

Salivary Transcriptome Diagnostics for Oral Cancer Detection

Yang Li,¹ Maie A. R. St. John,² Xiaofeng Zhou,¹ Yong Kim,¹ Uttam Sinha,⁶ Richard C. K. Jordan,⁷ David Eisele,⁸ Elliot Abemayor,² David Elashoff,³ No-Hee Park,^{1,4,5} and David T. Wong^{1,4,5}

¹School of Dentistry and Dental Research Institute, ²School of Medicine, ³School of Public Health, ⁴Jonsson Comprehensive Cancer Center, and ⁵Molecular Biology Institute, University of California-Los Angeles, Los Angeles, California; ⁶School of Medicine, University of Southern California, Los Angeles, California; and ⁷School of Dentistry, and ⁸Department of Otolaryngology-Head & Neck Surgery, University of California-San Francisco, San Francisco, California

ABSTRACT

Purpose: Oral fluid (saliva) meets the demand for non-invasive, accessible, and highly efficient diagnostic medium. Recent discovery that a large panel of human RNA can be reliably detected in saliva gives rise to a novel clinical approach, salivary transcriptome diagnostics. The purpose of this study is to evaluate the diagnostic value of this new approach by using oral squamous cell carcinoma (OSCC) as the proof-of-principle disease.

Experimental Design: Unstimulated saliva was collected from patients ($n = 32$) with primary T1/T2 OSCC and normal subjects ($n = 32$) with matched age, gender, and smoking history. RNA isolation was done from the saliva supernatant, followed by two-round linear amplification with T7 RNA polymerase. Human Genome U133A microarrays were applied for profiling human salivary transcriptome. The different gene expression patterns were analyzed by combining a *t* test comparison and a fold-change analysis on 10 matched cancer patients and controls. Quantitative polymerase chain reaction (qPCR) was used to validate the selected genes that showed significant difference ($P < 0.01$) by microarray. The predictive power of these salivary mRNA biomarkers was analyzed by receiver operating characteristic curve and classification models.

Results: Microarray analysis showed there are 1,679 genes exhibited significantly different expression level in saliva between cancer patients and controls ($P < 0.05$).

Seven cancer-related mRNA biomarkers that exhibited at least a 3.5-fold elevation in OSCC saliva ($P < 0.01$) were consistently validated by qPCR on saliva samples from OSCC patients ($n = 32$) and controls ($n = 32$). These potential salivary RNA biomarkers are transcripts of *IL8*, *IL1B*, *DUSP1*, *HA3*, *OAZ1*, *S100P*, and *SAT*. The combinations of these biomarkers yielded sensitivity (91%) and specificity (91%) in distinguishing OSCC from the controls.

Conclusions: The utility of salivary transcriptome diagnostics is successfully demonstrated in this study for oral cancer detection. This novel clinical approach could be exploited to a robust, high-throughput, and reproducible tool for early cancer detection. Salivary transcriptome profiling can be applied to evaluate its usefulness for other major disease applications as well as for normal health surveillance.

INTRODUCTION

More than 1.3 million new cancer cases are expected to be diagnosed in 2004 in the United States (1). Cancer will cause approximately 563,700 deaths of Americans this year, killing one person every minute. These numbers have been steadily increasing over the past 10 years, despite advances in cancer treatment. Moreover, for some cancers such as oral cavity cancer, the overall 5-year survival rates have not improved in the past several decades, remaining low at ~30 to 50% (2, 3). A critical factor in the lack of prognostic improvement is the fact that a significant proportion of cancers initially are asymptomatic lesions and are not diagnosed or treated until they reach an advanced stage. Early detection of cancer is the most effective means to reduce death from this disease.

The genetic aberrations of cancer cells lead to altered gene expression patterns, which can be identified long before the resulting cancer phenotypes are manifested. Changes that arise exclusively or preferentially in cancer, compared with normal tissue of the same origin, can be used as molecular biomarkers (4). Accurately identified, biomarkers may provide new avenues and constitute major targets for cancer early detection and cancer risk assessment. A variety of nucleic acid-based biomarkers have been demonstrated as novel and powerful tools for the detection of cancers (5–7). However, most of these markers have been identified either in cancer cell lines or in biopsy specimens from late invasive and metastatic cancers. We are still limited in our ability to detect cancer in its earliest stages with biomarkers. Moreover, the invasive nature of a biopsy makes it unsuitable for cancer screening in high-risk populations. This suggests an imperative need for developing new diagnostic tools that would improve early detection. The identification of molecular markers in bodily fluids that would predict the development of cancer in its earliest stage or in precancerous stage would constitute such a tool.

It has been shown that identical mutation present in the primary tumor can be identified in the bodily fluids tested from affected patients (8). Cancer-related nucleic acids in blood,

Received 6/16/04; revised 8/12/04; accepted 8/18/04.

Grant support: USPHS grants UO1 DE15018 and RO1 DE15970 and UCLA Jonsson Comprehensive Cancer Center grant (D. Wong); USPHS grant T32 DE07296-07 and Cancer Research Foundation of American fellowship (X. Zhou).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: David T. Wong, UCLA School of Dentistry, Dental Research Institute, 73–017 CHS, 10833 Le Conte Avenue, Los Angeles, CA 90095. Phone: (310) 206-3048; Fax: (310) 825-0921; E-mail: dtw@ucla.edu.

©2004 American Association for Cancer Research.

urine, and cerebrospinal fluid have been used as biomarkers for cancer diagnosis (9–11). More recently, mRNA biomarkers in serum or plasma have been targets for reverse transcription-PCR (RT-PCR)-based detection strategies in patients with cancers (12, 13). Parallel to the increasing number of such biomarkers in bodily fluids is the growing availability of technologies with more powerful and cost-efficient methods that enable mass screening for genetic alterations. Recent discovery by microarray technology that a large panel of human mRNA exists in saliva (14) suggests a novel clinical approach, salivary transcriptome diagnostics, for applications in disease diagnostics as well as for normal health surveillance. It is a high-throughput, robust, and reproducible approach to harness RNA signatures from saliva. Moreover, using saliva as a diagnostic fluid meets the demands for inexpensive, noninvasive, and accessible diagnostic methodology (15). In the present study, we tested the hypothesis that distinct mRNA expression patterns can be identified in saliva from cancer patients, and the differentially expressed transcripts can serve as biomarkers for cancer detection. The proof-of-principle disease in this study is oral squamous cell carcinoma (OSCC). The rationale is that oral cancer cells are immersed in the salivary milieu and genetic heterogeneity has been detected in saliva from patients with OSCC (16, 17).

PATIENTS AND METHODS

Patient Selection. OSCC patients were recruited from Medical Centers at University of California, Los Angeles (UCLA); University of Southern California (USC), Los Angeles, CA; and University of California San Francisco, San Francisco, CA. Thirty-two patients with documented primary T1 or T2 OSCC were included in this study. All of the patients had recently received diagnoses of primary disease and had not received any prior treatment in the form of chemotherapy, radiotherapy, surgery, or alternative remedies. An equal number of age- and sex-matched subjects with comparable smoking histories were selected as a control group (18). Among the two subject groups, there were no significant differences in terms of mean age: OSCC patients, 49.8 ± 7.6 years; normal subjects, 49.1 ± 5.9 years (Student's *t* test, $P > 0.80$); gender ($P > 0.90$); or smoking history ($P > 0.75$). No subjects had a history of prior malignancy, immunodeficiency, autoimmune disorders, hepatitis, or HIV infection. All of the subjects signed the institutional review board-approved consent form agreeing to serve as saliva donors for the experiments.

Saliva Collection and RNA Isolation. Unstimulated saliva samples were collected between 9 a.m. and 10 a.m. with previously established protocols (19). Subjects were asked to refrain from eating, drinking, smoking, or oral hygiene procedures for at least 1 hour before the collection. Saliva samples were centrifuged at $2,600 \times g$ for 15 minutes at 4°C . The supernatant was removed from the pellet and treated with RNase inhibitor (Superase-In, Ambion Inc., Austin, TX). RNA was isolated from 560 μL of saliva supernatant with QIAamp Viral RNA kit (Qiagen, Valencia, CA). Aliquots of isolated RNA were treated with RNase-free DNase (DNaseI-DNA-free, Ambion Inc.) according to the manufacturer's instructions. The quality of isolated RNA was examined by RT-PCR for three

cellular maintenance gene transcripts: glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), actin- β (*ACTB*), and ribosomal protein S9 (*RPS9*). Only those samples exhibiting PCR products for all three mRNAs were used for subsequent analysis.

Microarray Analysis. Saliva from 10 OSCC patients (7 male, 3 female; age, 52 ± 9.0 years) and from 10 gender- and age-matched normal donors (age, 49 ± 5.6 years) was used for a microarray study. Isolated RNA from saliva was subjected to linear amplification by RiboAmp RNA Amplification kit (ArcTurus, Mountain View, CA). The RNA amplification efficiency was measured by using control RNA of known quantity (0.1 μg) running in parallel with the 20 samples in five independent runs. Following previously reported protocols (14), the Human Genome U133A Array (HG U133A, Affymetrix, Santa Clara, CA) was applied for gene expression analysis.

The arrays were scanned and the fluorescence intensity was measured by Microarray Suit 5.0 software (Affymetrix, Santa Clara, CA); the arrays were then imported into DNA-Chip Analyzer software (<http://www.dchip.org>) for normalization and model-based analysis (20). S-plus 6.0 (Insightful, Seattle, WA) was used to carry out all statistical tests. We used three criteria to determine differentially expressed gene transcripts. First, we excluded probe sets on the array that were assigned as "absent" call in all samples. Second, a two-tailed Student's *t* test was used for comparison of average gene expression signal intensity between the OSCCs ($n = 10$) and controls ($n = 10$). The critical α level of 0.05 was defined for statistical significance. Third, fold ratios were calculated for those gene transcripts that showed statistically significant difference ($P < 0.05$). Only those gene transcripts that exhibited at least 2-fold change were included for further analysis.

Quantitative Polymerase Chain Reaction Validation. Quantitative polymerase chain reaction (qPCR) was performed to validate a subset of differently expressed transcripts identified by microarray analysis. Using MuLV reverse transcriptase (Applied Biosystems, Foster City, CA) and random hexamers as primer (Applied Biosystems), we synthesized cDNA from the original and unamplified salivary RNA. The qPCR reactions were performed in an iCycler PCR system with iQ SYBR Green Supermix (Bio-Rad, Hercules, CA). Primer sets were designed by using PRIMER3 software (<http://www.genome.wi.mit.edu>).⁹ All of the reactions were performed in triplicate with customized conditions for specific products. The initial amount of cDNA/RNA of a particular template was extrapolated from the standard curve as described previously (21). This validation completed by testing all of the samples ($n = 64$) including those 20 previously used for microarray study. Wilcoxon Signed-Rank test was used for statistical analysis.

Receiver Operating Characteristic Curve Analysis and Prediction Models. Using the qPCR results, we conducted receiver operating characteristic (ROC) curve analyses (22) by S-plus 6.0 to evaluate the predictive power of each of the biomarkers. The optimal cutpoint was determined for each biomarker by searching for those that yielded the maximum corresponding sensitivity and specificity. ROC curves were then

⁹ Internet address: <http://www.genome.wi.mit.edu>.

plotted on the basis of the set of optimal sensitivity and specificity values. Area under the curve was computed via numerical integration of the ROC curves. The biomarker that has the largest area under the ROC curve was identified as having the strongest predictive power for detecting OSCC.

Next, we constructed multivariate classification models to determine the best combination of salivary markers for cancer prediction. Firstly, using the binary outcome of the disease (OSCC) and nondisease (normal) as dependent variables, we constructed a logistic regression model controlling for patient age, gender, and smoking history. The backward stepwise regression (23) was used to find the best final model. We used leave-one-out cross-validation to validate the logistic regression model. The cross-validation strategy first removes one observation and then fits a logistic regression model from the remaining cases with all of the markers. Stepwise model selection is used for each of these models to remove variables that do not improve the model. Subsequently, we used the marker values for the case that was left out to compute a predicted class for that observation. The cross-validation error rate is then the number of samples predicted incorrectly divided by the number of samples. We then computed the ROC curve for the logistic model by a similar procedure, with the fitted probabilities from the model as possible cutpoints for computation of sensitivity and specificity.

Secondly, a tree-based classification model, classification and regression tree (CART), was constructed by S-plus 6.0 with the validated mRNA biomarkers as predictors. CART fits the classification model by binary recursive partitioning, in which each step involves searching for the predictor variable that results in the best split of the cancer *versus* the normal groups

(24). CART used the entropy function with splitting criteria determined by default settings for S-plus. By this approach, the parent group containing the entire samples ($n = 64$) was subsequently divided into cancer groups and normal groups. Our initial tree was pruned to remove all splits that did not result in sub-branches with different classifications.

RESULTS

On average, 54.2 ± 20.1 ng ($n = 64$) of total RNA was obtained from 560 μ L of saliva supernatant. There was no significant difference in total RNA quantity between the OSCC and matched controls (t test, $P = 0.29$, $n = 64$). RT-PCR results demonstrated that all of the saliva samples ($n = 64$) contained transcripts from three genes (*GAPDH*, *ACTB*, and *RPS9*), which were used as quality controls for human salivary RNAs (14). A consistent amplifying magnitude (658 ± 47.2 , $n = 5$) could be obtained after two rounds of RNA amplification. On average, the yield of biotinylated cRNA was 39.3 ± 6.0 μ g ($n = 20$). There were no significant differences of the cRNA quantity yielded between the OSCC and the controls (t test, $P = 0.31$, $n = 20$).

The HG U133A microarrays were used to identify the difference in salivary RNA profiles between cancer patients and matched normal subjects. Among the 10,316 transcripts included by the previously described criteria, we identified 1,679 transcripts with P value less than 0.05. Among these transcripts, 836 were up-regulated and 843 were down-regulated in the OSCC group. These transcripts observed were unlikely to be attributable to chance alone (χ^2 test, $P < 0.0001$), considering the false positives with $P < 0.05$. Using a predefined criteria of

Table 1 Salivary mRNA up-regulated (>3-fold, $P < 0.01$) in OSCC identified by microarray

Gene symbol	Gene name	GenBank accession no.	Locus	Gene functions
<i>B2M</i>	β -2-microglobulin	NM_04048	15q21-q22.2	Antiapoptosis; antigen presentation
<i>DUSP1</i>	Dual specificity phosphatase 1	NM_04417	5q34	Protein modification; signal transduction; oxidative stress
<i>FTH1</i>	Ferritin, heavy polypeptide 1	NM_02032	11q13	Iron ion transport; cell proliferation
<i>G0S2</i>	Putative lymphocyte G ₀ -G ₁ switch gene	NM_015714	1q32.2-q41	Cell growth and/or maintenance; regulation of cell cycle
<i>GADD45B</i>	Growth arrest and DNA-damage-inducible, β	NM_015675	19p13.3	Kinase cascade; apoptosis
<i>H3F3A</i>	H3 histone, family 3A	BE869922	1q41	DNA binding activity
<i>HSPC016</i>	Hypothetical protein HSPC016	BG167522	3p21.31	Unknown
<i>IER3</i>	Immediate early response 3	NM_003897	6p21.3	Embryogenesis; morphogenesis; apoptosis; cell growth and maintenance
<i>IL1B</i>	Interleukin 1, β	M15330	2q14	Signal transduction; proliferation; inflammation; apoptosis
<i>IL8</i>	Interleukin 8	NM_000584	4q13-q21	Angiogenesis; replication; calcium-mediated signaling pathway; cell adhesion; chemotaxis; cell cycle arrest; immune response
<i>MAP2K3</i>	Mitogen-activated protein kinase kinase 3	AA780381	17q11.2	Signal transduction; protein modification
<i>OAZ1</i>	Ornithine decarboxylase antizyme 1	D87914	19p13.3	Polyamine biosynthesis
<i>PRG1</i>	Proteoglycan 1, secretory granule	NM_002727	10q22.1	Proteoglycan
<i>RGS2</i>	Regulator of G-protein signaling 2, 24 kda	NM_002923	1q31	Oncogenesis; G-protein signal transduction
<i>S100P</i>	S100 calcium binding protein P	NM_005980	4p16	Protein binding; calcium ion binding
<i>SAT</i>	Spermidine/spermine N1-acetyltransferase	NM_002970	Xp22.1	Enzyme, transferase activity
	EST, highly similar ferritin light chain	BG537190		Iron ion homeostasis, ferritin complex

NOTE. The human Genome U133A microarrays were used to identify the difference in RNA expression patterns in saliva from 10 cancer patients and 10 matched normal subjects. Using a criteria of a change in regulation >3-fold in all 10 OSCC saliva specimens and a cutoff of P value < 0.01, we identified 17 mRNA, showing significant up-regulation in OSCC saliva.

Table 2 Quantitative PCR validation of selected nine transcripts in saliva (n = 64)

Gene symbol	Primer sequence (5' to 3')	Validated*	P value	Mean fold increase
<i>DUSP1</i>	F: CCTACCAGTATTATCCCGACG R: TTGTGAAGGCAGACACCTACAC	Yes	0.039	2.60
<i>H3F3A</i>	F: AAAGCACCCAGGAAGCAAC R: GCGAATCAGAAGTTCAGTGGAC	Yes	0.011	5.61
<i>IL1B</i>	F: GTGCTGAATGTGGACTCAATCC R: ACCCTAAGGCAGGCAGTTG	Yes	0.005	5.48
<i>IL8</i>	F: GAGGGTTGTGGAGAAGTTTTTG R: CTGGCATCTTCACTGATTCTTG	Yes	0.000	24.3
<i>OAZ1</i>	F: AGAGAGAGTCTTCGGGAGAGG R: AGATGAGCGAGTCTACGGTTC	Yes	0.009	2.82
<i>S100P</i>	F: GAGTTCATCGTGTTCGTGGCTG R: CTCCAGGGCATCATTTGAGTCC	Yes	0.003	4.88
<i>SAT</i>	F: CCAGTGAAGAGGGTTGGAGAC R: TGGAGGTTGTCATCTACAGCAG	Yes	0.005	2.98
<i>GADD45B</i>	F: TGATGAATGTGGACCCAGAC R: GAGCGTGAAGTGGATTTGC	No	0.116	
<i>RGS2</i>	F: CCTGCCATAAAGACTGACCTTG R: GCTTCTGATTCACTACCCAAC	No	0.149	

NOTE. qPCR were performed to validate the microarray findings on an enlarged sample size including saliva from 32 patients with OSCC and 32 matched control subjects. Nine potential salivary mRNA biomarkers were selected from the 17 candidates shown in Table 1. Seven of them were validated by qPCR ($P < 0.05$). Sample includes 32 saliva from OSCC patients and 32 from matched normal subjects.

* Wilcoxon's Signed Rank test: if $P < 0.05$, validated (Yes); if $P \geq 0.05$, not validated (No).

a change in regulation >3-fold in all 10 OSCC saliva specimens, and a more stringent cutoff of P value <0.01, we identified 17 transcripts as presented in Table 1. It should be noted that these 17 salivary mRNA are all up-regulated in OSCC saliva, whereas there are no mRNAs found down-regulated with the same filtering criteria. The biological functions of these genes and their products are presented in Table 1.

Quantitative PCR was performed to validate the microarray findings on an enlarged sample size including saliva from 32 OSCC patients and 32 matched controls. Nine candidates of salivary mRNA biomarkers: *DUSP1*, *GADD45B*, *H3F3A*, *IL1B*, *IL8*, *OAZ1*, *RGS2*, *S100P*, and *SAT* were selected based on their reported cancer association (Table 1). Table 2 presents their quantitative alterations in saliva from OSCC patients, determined by qPCR. The results confirmed that transcripts of 7 of the 9 candidate mRNA (78%), *DUSP1*, *H3F3A*, *IL1B*, *IL8*, *OAZ1*, *S100P*, and *SAT*, were significantly elevated in the saliva

of OSCC patient (Wilcoxon Signed-Rank test, $P < 0.05$). We did not detect the statistically significant differences in the amount of *RGS2* ($P = 0.149$) and *GADD45B* ($P = 0.116$) by qPCR. The validated seven genes could be classified in three ranks by the magnitude of increase: high up-regulated mRNA including *IL8* (24.3-fold); moderate up-regulated mRNAs including *H3F3A* (5.61-fold), *IL1B* (5.48), and *S100P* (4.88-fold); and low up-regulated mRNAs including *DUSP1* (2.60-fold), *OAZ1* (2.82-fold), and *SAT* (2.98-fold). The detailed statistics of the area under the receiver operator characteristics (ROC) curves, the threshold values, and the corresponding sensitivities and specificities for each of the seven potential salivary mRNA biomarkers for OSCC are listed in Table 3. The data showed *IL8* mRNA performed the best among the seven potential biomarkers for predicting the presence of OSCC. The calculated area under the ROC curve for *IL8* was 0.85. With a threshold value of 3.19E – 18 mol/L, *IL8* mRNA in saliva yields a sensitivity of

Table 3 Receiver operator characteristic (ROC) curve analysis of OSCC-associated salivary mRNA biomarkers

Biomarker	Area under ROC curve	Threshold/cutoff (M)	Sensitivity (%)	Specificity (%)	Selected references
<i>DUSP1</i>	0.65	8.35E-17	59	75	35
<i>H3F3A</i>	0.68	1.58E-15	53	81	55
<i>IL1B</i>	0.70	4.34E-16	63	72	45
<i>IL8</i>	0.85	3.19E-18	88	81	56
<i>OAZ1</i>	0.69	7.42E-17	100	38	38
<i>S100P</i>	0.71	2.11E-15	72	63	41
<i>SAT</i>	0.70	1.56E-15	81	56	36

NOTE. Utilizing the qPCR results, we conducted ROC curve analyses to evaluate the predictive power of each of the biomarkers. The optimal cutpoint was determined yielding the maximum corresponding sensitivity and specificity. The biomarker that has the largest area under the ROC curve was identified as having the strongest predictive power for detecting OSCC.

Table 4 Salivary mRNA biomarkers for OSCC selected by logistic regression model

Biomarker	Coefficient value	SE	P value
Intercept	-4.79	1.51	0.001
<i>IL1B</i>	5.10E +19	2.68E +19	0.062
<i>OAZ1</i>	2.18E +20	1.08E +20	0.048
<i>SAT</i>	2.63E +19	1.10E +19	0.020
<i>IL8</i>	1.36E +17	4.75E +16	0.006

NOTE. The logistic regression model was built based on the four of seven validated biomarkers (*IL1B*, *OAZ1*, *SAT*, and *IL-8*) that, in combination, provided the best prediction. The coefficient values are positive for these four markers, indicating that the synchronized increase in their concentrations in saliva increased the probability that the sample was obtained from an OSCC subject.

88% and a specificity of 81% to distinguish OSCC from the normal.

To demonstrate the utility of salivary mRNAs for disease discrimination, two classification/prediction models were examined. A logistic regression model was built based on the four of the seven validated biomarkers, *IL1B*, *OAZ1*, *SAT*, and *IL8*, which in combination provided the best prediction (Table 4). The coefficient values are positive for these four markers, indicating that the synchronized rise in their concentrations in saliva increased the probability that the sample was obtained from an OSCC subject. The leave-one-out cross-validation error rate based on logistic regression models was 19% (12 of 64). All but one (of the 64) of the models generated in the leave-one-out analysis used the same set of four markers found to be significant in the full data model specified in Table 4. The ROC curve was computed for the logistic regression model. Using a cutoff probability of 50%, we obtained a sensitivity of 91% and a specificity of 91%. The calculated area under the ROC curve was 0.95 for the logistic regression model (Fig. 1).

A second model, the “classification and regression trees (CART) model,” was generated (Fig. 2). Our fitted CART model used the salivary mRNA concentrations of *IL8*, *H3F3A*, and *SAT* as predictor variables for OSCC. *IL8*, chosen as the initial split, with a threshold of $3.14E - 18$ mol/L, produced two child groups from the parent group containing the total 64 samples. 30 samples with the *IL8* concentration $<3.14E - 18$ mol/L were assigned into “Normal-1,” whereas 34 with *IL8* concentration $\geq 3.14E - 18$ were assigned into “Cancer-1”. The “Normal-1” group was further partitioned by *SAT* with a threshold of $1.13E - 14$ mol/L. The resulting subgroups, “Normal-2” contained 25 samples with *SAT* concentration $<1.13E - 14$ mol/L, and “Cancer-2” contained 5 samples with *SAT* concentration $\geq 1.13E - 14$ mol/L. Similarly, the “Cancer-1” group was further partitioned by *H3F3A* with a threshold of $2.07E - 16$ mol/L. The resulting subgroups, “Cancer-3” contained 27 samples with *H3F3A* concentration $\geq 2.07E - 16$ mol/L, and “Normal-3” group contained 7 samples with *H3F3A* concentration $<2.07E - 16$ mol/L. Consequently, the 64 saliva samples involved in our study were classified into the “Cancer” group and the “Normal” group by CART analysis. The “Normal” group was composed of the samples from “Normal-2” and those from “Normal-3”. There are a total of 32 samples assigned in the “Normal” group, 29 from normal subjects and 3 from

cancer patients. Thus, by using the combination of *IL8*, *SAT*, and *H3F3A* for OSCC prediction, the overall sensitivity is 90.6% (29 of 32). The “Cancer” group was composed of the samples from “Cancer-2” and “Cancer-3.” There are a total of 32 samples assigned in the final “Cancer” group, 29 from cancer patients and 3 from normal subjects. Therefore, by using the combination of these three salivary mRNA biomarkers for OSCC prediction, the overall specificity is 90.6% (29 of 32).

DISCUSSION

The goal of a cancer-screening program is to detect tumors at a stage early enough that treatment is likely to be successful. Screening tools are needed that exhibit the combined features of high sensitivity and high specificity. Moreover, the screening tool must be sufficiently noninvasive and inexpensive to allow widespread applicability. Significant development of biotechnology and improvement in our basic understanding of the cancer initiation and progression now enable us to identify tumor signatures, such as oncogenes and tumor-suppressor gene alterations, in bodily fluids that drain from the organs affected by the tumor (8). The results obtained in this study will open new research directions supporting that salivary transcriptome diagnostics can be a suitable tool for the development of non-invasive diagnostic, prognostic, and follow-up tests for cancer.

Previous studies have shown that human DNA biomarkers can be identified in saliva and used for oral cancer detection (16, 17). The presence of human mRNA in saliva expands the repertoire of diagnostic analytes for translational and clinical applications. However, RNA is more labile than DNA and is presumed to be highly susceptible to degradation by RNases. Furthermore, RNase activity in saliva is reported to be elevated in patients with cancer (25). It has, thus, been commonly presumed that human mRNA could not survive extracellularly in saliva. Surprisingly, using RT-PCR, we can consistently detect human mRNA in saliva, thus opening the door to saliva-based expression profiling, as reported previously (14, 18). Using the described collection and processing protocols, we confirmed the presence of control RNAs in all saliva (patients and controls) by RT-PCR/qPCR. The quality of RNA could meet the demand for

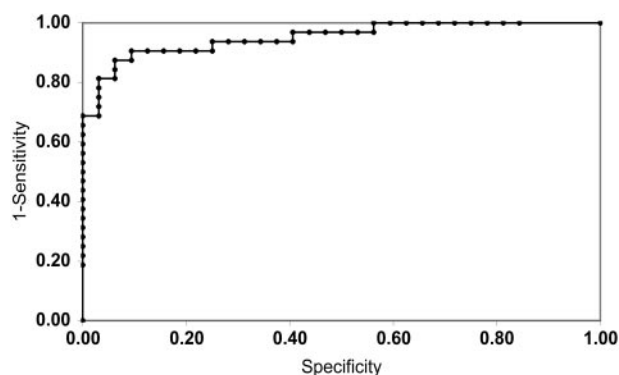
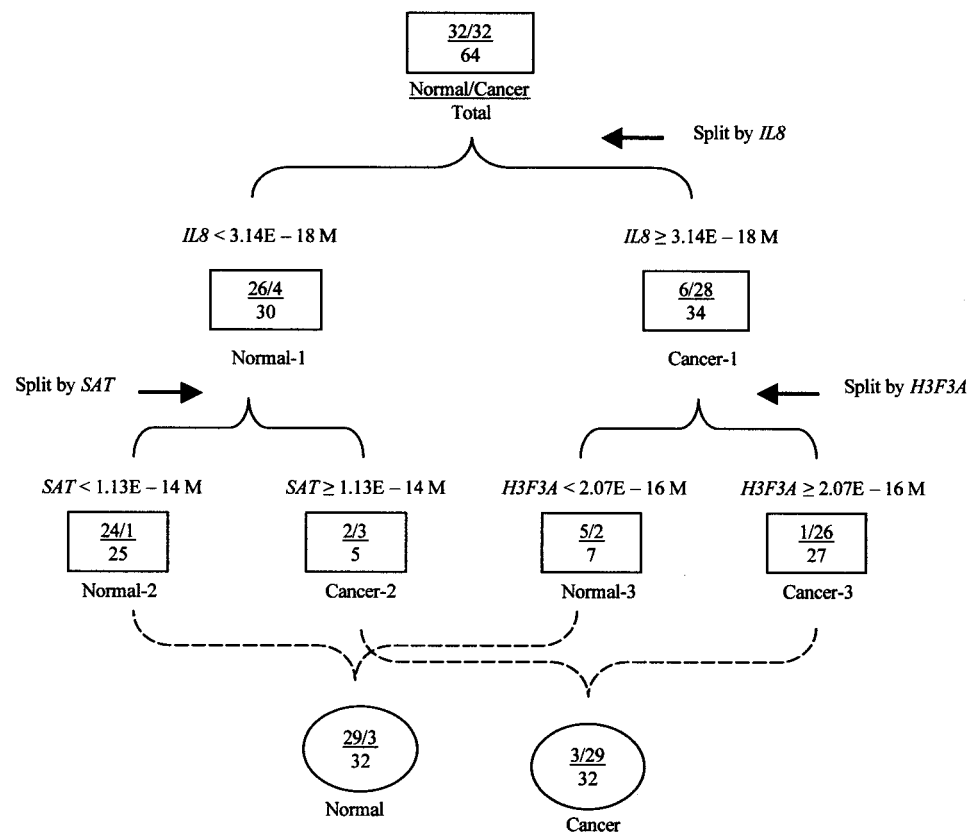


Fig. 1 ROC curve analysis for the predictive power of combined salivary mRNA biomarkers. The final logistic model included four salivary mRNA biomarkers, *IL1B*, *OAZ1*, *SAT*, and *IL8*. Using a cutoff probability of 50%, we obtained sensitivity of 91% and specificity of 91% by ROC. The calculated area under the ROC curve was 0.95.

Fig. 2 Classification and regression trees (CART) model assessing the salivary mRNA predictors for OSCC. *IL8* (cutoff value = $3.14E - 18$), chosen as the initial split, produced two child groups from the parent group containing the total 64 samples. Normal-1 group was further partitioned by *SAT* (cutoff value = $1.13E - 14$), whereas cancer-1 group was further partitioned by *H3F3A* (cutoff value = $2.07E - 16$). The 64 samples involved in this study were classified into the final cancer or normal group by CART. The overall sensitivity is 90.6% (29 of 32, in normal group) and specificity is 90.6% (29 of 32, in cancer group) for OSCC classification.



PCR, qPCR, and microarray assays. In this report, we used prompt addition of RNase inhibitors to freshly collected oral fluids followed by ultra low temperature storage (-80°C). Efforts are in progress to develop an ambient temperature saliva storage and RNA preservation protocol to facilitate clinical sample collection and ease of transportation.

Our reported findings will bring substantial interests to the field of cancer and disease diagnostics. The interests stem not only from the fact that a saliva-based diagnostic and screening test for cancer is a simple and attractive concept but also from the fact that conventional diagnostic cancer tests tend to be imperfect. With oral cancer as an example, the clearly disappointing survival rate may most probably be attributed to diagnostic delay (26). Because most oral cancers arise as asymptomatic small lesions at their early stage, only when the clinician or patient notes abnormal tissues do formal diagnosis procedures begin (2). Microscopic level for the progressive cancer is often too late for successful intervention (27). It is also impractical to use imaging techniques for cancer screening because they are time consuming and expensive. These techniques are typically used for confirmation because of their insensitivity for small lesions (28). Studies have demonstrated that good positive predictive value can be achieved by oral cancer tissue staining with toluidine blue (29). However, extensive experience is required in applying this technique and in interpreting its results. Exfoliative cytology may be a less invasive method for oral cancer detection (30). But exfoliated cancer cells tend to corre-

late with tumor burden, with lower rates of detection seen in those with minimal or early disease. The salivary mRNA biomarkers identified in this study provide a new avenue for OSCC detection. Salivary transcriptome diagnostics meets the demand for a noninvasive diagnostic tool with sufficient predictive power. Our results show much promise for salivary RNA-based clinical testing. However, we do understand that the results presented in this study have their limitations. First, the overall sample size is somewhat small ($n = 64$). Second, the use of an exploratory cohort (the 20 subjects used for the microarray experiment) in the validation cohort of 64 does introduce some bias into our findings. Third, whereas the results do demonstrate the utility of the salivary transcriptome to provide diagnostic information, they do not currently provide the basis for a population-level clinical screening test. Because of these concerns and the recommendations of the Early Detection Research Network (EDRN) of the National Cancer Institute (31), the next step will be to validate our results in an independent cohort with a larger sample size. There is also a need for additional exploratory research into salivary transcriptome diagnostics before its full clinical utility can be realized.

The cellular sources of the detected human salivary RNA is an interesting and important question to be addressed. For normal individuals, the salivary RNA sources are likely to be from one of the following three sources: salivary glands (parotid, submandibular, sublingual as well as minor glands), gingival crevicular fluids, and oral mucosal cells (lining or desqua-

mated). Efforts are in progress to obtain stratified oral fluids from these respective sources to reconstruct the salivary transcriptome in normal subjects. For oral cancer patients, the detected cancer-associated RNA signature is likely to originate from the matched tumor and/or a systemic response (local or distal) that further reflects itself in the whole saliva coming from each of the three major sources (salivary glands, gingival crevicular fluid, and oral mucosal cells). It is conceivable that disease-associated RNA can find its way into the oral cavity via the salivary gland or circulation through the gingival crevicular fluid. A good example is the elevated presence of HER-2 proteins in saliva of breast cancer patients (32). For oral cancer, the local tumor is the source of elevated salivary mRNAs. Early analysis of our data supports the matched tumor source of oral cancer salivary RNA signature. When we used a more restricted microarray (HG-U95A, Affymetrix, 12,627 probes) for previous oral tumor expression studies, *IL8*, *IL1B*, and ferritin polypeptide mRNAs were found to be significantly elevated in the saliva of oral cancer patients and are also significantly elevated in oral cancer tissues (30). In addition, because the tumor cells analyzed are procured by laser microdissection, the association should be definitive. It is gratifying to note the concordant elevation of these cellular RNAs in saliva and oral cancer tissues, via separate independent approaches. We have recently selected the most significantly elevated oral cancer tissue transcript, *IL8*, and have confirmed that its protein level (by ELISA) is also significantly elevated in saliva of oral cancer patients (18). Chen *et al.* (33) have previously independently demonstrated the elevation of IL8 protein expression in head and neck cancer tissues. These data jointly support the concordant alteration of oral cancer-associated expression changes in the tumor tissues and saliva, at the mRNA and protein levels.

In addition to *IL8*, we have identified six other cancer-associated genes as being up-regulated in saliva from oral cancer patients, such as *DUSP*, *H3F3A*, *OAZ1*, *SAT*, *S100P*, and *IL1B*. *DUSP1* gene encodes a dual specificity phosphatase and has been implicated as a mediator of tumor suppressor PTEN signaling pathway (34). The expression of *DUSP1* has been shown to decrease in ovarian tumors and a novel single-nucleotide polymorphism in the *DUSP1* gene has been identified (35). *H3F3A* mRNA is commonly used as a proliferative marker, and its level has been shown to be up-regulated in prostate cancers and colon cancers (36, 37). *OAZ1* is predicted as a tumor suppressor based on its known inhibitory function to ornithine decarboxylase (38). However, it has been reported that *OAZ1* mRNA is up-regulated in prostate cancers (36). Interestingly, the expression of *SAT*, which is also involved in polyamine metabolism, has been shown to be significantly higher in prostate cancers (36). *S100P* is known to be associated with prostate cancer progression, and its overexpression is associated with an immortalization of human breast epithelial cells *in vitro* and early stages of breast cancer development *in vivo* (39–42). Recent study shows that differential expression of *S100P* is associated with pancreatic carcinoma (43, 44). The expression of *IL1B* is also associated with cancers. The serum level of *IL1B* has been shown to be higher in patients with squamous cell carcinoma of the oral cavity (45). Also, it has been reported that the level of *IL1B* is significantly increased in the ascitic fluid of women with ovarian cancer (46). Genetic polymorphisms of

IL1B have been reported to have potential associations with the risk of diseases, such as gastric cancer and breast cancer (47, 48). It remains to be explored whether the aberrant expressions of these genes functionally contribute to the development of human OSCC. The biological significance of differential expression of these genes in head and neck/oral cancer should be determined. Identification of cancer-associated genes that are consistently changed in cancer patients will provide us not only with diagnostic markers but also with insights about molecular profiles involved in head and neck cancer development.

Understanding the profile of molecular changes in any particular cancer will be extremely useful because it will become possible to correlate the resulting phenotype of that cancer with molecular events. One of our goals is to construct risk models to facilitate assigning the appropriate salivary transcriptome-based diagnosis for patients' specific cancer risk. The multifactorial nature of oncogenesis and the heterogeneity in oncogenic pathways make it unlikely that a single biomarker will detect all cancer of a particular organ with high specificity and sensitivity. To overcome these difficulties, multiple statistical strategies were used for our prediction model to identify combinations of biomarkers that can identify OSCC patients in our samples. Although promising, the sensitivity (91%) and specificity (91%) cannot meet the demands for being a clinical tool for disease screening. Efforts are under way to validate other candidate markers and to combine them to generate a higher power for oral cancer discrimination and prediction. If we can take advantage of high-throughput RNA-based methodologies including microarray and PCR, it will be possible to multiplex biomarkers detection for disease diagnostics. The concern that informative mRNA in saliva is present in lower amounts than in cells has been addressed by using RNA-based linear amplification methodologies (49).

Saliva is increasingly being used as an investigational aid in the diagnosis of systemic diseases, such as HIV (50), diabetes mellitus (51), and breast cancer (32). Most importantly, the concepts, techniques and approach of multiple biomarkers applied in our study could easily be modified to screen and monitor other diseases. Although specific modifications may be necessary in specific applications, our proposal will provide a framework that should be addressed during the development of salivary transcriptome diagnostics. For oral cancer, one of the most important applications of the salivary transcriptome diagnostics approach is to detect the cancer conversion of oral premalignant lesions. The overall malignant transformation rates range from 11 to 70.3% (52, 53). Analysis of the DNA content in cells of oral leukoplakia was demonstrated to be useful for predicting the risk of oral cancer (54). However, it is still a post-biopsy methodology. We are currently enrolling patients with oral premalignancy to determine the specific mRNA signature in saliva as diagnostic markers. When fully explored, this innovative approach, "salivary transcriptome diagnostics," will provide new opportunities for early diagnostics of oral cancer and other human diseases.

ACKNOWLEDGMENTS

We thank Dr. Janet. A. Warrington from Affymetrix, Inc. and Microarray Facility Core in UCLA for supporting the microarray studies.

REFERENCES

1. Cancer facts and figures 2004. Atlanta: American Cancer Society; 2004.
2. Epstein JB, Zhang L, Rosin M. Advances in the diagnosis of oral premalignant and malignant lesions. *J Can Dent Assoc* 2002;68:617-21.
3. Mao L, Hong WK, Papadimitrakopoulou VA. Focus on head and neck cancer. *Cancer Cell* 2004;5:311-6.
4. Sidransky D. Emerging molecular markers of cancer. *Nat Rev Cancer* 2002;2:210-9.
5. Hollstein M, Sidransky D, Vogelstein B, Harris CC. p53 mutations in human cancers. *Science (Wash DC)* 1991;253:49-53.
6. Liu T, Wahlberg S, Burek E, Lindblom P, Rubio C, Lindblom A. Microsatellite instability as a predictor of a mutation in a DNA mismatch repair gene in familial colorectal cancer. *Genes Chromosomes Cancer* 2000;27:17-25.
7. Groden J, Thliveris A, Samowitz W, et al. Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* 1991;66:589-600.
8. Sidransky D. Nucleic acid-based methods for the detection of cancer. *Science (Wash DC)* 1997;278:1054-9.
9. Anker P, Mulcahy H, Chen XQ, Stroun M. Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients. *Cancer Metastasis Rev* 1999;18:65-73.
10. Rieger-Christ KM, Mourtzinou A, Lee PJ, et al. Identification of fibroblast growth factor receptor 3 mutations in urine sediment DNA samples complements cytology in bladder tumor detection. *Cancer (Phila)* 2003;98:737-44.
11. Wong LJ, Lueth M, Li XN, Lau CC, Vogel H. Detection of mitochondrial DNA mutations in the tumor and cerebrospinal fluid of medulloblastoma patients. *Cancer Res* 2003;63:3866-71.
12. Kopreski MS, Benko FA, Gocke CD. Circulating RNA as a tumor marker: detection of 5T4 mRNA in breast and lung cancer patient serum. *Ann N Y Acad Sci* 2001;945:172-8.
13. Bunn PJ Jr. Early detection of lung cancer using serum RNA or DNA markers: ready for "prime time" or for validation? *J Clin Oncol* 2003;21:3891-3.
14. Li Y, Zhou X, St. John MAR, Wong DT. RNA profiling of cell-free saliva using microarray technology. *J Dent Res* 2004;83:199-203.
15. Lawrence HP. Salivary markers of systemic disease: noninvasive diagnosis of disease and monitoring of general health. *J Can Dent Assoc* 2002;68:170-4.
16. El-Naggar AK, Mao L, Staerkel G, et al. Genetic heterogeneity in saliva from patients with oral squamous carcinomas: implications in molecular diagnosis and screening. *J Mol Diagn* 2001;3:164-70.
17. Liao PH, Chang YC, Huang MF, Tai KW, Chou MY. Mutation of p53 gene codon 63 in saliva as a molecular marker for oral squamous cell carcinomas. *Oral Oncol* 2000;36:272-6.
18. St. John MA, Li Y, Zhou X, et al. Interleukin 6 and interleukin 8 as potential biomarkers for oral cavity and oropharyngeal squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg* 2004;130:929-35.
19. Navazesh M. Methods for collecting saliva. *Ann N Y Acad Sci* 1993;694:72-7.
20. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci USA* 2001;98:31-6.
21. Ginzinger DG. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Hematol* 2002;30:503-12.
22. Grunkemeier GL, Jin R. Receiver operating characteristic curve analysis of clinical risk models. *Ann Thorac Surg* 2001;72:323-6.
23. Renger R, Meadows LM. Use of stepwise regression in medical education research. *Acad Med* 1994;69:738.
24. Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W. Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Ann Behav Med* 2003;26:172-81.
25. Kharchenko SV, Shpakov AA. [Regulation of the RNase activity of the saliva in healthy subjects and in stomach cancer]. *Izv Akad Nauk SSSR Biol* 1989;58-63.
26. Wildt J, Bundgaard T, Bentzen SM. Delay in the diagnosis of oral squamous cell carcinoma. *Clin Otolaryngol* 1995;20:21-5.
27. Fong KM, Srivastava S, Gopal-Srivastava R, Kramer BS. Molecular genetic basis for early cancer detection and cancer susceptibility. In: Srivastava SHD, Gazdar AF, editors. *Molecular pathology of early Cancer*. Amsterdam, The Netherlands, IOS Press; 1998. p. 13-26.
28. Myers LL, Wax MK. Positron emission tomography in the evaluation of the negative neck in patients with oral cavity cancer. *J Otolaryngol* 1998;27:342-7.
29. Mashberg A, Samit A. Early diagnosis of asymptomatic oral and oropharyngeal squamous cancers. *CA Cancer J Clin* 1995;45:328-51.
30. Rosin MP, Epstein JB, Berean K, et al. The use of exfoliative cell samples to map clonal genetic alterations in the oral epithelium of high-risk patients. *Cancer Res* 1997;57:5258-60.
31. Sullivan Pepe M, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst (Bethesda)* 2001;93:1054-61.
32. Streckfus C, Bigler L, Dellinger T, Dai X, Kingman A, Thigpen JT. The presence of soluble c-erbB-2 in saliva and serum among women with breast carcinoma: a preliminary study. *Clin Cancer Res* 2000;6:2363-70.
33. Chen Z, Malhotra PS, Thomas GR, et al. Expression of proinflammatory and proangiogenic cytokines in patients with head and neck cancer. *Clin Cancer Res* 1999;5:1369-79.
34. Unoki M, Nakamura Y. Growth-suppressive effects of BPOZ and EGR2, two genes involved in the PTEN signaling pathway. *Oncogene* 2001;20:4457-65.
35. Suzuki C, Unoki M, Nakamura Y. Identification and allelic frequencies of novel single-nucleotide polymorphisms in the DUSP1 and BTG1 genes. *J Hum Genet* 2001;46:155-7.
36. Bettuzzi S, Davalli P, Astancolle S, et al. Tumor progression is accompanied by significant changes in the levels of expression of polyamine metabolism regulatory genes and clusterin (sulfated glycoprotein 2) in human prostate cancer specimens. *Cancer Res* 2000;60:28-34.
37. Torelli G, Venturelli D, Colo A, et al. Expression of c-myc protooncogene and other cell cycle-related genes in normal and neoplastic human colonic mucosa. *Cancer Res* 1987;47:5266-9.
38. Tsuji T, Usui S, Aida T, et al. Induction of epithelial differentiation and DNA demethylation in hamster malignant oral keratinocyte by ornithine decarboxylase antizyme. *Oncogene* 2001;20:24-33.
39. Gribenko A, Lopez MM, Richardson JM III, Makhatadze GI. Cloning, overexpression, purification, and spectroscopic characterization of human S100P. *Protein Sci* 1998;7:211-5.
40. Guerreiro Da Silva ID, Hu YF, Russo IH, et al. S100P calcium-binding protein overexpression is associated with immortalization of human breast epithelial cells in vitro and early stages of breast cancer development in vivo. *Int J Oncol* 2000;16:231-40.
41. Mousses S, Bubendorf L, Wagner U, et al. Clinical validation of candidate genes associated with prostate cancer progression in the CWR22 model system using tissue microarrays. *Cancer Res* 2002;62:1256-60.
42. Mackay A, Jones C, Dexter T, et al. cDNA microarray analysis of genes associated with ERBB2 (HER2/neu) overexpression in human mammary luminal epithelial cells. *Oncogene* 2003;22:2680-8.
43. Logsdon CD, Simeone DM, Binkley C, et al. Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer. *Cancer Res* 2003;63:2649-57.
44. Crnogorac-Jurcevic T, Missiaglia E, Blaveri E, et al. Molecular alterations in pancreatic carcinoma: expression profiling shows that dysregulated expression of S100 genes is highly prevalent. *J Pathol* 2003;201:63-74.

45. Jablonska E, Piotrowski L, Grabowska Z. Serum Levels of IL-1b, IL-6, TNF-a, sTNF-RI and CRP in patients with oral cavity cancer. *Pathol Oncol Res* 1997;3:126–9.
46. Chen CK, Wu MY, Chao KH, Ho HN, Sheu BC, Huang SC. T lymphocytes and cytokine production in ascitic fluid of ovarian malignancies. *J Formos Med Assoc* 1999;98:24–30.
47. Hamajima N, Yuasa H. [Genetic polymorphisms related to interleukin-1 beta production and disease risk]. *Nippon Kosho Eisei Zasshi* 2003;50:194–207.
48. El-Omar EM, Rabkin CS, Gammon MD, et al. Increased risk of noncardia gastric cancer associated with proinflammatory cytokine gene polymorphisms. *Gastroenterology* 2003;124:1193–201.
49. Alevizos I, Mahadevappa M, Zhang X, et al. Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene* 2001;20:6196–204.
50. Malamud D. Oral diagnostic testing for detecting human immunodeficiency virus-1 antibodies: a technology whose time has come. *Am J Med* 1997;102:9–14.
51. Guven Y, Satman I, Dincag N, Alptekin S. Salivary peroxidase activity in whole saliva of patients with insulin-dependent (type-1) diabetes mellitus. *J Clin Periodontol* 1996;23:879–81.
52. Lee JJ, Hong WK, Hittelman WN, et al. Predicting cancer development in oral leukoplakia: ten years of translational research. *Clin Cancer Res* 2000;6:1702–10.
53. Silverman S Jr, Gorsky M. Proliferative verrucous leukoplakia: a follow-up study of 54 cases. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 1997;84:154–7.
54. Sudbo J, Kildal W, Risberg B, Koppang HS, Danielsen HE, Reith A. DNA content as a prognostic marker in patients with oral leukoplakia. *N Engl J Med* 2001;344:1270–8.
55. Berta GN, Ghezzi F, D'Avolio A, et al. Enhancement of calcyclin gene RNA expression in squamous cell carcinoma of the oral mucosa, but not in benign lesions. *J Oral Pathol Med* 1997;26:206–10.
56. Watanabe H, Iwase M, Ohashi M, Nagumo M. Role of interleukin-8 secreted from human oral squamous cell carcinoma cell lines. *Oral Oncol* 2002;38:670–9.