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. Wong1,4,5

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,
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 × 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 (DNasel-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.dchp.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) and random hexamers as primer (Applied Biosystems). 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 each step involved in 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 proves the model. Subsequently, we used the marker values for 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 μg (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 (χ² test, P < 0.0001), considering the false positives with P < 0.05. Using a predefined criteria of

### RESULTS

On average, 54.2 ± 20.1 ng (n = 64) of total RNA was obtained from 560 μL 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 salivary 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 μg (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 (χ² test, P < 0.0001), considering the false positives with P < 0.05. Using a predefined criteria of

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>GenBank accession no.</th>
<th>Locus</th>
<th>Gene functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2M</td>
<td>β-2-microglobulin</td>
<td>NM_04048</td>
<td>15q21-q22.2</td>
<td>Antiapoptosis; antigen presentation</td>
</tr>
<tr>
<td>DUSP1</td>
<td>Dual specificity phosphatase 1</td>
<td>NM_04417</td>
<td>5q34</td>
<td>Protein modification; signal transduction; oxidative stress</td>
</tr>
<tr>
<td>FTH1</td>
<td>Ferritin, heavy polypeptide 1</td>
<td>NM_02032</td>
<td>11q13</td>
<td>Iron ion transport; cell proliferation</td>
</tr>
<tr>
<td>G0S2</td>
<td>Putative lymphocyte G0-G1 switch gene</td>
<td>NM_015714</td>
<td>1q32.2-q41</td>
<td>Cell growth and/or maintenance; regulation of cell cycle</td>
</tr>
<tr>
<td>GADD45B</td>
<td>Growth arrest and DNA-damage-inducible, β</td>
<td>NM_015675</td>
<td>1p13.3</td>
<td>Kinase cascade; apoptosis</td>
</tr>
<tr>
<td>H3F3A</td>
<td>H3 histone, family 3A</td>
<td>BE869922</td>
<td>1q41</td>
<td>DNA binding activity</td>
</tr>
<tr>
<td>HSPC016</td>
<td>Hypothetical protein HSPC016</td>
<td>BG167522</td>
<td>3p21.31</td>
<td>Unknown</td>
</tr>
<tr>
<td>IER3</td>
<td>Immediate early response 3</td>
<td>NM_003897</td>
<td>6q21.3</td>
<td>Embryogenesis; morphogenesis; apoptosis; cell growth and maintenance</td>
</tr>
<tr>
<td>IL1B</td>
<td>Interleukin 1, β</td>
<td>M15330</td>
<td>2q14</td>
<td>Signal transduction; proliferation; inflammation; apoptosis</td>
</tr>
<tr>
<td>IL8</td>
<td>Interleukin 8</td>
<td>NM_000584</td>
<td>4q13-q21</td>
<td>Angiogenesis; replication; calcium-mediated signaling pathway; cell adhesion; chemotaxis; cell cycle arrest; immune response</td>
</tr>
<tr>
<td>MAP2K3</td>
<td>Mitogen-activated protein kinase kinase 3</td>
<td>AA780381</td>
<td>17q11.2</td>
<td>Signal transduction; protein modification</td>
</tr>
<tr>
<td>OAZ1</td>
<td>Ornithine decarboxylase antizyme 1</td>
<td>D87914</td>
<td>19p13.3</td>
<td>Polyamine biosynthesis</td>
</tr>
<tr>
<td>PRG1</td>
<td>Proteoglycan 1, secretory granule</td>
<td>NM_002727</td>
<td>10q22.1</td>
<td>Proteoglycan</td>
</tr>
<tr>
<td>RG52</td>
<td>Regulator of G-protein signaling 2, 24 kda</td>
<td>NM_002923</td>
<td>1q31</td>
<td>Oncogenesis; G-protein signal transduction</td>
</tr>
<tr>
<td>S100P</td>
<td>S100 calcium binding protein P</td>
<td>NM_005980</td>
<td>4p16</td>
<td>Protein binding; calcium ion binding</td>
</tr>
<tr>
<td>SAT</td>
<td>Spermideine/spermine N1-acetyltransferase</td>
<td>NM_002970</td>
<td>Xp22.1</td>
<td>Enzyme, transferase activity</td>
</tr>
</tbody>
</table>

**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.
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 mRNAs 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 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.

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

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Primer sequence (5’ to 3’)</th>
<th>Validated*</th>
<th>$P$ value</th>
<th>Mean fold increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUSP1</td>
<td>F: CTCACCAGTATTATTCCCGACG</td>
<td>Yes</td>
<td>0.039</td>
<td>2.60</td>
</tr>
<tr>
<td></td>
<td>R: GTGGAAGGCAGACCTACAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3F3A</td>
<td>F: AAAGCCACCCAGGAGAGCAAC</td>
<td>Yes</td>
<td>0.011</td>
<td>5.61</td>
</tr>
<tr>
<td></td>
<td>R: GCGAATCAGAAGTCTGGGAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL1B</td>
<td>F: GTGCTGAATGTTGAGCATTACGCC</td>
<td>Yes</td>
<td>0.005</td>
<td>5.48</td>
</tr>
<tr>
<td></td>
<td>R: ACCCTAGGACAGGCAGGCTTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL8</td>
<td>F: GAGGGTGTGAGAGAAGTTCGTTT</td>
<td>Yes</td>
<td>0.000</td>
<td>24.3</td>
</tr>
<tr>
<td></td>
<td>R: CTGGGCATCTCTCAGGTTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OAZ1</td>
<td>F: AGAGAGAGCCTTGGGAGAGG</td>
<td>Yes</td>
<td>0.009</td>
<td>2.82</td>
</tr>
<tr>
<td></td>
<td>R: AGATGAGCCTGCTACGGGTTC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S100P</td>
<td>F: GAGGCTACGTTGGCTGGGCTTG</td>
<td>Yes</td>
<td>0.003</td>
<td>4.88</td>
</tr>
<tr>
<td></td>
<td>R: CTCAGGGGCGATCATTGAGACCAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAT</td>
<td>F: CCAGTGAAGAAGGTTGAGAAGC</td>
<td>Yes</td>
<td>0.005</td>
<td>2.98</td>
</tr>
<tr>
<td></td>
<td>R: TGGAGGGTGTTCATCAAGCAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GADD45B</td>
<td>F: TGATGAAATGTTGGGACCGAC</td>
<td>No</td>
<td>0.116</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>R: GAGCGTGAGAAGTGTGGC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGS2</td>
<td>F: CCTGCCATAAAGACTGACCTTG</td>
<td>No</td>
<td>0.149</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>R: GCTTCCTGATTCCACTACCAAAC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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).

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

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

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

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Coefficient value</th>
<th>SE</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercep</td>
<td>-4.79</td>
<td>1.51</td>
<td>0.001</td>
</tr>
<tr>
<td>IL1B</td>
<td>5.10E +19</td>
<td>2.68E +19</td>
<td>0.062</td>
</tr>
<tr>
<td>OAZ1</td>
<td>2.18E +20</td>
<td>1.08E +20</td>
<td>0.048</td>
</tr>
<tr>
<td>SAT</td>
<td>2.63E +19</td>
<td>1.10E +19</td>
<td>0.020</td>
</tr>
<tr>
<td>IL8</td>
<td>1.36E +17</td>
<td>4.75E +16</td>
<td>0.006</td>
</tr>
</tbody>
</table>

NOTE. The logistic regression model was built based on the four of seven validated biomarkers (IL1B, OAZ1, SAT, and IL8) 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 increase 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 samples. 30 samples with the

18 were assigned into “Cancer-1”. 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 sensitivity 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...
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 correlate 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 saliva 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, OA31, SAT, SI100P, 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). SI100P 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 SI100P is associated with pancreatic carcinoma (43, 44). The expression of IL1B is also associated with cancers. The serum level of IL1B has 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


Salivary Transcriptome Diagnostics for Oral Cancer Detection

Yang Li, Maie A. R. St. John, Xiaofeng Zhou, et al.


Updated version
Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/10/24/8442

Cited articles
This article cites 50 articles, 14 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/10/24/8442.full#ref-list-1

Citing articles
This article has been cited by 43 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/10/24/8442.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.