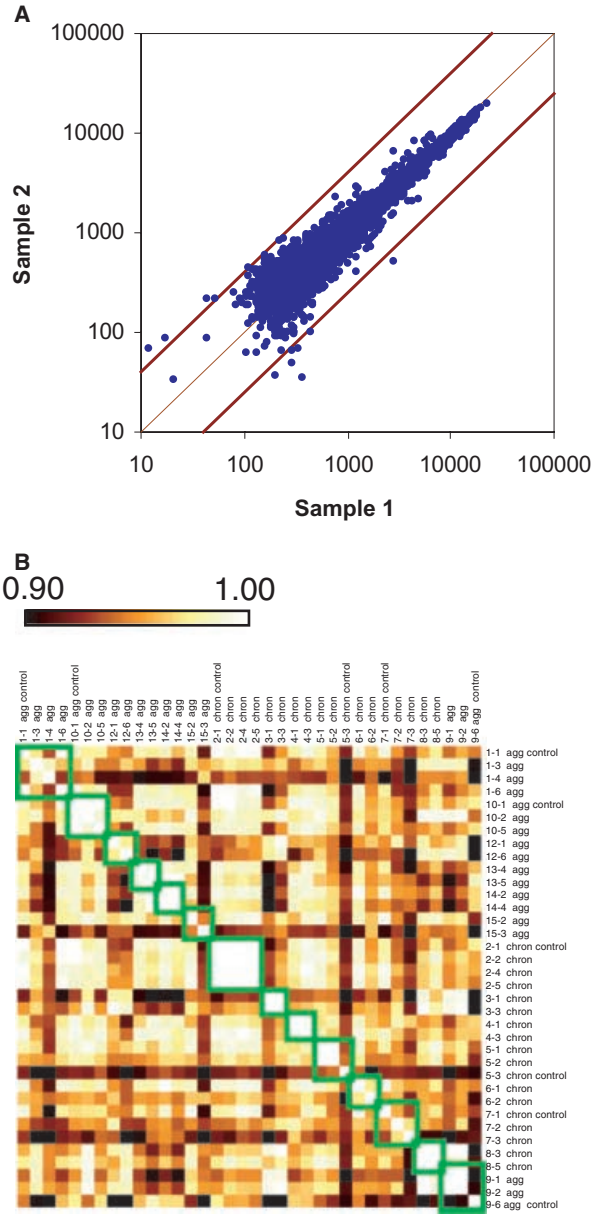

Fig. 1. Within subject variance of gene expression. (A) Plot of expression values for two samples from the same subject. The red lines indicate fourfold differences. (B) Correlation matrix for the 36 microarrays processed. As indicated by the highly compressed scale, correlations below 0.9 are indicated in black, while correlations of 1.0 are white. The lowest correlation in the data set was *c*. 0.85. Green boxes along the diagonal illustrate within subject correlations. Note that although 14 patients are included in the analysis, patient numbers range from 1 to 15.

white. The lowest correlation observed in this particular data set was 0.85. The green boxes along the diagonal of the graph highlight comparisons among samples obtained from the same subject. For the most part, very high correlations were observed within the green boxes (illustrated by the almost white color), indicating a high agreement among samples originating from the same subject. However, greater heterogeneity can be observed in a few subjects, such as in subjects 1, 7 and 15.

We next examined differences in gene expression between control and diseased sites. Using data from six control sites assayed among the 35 in total, only limited differences emerged between control and diseased sites. Based on a two-tailed Student's *t*-test, at a false discovery rate of approximately 0.1, there were only 31 genes showing significant differential expression (data not shown).

We next addressed the question of whether there are differences in expression profiles in samples taken from subjects with aggressive vs. chronic periodontitis. In this, and all further analyses, samples from control sites were excluded, and expression profiles from diseased sites were averaged within each subject. Using a two-tailed *t*-test with 12 degrees of freedom, we found no genes that were significantly differentially expressed between the two diagnostic categories. Corroborating this observation, our support vector machine supervised learning procedure was unable to 'learn' the difference between aggressive and chronic periodontitis from these data. Indeed, only six out of 14 predictions were correct, indicating a performance at the level of chance.

Our subsequent analysis attempted to identify a novel grouping of periodontitis patients based on gene expression profiles that may reflect hitherto unrecognized disease categories. Our approach was to analyse the patient-average expression profiles using a hierarchical clustering algorithm (Fig. 2). As seen, our earlier observation of no pronounced gene expression differences between chronic and aggressive periodontitis was reinforced by the fact that the chronic and aggressive periodontitis patients did not cluster together. However, two main clusters emerged that were fairly robust, as illustrated both by the height of the bars separating them from the parent branch, indicating the relative dissimilarity of the clusters, and by the bootstrap results. In particular, the bootstrap results suggested a group of five samples (all but one of those labeled in red in Fig. 2) that were tightly clustered, and a somewhat more diffuse cluster of 8 samples.

Using a *t*-test, we could identify at least 200 genes that appeared to be among those most responsible for the appearance of the two clusters shown in Fig. 2. The top 20 of these genes are listed in Fig. 3. To explore whether these clusters would be biologically or clinically relevant, we first examined the discriminatory genes for their functional roles. Among the differentially regulated genes, a number have functions in the immune system.

Next, we explored differences between the subjects in Clusters 1 and 2 with respect to clinical periodontal parameters. As shown in Table 2, patients in Clusters 1 and 2 were quite similar with respect to clinical

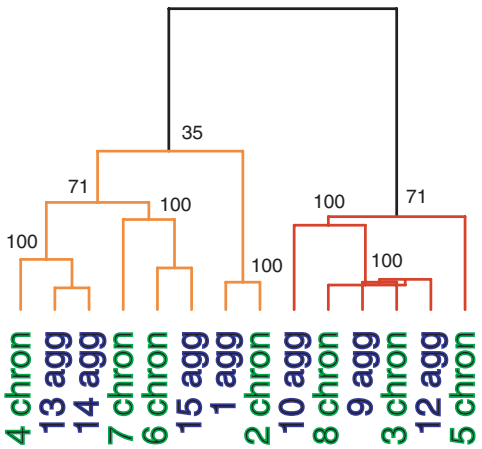

Fig. 2. Clustering of expression profiles from all patients. Patient expression profiles are averages across diseased sites. The chronic periodontitis patients (green) and the aggressive periodontitis patients (blue) do not cluster together overall, but two fairly robust clusters do emerge (red and orange). The numbers indicate the count of bootstrap trials (out of 100) during which each cluster remained identical. Thus, the red cluster was identical in 71/100 trials. Apparently, in 29 trials, some sample(s) (presumably the sample at the far right, as the remaining cluster is stable) were moved out of the cluster.

periodontal status, as expressed by their average percentage of bleeding of probing, their mean probing depth and attachment level, and the number of deep pockets (≥ 6 mm) per subject, or the number of sites per subject with attachment loss of ≥ 6 mm.

When we examined the mean bacterial load per subject in Clusters 1 and 2 with respect to 12 bacterial species, we observed largely overlapping colonization patterns between the two clusters (data not shown). We then proceeded with an analysis of serum antibody responses and infection ratios in the two clusters. These findings are summarized in Table 3. Interestingly, subjects in Cluster 2 had statistically significantly lower titers against two of the three established causative bacteria for periodontitis (*Porphyromonas gingivalis* and *Tannerella forsythensis*) as well as to the putative pathogen *Campylobacter rectus*. The same observation with respect to these three species was made when the homologous bacterial colonization was concomitantly taken into consideration. Indeed, infection ratios for *P. gingivalis, T. forsythensis,* and *C. rectus* were statistically significantly lower in patients of Cluster 2, suggesting an overall lower responsiveness of these subjects to important periodontal pathogens.

Discussion

To our knowledge, this is the first study that examines gene expression in affected gingival tissues in chronic and aggressive periodontitis by means of microarray technology. We intend to use the pilot data presented above in a larger study that will explore a novel classification of periodontal diseases based on similarities of gene expression in pathological periodontal tissues. The findings obtained so far can be summarized as follows:

Probe	p value	Description
209829 at	1.41E-08	chromosome 6 open reading frame 32 | C6orf32
218871 x at	1.29E-07	chondroitin sulfate GalNAcT-2 | GALNACT-2
206555 s at	2.09E-07	hypothetical protein FLJ20274 | FLJ20274
213186 at	2.40E-07	KIAA0675 gene product | KIAA0675
202422 s at	3.39E-07	fatty-acid-Coenzyme A ligase, long-chain 4 | FACL4
212737 at	4.79E-07	GM2 ganglioside activator protein | GM2A
218294 s at	5.75E-07	nucleoporin 50kDa | NUP50
202596 at	5.89E-07	endosulfine alpha | ENSA
217738 at	6.92E-07	pre-B-cell colony-enhancing factor | PBEF
202303_x_at	7.24E-07	SWI/SNF related, matrix associated, actin dependent regulator of chromatin | SMARCA5
210815 s at	8.13E-07	calcitonin receptor-like | CALCRL
202901 x at	8.71E-07	cathepsin S | CTSS
206636 at	9.12E-07	RAS p21 protein activator 2 | RASA2
214678 x at	1.00E-06	zinc finger protein, X-linked | ZFX
218193 s at	1.66E-06	CGI-141 protein | CGI-141
210148 at	1.66E-06	homeodomain interacting protein kinase 3 | HIPK3
200830_at	1.78E-06	proteasome (prosome, macropain) 26S subunit, non-ATPase, 2 | PSMD2
210014 x at	1.86E-06	NAD+-specific isocitrate dehydrogenase beta
216915 s at	1.95E-06	protein tyrosine phosphatase (PTPG1) mRNA
200789 at	2.00E-06	enoyl Coenzyme A hydratase 1, peroxisomal | ECH1

Fig. 3. Top 20 genes that distinguish the clusters shown in Fig. 2. At left, each column indicates one patient-mean profile, while each row represents one probe set (gene). The two clusters are labeled as 'Cl1' and 'Cl2'. Brighter colors in the heat map indicate higher relative levels of expression. *P*-values are obtained from a two-tailed *t*-test comparing the two groups. Genes in red have clear roles in the immune system. A number of genes (such as the top gene) are poorly annotated and have unknown functions.

Table 2

Clinical periodontal status in Clusters 1 and 2

	BoP (%)	PD (mm)	No. of sites with PD ≥6 mm	AL (mm)	No. of sites with AL ≥6 mm
Cluster 1	56 (32)	3.9 (0.7)	30.5 (18.9)	4.2 (1.5)	43.8 (39.2)
Cluster 2	46 (37)	4.2 (0.7)	39.6 (28.9)	4.6 (0.7)	51.4 (31.3)
P-value	0.55	0.50	0.49	0.56	0.70

Data are mean with SD in parenthesis; $n = 6$, Cluster 1; $n = 8$, Cluster 2.
BoP, bleeding on probing; PD, pocket depth; AL, attachment loss.

(i) there is good overall agreement of gene expression data collected from multiple diseased sites from a single patient, but there is sufficient variability to justify analysing two gingival tissue specimens per patient; (ii) in this limited-size sample, no significant differences in gene expression were detected between gingival tissues of subjects with chronic and aggressive periodontitis; and (iii) gene expression signature similarities formed two subgroups of patients that differed with respect to serum antibody responses and infection ratios for important periodontal pathogens.

Some further discussion of the above main points is pertinent. Attempting to classify periodontitis subtypes based on gene expression will be feasible and meaningful only if the variance in gene expression between several affected sites within the same oral cavity is relatively low and substantially lower than that between subjects. Our findings indicate that more than one sample per patient need to be analysed in order to increase the reliability of the results, but also that processing of two microarrays

Table 3

Log-transformed serum IgG titers and log-transformed infection ratios in Clusters 1 and 2

	Titers (ng ml^{-1})			Infection ratios (ng ml^{-1} 10^{-6} bacterial cells)		
Species*	Cluster 1	Cluster 2	*P*	Cluster 1	Cluster 2	*P*
Aa	5.81 (0.47)	6.08 (0.44)	0.29	0.88 (0.07)	0.91 (0.06)	0.45
Pg	6.32 (0.40)	5.57 (0.37)	0.003	0.90 (0.03)	0.80 (0.04)	0.001
Tf	5.32 (0.54)	4.82 (0.31)	0.04	0.77 (0.06)	0.67 (0.07)	0.03
Td	4.90 (0.46)	4.72 (0.61)	0.54	0.65 (0.05)	0.64 (0.07)	0.65
Pi	5.64 (0.25)	5.63 (0.44)	0.95	0.71 (0.03)	0.71 (0.07)	0.90
Fn	3.86 (1.93)	3.55 (1.58)	0.75	0.50 (0.24)	0.46 (0.21)	0.77
Mm	4.82 (0.40)	4.81 (0.43)	0.96	0.62 (0.04)	0.63 (0.07)	0.89
Cr	5.74 (0.71)	4.97 (0.33)	0.01	0.86 (0.12)	0.73 (0.05)	0.02
En	5.56 (1.05)	5.81 (0.90)	0.64	0.93 (0.17)	0.95 (0.14)	0.80
Ec	5.44 (0.15)	5.48 (0.28)	0.70	0.82 (0.02)	0.81 (0.05)	0.56
Vp	3.99 (1.02)	3.73 (1.04)	0.65	0.58 (0.13)	0.52 (0.15)	0.47
An	4.02 (2.09)	4.63 (0.85)	0.52	0.52 (0.27)	0.59 (0.10)	0.58

Data are mean with SD in parenthesis.
* Species abbreviations: Aa, *Actinobacillus actinomycetemcomitans*; Pg, *Porphyromonas gingivalis*; Tf, *Tannerella forsythensis*; Td, *Treponema denticola*; Pi, *Prevotella intermedia*; Fn, *Fusobacterium nucleatum*; Mm, *Micromonas micros*; Cr, *Campylobacter rectus*; En, *Eubacterium nodatum*; Ec, *Eikenella corrodens*; Vp, *Veillonella parvula*; An, *Actinomyces naeslundii*.

per subject is sufficient to identify and address possible 'outlier' samples. The same conclusion was reinforced by a clustering analysis (not shown), which demonstrated that samples from one subject usually, but not always, cluster together very closely. Importantly, as illustrated by points away from the diagonal in Fig. 1B, more pronounced differences were observed between than within subjects.

Interestingly, our preliminary results indicate that relatively limited differences in gene profiling were detectable between diseased and healthy sites of the same patient. These differences were of much lower magnitude than those observed in earlier experimental studies where inflamed uterine tissue was compared with non-inflamed control tissue (P. Pavlidis, unpubl. observations). Our provisional interpretation of this observation is that the control sites, which were collected from periodontitis-affected individuals at locations proximal to the diseased sites, may reflect the effects of disease to a much greater extent than is apparent from the clinical appearance of the gingival tissue.

Similarly, our pilot study failed to detect any significant difference in gene expression between affected gingival tissues of subjects with chronic or aggressive periodontitis. This finding, based on an admittedly limited subject sample, suggests that any gene expression differences between these two disease types is likely to be subtle. Nevertheless, our unsupervised analysis did identify two fairly robust groups of patients that displayed distinct gene expression signatures. Although many of those differentially regulated genes are poorly annotated, some have functions in the immune system. For example, cathepsin S is a cysteine protease that may participate in the degradation of antigenic proteins to peptides for presentation on major histocompatibility complex (MHC) class II molecules in response to infection. Similarly, the proteasome, represented in our list by PSMD2, is involved in antigen presentation. A third immune-response gene, *PBEF* (pre-B-cell colony enhancing factor), the expression of which is upregulated in activated lymphocytes, codes for a nicotinamide phosphoribosyltransferase, a cytosolic enzyme recently shown to be involved in nicotinamide adenine dinucleotide (NAD) biosynthesis (29). Intriguingly, in our data a strong pattern emerges involving NAD, including the NAD+ isocitrate dehydrogenase beta gene (Fig. 3), as well as a number of other genes not listed, such as NADH dehydrogenase complex members *NDUFB8*, *NDUFC1* and *NDUFB2*. Thus, the two clusters identified may have revealed a pattern of expression representing differences in responses to infection among subpopulations of periodontitis patients. The findings in Table 3 appear to corroborate the above postulate, as Clusters 1 and 2 displayed a different 'responsiveness' to colonization by important periodontal pathogens, such as *P. gingivalis*, *T. forsythensis*, and *C. rectus*.

Given the small sample size in our pilot study, it is clear that we have gained only a little insight into the actual properties of expression profiles in periodontitis and their biological significance. It must be emphasized that the reported findings on the differential gene expression between the clusters identified or on the phenotypic characteristics of the clusters are not definitive, and may or may not hold true if a larger patient sample is analysed. Nevertheless, the primary value of these pilot data lies in their ability to illustrate the potential of this multilevel genomics and bioinformatics approach in the study of the pathobiology of periodontitis. As exemplified above, this strategy can facilitate the identification of homogeneous subtypes of periodontitis and can provide a detailed description of the distinct gene expression profiles that characterize each subtype. Importantly, by concomitantly exploring the disease phenotype in the different subtypes and identifying surrogate markers for each (e.g. as illustrated in our analysis, by linking responsiveness to infection to a particular cluster), this approach can tie the molecular diagnosis with clinical and laboratory findings.

Acknowledgement – This study was supported in part by an unrestricted gift by Colgate-Palmolive Inc., USA.

References

1. 1999 International workshop for a classification of periodontal diseases and conditions Oak Brook, Illinois. *Ann Periodontol* 1999; **4**: 1–112.
2. Armitage GC. Development of a classification system for periodontal diseases and conditions. *Ann Periodontol* 1999; **4**: 1–6.
3. Armitage GC. Classifying periodontal diseases – a long-standing dilemma. *Periodontol 2000*, 2002; **30**: 9–23.
4. Chung CH, Bernard PS, Perou CM. Molecular portraits and the family tree of cancer. *Nat Genet* 2002; **32**: 533–540.
5. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; **415**: 530–536.
6. van de Vijver MJ, He YD, van't VEERLJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; **347**: 1999–2009.
7. Schubert CM. Microarray to be used as routine clinical screen. *Nat Med* 2003; **9**: 9.
8. Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, Allen JC, Zagzag D, Olson JM, Curran T, Wetmore C, Biegel JA, Poggio T, Mukherjee S, Rifkin R, Califano A, Stolovitzky G, Louis DN, Mesirov JP, Lander ES, Golub TR. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 2002; **415**: 436–442.
9. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J JR, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; **403**: 503–511.
10. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D.

Molecular portraits of human breast tumours. *Nature* 2000; **406**: 747–752.
11. Haslett JN, Kunkel LM. Microarray analysis of normal and dystrophic skeletal muscle. *Int J Dev Neurosci* 2002; **20**: 359–365.
12. Colangelo V, Schurr J, Ball MJ, Pelaez RP, Bazan NG, Lukiw WJ. Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling. *J Neurosci Res* 2002; **70**: 462–473.
13. Thornton S, Sowders D, Aronow B, Witte DP, Brunner HI, Giannini EH, Hirsch R. DNA microarray analysis reveals novel gene expression profiles in collagen-induced arthritis. *Clin Immunol* 2002; **105**: 155–168.
14. Burke W. Genomics as a probe for disease biology. *N Engl J Med* 2003; **349**: 969–974.
15. Consensus report: periodontitis as a manifestation of systemic diseases. *Ann Periodontol* 1999; **4**: 64.
16. Consensus Report. Aggressive periodontitis. *Ann Periodontol* 1999; **4**: 53.
17. Consensus Report. Chronic Periodontitis. *Ann Periodontol* 1999; **4**: 38.
18. Papapanou PN, Neiderud A-M, Papadimitriou A, Sandros J, Dahlén G. 'Checkerboard' assessments of periodontal microbiota and serum antibody responses: a case-control study. *J Periodontol* 2000; **71**: 885–897.
19. Socransky SS, Smith C, Martin L, Paster BJ, Dewhirst FE, Levin AE. 'Checkerboard' DNA–DNA, hybridization. *Biotechniques* 1994; **17**: 788–792.
20. Sakellari D, Socransky SS, Dibart S, Eftimiadi C, Taubman MA. Estimation of serum antibody to subgingival species using checkerboard immunoblotting. *Oral Microbiol Immunol* 1997; **12**: 303–310.
21. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc* 1995; **57**: 289–300.
22. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; **286**: 531–537.
23. Dyrskjot L, Thykjaer T, Kruhoffer M, Jensen JL, Marcussen N, Hamilton-Dutoit S, Wolf H, Orntoft TF. Identifying distinct classes of bladder carcinoma using microarrays. *Nat Genet* 2003; **33**: 90–96.
24. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 2000; **16**: 906–914.
25. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 2001; **98**: 15149–15154.
26. Segal NH, Pavlidis P, Antonescu CR, Maki RG, Noble WS, Desantis D, Woodruff JM, Lewis JJ, Brennan MF, Houghton AN, Cordon-Cardo C. Classification and subtype prediction of adult soft tissue sarcoma by functional genomics. *Am J Pathol* 2003; **163**: 691–700.
27. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; **95**: 14863–14868.
28. Felsenstein J. Confidence limits on phylogenies: an Approach using bootstrap. *Evolution* 1985; **39**: 783–791.
29. Rongvaux A, Shea RJ, Mulks MH, Gigot D, Urbain J, Leo O, Andris F. Pre-B-cell colony-enhancing factor, whose expression is up-regulated in activated lymphocytes, is a nicotinamide phosphoribosyltransferase, a cytosolic enzyme involved in NAD biosynthesis. *Eur J Immunol* 2002; **32**: 3225–3234.