Prevalidation of Salivary Biomarkers for Oral Cancer Detection


Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-11-1093

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2012/01/31/1055-9965.EPI-11-1093.DC1.html

Cited Articles
This article cites by 20 articles, 13 of which you can access for free at:
http://cebp.aacrjournals.org/content/21/4/664.full.html#ref-list-1

Citing articles
This article has been cited by 6 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/21/4/664.full.html#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.
Prevalidation of Salivary Biomarkers for Oral Cancer Detection

David Elashoff1, Hui Zhou2, Jean Reiss2, Jianghua Wang4, Hua Xiao4, Bradley Henson4, Shen Hu4, Martha Arellano5, Uttam Sinha5, Anh Le3, Diana Messadi4, Marilene Wang3, Vishad Nabili3, Mark Lingen8, Darly Morris9, Timothy Randolph9, Ziding Feng4, David Akin4, Dragana A. Kastratovic10, David Chia4, Elliot Abemayor3, and David T.W. Wong4,5

Abstract

Background: Oral cancer is the sixth most common cancer with a 5-year survival rate of approximately 60%. Presently, there are no scientifically credible early detection techniques beyond conventional clinical oral examination. The goal of this study is to validate whether the seven mRNAs and three proteins previously reported as biomarkers are capable of discriminating patients with oral squamous cell carcinomas (OSCC) from healthy subjects in independent cohorts and by a National Cancer Institute (NCI)-Early Detection Research Network (EDRN)-Biomarker Reference Laboratory (BRL).

Methods: Three hundred and ninety-five subjects from five independent cohorts based on case controlled design were investigated by two independent laboratories, University of California, Los Angeles (Los Angeles, CA) discovery laboratory and NCI-EDRN-BRL.

Results: Expression of all seven mRNA and three protein markers was increased in OSCC versus controls in all five cohorts. With respect to individual marker performance across the five cohorts, the increase in interleukin (IL)-8 and subcutaneous adipose tissue (SAT) was statistically significant and they remained top performers across different cohorts in terms of sensitivity and specificity. A previously identified multiple marker model showed an area under the receiver operating characteristic (ROC) curve for prediction of OSCC status ranging from 0.74 to 0.86 across the cohorts.

Conclusions: The validation of these biomarkers showed their feasibility in the discrimination of OSCCs from healthy controls. Established assay technologies are robust enough to perform independently. Individual cutoff values for each of these markers and for the combined predictive model need to be further defined in large clinical studies.

Impact: Salivary proteomic and transcriptomic biomarkers can discriminate oral cancer from control subjects. Cancer Epidemiol Biomarkers Prev; 21(4); 664–72. ©2012 AACR.

Introduction

Oral cancer [more than 90% are oral squamous cell carcinomas (OSCC)] is the sixth most common cancer worldwide with an average 5-year survival rate of approximately 60% (1). This poor survival rate has not improved in the past 3 decades despite improvements in therapeutic strategies. The key challenge to reduce the mortality and morbidity of this disease is to develop strategies to identify and detect OSCC when it is at very early stage, which will enable effective intervention and therapy. Up to now, beyond conventional clinical oral examination, there are no scientifically credible, reliable early detection techniques available (1).

Authors’ Affiliations: Departments of 1Medicine and Biostatistics and 2Pathology and Laboratory Medicine, 3Division of Head and Neck Surgery/Otolaryngology, David Geffen School of Medicine, 4School of Dentistry and Dental Research Institute, 5Jonsson Comprehensive Cancer Center, University of California at Los Angeles; 6Otolaryngology/Head and Neck, University of Southern California, Keck’s School of Medicine; 7Division of Endodontics, Oral and Maxillofacial Surgery and Orthodontics, Center for Craniofacial Molecular Biology, Herman Ostrow School of Dentistry of University of Southern California, Los Angeles, California; 8Department of Pathology, University of Chicago, Chicago, Illinois; 9Bio-statistics Program, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington; and 10Clinical Center of Serbia, Pasterova, Belgrade, Serbia

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

D. Elashoff, H. Zhou, J. Reiss, and J. Wang are joint first authors.

D. Chia, E. Abemayor, and D.T.W. Wong are joint senior authors.

Hui Zhou’s current affiliation: Genome Sciences Resource, Vanderbilt University Medical Center, Nashville, TN. Email: usczhou@gmail.com.

Corresponding Author: D.T.W. Wong, School of Dentistry, University of California at Los Angeles, 10833 Le Conte Avenue, 73-017 CHS, Los Angeles, CA 90095. Phone: 310-206-3048; Fax: 310-825-7609; E-mail: dtww@ucla.edu

doi: 10.1158/1055-9965.EPI-11-1093

©2012 American Association for Cancer Research.
Saliva is a valuable body fluid for disease diagnosis, due to its noninvasive nature, and has been increasingly used as a source for discovery of oral cancer biomarkers (2–6). The ability to detect molecules in saliva from patients with head and neck cancer (7–9) provides unique opportunities to develop noninvasive diagnostics. Salivary molecules have already been proposed as potential oral cancer biomarkers. For example, salivary soluble CD44 (3, 10), salivary Cyfra 21-1, tissue polypeptide anti-gene, and CA125 have been proposed as oral cancer markers (6). Nevertheless, no single biomolecule has been shown to meet the real-world requirement for high accuracy in identifying early disease onset, suggesting that multiple biomarker candidates are needed for high accuracy and sensitivity in detecting OSCCs. In addition, extensive and rigorous biomarker validation will be crucial to the acceptance of newly discovered biomarker candidates before adoption for clinical use.

Global profiling of disease-associated molecules, such as proteins, DNA, mRNA, microRNA, and metabolites is becoming the state-of-the-art method to provide promising disease biomarker candidates. The ability to globally profile these molecules in saliva, through transcriptomic (7) and proteomic (5, 11) approaches, as well as the ability to detect specific molecules in saliva will greatly enhance the opportunities to identify reliable oral cancer biomarkers. Our prior salivary transcriptomic studies have discovered 7 OSCC-associated salivary RNAs (IL-8, SAT, IL-1B, OAZ1, H3F3A, DUSP, S100P). Initially, the levels of these RNAs were measured in a training set of 32 OSCC/32 control by quantitative PCR, and a logistic regression model including 4 markers (IL-8, SAT, IL-1B, and OAZ1) achieved a cross-validation prediction accuracy rate of 81% showing their potential as biomarkers for oral cancer detection (7). This has opened a new avenue for salivary biomarker discovery for human diseases. Here, we first report our follow-up validations of these oral cancer biomarkers in 2 independent cohorts in our laboratory. We also described our efforts in standardizing methods used in the validation process of these salivary biomarkers. In addition, we have detailed the outcomes of another 3 independent validations conducted in collaboration with the National Cancer Institute’s Early Detection Research Network (NCI-EDRN)’s Biomarker Reference Laboratory (BRL). The entire salivary biomarker assays for mRNAs and proteins have been streamlined and automated to minimize interlaboratory variance.

Materials and Methods

Subjects

Patients and control subjects were recruited at the Medical Centers at the University of California at Los Angeles (UCLA; Los Angeles, CA) and University of Southern California (USC; Los Angeles, CA) and Veteran Hospital in greater Los Angeles (VAGLA; Los Angeles, CA) from year 2004 to 2007. We enrolled patients diagnosed with primary OSCC stage I–IV (T1–T4) according to the TNM (tumor-node-metastasis) criteria without previous chemotherapy or radiotherapy. All the protocols used in this study were approved by the Ethics Review Boards, Institutional Review Board at UCLA, and at all of the participating institutions. All participants provided written informed consent before saliva collection.

We sequentially recruited 5 retrospective case–control groups in which the disease status was known at time of subject recruitment. These 5 groups are described as cohort 1 through cohort 5 throughout. Cohorts 1 and 2 were configured with matched sex distributions among the subjects. Cohorts 3–5 were distributionally matched for age, sex, ethnicity, and smoking history. Controls were selected on the basis of matched criteria from enrolled populations from “spit drives” at the UCLA Medical Center without a prior history and diagnosis of head and neck cancer.

Saliva collection

Unstimulated whole saliva was collected according to our published protocol (7), with modifications and supernatant saliva processed the same way as reported (12) and stored as SUPERASE-In or protease inhibitor–preserved salivary supernatant.

Primer design

Primers (set 1) designed for real-time PCR has been described in the work of Li and colleagues (7, 12) and was used in study cohorts 1 and 2. Primers (set 2) were used in generating templates for in vitro transcription of RNA to produce standard RNA curves. The sequence were based on set 1 but extended to both 5’ and 3’ ends with T7 promoter sequences fused to the 5’ end of each sense primer. The primers were designed by using Primer3. All primers were BLAST-searched against GenBank to confirm the gene specificity. Primers were synthesized by Sigma-Aldrich.

Generation of in vitro transcribed RNA for standard curves of the 7 salivary RNA oral cancer markers

To generate templates for in vitro transcription for the 7 salivary oral cancer markers, conventional real-time PCR (RT-PCR) was carried out using total RNA from human OSCC cell line as templates. The cDNA was synthesized in 20 μL of reverse transcription reaction mix with 50 U MuLV reverse transcriptase (Applied Biosystems), 20 U RNAse inhibitor (Applied Biosystems), 10 mmol/L dNTPs, and 5 mmol/L random hexamers. One-microliter aliquots of cDNA were used in a 20-μL PCR reaction with 400 nmol/L set 2 primers at the PCR condition of 95°C for 3 minutes followed by 40 cycle of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 30 seconds, and final extension at 72°C for 7 minutes. RT-PCR products were confirmed on a 2% agarose gel stained with ethidium bromide. In vitro transcription was conducted.
using T7 MEGAsprint transcribe kit (Ambion). The resultant RNA transcripts were purified with column-based RNA clean up module (Arcturus). The cRNAs were quantified with Nanodrop spectrometry ND-1000 for quantity and quality ($A_{260}/A_{280}$ ratio). Aliquots from the same standard stock for each of the salivary RNA biomarkers were used in all qPCR reactions.

**Total salivary RNA extraction, cDNA synthesis**

Total RNA was extracted according to the following procedures for the first 2 cohorts: frozen RNA later preserved as saliva (containing equal volume of raw saliva) was thawed on ice and total RNA was extracted using Viral Mini Kit (Qiagen) according to manufacturer’s instructions. The resultant total RNA was eluted in 40 µL of elution buffer and was treated with DNase I (Ambion) according to manufacturer’s instruction for 30 minutes at 37°C to remove any genomic DNA contamination. After inactivating DNase, purified total RNA aliquots were stored at -80°C until use.

Starting from cohort 3 and all the EDRN studies, SUPERASE-preserved supernatant saliva was used with a semi-automated robot (KingFisher, Thermo Scientific) and MagMix viral RNA mini kit (Ambion) in total RNA isolation. Briefly, 300 µL samples were mixed with magnet beads and binding enhancer (1:1, total 20 µL) and then with lysis buffer for 10 minutes at room temperature before adding 800 µL isopropanol. The total lysates were then transferred into the plastic holder and the automated RNA isolation was started. The resulting purified RNAs were eluted into 100 elution buffer with DNase I (Ambion) treatment as described above.

For cDNA synthesis, 1 µL of RNA standards (containing 10⁷ to 10⁸ copy number of IVT RNA), 1 µL of salivary RNA was reverse transcribed with 50 U of MuLV reverse transcriptase (Applied Biosystems), using gene-specific primers (downstream primer of set 1) in a total volume of 20 µL at the following condition: incubation at 25°C for 10 minutes, reverse transcribed at 42°C for 45 minutes followed by inactivation at 95°C for 5 minutes. The resulting cDNAs were stored at -20°C until later use.

**SYBR-based quantitative real-time PCR**

Quantitative PCR was carried out in triplicate for each sample in a 3-step protocol using melting curve method. PCR amplification of the 7 salivary marker mRNAs was carried out in a 20 µL reaction mix containing a final concentration of 400 nmol/L sense and antisense primers, 3 µL cDNA (salivary cDNA) or 2 µL standard cDNA, and mixed with 10 µL 2 x SYBR Green I Master mix (Applied Biosystems). All PCR reactions were carried out on ABI 7500 Fast Thermal Cycler.

**Methods for NCI-EDRN-BRL validation**

**Development of standardized protocols and automation to enhance biomarker validation.** The NCI-EDRN is a dedicated entity committed to make molecular diagnostics for cancer a clinical reality. We subjected our salivary oral cancer biomarkers to a phase II validation study with an EDRN-BRL to assess the reproducibility of the oral cancer salivary biomarkers. Initially, technical and methodologic information was transferred, via hands-on training, to the operator in an NCI-EDRN-sponsored laboratory. Experimental protocols were optimized in the following ways. First, by incorporating semiautomation into the RNA isolation step, operator variance was reduced and reproducibility in RNA recovery efficiency was improved. Second, high-throughput multiplex RT and a novel one-step RT-PCR preamplification method developed in our laboratory (13) to analyze multiple targets in a single reaction were used. Third, the automated liquid handling robot in the PCR setup procedures was applied. With these technical improvements in place, the NCI-EDRN laboratory conducted independent assays for the salivary oral cancer biomarkers on 3 different independent cohorts. Table 1 provides an overview of different experimental methods we used to assay the cohorts in each laboratory (cohorts 3–5 were assayed at EDRN-BRL).

**RT-PCR preamplification**

We have optimized a preamplification procedure for salivary mRNA for oral cancer target detection, and

### Table 1. Comparison of materials and methods used in the 5 OSCC salivary mRNA biomarker validation studies

<table>
<thead>
<tr>
<th>Data set</th>
<th>Sample size (OSCC/control)</th>
<th>Type of saliva</th>
<th>RNA isolation</th>
<th>Concentration determination</th>
<th>Priming</th>
<th>RT and PCR strategy</th>
<th>Equipment including</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohorts 1, 2</td>
<td>48/48; 24/24</td>
<td>Whole saliva</td>
<td>QIAamp Viral RNA Mini-Kit (Qiagen)</td>
<td>Cq</td>
<td>Mixed oligo random</td>
<td>Multiplex RT, two-step RT-PCR</td>
<td>BioRad iQ5 Manual baseline</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,120 µL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohorts 3, 4, 5</td>
<td>30/30; 54/36; 70/31</td>
<td>Supernatant</td>
<td>Ambion Viral RNA Kit</td>
<td>Cq</td>
<td>Nested gene-specific primers</td>
<td>Preamplification RT and PCR</td>
<td>ABI 7500 fast</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(300 µL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
detailed methods for primer design, one-step RT and PCR, excess primer removal, and quantitative PCR programming have been described previously (13). Primers for preamplification and quantitative PCR were designed as intron-span and sequence can be found in Supplementary Table S2. The reactions were set up using the BioMek 3000 Liquid Handling Platform into 96-well plates on PCR plate cooler.

Salivary protein detection
Salivary interleukin (IL)-8, IL-1B and M2BP proteins were detected with specific ELISA kits (Pierce) according to manufacturer’s instruction. M2BP is a tumor antigen (14–16) found to be significantly upregulated in nasopharyngeal carcinoma. The serum levels of M2BP were also significantly elevated in both patients with nasopharyngeal carcinoma and nasopharyngeal carcinoma nude mice model compared with healthy people or tumor-free mice (17).

Optimal dilutions for saliva are 1:8 for IL-8, 1:3 for IL-1B, and 1:20 for M2BP. All samples were assayed in same plate in duplicates. Results were obtained by microplate reader and converted into picogram per milliliter according to the standard curve with the standards provided.

Statistical methods
All the statistical analyses were conducted using the statistical software R 2.9.1 and Bioconductor 2.5 (packages: ROC, meta). Means and SDs were calculated for both OSCCs and control samples for each marker in each cohort. Differential expression of each marker was assessed within each cohort by using the Mann–Whitney U test to determine statistical significance and by constructing a receiver operating characteristic (ROC) curves. The area under the curve (AUC) of the each of the ROC curves was obtained by numerical integration. Sensitivity and specificity were assessed for each marker in each cohort by using an expression cutoff point corresponding to the value of the pth percentile for the marker, where p is the proportion of OSCC cases in the cohort under consideration.

For meta-analysis of the individual markers, the estimates of sensitivity and specificity were used to generate a combined estimate using both fixed- and random-effects inverse variance meta-analysis models (18). These models were constructed separately for the sensitivity and specificity and were weighted proportionally to the relevant sample sizes in each study. The individual cohort sensitivities and specificities were used instead of the original marker values because of the differences in the measurement scales across the experiments and the differences in the proportions of OSCC cases in each cohort. We also conducted a meta-analysis to compare marker expression between cancer and controls. This analysis used a z-transformation within each group to normalize the expression levels across cohorts and then used a mixed-effect model to compare the z-transformed values between cancer and controls.

For the classification model evaluation, logistic regression was used to examine the use of combinations of markers for OSCC classification. For each cohort, we constructed 2 models. First, we used the same 4 markers (IL-8 + IL-1B + SAT + OAZ1) as in our original report (7) in each cohort. Next, we used forward stepwise regression within each cohort to identify the best 4 markers for each individual cohort. The AUC was computed by constructing an ROC curve using the predicted probabilities from each logistic model and then numerically integrating the curve. Sensitivity and specificity for the models were computed using the same method as for the individual marker models.

Results

Technical variations between the validation studies

Table 1 summarizes the technical details, modifications, and improvements that were made to improve the robustness and reproducibility of the salivary assays. A number of experimental parameters were evaluated. For RNA isolation, the Qiagen Viral RNA Mini-kit was first used for the first 2 cohorts. Upon working with the NCI-EDRN group, it became clear that to minimize assay variance, automated RNA isolation and liquid handling were necessary. We found the KingFisher automated RNA extractor and the BioMek 300 liquid handler to be of greatest value. Upon working with the EDRN-BRL, we also switched from the BioRad iQ5 to the ABI 7500 fast equipment for qPCR. Primer designs and qPCR strategies started with mixed oligonucleotide random primers. Gene-specific priming improved the specificity and this was further modified by our finding that preamplification of RNA targets using gene-specific primers followed by PCR significantly increased the robustness and the ability to assay small amount of salivary RNA.

Population cohorts
Table 2 reports the demographic characteristics of the subject cohorts. The average age for healthy controls was significantly lower than OSCC subjects for cohorts 1 and 2. For example in cohort 1, controls had an average age of 31 years whereas patients with OSCC averaged age of 63 years. Age was not significantly different in the later 3 cohorts due to distributional age matching. We have staging for 55 of the subjects from the later 3 cohorts. For these subjects, 13 (24%) were stage I, 9 (16%) were stage II, 21 (38%) were stage III, and 12 (22%) were stage IV.

Validation of the biomarkers in five independent cohorts
Since the publication of our initial discovery of 7 candidate mRNA biomarkers for OSCC detection almost 7 years ago (7), the behavior of the 7 oral cancer mRNA candidate biomarkers has been evaluated in 5 population-based, case-control studies totaling 395 subjects. All 7
markers were increased in OSCC saliva versus controls in all 5 cohorts (Table 3). The number of salivary RNA markers showing a significant increase ($p < 0.05$) varied across the cohorts from a minimum of 2 in cohorts 1 to a maximum of 7 in cohort 2.

Cohorts 1 (48 controls/48 OSCC) and 2 (24/24) were validation studies done at UCLA. The technical procedures were similar between the 2 experimental sets with the exception that gene-specific primers were used for the qPCR assays. Cohorts 3 (30/30), 4 (54/36), and 5 (70/31) were assayed at the NCI-EDRN-BRL using samples procured for the independent assay before the independent assay was carried out at EDRN-BRL, a series of assay transfer and standardization steps were taken and installed as standard operating procedure. These included saliva collection, processing, stabilization, and storage (Supplementary Information S1). Using these as standard operating procedures and assaying the same clinical saliva samples at both the UCLA and NCI-EDRN-BRL, we found that the correlation of expression results across the cohorts from a minimum of 2 in cohorts 1 to a maximum of 7 in cohort 2.

Cohorts 3 (30/30), 4 (54/36), and 5 (70/31) were assayed at the NCI-EDRN-BRL using samples procured for the independent assay before the independent assay was carried out at EDRN-BRL, a series of assay transfer and standardization steps were taken and installed as standard operating procedure. These included saliva collection, processing, stabilization, and storage (Supplementary Information S1). Using these as standard operating procedures and assaying the same clinical saliva samples at both the UCLA and NCI-EDRN-BRL, we found that the correlation of expression results across the cohorts from a minimum of 2 in cohorts 1 to a maximum of 7 in cohort 2.

**Table 2.** Demographic information (gender, ethnicity, and age) for individual biomarkers across the 5 independent cohorts

<table>
<thead>
<tr>
<th>Gender</th>
<th>Cohort 1 (N = 48)</th>
<th>Cohort 2 (N = 24)</th>
<th>Cohort 3 (N = 30)</th>
<th>Cohort 4 (N = 54)</th>
<th>Cohort 5 (N = 31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>33</td>
<td>14</td>
<td>20</td>
<td>50</td>
<td>61</td>
</tr>
<tr>
<td>Female</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>29</td>
<td>15</td>
<td>16</td>
<td>42</td>
<td>55</td>
</tr>
<tr>
<td>Asian</td>
<td>12</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>African-American</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Age, y</td>
<td>Mean (SD)</td>
<td>31.4 (12.7)</td>
<td>62.7 (12.1)</td>
<td>41.1 (13.4)</td>
<td>64.9 (15.2)</td>
</tr>
</tbody>
</table>

The performance of the original previously described (7) 4-marker model (IL-8 + IL-1B + SAT + OAZ1) was evaluated across the independent cohorts. The AUCs derived from the fitted probability estimates from the logistic regression model ranged from 0.74 to 0.86 across the cohorts. The level of model performance was comparable with the best panel of markers (using stepwise

**Meta-analysis of performance of the markers across cohorts**

Performance of individual markers was evaluated across the 5 independent cohorts, and individual estimates of sensitivity and specificity were extracted from each study. IL-8, IL-1B, and SAT are the top performers across different cohorts in terms of sensitivity and specificity, from 0.61 to 0.79. Combined estimates were calculated using fixed- and random-effects models. The 2 types of models gave similar results and we report the results of the random-effects model in Table 4. Overall, there was not a significant amount of heterogeneity across the cohorts (Cochran’s Q value was generally smaller than the degrees of freedom and the estimate of t-squared was zero). Combined sensitivities from the random-effects model were 0.68, 0.65, and 0.66 and specificities were 0.64, 0.60, and 0.63 for IL-8, IL-1B and SAT, respectively (Table 4).

In addition, we conducted a meta-analysis to assess an overall significance level of the markers across the cohorts. We found that all 7 RNA markers were significantly differentially expressed between cancer and controls across the cohorts (all $p < 0.02$). In addition, we found that IL-8 and M2BP proteins were also significantly different ($p < 0.02$) and that IL-1B protein was marginally so ($p = 0.053$).

**Combination models for OSCC classification across the cohorts**

The performance of the original previously described (7) 4-marker model (IL-8 + IL-1B + SAT + OAZ1) was evaluated across the independent cohorts. The AUCs derived from the fitted probability estimates from the logistic regression model ranged from 0.74 to 0.86 across the cohorts. The level of model performance was comparable with the best panel of markers (using stepwise
Table 3. Statistics [mean (SD) and AUC] for individual markers across 5 independent cohorts

<table>
<thead>
<tr>
<th>Marker</th>
<th>Control (N = 48)</th>
<th>OSCC (N = 30)</th>
<th>Control (N = 70)</th>
<th>OSCC (N = 36)</th>
<th>Control (N = 24)</th>
<th>OSCC (N = 1-4)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA (qPCR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DUSP1</td>
<td>35 (2.9)</td>
<td>34.3 (3.3)</td>
<td>30.8 (3.3)</td>
<td>30.1 (3.3)</td>
<td>33.1 (2.2)</td>
<td>34.9 (2.0)</td>
<td>0.58</td>
</tr>
<tr>
<td>H3F3A</td>
<td>34 (2.3)</td>
<td>34.4 (3.6)</td>
<td>30.2 (3.3)</td>
<td>30.7 (3.5)</td>
<td>33.8 (2.3)</td>
<td>35.2 (2.0)</td>
<td>0.61</td>
</tr>
<tr>
<td>IL-1B</td>
<td>33 (2.4)</td>
<td>32.8 (4.1)</td>
<td>30.5 (2.5)</td>
<td>30.7 (3.5)</td>
<td>32.5 (2.5)</td>
<td>33.9 (2.0)</td>
<td>0.64</td>
</tr>
<tr>
<td>OAQ1</td>
<td>35 (2.4)</td>
<td>34.7 (3.5)</td>
<td>30.1 (3.0)</td>
<td>30.4 (3.5)</td>
<td>33.4 (2.4)</td>
<td>35.0 (2.0)</td>
<td>0.59</td>
</tr>
<tr>
<td>SAT</td>
<td>32 (2.5)</td>
<td>31.9 (3.8)</td>
<td>30.7 (2.5)</td>
<td>30.4 (3.5)</td>
<td>32.4 (2.5)</td>
<td>33.6 (2.0)</td>
<td>0.66</td>
</tr>
<tr>
<td>Protein (ELISA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL-8</td>
<td>31.7 (2.2)</td>
<td>30.3 (3.6)</td>
<td>29.2 (2.1)</td>
<td>29.6 (3.6)</td>
<td>31.4 (2.5)</td>
<td>33.0 (2.1)</td>
<td>0.60</td>
</tr>
<tr>
<td>M2BP</td>
<td>32 (2.4)</td>
<td>32.7 (3.6)</td>
<td>30.5 (2.5)</td>
<td>30.7 (3.5)</td>
<td>32.2 (2.5)</td>
<td>34.0 (2.0)</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Discussion

To substantiate the development of salivary biomarkers, we assessed the original putative OSCC markers in 395 subjects, in 5 independent validation cohorts. In this article, we describe 2 important validation steps: first, we independently validated the behavior of these biomarker candidates in multiple cohorts and second, we showed the reproducibility and robustness of the assays in an outside reference laboratory.

Patients within each cohort consisted of stage I to IV oral cancers, which represent the entire spectrum of OSCCs (19). The expression levels of these markers were not significantly associated with tumor stage in the cohorts tested (data not shown); therefore, it appears that this panel of 7 OSCC biomarkers may not be indicative of disease progression. This finding might be explained by the hypothesis that the mRNA signatures do not come from a tumor cell origin but rather a tumor-induced response (20). Further studies testing this hypothesis would indeed be intriguing. Nevertheless, these data support the important assertion that these biomarkers are associated with oral cancer. For the validated salivary biomarkers, the shown high sensitivity, specificity, discriminatory value, and low-cost support their potential use for disease screening purposes in at-risk populations.

We conducted multiple validation studies in which slightly modified technological adjustments were made to decrease interoperator and intersample variations. For example, to minimize the sample loading variations between individual PCR setups, we used robot liquid handling, which greatly improved the sample loading accuracy. The operator conducting the experimental procedures, protocol, and even equipment can play an important role in data compatibility. To show the robustness of the assay and minimize the variations brought by these factors, we standardized a number of controllable factors, such as operator, primers, reagents, equipments, and protocols. The technological adjustments made to our protocol were the natural progression from advancing our understanding of handling and evaluating salivary mRNA. Despite some of these differences in techniques between each cohort, we found that the association of these biomarker candidates with OSCC was highly consistent throughout all of the validation studies with IL-8 and SAT increased in patients with OSCC in all the cohort studies regardless of the specific method used.
We have noted an operator training/experience effect in the ability to validate the salivary biomarkers. For cohorts 1 and 2, a new operator (J. Wang) conducted the validation assays. Cohort 1 validated 2 of 7 markers whereas in cohort 2, 7 of 7 markers were validated. Cohorts 3 to 5 were assayed at the NCI-EDRN-BRL, with a new operator (J. Reiss) conducting the assays. Cohort 3 validated 3 of 7 markers; cohort 4 validated 4 of 7 markers and cohort 5 validated 6 of 7 markers. This steady increase in the ability of operators to validate these markers points to the importance of technical training and competency.

Defining abnormal values is a critical step before the clinical implementation of any biomarker set. Our next step in developing these mRNA signatures into clinically useful biomarkers includes standardizing measurements and characterizing the markers’ distribution in the general population with regard to age, gender, smoking history, and ethnicity. The ability to screen individuals at risk for OSCC, using the biomarkers presented here with confirmation of diagnosis through imaging techniques and biopsy, will improve the clinicians’ ability to detect tumors at a stage where therapeutic interventions still have a reasonable curative potential. Large-scale, prospective clinical trials using the PRoBE (21) design framework (using a fixed biomarker algorithm applied to subjects in an appropriate diagnostic clinical scenario) will ultimately determine the ability for these biomarkers to assess disease status in individual patients.

We are cognizant of the limitations of our current study in that these were case–control biomarker development studies. Midstream of our validation studies, in October 2008, the NCI-EDRN published a set of study design guidelines that are critical to be in place to eliminate bias for biomarker research (21). Four essential aspects of the study design that relate to the clinical context, biomarker performance criteria, the biomarker test, and study size were prescribed and must be vigorously followed. We did not incorporate these criteria into the design in the 5 case–control studies in the sense that the sample collection from patients with OSCC was retrospective after diagnosis and controls were healthy people. Collecting salivary specimens before biopsy for patients with OSCC will eliminate potential bias that the case–control study will inherit. Controls in our study

<table>
<thead>
<tr>
<th>Table 4. Meta-analysis of the sensitivity and specificity individual markers across 5 independent cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random-effects model</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>DUSP1</td>
</tr>
<tr>
<td>H3F3A</td>
</tr>
<tr>
<td>IL-1B</td>
</tr>
<tr>
<td>IL-8</td>
</tr>
<tr>
<td>OAZ1</td>
</tr>
<tr>
<td>S100P</td>
</tr>
<tr>
<td>SAT</td>
</tr>
<tr>
<td>IL-8</td>
</tr>
<tr>
<td>M2BP</td>
</tr>
</tbody>
</table>

Figure 1. Correlation of the 7 individual mRNA markers between UCLA laboratory and NCI-EDRN-BRL.
would be appropriate if the clinical context is a general population screening on people who do not have an oral lesion. However, if the clinical context is to interpret oral lesions and identify who needs biopsies, a different type of controls, that is, who had oral lesions but without OSCC would be required. In addition, adding subjects with other types of oral pathology, such as periodontitis, leukoplaikia, or caries, would help establish the specificity of the markers for OSCC rather than oral pathology generally. Therefore, we view our study as a major step in validating that saliva biomarkers provide optimism for the future of saliva diagnostics for oral cancer.

Role of the Sponsor
The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interests were disclosed.

Authors’ Contributions
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): D. Elashoff, H. Zhou, M. Wang, E. Abemayor, D.T.W. Wong.

Grant Support
This study was supported by the NIH (R01DE017170, U01DE016275, and R21CA126733), the Barnes Family Trust Fund, and the Felix and Mildred Yip’s Endowed Professorship.

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.

Received November 21, 2011; revised January 9, 2012; accepted January 23, 2012; published OnlineFirst February 1, 2012.

Table 5. Predictive model comparisons for original 4 markers reported in the work of Li and colleagues (7) and stepwise selection across different cohorts

<table>
<thead>
<tr>
<th>Original 4 markers</th>
<th>Stepwise marker selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Markers</td>
<td>AUC</td>
</tr>
<tr>
<td>Cohort 1 IL-8 + IL-1B + SAT + OAZ1</td>
<td>0.74</td>
</tr>
<tr>
<td>Cohort 2 IL-8 + IL-1B + SAT + OAZ1</td>
<td>0.86</td>
</tr>
<tr>
<td>Cohort 3 IL-8 + IL-1B + SAT + OAZ1</td>
<td>0.82</td>
</tr>
<tr>
<td>Cohort 4 IL-8 + IL-1B + SAT + OAZ1</td>
<td>0.75</td>
</tr>
<tr>
<td>Cohort 5 IL-8 + IL-1B + SAT + OAZ1</td>
<td>0.74</td>
</tr>
</tbody>
</table>

References