

# Noninvasive Evaluation of Oral Lesions Using Depth-sensitive Optical Spectroscopy

Richard A. Schwarz, MS<sup>1</sup>, Wen Gao, PhD<sup>1</sup>, Crystal Redden Weber, BS<sup>2</sup>, Cristina Kurachi, DDS, PhD<sup>3</sup>, J. Jack Lee, PhD<sup>4</sup>, Adel K. El-Naggar, MD, PhD<sup>5</sup>, Rebecca Richards-Kortum, PhD<sup>1</sup>, and Ann M. Gillenwater, MD<sup>6</sup>

**BACKGROUND:** Optical spectroscopy is a noninvasive technique with potential applications for diagnosis of oral dysplasia and early cancer. In this study, we evaluated the diagnostic performance of a depth-sensitive optical spectroscopy (DSOS) system for distinguishing dysplasia and carcinoma from non-neoplastic oral mucosa. **METHODS:** Patients with oral lesions and volunteers without any oral abnormalities were recruited to participate. Autofluorescence and diffuse reflectance spectra of selected oral sites were measured using the DSOS system. A total of 424 oral sites in 124 subjects were measured and analyzed, including 154 sites in 60 patients with oral lesions and 270 sites in 64 normal volunteers. Measured optical spectra were used to develop computer-based algorithms to identify the presence of dysplasia or cancer. Sensitivity and specificity were calculated using a gold standard of histopathology for patient sites and clinical impression for normal volunteer sites. **RESULTS:** Differences in oral spectra were observed in: (1) neoplastic versus nonneoplastic sites, (2) keratinized versus nonkeratinized tissue, and (3) shallow versus deep depths within oral tissue. Algorithms based on spectra from 310 nonkeratinized anatomic sites (buccal, tongue, floor of mouth, and lip) yielded an area under the receiver operating characteristic curve of 0.96 in the training set and 0.93 in the validation set. **CONCLUSIONS:** The ability to selectively target epithelial and shallow stromal depth regions appeared to be diagnostically useful. For nonkeratinized oral sites, the sensitivity and specificity of this objective diagnostic technique were comparable to that of clinical diagnosis by expert observers. Thus, DSOS has potential to augment oral cancer screening efforts in community settings. *Cancer* 2009;115:1669-79. © 2009 American Cancer Society.

**KEY WORDS:** spectroscopy, diagnosis, cancer, oral carcinoma, fluorescence, reflectance.

**Oral** cancer ranks as the 11th most common cancer in the world, with 390,000 new cases estimated to occur annually worldwide.<sup>1</sup> In the US, cancers of the oral cavity and pharynx are predicted to account for over 35,000 new cases and more than 7500 deaths this year.<sup>2</sup> Despite advances in treatment methods, the 5-year survival rate for oral cancer has not increased substantially during the past several decades.<sup>3</sup> Treatment is more effective in patients with early disease; however, most patients present with advanced tumors

**Corresponding author:** Ann M. Gillenwater, MD, Department of Head and Neck Surgery, #441, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030; Fax: (713) 794-4662; agillenw@mdanderson.org

<sup>1</sup>Department of Bioengineering, Rice University, Houston, Texas; <sup>2</sup>Department of Chemistry, Rice University, Houston, Texas; <sup>3</sup>Institute of Physics of São Carlos, University of São Paulo, São Carlos, São Paulo, Brazil; <sup>4</sup>Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas; <sup>5</sup>Department of Pathology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas; <sup>6</sup>Department of Head and Neck Surgery, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

**Received:** July 2, 2008; **Revised:** September 9, 2008; **Accepted:** October 15, 2008

**Published online:** January 23, 2009 © 2009 American Cancer Society

**DOI:** 10.1002/cncr.24177, www.interscience.wiley.com

for which treatment is less successful and may cause severe deficits in speech, swallowing, facial appearance, and quality of life.<sup>4</sup> Detection and diagnosis of early neoplastic changes may be the best way to improve patient outcomes.

During carcinogenesis in the oral cavity, structural and biochemical changes in both the epithelium and stroma alter the optical properties of dysplastic and cancerous tissue. Increased nuclear size and nuclear to cytoplasmic ratio, increased microvascularization, degradation of stromal collagen, and changes in the concentration of mitochondrial fluorophores such as reduced nicotinamide adenine dinucleotide (NADH) and flavin adenine dinucleotide (FAD) lead to changes in optical scattering, absorption, and autofluorescence characteristics within the tissue.<sup>5-8</sup> Several groups have reported that these alterations can be detected using spectroscopic techniques.<sup>9-15</sup> Optical spectroscopy may, therefore, be a useful noninvasive and objective clinical tool to help improve early detection and diagnosis of oral neoplasia.

In current practice, oral premalignant lesions and cancer are diagnosed by visual inspection and palpation, identification of areas that appear clinically abnormal, and invasive biopsy and histologic examination of the removed tissue. However, visual identification of early lesions can be difficult even for experienced clinicians,<sup>16</sup> and many less experienced examiners such as community dentists, primary care physicians, and health care workers believe themselves to be insufficiently trained to perform this important task.<sup>17</sup> To address the challenge of early detection and diagnosis of oral cancer, several alternative diagnostic techniques and visualization aids for examining the oral cavity have recently become commercially available. These innovations include the OralCDx BrushTest (OralCDx Laboratories, Inc., Suffern, NY), an oral brush cytology test; ViziLite Plus (Zila Pharmaceuticals, Inc., Phoenix, Ariz), a direct tissue visualization technique using acetic acid and a blue light source; and the VELscope (LED Dental, Inc., White Rock, BC, Canada), a handheld device for direct visualization of tissue fluorescence.

Clinical studies to evaluate the performance of new diagnostic aids for oral precancer and cancer have been reviewed and critiqued elsewhere.<sup>18,19</sup> Although promising results have been reported for several diagnostic technologies including optical spectroscopy, none has been

definitively proven to improve diagnostic yields over conventional oral examination.<sup>18</sup> Autofluorescence spectroscopy has been reported to be accurate for distinguishing malignant tumors from healthy oral mucosa, but less reliable for distinguishing between benign lesions, such as inflammation, and dysplastic or malignant lesions.<sup>19</sup> Indeed, the presence of inflammation may be a complicating factor in spectroscopic diagnosis of oral lesions; reduced autofluorescence due to inflammation may be difficult to distinguish from reduced autofluorescence due to neoplasia.<sup>20</sup> Because inflammation primarily affects the stroma whereas dysplastic changes occur in the epithelium, depth-sensitive spectral data, particularly that obtained from more superficial layers, may provide more useful information for discriminating benign inflammatory lesions from dysplastic or malignant lesions.

We have previously reported the development of a clinical spectroscopy system with a depth-sensitive, ball lens coupled fiber-optic probe for noninvasive in vivo measurement of oral autofluorescence and diffuse reflectance spectra.<sup>21</sup> Here, we describe results obtained using this depth-sensitive optical spectroscopy (DSOS) system to measure oral sites in 124 subjects. The goal of this study was to investigate 3 questions regarding depth-sensitive optical spectroscopy: 1) whether spectral differences are observed in the signal collected from different depth regions; 2) whether the ability to collect signal from different depth regions enhances diagnostic performance; and 3) how the diagnostic performance of depth-sensitive spectroscopy compares to other diagnostic methods.

## MATERIALS AND METHODS

### *Study Population*

This study, conducted at the University of Texas M.D. Anderson Cancer Center (UTMDACC) and Rice University, was approved by the Institutional Review Boards at both institutions. Patients with lesions of the oral mucosa and normal volunteers aged  $\geq 18$  years were recruited to participate. Persons with previous squamous cell carcinoma of the oral cavity, previous radiation therapy to the head and neck region, chemotherapy within the previous 6 months, use of smokeless tobacco, or current oral cavity lesions were excluded from the normal volunteer



**FIGURE 1.** Spectroscopic probe in contact with a measurement site.

pool. Written informed consent was obtained from all subjects.

### **Protocol**

Spectroscopic measurements of patients were performed at UTMDACC in the operating room immediately before surgery, or in the clinic. Measurements of normal volunteers were performed at UTMDACC and Rice University. No oral rinse or other prior preparation of the oral cavity was required. The oral cavity was inspected by conventional visual examination. Several sites in each subject were selected by the clinician for spectroscopic measurement, including clinically suspicious lesions if present and at least 1 contralateral site with a normal clinical appearance. The clinical appearance of each measured site was categorized by a single expert observer as Normal, Abnormal Low Risk, Abnormal High Risk, or Cancer. Seven expert observers took part in the study, with 2 of the participating experts providing >96% of the clinical evaluations. The probe was placed in gentle contact with the mucosal surface and held in place by the clinician for the duration of the measurement (Fig. 1). Spectroscopic measurements were performed in a darkened room to minimize the effects of ambient light. The measurements were collected over a period of 21 months with measurements of normal volunteers and patients interspersed throughout that time period.

The spectroscopic instrumentation used in this study, including the depth-sensitive fiber-optic probe, methods used for calibration and data processing, and

examples of measured spectra from individual sites have been described previously.<sup>21</sup> Briefly, autofluorescence spectra at 12 excitation wavelengths ranging from 300-470 nanometers (nm) and a diffuse reflectance spectrum under white light illumination were collected through each of 4 probe channels with different depth responses, for a total of 52 spectra collected in each 90-second measurement. The shallow channel has a depth response weighted toward the epithelial tissue layer; the medium channel interrogates both epithelium and shallow stroma; and the 2 deep channels collect signal primarily from the stroma.<sup>21</sup> Wavelength calibration, power calibration, and standards measurements were performed daily before or after patient measurements. The probe was disinfected before and after each patient with Cidex OPA (Advanced Sterilization Products, Johnson & Johnson Gateway, LLC).

Upon completion of the optical measurements, tissue specimens were collected from the measured sites for histopathologic evaluation. Usually, 4-mm punch biopsies were performed immediately after spectroscopy; however, in some cases, measured sites within a region of tissue to be resected were marked and later identified on the resected tissue. Specimens were placed in fixative and analyzed by the study pathologist. The study pathologist's diagnoses were categorized as Normal/Benign, Mild Dysplasia, Moderate to Severe Dysplasia, or Cancer. For histologic diagnosis, Normal/Benign was defined as normal, hyperkeratosis, hyperplasia, and/or inflammation without dysplasia or with only focal mild dysplasia. There were no benign tumors in the study. In normal volunteers biopsies were not performed but the clinical appearance of measured sites was noted.

### **Data Analysis**

A quality control check was performed for each spectroscopic measurement before analyzing the data. Measurements for which 1 or more spectra were missing or otherwise unsatisfactory due to instrument malfunction or flawed measurement conditions (such as excessive ambient light) were excluded from the analysis. Each measurement site was assigned either a "hard" or "soft" gold standard diagnosis following the terminology used by Lingen et al.<sup>18</sup> For patient sites, the diagnosis assigned by the study pathologist was used as the "hard" gold standard.

**Table 1.** Oral Tissue Measurements In Vivo of the Study Populations

|   |                       | Patients | Normal Volunteers | Total |
|---|-----------------------|----------|-------------------|-------|
| Performed   | Subjects              | 73       | 66                | 139   |
|   | Sites                 | 232      | 283               | 515   |
|   | Measurements          | 411      | 284               | 695   |
| Passed quality control check                                      | Subjects              | 73       | 64                | 137   |
|   | Sites                 | 231      | 270               | 501   |
|   | Measurements          | 405      | 271               | 676   |
| Passed quality control check, with valid histopathology diagnosis | Subjects              | 60       | 0*                | 60    |
|   | Sites                 | 154      | 0*                | 154   |
|   | Measurements          | 281      | 0*                | 281   |
| Final data set for analysis                                       | Subjects              | 60       | 64                | 124   |
|   | Sites                 | 154      | 270               | 424   |
|   | Measurements          | 281      | 271               | 552   |
| Subset: Nonkeratinized tissue, training set                       | Subjects represented† | 31       | 38                | 69    |
|   | Sites                 | 70       | 121               | 191   |
|   | Measurements          | 131      | 122               | 253   |
| Subset: Nonkeratinized tissue, validation set                     | Subjects represented† | 21       | 25                | 46    |
|   | Sites                 | 45       | 74                | 119   |
|   | Measurements          | 81       | 74                | 155   |
| Subset: Keratinized tissue  | Subjects represented† | 17       | 47                | 64    |
|   | Sites                 | 39       | 75                | 114   |
|   | Measurements          | 69       | 75                | 144   |

\* No biopsies were taken in normal volunteers.

† Some subjects are represented in both the keratinized subset and a nonkeratinized subset.

For normal volunteer sites, a “soft” gold standard of Normal/Benign was used based on expert clinical impression. All measurements that passed the quality control check and had a corresponding gold standard diagnosis were included in the data set.

Data were separated into training and validation sets, with all spectroscopic measurements from a single patient randomly assigned to either the training set or the validation set. The training set was used to define a set of spectral features; to reduce the data to a diagnostically relevant subset of the spectral features; and to develop a diagnostic classification algorithm to classify tissue sites based on the identified subset of spectral features. The algorithm was developed using linear discriminant analysis with automated forward stepwise feature selection. For all calculations of sensitivity and specificity, a binary diagnosis of Normal to Mild Dysplasia (negative) versus Moderate Dysplasia to Cancer (positive) was used. Equal prior probabilities were assigned for positive and negative diagnosis categories. The algorithm was used to calculate the posterior probability of disease for each measurement. In cases for which a tissue site was measured more than once, the highest posterior probability from the measurements at that site was used. Diagnostic predictions were made as

the threshold was varied from 0 to 1 to generate a receiver operating characteristic (ROC) curve. The classification algorithm was chosen to maximize the area under the ROC curve in the training set. For the final algorithm, an arbitrary limit of 6 spectral features was imposed to reduce the risk of overtraining. This limit was chosen in part because it is unlikely that >6 endogenous biologic fluorophores and chromophores make major contributions to the measured signal.<sup>5-8</sup> An operating point on the ROC curve, corresponding to a specific threshold, was established using the training set data. The algorithm and threshold were then applied to the validation set to evaluate the diagnostic performance of spectroscopy with respect to the gold standard.

## RESULTS

Table 1 summarizes the numbers of subjects, oral sites, and spectroscopic measurements involved in the study. A total of 695 in vivo measurements of oral tissue were collected from 515 sites in 139 subjects. Of the 695 measurements performed, 676 (97%) passed the quality control check. There were 405 measurements of sites in patients that passed the quality control check; of these, 281 (69%)

**Table 2.** Histopathology Versus Clinical Diagnosis of Measured Oral Sites in Patients

| Data Set   | Histopathology Diagnosis     | Expert Clinical Impression |                   |                    |        | Total |
|--|------------------------------|----------------------------|-------------------|--------------------|--------|-------|
|  |                              | Normal                     | Abnormal Low Risk | Abnormal High Risk | Cancer |       |
| All sites in patients                                    | Normal/benign                | 53                         | 4                 | 6                  | 3      | 66    |
|  | Mild dysplasia               | 9                          | 1                 | 8                  | 7      | 25    |
|  | Moderate or severe dysplasia | 4                          | 0                 | 11                 | 4      | 19    |
|  | Cancer                       | 0                          | 0                 | 4                  | 40     | 44    |
|  | Total                        | 66                         | 5                 | 29                 | 54     | 154   |
| Nonkeratinized tissue, training set, sites in patients   | Normal/benign                | 20                         | 1                 | 3                  | 0      | 24    |
|  | Mild dysplasia               | 7                          | 0                 | 3                  | 4      | 14    |
|  | Moderate or severe dysplasia | 1                          | 0                 | 5                  | 4      | 10    |
|  | Cancer                       | 0                          | 0                 | 2                  | 20     | 22    |
|  | Total                        | 28                         | 1                 | 13                 | 28     | 70    |
| Nonkeratinized tissue, validation set, sites in patients | Normal/benign                | 17                         | 2                 | 2                  | 1      | 22    |
|  | Mild dysplasia               | 1                          | 1                 | 2                  | 2      | 6     |
|  | Moderate or severe dysplasia | 1                          | 0                 | 3                  | 0      | 4     |
|  | Cancer                       | 0                          | 0                 | 0                  | 13     | 13    |
|  | Total                        | 19                         | 3                 | 7                  | 16     | 45    |
| Keratinized tissue, sites in patients                    | Normal/benign                | 16                         | 1                 | 1                  | 2      | 20    |
|  | Mild dysplasia               | 1                          | 0                 | 3                  | 1      | 5     |
|  | Moderate or severe dysplasia | 2                          | 0                 | 3                  | 0      | 5     |
|  | Cancer                       | 0                          | 0                 | 2                  | 7      | 9     |
|  | Total                        | 19                         | 1                 | 9                  | 10     | 39    |

had a corresponding biopsy available from the same tissue site that produced a valid histopathology diagnosis (a “hard” gold standard). There were 271 measurements of sites in normal volunteers that passed the quality control check. Of these, all were considered to have a “soft” gold standard of Normal/Benign based on expert clinical impression, including 4 sites at which inflammation was noted. The final data set consisted of 552 measurements from 424 sites in 124 subjects.

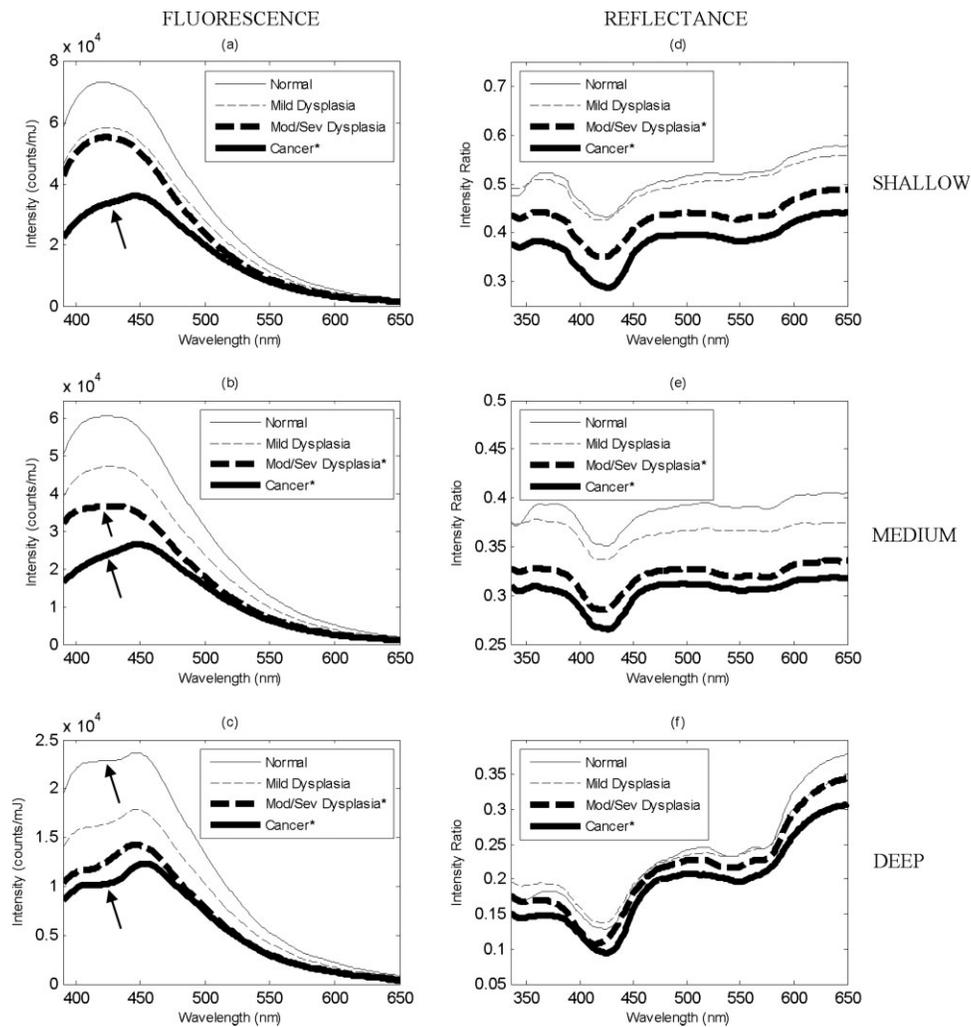
Table 2 shows a comparison between the conventional pathologic and expert clinical diagnosis for sites in patients. The clinical impression category Abnormal Low Risk includes sites described clinically as inflammation, lichen planus, or scar tissue. The clinical impression category Abnormal High Risk includes sites described clinically as leukoplakia, erythroplasia, pre-cancer, dysplasia, or verrucous lesion. In Table 2, with clinical impression categories grouped as Normal to Abnormal Low Risk versus Abnormal High Risk to Cancer, and histopathologic categories grouped as Normal to Mild Dysplasia versus Moderate Dysplasia to Cancer, expert clinical diagnosis of all patient sites correlated with pathologic diagnosis with a sensitivity of 94% and a specificity of 74%.

Anatomic sites measured include buccal mucosa, tongue (predominantly lateral tongue), floor of mouth,

lip, gingiva, and palate (predominantly hard palate). The spectra of normal gingiva and palate sites were observed to differ from the spectra of other normal anatomic sites in both intensity and variability. Based on these observations, and on the results of a separate recent study in which fluorescence microscopy indicated differences in gingiva and palate compared with other oral sites,<sup>20</sup> the 424 sites were divided into 2 groups for analysis: nonkeratinized tissues, including buccal, tongue, floor of mouth, and lip (310 sites); and keratinized tissues, including gingiva and palate (114 sites). The 2 groups were analyzed separately and a different classification algorithm was developed for each group.

Figure 2 shows average spectra collected at 350-nm excitation from nonkeratinized tissues for each diagnostic category. The figure illustrates the progressive reduction in blue-green fluorescence intensity that is observed in dysplastic and cancerous oral tissue compared with normal tissue. This loss of fluorescence has been reported by many investigators and serves as the basis for the operation of the VELscope.<sup>8</sup> The reduced fluorescence associated with neoplasia was observed across a wide range of excitation wavelengths from 330 nm to 470 nm in this study.

As Figure 2 indicates, spectral differences were observed in the signal collected from different depth



**FIGURE 2.** Average spectra of nonkeratinized tissue by diagnosis, illustrating differences in data obtained at different depths. *Left column:* Fluorescence spectra at 350-nanometer (nm) excitation; arrows indicate absorption of fluorescent light by hemoglobin. *Right column:* Reflectance spectra with white light illumination. *Top, middle, and bottom:* Shallow, medium, and deep probe channels, respectively. An asterisk (\*) next to a diagnostic category indicates that differences in the mean intensities of normal tissue and that diagnostic category were statistically significant (2-tailed Student *t* test,  $P < .05/6$ , correcting for 6 comparisons per panel). Peak fluorescence intensity in the 390- to 650-nm region and reflectance intensity ratio at 420 nm were used in the statistical comparisons.

regions through the separate probe channels. The greatest depth-dependent differences occurred in fluorescence measurements at 330- to 350-nm excitation and in diffuse reflectance measurements. At 350-nm excitation at shallow depths, the average fluorescence spectra from normal and dysplastic sites have smooth, rounded peaks, whereas in cancer sites, the peak emission wavelength is shifted to the right (red shifted) and a valley begins to appear in the 420-nm region. At medium depths, the 420-nm valley is evident in the spectra from moderate/severe dysplasia sites

as well. Finally, in the deep channel measurements the 420-nm valley appears prominently in all the spectra regardless of diagnosis.

Depth-dependent differences in spectral shape are also evident in the reflectance data shown in Figure 2. The slope of the reflectance spectrum in the 500- to 650-nm region is relatively flat in the shallow and medium depth data, but increases in measurements from deeper regions. Greater differences in intensity between the diagnostic categories of Normal to Mild Dysplasia and Moderate

Dysplasia to Cancer are observed in shallow and medium depth reflectance measurements than in deep reflectance measurements. Note, however, that the general trend of decreasing fluorescence intensity and reflectance intensity with disease progression is observed in all measured depth regions.

A total of 160 distinct spectral features of interest were defined from the measured spectra, including such quantities as peak emission intensity and peak emission wavelength at each excitation wavelength and depth channel. A subset of diagnostically useful features was identified in 2 steps. In the first step, the 160 features were examined and reduced to an intermediate subset of features. The diagnostic performance of each individual feature was first evaluated independently using data from the training set and features were ranked accordingly. Spectra and classification results associated with high-performing individual features were inspected. Features that were derived from spectra that appeared qualitatively similar and that produced similar classification results (similar sensitivity and specificity, and misclassified mostly the same sites) were considered to be correlated. For correlated features, a representative feature with optimal individual performance was included and the other correlated features were excluded. For nonkeratinized tissues, the data were reduced to an intermediate subset of 16 features associated with fluorescence at 380-nm excitation, fluorescence at 470-nm excitation, and reflectance, all using the shallow and medium channels. For keratinized tissues, the data were reduced to an intermediate subset of 36 features including fluorescence at 300- to 330-nm excitation, fluorescence at 470-nm excitation, and reflectance, using the shallow, medium, and deep channels. In the second step, the intermediate set of features was used as input for algorithm development. An algorithm was developed using linear discriminant analysis with automated forward stepwise feature selection, resulting in a final set of 6 features chosen to maximize the area under the ROC curve.

Table 3 summarizes the set of optimal spectral features selected for diagnostic classification of nonkeratinized tissues. The 380 nm/472 nm fluorescence excitation/emission combination using the medium depth channel proved to be the most diagnostically useful single feature for nonkeratinized tissues. This finding is consistent with results reported by Heintzelman et al.<sup>22</sup> The 6 features in the set selected for nonkeratinized tissues were all associ-

**Table 3.** Feature Set for Diagnostic Classification of Nonkeratinized Sites of the Oral Cavity

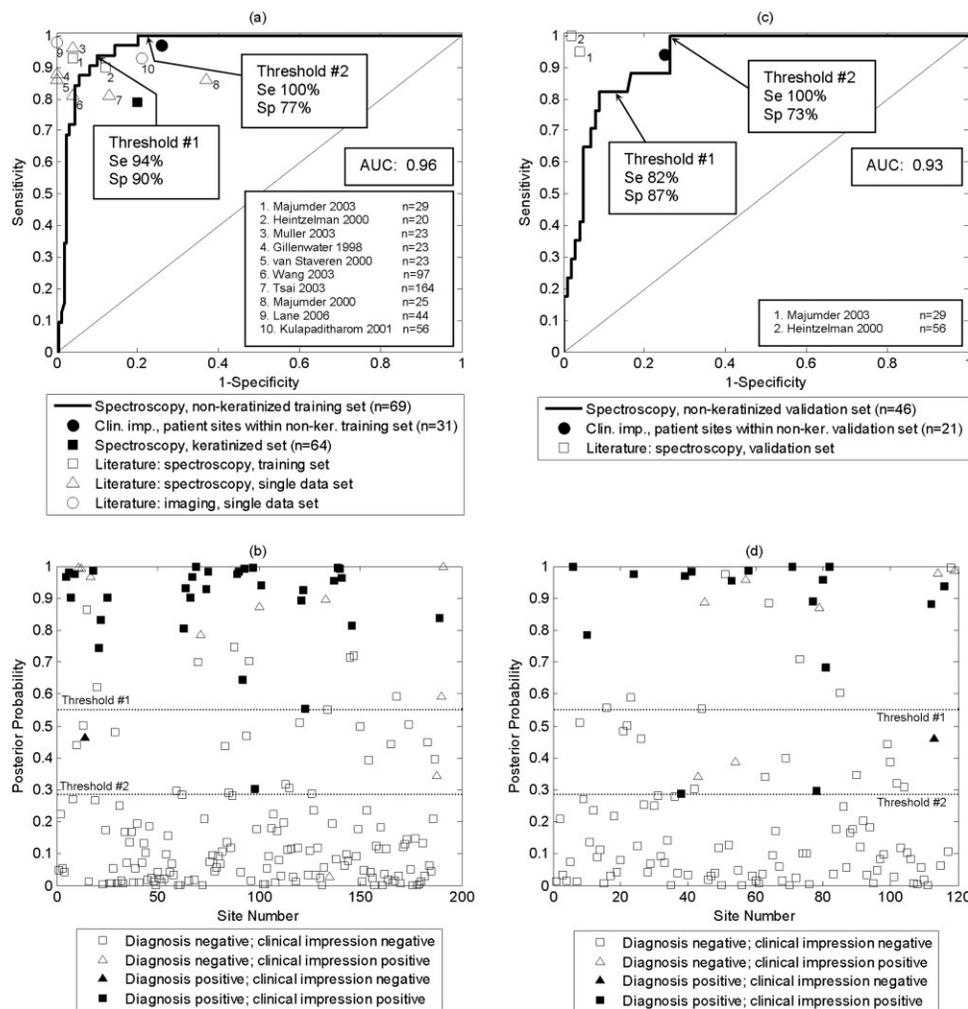
| No. | Spectral Feature  | Depth Channel |
|-----|---|---------------|
| 1   | Fluorescence: 380-nm excitation, 472-nm emission                        | Medium        |
| 2   | Reflectance: 650-nm/500-nm intensity ratio                              | Medium        |
| 3   | Reflectance: 500-nm/420-nm intensity ratio                              | Medium        |
| 4   | Reflectance: 500-nm intensity   | Medium        |
| 5   | Fluorescence: 380-nm excitation, 478-nm/458-nm emission intensity ratio | Medium        |
| 6   | Reflectance: 420-nm intensity   | Shallow       |

nm indicates nanometer.

ated with fluorescence at 380 nm excitation or diffuse reflectance; 5 were obtained using the medium depth channel and 1 was obtained using the shallow channel. For keratinized tissues, diffuse reflectance spectra obtained using the medium and deep channels were optimal for diagnostic classification; fluorescence spectra at 300-330 and 470-nm excitation, obtained using the shallow, medium, and deep channels, were of secondary importance.

Figure 3 shows the diagnostic performance of depth-sensitive spectroscopy for nonkeratinized tissues. In the posterior probability plots, the data value on the vertical axis represents the posterior probability of disease at each site according to spectroscopy, and the horizontal lines indicate thresholds established for a positive result. The area under the ROC curve was 0.96 in the training set and 0.93 in the validation set. Using the training set data, Threshold 1 was selected as the operating point. Using Threshold 1, the sensitivity and specificity, respectively, were 94% and 90% in the training set (191 sites in 69 subjects), and 82% and 87% in the validation set (119 sites in 46 subjects). For comparison, Threshold 2 is also shown as an alternate operating point. Using Threshold 2, the sensitivity and specificity, respectively, were 100% and 77% in the training set, and 100% and 73% in the validation set.

Figure 3a and 3c also show the diagnostic performance of clinical impression as judged by an expert observer. Note that the data set for clinical impression is a smaller subset of the data set for spectroscopy; only sites in patients were included in calculations of sensitivity and specificity of clinical impression, because in normal volunteers the gold standard itself is based on clinical



**FIGURE 3.** Diagnostic classification results. *Left column:* Nonkeratinized training set (191 sites in 69 subjects). *Right column:* Nonkeratinized validation set (119 sites in 46 subjects). *Top row:* receiver operating characteristic (ROC) curves for depth-sensitive spectroscopy; sensitivity and specificity values for clinical impression; and sensitivity and specificity values reported in the literature. *Bottom row:* Posterior probability values corresponding to the plotted ROC curves. Sensitivity and specificity of spectroscopy in the keratinized set (114 sites in 64 subjects) are also shown in (a). n, number of subjects; AUC, area under ROC curve.

impression. For clinical impression, a binary grouping of Normal to Abnormal Low Risk (negative) versus Abnormal High Risk to Cancer (positive) was used. For nonkeratinized sites in patients, expert clinical impression had a sensitivity of 97% and a specificity of 74% within the training set (70 sites in 31 patients), and a sensitivity of 94% and a specificity of 75% within the validation set (45 sites in 21 patients).

Figure 3b and 3d compares the performance of depth-sensitive spectroscopy and expert clinical impression at individual sites. In the training set, spectroscopy (using Threshold 1) misclassified 2 of 32 positive sites and clinical impression misclassified 1 of 32 positive sites,

with 1 site misclassified by both; and spectroscopy misclassified 16 of 159 negative sites and clinical impression misclassified 10 of 38 negative sites, with 8 sites misclassified by both. In the validation set, spectroscopy misclassified 3 of 17 positive sites and clinical impression misclassified 1 of 17 positive sites, with 1 site misclassified by both; and spectroscopy misclassified 13 of 102 negative sites and clinical impression misclassified 7 of 28 negative sites, with 5 sites misclassified by both. Also shown in Figure 3a and 3c are sensitivity and specificity values reported in the literature<sup>6-8,22-28</sup> for spectroscopy and imaging studies based on measurements of blue-green autofluorescence and/or diffuse reflectance. Values reported from a

single data set or from a training set are shown in Figure 3a. Values reported from an independent validation set are shown in Figure 3c.

For keratinized tissue, there were insufficient sites for analysis using separate training and validation sets. Five-fold cross-validation was used instead to evaluate the diagnostic performance of depth-resolved spectroscopy. In the cross-validation set of 114 keratinized sites, the area under the ROC curve was 0.76; at a selected operating point the sensitivity was 79% and the specificity was 80%, as shown in Figure 3a.

## DISCUSSION

Our results demonstrate the diagnostic potential of optical spectroscopy for objectively and noninvasively distinguishing dysplastic and cancerous oral sites from benign lesions and normal mucosa. Furthermore, these findings support the use of a depth-sensitive spectroscopy system to enhance diagnostic performance of optical spectroscopy.

The first question addressed by this study is whether spectral differences are observed in the signal collected from different depths in oral tissue. As shown in Figure 2, spectra collected using the shallow, medium, and deep channels of the depth-sensitive probe have distinctive characteristics. Fluorescence spectra collected from deeper in the tissue display a different spectral shape than spectra collected from superficial tissue, including a more pronounced valley in the 420-nm wavelength region. This depth-dependent variation in fluorescence spectral shape is attributed partly to hemoglobin, which is present within vascular spaces in the stromal layer and absorbs a portion of the emitted fluorescence; this finding is particularly noticeable at 330- to 350-nm excitation, where the peak fluorescence emission lies near the hemoglobin peak absorption wavelength of 420 nm. The depth-dependent distribution of fluorophores in the tissue also plays a role, with epithelial fluorophores such as NADH contributing primarily to the signal measured from shallow depths, and stromal fluorophores such as collagen contributing primarily to the signal measured from deeper regions. Depth-dependent variations are also observed in the reflectance spectra. Whereas the characteristic hemoglobin absorption spectrum is evident in the reflectance spectra from all depth channels, the reflectance spectra from shallow and medium depths show greater intensity differences

among diagnostic categories than those from deeper in the tissue. This finding suggests that the shallow and medium channels may be more sensitive than the deep channels to alterations in the optical scattering properties of the epithelium, where early changes in nuclear size and nuclear to cytoplasmic ratio occur.

The second question is whether signal collection from different depth regions enhances diagnostic performance. In this study, the ability to collect and analyze data from specific depth channels did prove to be diagnostically useful. For nonkeratinized tissue, optimum diagnostic performance was achieved using only spectra from shallow and medium depths. It is interesting that the medium depth channel, which interrogates epithelial and shallow stromal regions, provided the best diagnostic performance of any channel alone. It provided slightly better discrimination between normal and abnormal sites than the shallow channel, which is strongly weighted toward the epithelial layer and minimizes the effects of hemoglobin absorption. The medium channel also outperformed the deep channels, which primarily interrogate stromal regions and are strongly affected by hemoglobin absorption. Thus, it appears that the most diagnostically significant tissue alterations detected by optical spectroscopy occur in the deeper portion of the epithelium and in the shallow region of the stroma. These diagnostically important alterations are likely to include changes in nuclear size and nuclear to cytoplasmic ratio in the epithelium, and increased hemoglobin absorption and breakdown of collagen crosslinks in the shallow stroma.<sup>5-8</sup> This region is also critically important for identification of dysplasia and invasive carcinoma using standard histopathologic evaluation. Indeed, pathologic grading of dysplasia is also imperfect and subject to interobserver variability.<sup>29</sup>

The third question addressed by this study is how the performance of depth-sensitive optical spectroscopy compares to that published for other diagnostic methods for distinguishing neoplastic from non-neoplastic oral mucosa. A wide range of sensitivity and specificity values for various oral cancer detection techniques has been reported in the literature using a variety of study designs and subject populations. The OralCDx test has been reported to perform as well as 100% sensitivity and 93% specificity,<sup>30</sup> and as poorly as 71% sensitivity and 32% specificity.<sup>31</sup> Two studies have reported that the ViziLite system has a high sensitivity (100%) but extremely low

specificity (0%-14%).<sup>32,33</sup> A single pilot study of the VELscope reported a sensitivity of 98% and a specificity of 100%.<sup>8</sup>

In this study of DSOS, we report a sensitivity of 94% and a specificity of 90% in the training set and a sensitivity of 82% and a specificity of 87% in the validation set. As Figure 3a indicates, the results obtained for the training set are comparable to results reported in the literature for other spectroscopy and imaging studies using either a training set or a single data set. Spectroscopy and imaging studies that include an independent validation set are infrequent. Heintzelman reported a sensitivity of 100% and a specificity of 98% in an independent validation set of 281 sites in 56 subjects, but only 4 of the sites were abnormal.<sup>22</sup> Majumder et al. reported a sensitivity of 95% and a specificity of 96% in a validation set derived from measurements of 29 subjects<sup>23</sup>; however, as noted by De Veld et al.,<sup>19</sup> the analysis included a preprocessing step that introduced information regarding lesion type. Here, we report a sensitivity of 82% and a specificity of 87% in a validation set of 119 sites, 17 of which are abnormal, in 46 subjects.

The diagnostic performance of DSOS in nonkeratinized sites approaches that of expert clinical diagnosis (Fig. 3), and the 2 methods tend to err on many of the same sites. This finding suggests that this objective technique may improve the ability of clinicians to diagnose early oral neoplasia, regardless of their experience, especially given that this particular study population was quite a difficult one; it included many patients with extensive field cancerization as well as individuals who had received previous surgery and/or radiation treatment. The diagnostic performance of DSOS in keratinized sites (gingiva and hard palate) is lower than in nonkeratinized sites, and different algorithms appear to be needed to obtain optimal performance in keratinized tissue. Although the majority of cancers in the oral cavity arise in nonkeratinized tissue (>80%),<sup>2</sup> this is an important consideration for future study.

The context of this study should be considered when interpreting its results. A large fraction of the study population consisted of patients with oral cancer or a history of oral cancer. The diagnostic performance of expert clinical impression in this study is likely to have been inflated because participating patients had all been referred to a tertiary care cancer center, many with previous biopsies indicating dysplasia or invasive carcinoma. Further work is needed

to characterize the performance of DSOS in a community setting with a more general population, in which most patients have no disease or inflammatory disease.

Overall, the results of this study indicate that DSOS can be a useful technique for noninvasive evaluation of oral lesions, especially in situations in which the oral screening examination is performed by a community dentist or health care worker rather than an expert clinical observer. In practice, it is anticipated that the depth-sensitive point probe system would be paired with a wide-field imaging device<sup>34</sup>; the imaging device would identify regions of interest, and the depth-sensitive point probe would be used for noninvasive evaluation of those sites. The identification of diagnostic algorithms based on a limited number of excitation wavelengths and depth channels should enable fabrication of simplified DSOS devices that are portable and comparatively inexpensive. This advance should facilitate transition of these noninvasive and objective techniques from the laboratory to community and low-resource settings, for which they are needed most to aid diagnosis of oral dysplasia and cancer.

### Conflict of Interest Disclosures

Supported by National Cancer Institute Grant R01-CA095604.

Dr. Richards-Kortum serves as an unpaid scientific advisor to Remicalm LLC, holds patents related to optical diagnostic technologies that have been licensed to Remicalm LLC, and holds minority ownership in Remicalm LLC.

Dr. Richards-Kortum, Dr. Gillenwater, and Mr. Schwarz hold patents related to optical diagnostics of precancer.

### References

1. Stewart BW, Kleihues P, eds. World Cancer Report. Lyon, France: IARC Press; 2003.
2. Ries LAG, Melbert D, Krapcho M, et al. eds. SEER Cancer Statistics Review, 1975-2005, National Cancer Institute. Available at: [http://seer.cancer.gov/csr/1975\\_2005/](http://seer.cancer.gov/csr/1975_2005/), based on November 2007 SEER data submission, posted to the SEER web site, 2008.
3. Neville BW, Day TA. Oral cancer and precancerous lesions. *CA Cancer J Clin*. 2002;52:195-215.
4. Chen AY, Myers JN. Cancer of the oral cavity. *Dis Mon*. 2001;47:275-361.
5. Fryen A, Glanz H, Lohmann W, Dreyer T, Bohle RM. Significance of autofluorescence for the optical demarcation

- of field cancerisation in the upper aerodigestive tract. *Acta Otolaryngol.* 1997;117:316-319.
6. Gillenwater A, Jacob R, Ganeshappa R, et al. Noninvasive diagnosis of oral neoplasia based on fluorescence spectroscopy and native tissue autofluorescence. *Arch Otolaryngol Head Neck Surg.* 1998;124:1251-1258.
  7. Muller MG, Valdez TA, Georgakoudi I, et al. Spectroscopic detection and evaluation of morphologic and biochemical changes in early human oral carcinoma. *Cancer.* 2003;97:1681-1692.
  8. Lane PM, Gilhuly T, Whitehead P, et al. Simple device for the direct visualization of oral-cavity tissue fluorescence. *J Biomed Opt.* 2006;11:024006.
  9. Kolli VR, Savage HE, Yao TJ, Schantz SP. Native cellular fluorescence of neoplastic upper aerodigestive mucosa. *Arch Otolaryngol Head Neck Surg.* 1995;121:1287-1292.
  10. Dhingra JK, Perrault DF Jr, McMillan K, et al. Early diagnosis of upper aerodigestive tract cancer by autofluorescence. *Arch Otolaryngol Head Neck Surg.* 1996;122:1181-1186.
  11. Betz CS, Mehlmann M, Rick K, et al. Autofluorescence imaging and spectroscopy of normal and malignant mucosa in patients with head and neck cancer. *Lasers Surg Med.* 1999;25:323-334.
  12. Badizadegan K, Backman V, Boone CW, et al. Spectroscopic diagnosis and imaging of invisible pre-cancer. *Faraday Discuss.* 2004;126:265-279.
  13. De Veld DCG, Skurichina M, Witjes MJH, Duin RPW, Sterenberg HJCM, Roodenburg JLN. Clinical study for classification of benign, dysplastic, and malignant oral lesions using autofluorescence spectroscopy. *J Biomed Opt.* 2004;9:940-950.
  14. De Veld DCG, Skurichina M, Witjes MJH, Duin RPW, Sterenberg HJCM, Roodenburg JLN. Autofluorescence and diffuse reflectance spectroscopy for oral oncology. *Lasers Surg Med.* 2005;36:356-364.
  15. Majumder SK, Gupta A, Gupta S, Ghosh N, Gupta PK. Multi-class classification algorithm for optical diagnosis of oral cancer. *J Photochem Photobiol B.* 2006;85:109-117.
  16. Gillenwater A, Papadimitrakopoulou V, Richards-Kortum R. Oral premalignancy: new methods of detection and treatment. *Curr Oncol Rep.* 2006;8:146-154.
  17. Yellowitz JA, Horowitz AM, Drury TF, Goodman HS. Survey of U.S. dentists' knowledge and opinions about oral pharyngeal cancer. *J Am Dent Assoc.* 2000;131:653-661.
  18. Lingen MW, Kalmar JR, Karrison T, Speight PM. Critical evaluation of diagnostic aids for the detection of oral cancer. *Oral Oncol.* 2008;44:10-22.
  19. De Veld DCG, Witjes MJH, Sterenberg HJCM, Roodenburg JLN. The status of in vivo autofluorescence spectroscopy and imaging for oral oncology. *Oral Oncol.* 2005;41:117-131.
  20. Pavlova I, Williams M, El-Naggar A, Richards-Kortum R, Gillenwater A. Understanding the biological basis of autofluorescence imaging for oral cancer detection: high-resolution fluorescence microscopy in viable tissue. *Clin Cancer Res.* 2008;14:2396-2404.
  21. Schwarz RA, Gao W, Daye D, Williams MD, Richards-Kortum R, Gillenwater AM. Autofluorescence and diffuse reflectance spectroscopy of oral epithelial tissue using a depth-sensitive fiber-optic probe. *Appl Opt.* 2008;47:825-834.
  22. Heintzelman DL, Utzinger U, Fuchs H, et al. Optimal excitation wavelengths for in vivo detection of oral neoplasia using fluorescence spectroscopy. *Photochem Photobiol.* 2000;72:103-113.
  23. Majumder SK, Ghosh N, Kataria S, Gupta PK. Nonlinear pattern recognition for laser-induced fluorescence diagnosis of cancer. *Lasers Surg Med.* 2003;33:48-56.
  24. Van Staveren HJ, van Veen RLP, Speelman OC, Witjes MJH, Star WM, Roodenburg JLN. Classification of clinical autofluorescence spectra of oral leukoplakia using an artificial neural network: a pilot study. *Oral Oncol.* 2000;36:286-293.
  25. Majumder SK, Mohanty SK, Ghosh N, Gupta PK, Jain DK, Khan F. A pilot study on the use of autofluorescence spectroscopy for diagnosis of the cancer of human oral cavity. *Curr Sci.* 2000;79:1089-1094.
  26. Kulapaditharom B, Boonkitticharoen V. Performance characteristics of fluorescence endoscope in detection of head and neck cancers. *Ann Otol Rhinol Laryngol.* 2001;110:45-52.
  27. Wang CY, Tsai T, Chen HM, Chen CT, Chiang CP. PLS-ANN based classification model for oral submucous fibrosis and oral carcinogenesis. *Lasers Surg Med.* 2003;32:318-326.
  28. Tsai T, Chen HM, Wang CY, Tsai JC, Chen CT, Chiang CP. In vivo autofluorescence spectroscopy of oral premalignant and malignant lesions: distortion of fluorescence intensity by submucous fibrosis. *Lasers Surg Med.* 2003;33:40-47.
  29. Fischer DJ, Epstein JB, Morton TH, Schwartz SM. Inter-observer reliability in the histopathologic diagnosis of oral pre-malignant and malignant lesions. *J Oral Pathol Med.* 2004;33:65-70.
  30. Sciubba JJ. Improving detection of precancerous and cancerous oral lesions. Computer-assisted analysis of the oral brush biopsy. *J Am Dent Assoc.* 1999;130:1445-1457.
  31. Poate TW, Buchanan JA, Hodgson TA, et al. An audit of the efficacy of the oral brush biopsy technique in a specialist Oral Medicine unit. *Oral Oncol.* 2004;40:829-834.
  32. Ram S, Siar CH. Chemiluminescence as a diagnostic aid in the detection of oral cancer and potentially malignant epithelial lesions. *Int J Oral Maxillofac Surg.* 2005;34:521-527.
  33. Farah CS, McCullough MJ. A pilot case control study on the efficacy of acetic acid wash and chemiluminescent illumination (ViziLite™) in the visualization of oral mucosal white lesions. *Oral Oncol.* 2007;43:820-824.
  34. Roblyer D, Richards-Kortum R, Sokolov K, et al. Multi-spectral optical imaging device for in vivo detection of oral neoplasia. *J Biomed Opt.* 2008;13:024019.