Laboratory Investigation (2010) 90, 589–598

Cytopathology by optical methods: spectral cytopathology of the oral mucosa

Kostas Papamarkakis[1], Benjamin Bird[1], Jennifer M Schubert[1], Miloš Miljković[1], Richard Wein[2], Kristi Bedrossian[3], Nora Laver[3] and Max Diem[1]

Spectral cytopathology (SCP) is a novel approach for diagnostic differentiation of disease in individual exfoliated cells. SCP is carried out by collecting information on each cell's biochemical composition through an infrared micro-spectral measurement, followed by multivariate data analysis. Deviations from a cell's natural composition produce specific spectral patterns that are exclusive to the cause of the deviation or disease. These unique spectral patterns are reproducible and can be identified and used through multivariate statistical methods to detect cells compromised at the molecular level by dysplasia, neoplasia, or viral infection. In this proof of concept study, a benchmark for the sensitivity of SCP is established by classifying healthy oral squamous cells according to their anatomical origin in the oral cavity. Classification is achieved by spectrally detecting cells with unique protein expressions: for example, the squamous cells of the tongue are the only cell type in the oral cavity that have significant amounts of intracytoplasmic keratin, which allows them to be spectrally differentiated from other oral mucosa cells. Furthermore, thousands of cells from a number of clinical specimens were examined, among them were squamous cell carcinoma, malignancy-associated changes including reactive atypia, and infection by the herpes simplex virus. Owing to its sensitivity to molecular changes, SCP often can detect the onset of disease earlier than is currently possible by cytopathology visualization. As SCP is based on automated instrumentation and unsupervised software, it constitutes a diagnostic workup of medical samples devoid of bias and inconsistency. Therefore, SCP shows potential as a complementary tool in medical cytopathology.
Laboratory Investigation (2010) **90,** 589–598; doi:10.1038/labinvest.2010.1; published online 8 February 2010

KEYWORDS: cytopathology; infrared micro-spectroscopy; malignancy-associated changes; multivariate statistics; squamous cell carcinoma; spectral cytopathology

A significant amount of attention and awareness has converged on oral cancer in recent years, a response by the medical community to the fact that it has become the sixth most common cancer worldwide. Of those newly diagnosed with oral cancer, only half will survive within 5 years, a characteristic attributed to oral cancer's high recurrence rates and metastasis relative to most other cancers. In fact, it is well established that for oral cancers, patients who survive their first encounter with the disease have up to a 20 times higher risk of developing a second cancer, a characteristic attributed to 'field cancerization' or malignancy-associated changes (MACs), which represent a biochemical change shown to be diagnostic in this paper. Furthermore, fatalities by oral cancers have not decreased in decades; earlier diagnosis and treatment modalities are needed. The medical community recognizes that without the implementation of new standardized screening procedures, oral cancers are found too late at metastatic stages in which the cancer has progressed to involve lymph nodes of the neck. Prognosis and treatment options at later stages are significantly worse than those for cancers detected early, in which the tumor has not had time to invade deep into local structures.[1]

Emerging new technology in the form of diagnostic vibrational spectroscopy may offer the needed advancements in the fields of microscopy and cytopathology to quickly and accurately screen for oral cancer. Although conventional cytopathology has greatly improved the early detection of disease in individual exfoliated cells, this methodology is inherently limited to subjective interpretations of nuclear integrity, cellular size and structure, and stain uptake. Diagnosis by analysis of morphology is further complicated by low-grade dysplasia or pre-cancerous samples, in which the

[1]Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA, USA; [2]Department of Otolaryngology, Tufts Medical Center, Boston, MA, USA and [3]Department of Pathology, Tufts Medical Center, Boston, MA, USA
Correspondence: K Papamarkakis, Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Avenue, 102 Hurtig Hall, Boston, MA 02115, USA. E-mail: Kpapamar@hotmail.com

Received 17 June 2009; revised 12 October 2009; accepted 12 November 2009

majority of the cells have yet to take on the structural characteristics of disease necessary for rendering a diagnosis. High-grade dysplastic samples are also complicated, as diagnosis is dependent on eyeing the presence of very few morphologically abnormal cells among a large number of cells with relatively normal morphology. Spectral cytopathology (SCP), however, provides a non-destructive photonic approach, which takes a rapid measurement of a sample's biochemistry and identifies variations that occur between healthy and diseased specimens. SCP uses compositional molecular signatures, rather than the distribution of stains and cellular or nuclear morphology for diagnosis. The molecular signatures are captured through a global spectroscopic measurement, and the changes in spectral patterns are analyzed through multivariate statistical methods. The advantage that SCP has over conventional microscopy is the ability to provide an objective and reproducible diagnosis, independent of fatigue, experience, inter-observer variability, and vague morphological changes.

The limitations in the microscopy of stained medical specimens have compromised the accuracy and reproducibility of cytopathological diagnostics, which have been well documented. In a study regarding the accuracy of urinary cytology in daily practice, it was confirmed that the diagnostic sensitivity and specificity toward high-grade transitional cell neoplasms were roughly 79 and 95%, respectively.[2] These operational statistics are considered very high, but are so only in such high-grade cases in which disease has progressed to stages requiring expensive therapies. As for the sensitivity in low-grade transitional cell neoplasms, the sensitivity of daily practice was confirmed at about 26%.[2] A significant decrease in accuracy is common in such low-grade cases, in which the majority of these samples are overwhelmed with disaggregated cells, which usually lack recognizable features of dysplasia or neoplasia. Attempts to classify such cases using conventional cytopathology result in low diagnostic yield. These low yields are not confined to urinary diseases; they span all medical cases and abnormalities. The reproducibility of cervical cytologic and histologic diagnosis is at 46%.[3] The lack of reproducibility of pathology was most evident, again, for less severe cases. The accuracy of clinical diagnosis was also studied in lung diseases. It was concluded that diagnoses of interstitial lung diseases of patients who underwent surgical lung biopsy were 62% accurate.[4] The accuracy rate is a direct result of analysis dependent on morphological features.

This evidence shows that restricting cytopathological analysis to morphological interpretations is not sufficient for disease diagnosis or for preventative diagnostics. Changes in cellular morphology characteristic of progressive disease are simply delayed responses to the preceding compositional disturbance of the disease. There is value, therefore, in examining the biochemical composition of tissues and cells by SCP. New advancements in instrument technology permit the collection of a spectrum from an individual cell, making it possible to study the internal biochemical composition of cells, and also to detect variations in the composition brought about by disease, before such effects are manifested morphologically. Here, we introduce SCP to both monitor the total biochemical composition of individual cells and analyze the spectral information through multivariate methods such as principal component analysis (PCA).[5] This mathematical method provides an unbiased, reproducible approach to aid cytopathologists and to improve patient care.

SCP is carried out on unstained cells, as it uses an intrinsic measurement of the cell's biochemical composition for analysis and uses principles of vibrational spectroscopy for the analysis. As the biochemical components of cells absorb infrared (IR) radiation at frequencies in resonance with the vibrating bonds, IR spectra are produced in which the frequency of vibration can be correlated to particular bond type or characteristic functional groups. In addition, the absorption intensity can be correlated to the abundance of the bond type or functional group. Biochemical compounds (peptides, nucleic acids, phospholipids) have uniquely defined absorption bands within the 1800–900 cm^{-1} range, commonly referred to as their spectral fingerprint.[6]

In the authors' laboratory, SCP has been applied to a number of model studies to establish the sensitivity of this new method, such as in urinary cytology. In this study, two distinct cell types were observed after Pap staining from voided urine: squamous epithelial and transition urothelial cells. SCP showed statistically significant spectral variance between unstained cells of these two cell types, as well as spectral variations in the glycogen content of squamous cells.[7] In another recent study, squamous cells were exfoliated from the cervix of spayed dogs. Statistical analysis spectrally differentiated cervical cells collected from canines that were in various states of estrous. The observed differences were attributed to changes in the nucleus-to-cytoplasm ratios caused by hormones that initiate maturation.[8] Studies have also established spectral patterns contributed by menopausal status and cellular maturity (superficial and intermediate) within pre- and post-menopausal individual exfoliated human cervical cells.[5]

The usage of spectrally differentiating individual cells is vast, and these studies support the use of SCP to spectrally characterize individual cells. This approach, in which a classification is based on physical spectral measurements of a cell's distinct biochemistry, can provide a complimentary method to conventional cytopathology. In this proof of principle study, a benchmark for the sensitivity of SCP in the oral cavity is established for the first time. Tens of thousands of oral cells were interrogated and spectrally classified by anatomical position, drug metabolism, MACs including reactive atypia, dysplasia, squamous cell carcinoma (SCC), and viral infection. Though a small number of samples were included in these initial studies, thousands of cells from each of the samples were scrutinized by SCP (see Table 1). The biochemical signatures of disease found are reproducible

Table 1 Samples studied by spectral cytopathology

Oral cell origin	Pathology	No. of cells studied	Patient donors
Mouth floor	Normal	910	4
Tongue	Normal	1042	4
Cheek, palate, and gums	Normal	966	6
Mouth floor w/ Ibuprofen	Normal	986	4
Palate	MACs; patient history of SCC	7752	2
Palate	SCC	4490	1
Tongue	MACs; patient history of SCC	4360	1
Tongue	SCC	2846	1
Cheek	Herpes viral infection	2644	1

The anatomical origin, the original cytopathological diagnosis, the number of individual cells, and the number of patient donors for each case study are included. Patients' relevant medical histories are also listed. For listings in which there is more than one patient donor, relatively equal numbers of cells were collected from each donor.

throughout the majority of cells from each sample. The purpose of these initial studies is to establish the concept that quantifiable spectral changes can be used to classify cells according to specific disease processes.

MATERIALS AND METHODS

Sample Preparation

All clinical oral samples were obtained from the Cytopathology Division of the Pathology Department at Tufts Medical Center (Boston, MA, USA) after routine testing and follow-up had been performed. Samples on cytological brushes were stored in SurePath® solution (Burlington, NC, USA). Subsequently, cells were vortexed off the brushes, filtered to remove debris, and deposited onto low-e microscope slides (Kevley Technologies, Chesterland, OH, USA) using cytocentrifugation (CytoSpin, Thermo, Waltham, MA, USA).

Normal oral cytology samples were collected from laboratory volunteers at Northeastern University under a local institutional review board (IRB). Cells from the oral cavity were exfoliated from five regions of the mouth to correlate specific spectral changes contributed by anatomical origin: cheeks, tongue, hard palate, gums, and floor of the mouth. Before sampling, the subjects rinsed their mouth with water to rid the cavity of any debris. Subsequently, oral mucosa cells were obtained by a 30 s swabbing of the area of interest using a Fisherbrand sterilized polyester swab. In drug metabolite experiments, oral mucosa cells were collected in a similar manner, 1 h after ingestion of 600 mg of isobutyl propanoic phenolic acid (Ibuprofen). All cells were immediately fixed in SurePath fixative solution and prepared onto low-e microscope slides in a similar manner as described above.

Ethical approval for this study was provided by a local IRB by the National Institute of Health.

Clinical Diagnosis

All clinical samples were diagnosed using standard cytopathology methods and were paired with tissue biopsy sample histopathological diagnosis. After standard cytopathological evaluation and follow-up had been performed, the residual sample was prepared for SCP analysis (unstained, on low-e microscope slides). The Cytopathology Division at Tufts Medical Center provided both the clinical samples for SCP analysis and the original cytopathological diagnoses for comparison purposes. All clinical samples in this study had present and earlier positive histopathological diagnoses.

Data Collection

Spectral data acquisition was carried out by imaging a $4.0\,\text{mm} \times 4.0\,\text{mm}$ spot on a low-e microscope slide containing the deposited cells. This was accomplished by raster, scanning the substrate through the focus ($6.25\,\mu\text{m} \times 6.25\,\mu\text{m}$ in size) of a beam of IR light using one of two PerkinElmer Spotlight 400 FTIR Imaging Systems (Perkin Elmer, Shelton, CT, USA) in the Laboratory for Spectral Diagnosis at Northeastern University. The instrument optical bench, the IR microscope, and an external microscope enclosure box were purged with a continuous stream of dry air (-40°C dew point) to reduce atmospheric water vapor spectral contributions. Data were acquired in reflectance mode using the following parameters: $4\,\text{cm}^{-1}$ spectral resolution, Norton-Beer apodization, and 1 level of zero filling. Two co-added interferograms for each pixel were Fourier transformed to yield spectral vectors, each with a range of 4000–700 cm^{-1} at $2\,\text{cm}^{-1}$ intervals. Background spectra for all 16 detector elements were collected using 128 co-added interferograms. Raw datasets consist of 409 600 spectra and occupy about 2.54 GB each.

Image Processing

Raw data sets from the IR instruments were imported into a program written in-house referred to as PapMap.[9] This program is written in 64-bit MATLAB (The Mathworks, Natick, MA, USA) to accommodate the large data matrices. PapMap reconstructs the spectra of individual cells, collected in mapping mode, between 9 and 100 individual pixel spectra for each cell. It does so by establishing which pixel spectra belong to a given cell in the image map by constructing a binary mask in which contiguous regions belonging to individual cells are identified. This mask is established by defining a threshold for the amide I intensity. For each contiguous area occupied by a cell, the cellular spectrum is calculated, starting from the spectrum with the largest amide I intensity. This spectrum is presumably from the nucleus of the cell, which always exhibits the strongest protein intensity.

All spectra within the areas indicated by the binary mask are subsequently co-added and subject to several constraints to ensure excellent spectral quality. These criteria are imposed to prevent the co-addition of very weak spectra with poor signal-to-noise to contaminate the cell spectrum, and spectra from the edges of a cell that may be contaminated by a scattering artifact.[10,11]

The co-added cellular spectra, as well as the coordinates of each cell, are then exported for further data analysis. After IR data collection, the cells on a slide are manually stained using standard cytological stain combinations first developed by Papanicolaou:[12] Protocol OG6 (Fisher Scientific, Kalamazoo, MI, USA); EA-50 (Surgipath Medical Industries, Richmond, IL, USA); Hemotoxylin 1, Clarifier 1, and Bluing (Richard-Allan Scientific, Kalamazoo, MI, USA). Tap water and solutions of ethanol are used in the washing steps. Finally, to avoid degradation, slides are dipped in xylene and coverslipped for cytological analysis. Next, visual images at $40\times$ magnification of each stained cell are collected at the coordinates indicated by the PapMap algorithm, using an Olympus BX40 microscope fitted with a computer-controlled stage and a QImaging GO3 3 MB digital color camera. The images and cellular spectra are linked and stored in a database for easy identification. The cell images are diagnosed by a cytolopathologist, and the resulting medical diagnosis is correlated to spectral and cytologic data.

Data Analysis

All data analysis was performed using an unsupervised method of multivariate analysis, PCA.[13] This method is used to determine whether or not there are systematic changes in cellular spectra. PCA was carried out using the PLS Toolbox 402 in MATLAB (Eigenvector Research, Wenatchee, WA, USA) on a restricted spectral range (1800–900 cm^{-1}) after second derivative calculation and subsequent vector normalization.

RESULTS

The oral cavity is comprised of several defined regions that include the buccal mucosa (mucous membrane of the cheeks), tongue, gingiva (gums), floor of the mouth, hard and soft palate, lips, and the uvula. Although these structures appear and feel unique, the morphology of their respective cells is very similar. The biochemical composition of normal mucosa cells, however, subtly varies from region to region. For example, the tongue and floor of the mouth, two of the most common sites of oral cancer,[14] have compositional differences, which can only be observed spectrally, as shown in Figure 1. These compositional changes are attributed to the unique expression of esophageal-type keratins and the abundant expression of collagen in tongue and floor of mouth cells, respectively.[14] These intrinsic biochemical differences that are exclusive to these mucosal origins cannot be distinguished morphologically (Figure 1c–e). By means of SCP, however, these cells can be differentiated on a biochemical level.

Figure 1a shows the PCA 'scores plot' of the significant spectral differences among normal oral cells from the five anatomical sites studied in the oral cavity. In such a scores plot, each symbol represents the spectrum of one cell, plotted in a coordinate system that indicates the contribution of 'principal components' (PCs) to a given spectrum. The PCs are 'basis spectra' calculated from the variance within the data set and are obtained by well-established methods of linear algebra.[13]

As expected by the morphology, most normal oral cavity cells are spectrally similar. The blue squares represent approximately 1000 spectra of cells collected from the cheeks, palate, and gums from a total of six subjects. These three origins were combined into one dataset, as their respective cells did not separate in the PC2 *vs* PC3 scores plots because they have very similar biochemical compositions. However, it is observed that the biochemistry of tongue cells (red triangles) and floor of mouth cells (green asterisks) are different from the other cells of the oral cavity. The mean spectral patterns of the three clusters are overlaid in Figure 1b. Each spectral trace shown is the mean of the second derivative, vector normalized spectra in each dataset. They represent 'snapshots' of the unique biochemistry of these oral cell types.

Cells exfoliated from the tongue have a very distinct IR spectrum and display frequency shifts and intensity variations to both the amide I (*ca*. 1610–1690 cm^{-1}) and amide II (1590–1480 cm^{-1}) absorption regions compared with cells exfoliated from the cheeks/palate/gums. The amide I band has a peak maximum at *ca*. 1648 cm^{-1}, the highest of the cell types analyzed, and the most intense low wave number shoulder of the amide II is located at 1514 cm^{-1}. However, the most notable spectral change is the amide I/amide II peak height ratio, which is significantly smaller in the tongue spectra compared with the other cell types. This peak height ratio in turn highlights the more intense bands located at 1450, 1390, and 1230 cm^{-1}, which are characteristic keratin absorption bands as shown in the pure keratin spectrum in Figure 1b. Furthermore, an amide I peak position shift to higher wave numbers is observed. These spectral bands and shifts observed in cells exfoliated from the tongue are contributed by an abundance of keratin. This spectral identity attributed to cells of the tongue is also observed in autofluorecence spectroscopy,[15–19] where it was hypothesized that unique content of keratin in tongue keratinocytes contributes to observed spectral distinction.

Similarly, a change in the amide I to amide II peak height ratio between cheek/palate/gums and mouth floor cells is evident. This has been observed earlier and attributed to the nucleus-to-cytoplasm ratio variations in oral epithelial cells.[20] However, collagen abundance in the connective tissue unique to cells of the mouth floor may also contribute to this unique spectral protein pattern. These intrinsic, normal

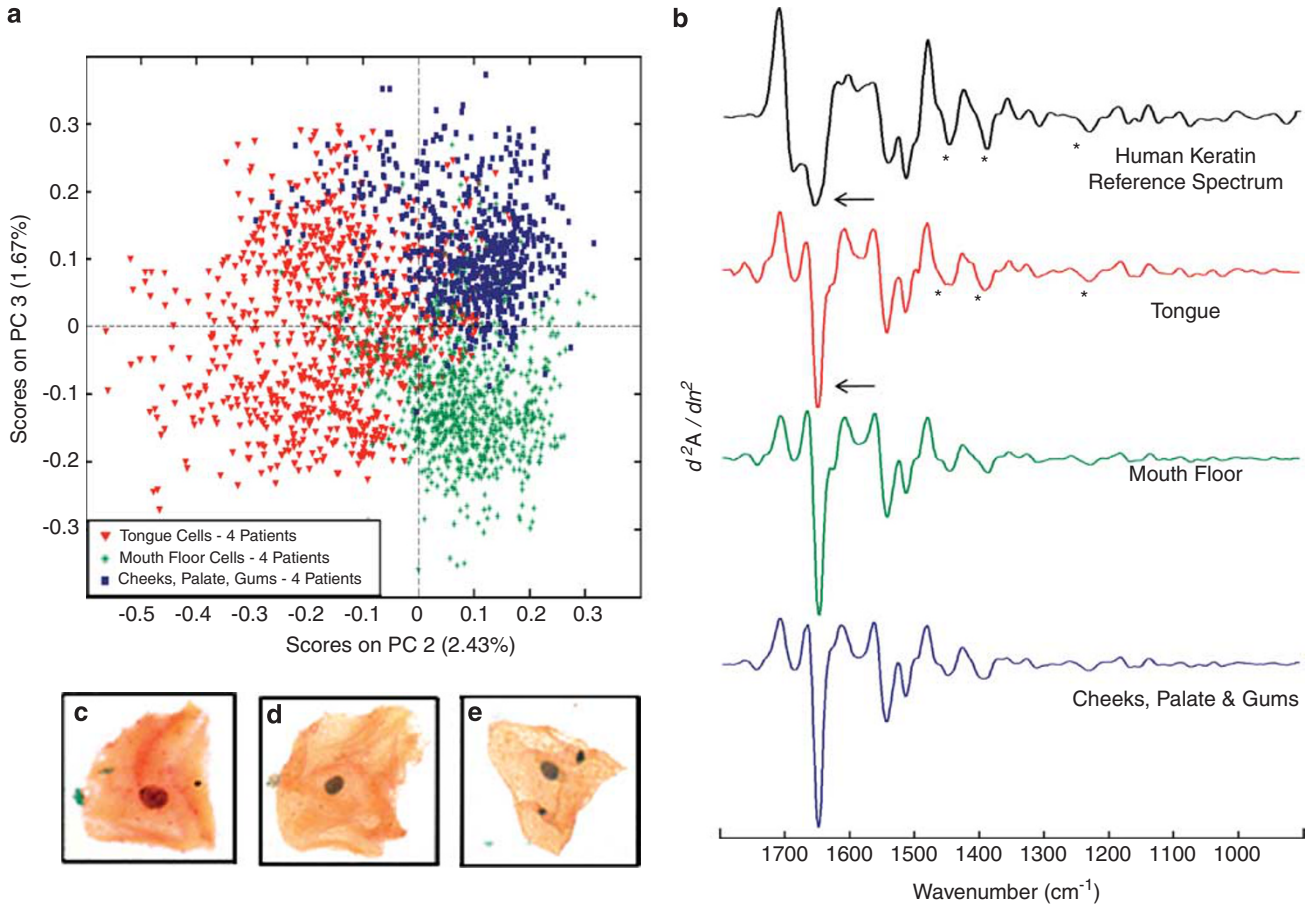

Figure 1 Spectral heterogeneity between anatomical origins in the oral cavity. (**a**) PCA scores plot of the significant spectral differences among normal oral cells from five distinct origins of the oral cavity: cheeks, gums, palate, tongue, and mouth floor. (**b**) Mean second derivative, vector normalized spectra representative of oral cells from the various anatomical origins and a reference spectrum of human keratin. The asterisks mark the spectral patterns contributed by keratin. The arrows signify an amide I shift to higher wave numbers. (**c**–**e**) 40 × visual images of epithelial cells from the cheek, mouth floor, and tongue, respectively.

variations at the biochemical level contribute no discernable information to the cell morphology, as shown in Figure 1c–e. Thus, it seems that SCP is differentiating cells by a few, unique biomolecules.

Further investigations were carried out on the cells of the mouth floor regarding spectral influences of drug metabolites (Figure 2). Mouth floor cells from four normal volunteers were collected pre- and post-ingestion of 600 mg of Ibuprofen (pre- and post-ingestion samples were collected 1 week apart to ensure the same epithelial layer was exfoliated). As shown in Figure 2a, there is a distinct spectral pattern contributed by drug metabolites circulating the blood stream post-ingestion. However, significant spectral changes are not seen in any of the other cells of the oral mucosa (cheek, palate, gums, or tongue). These results are likely observed only in the floor of the mouth, as it is the only structure in the mucosa, which has a thin, perfuse mucous lining loosely bound by connective tissue.[14] The spectral patterns distinguishing floor of mouth cells before and after Ibuprofen ingestion are shown in Figure 2b. A shoulder variation on the low wave number side of the amide I band is observed that is accompanied by subtle variations in the frequency range of 1465–1360 cm^{-1}. We postulate that the variations observed in the frequency range of 1465–1360 cm^{-1} are contributed by the carboxylic acid functional group of isobutyl propanoic phenolic acid. Contributions of drug metabolites, however, are not identifiable morphologically (Figure 2c and d).

The analysis of the biochemical composition of cells is a novel approach for cytopathology and has vast implications. The impact of detecting cellular keratins, for example, the expressions of which are altered on early malignant transformations of human oral cancers, may have widespread diagnostic value.[21] With this in mind, SCP was tested in the authors' laboratory as a method for cancer detection. In fact, human uterine cervical screening SCP results are completely analogous to the oral results reported here, and are based on much larger datasets and more robust sample sizes.[22]

The potential for SCP to accurately and reproducibly detect diseased cells at a variety of progressive states is presented

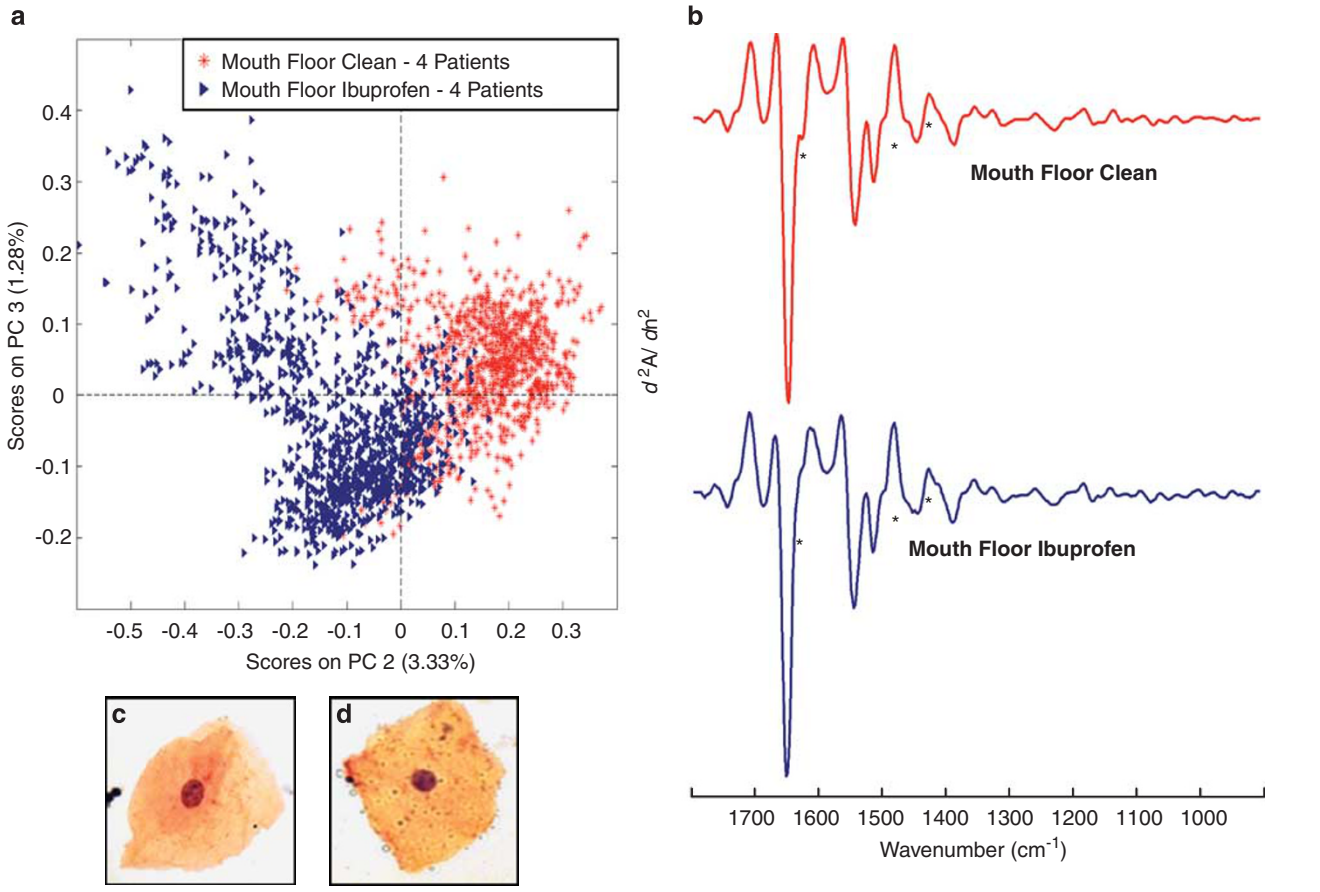

Figure 2 Drug metabolite influence on mouth floor cells. (**a**) PCA scores plot of the significant spectral differences between normal epithelial mouth floor cells before and after Ibuprofen ingestion. (**b**) Mean second derivative, vector normalized spectra representative of mouth floor cells before (red) and after (blue) Ibuprofen influence. (**c**, **d**) 40 × visual images of epithelial mouth floor cells before and after ingestion of Ibuprofen, respectively.

in Figure 3. In this figure, the diseased cells are compared with control groups of the same anatomical origin in the oral cavity to ensure that the spectral changes are contributed solely by disease rather than origin, as was seen in Figure 1. The sensitivity of SCP is most dramatically seen in Figure 3a. Here, spectral data of cells from the palate of two patients diagnosed by cytopathology with reactive atypical changes or MACs and of one patient diagnosed with SCC are compared with spectral data from the palate cells of six normal volunteers. Despite the few number subjects, this part of the study contains a sample size of $>13\,000$ individual cells (Table 1).

The two patients diagnosed with reactive atypia have a biopsy-confirmed medical history of SCC of the oral mucosa. Thus, their present atypical cytological diagnosis (MAC) implies that their cells have some residual nuclear changes associated to an earlier or transitioning malignancy. Although the majority of cells from the three medical samples seems morphologically normal (Figure 3d), they share some compositional variations, which deviate from the biochemistry of normal cells (Figure 3a). In fact, all cells from the two reactive atypical samples have non-cancerous morphology and cluster with the majority of the cells from the sample diagnosed with SCC (Figure 3a).

The fact that reactive atypia cells cluster separately from normal cells, but together with cancer cells, is attributed to the phenomenon of MACs, which broadly can be defined as nuclear differences in cells with otherwise normal morphology. These cells with reactive atypia or MACs are from patients with present or earlier carcinomas at a localized site. As these cells produce spectral patterns similar to those of cancer samples, we may infer some consistent biochemical trends throughout all stages of malignant transformation.[23]

The most significant usage of SCP may be in its sensitivity to detect MACs, which can correlate to compositional states influenced by pre-cancerous differentiation. Interpretations of a cellular sample would no longer depend on the detection of a few high-grade dysplastic cells, which may or may not happen to be collected or prepared on the microscope slide. Instead, the sensitivity of SCP would allow for cytopathological interpretations to be accurately and reproducibly made throughout the entire transformed sample.

In samples diagnosed with SCC, cells with cancerous morphology (Figure 3e and g) often make up the minority of

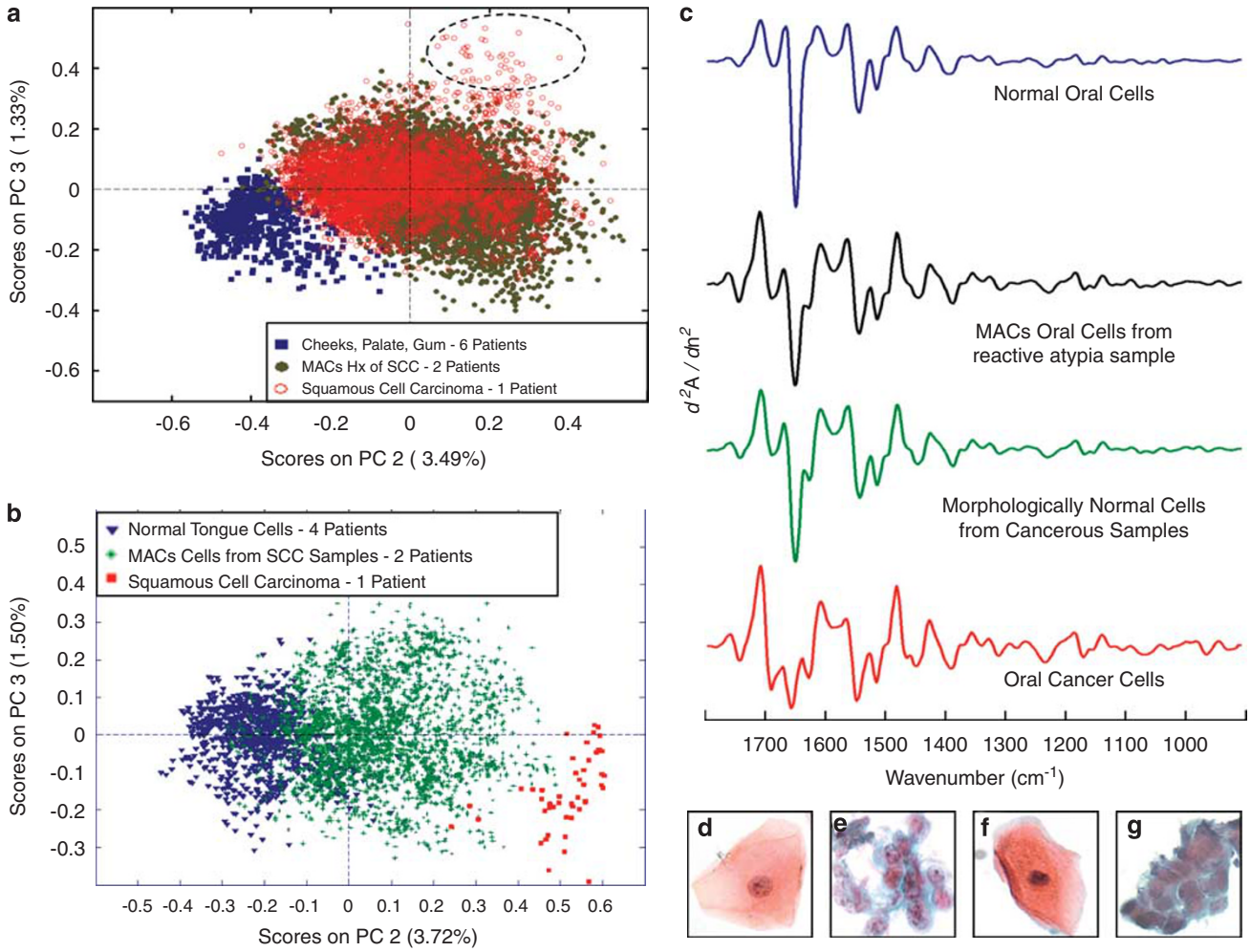

Figure 3 MACs in SC of the tongue and palate. PCA scores plot of the significant spectral differences between normal and diseased oral cells of (**a**) the palate and (**b**) the tongue. (**c**) Second derivative, vector normalized spectra representative of oral cells in successive states of abnormality. ×40 visual images of epithelial palate cells from diagnosed (**d**) reactive and (**e**) carcinoma samples. (**f**, **g**) ×40 visual images of morphologically normal epithelial tongue cells and morphologically cancerous epithelial tongue cancer cells, both from diagnosed cancer biopsies.

cells. As a result, as few as one to three cells with cancerous morphology may be observed in a pre-malignant sample, and they would have to be sufficient for a cytopathological diagnosis. In conventional cytopathology, the detection of few cells in thousands means the difference between an accurate cancer diagnosis and a false negative one. This is the main advantage of SCP over conventional methodology: providing a diagnostic workup, which tracks biomolecular changes and trends preserved throughout dysplasia and into neoplasia, evident well before morphological compromise.

The ellipse drawn in Figure 3a highlights the cells of cancerous morphology that can be diagnosed by cytopathology. These few cells, an example of one shown in Figure 3e, cluster furthest away from the normal cells and indicate the largest compositional variation from normal biochemistry. Our unsupervised implementation of SCP has blindly distinguished cells by their progression states of disease and accurately detected the minority of cells with cancerous morphology. Furthermore, the sensitivity of SCP has provided spectral patterns manifested by cancer, consistent throughout all the cells of the clinical samples regardless of morphology (Figure 3c).

Figure 3b continues with a similar case study for oral cells originating from the tongue. The PCA scores plot presents tongue cancer cells from two patients diagnosed with SCC plotted with normal tongue cells from four healthy volunteers. Although only a small number of cells (red squares) from the entirety of the two cancerous samples (over 7000 cells analyzed in total) were diagnosed with cancerous morphology by conventional cytopathology, SCP identifies intrinsic compositional changes unique to all of the cells from the SCC samples, which vary significantly from the composition of normal tongue cells, Figure 3c. These compositional changes may be a result of MACs, from the presence of present or earlier malignancy. Figure 3b shows the successful detection of cells transitioning from normal

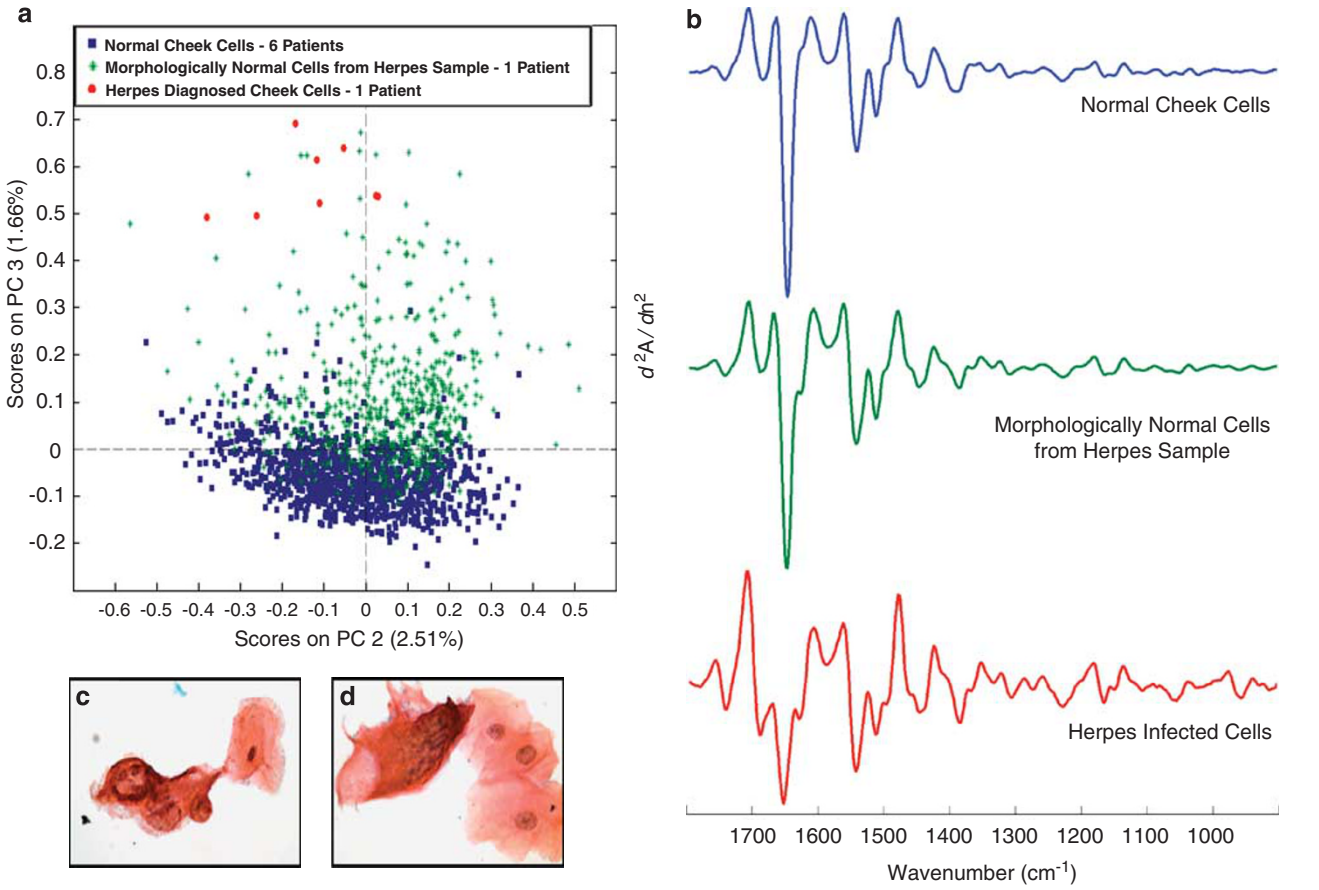

Figure 4 Herpes simplex viral infection. (**a**) PCA scores plot of the significant spectral differences between normal epithelial cheek cells and virally infected epithelial cheek cells. (**b**) Second derivative, vector normalized spectra of normal (blue) and *Herpes* infected (green) cheek cells. (**c**, **d**) 40 × visual images of epithelial cheek cells diagnosed with *Herpes* viral infection.

state to reactive or pre-cancerous state, and continuing to low and high grades of carcinoma of the tongue. These results here correlate very well with those of the palate (Figure 3a).

Figure 3c depicts a stack plot of the averaged second derivate, normalized spectra representative of samples in states ranging from normal to disease. There is a significant trend in intensities of the amide I and phosphate bands (we refer to the vibrations of the phosphodiester group, $-O-(PO_2)^{-}-O-$, at 1080 and 1230 cm^{-1}, as the 'phosphate bands', in accordance with general biochemical nomenclature): as disease progresses, the amide I decreases in intensity as the phosphate bands increase. This amide I trend may be associated with the degradation of common proteins and the expression of different proteins as cells respond to the disease. Phosphate band intensities may increase as cancerous cells have increased rates of replication. In addition, the formation of new proteins as disease progresses is indicated by the appearance of a low wave number shoulder of the amide I band at 1618 cm^{-1}, which is reproducible in all spectra of diseased cells. Furthermore, the mean second derivate, normalized spectra of cells with MACs and that of morphologically looking cells from cancerous samples are virtually identical, supporting SCP's potential for detecting MACs throughout the entire sample.

SCP was also used to investigate spectral changes in oral cells infected with *herpes simplex* virus. *Herpes simplex* virus is responsible for the outbreak of sores in and around the oral cavity.[24] The clinical sample was collected by harvesting the cells directly from the sores on the cheeks. The PCA scores plot in Figure 4a comprises spectra from over 1000 cheek cells from six normal volunteers (blue squares) in addition to spectra from over 2600 cells from one patient with the virus (green asterisks/red circles). The majority of cells exfoliated from the virally infected patient cluster above the normal cells along PC 3, the majority of which had normal looking morphology (green asterisks), Figure 4c. Only a small number of cells within the entire sample could be diagnosed by cytopathology as exhibiting viral infection (red circles), Figure 4d.

Although the majority of the cells in the virally infected sample are morphologically normal, they exhibit biochemical compositional changes, which can be observed in the averaged second derivative spectra, Figure 4b. Again, the observed spectral trends indicate a progression from normal to virally

infected states. There is a shift to higher wave numbers in the amide I and II bands, accompanied by an increase in the intensity of the phosphate bands in the virally infected cells. This trend is likely associated with the degradation of host proteins and production of viral proteins coupled with the integration of viral genome with that of the host's. This presents strong evidence that SCP can distinguish virally infected cells in early and terminal stages.

The spectral changes observed between cancerous progression and viral infection bear some similarities that may suggest that the primary spectral differentiation in both cases may be due to an underlying viral infection. This is particularly significant as viral infections (by the Epstein–Barr and the HPV virus) have been implemented in oral cancers.[24]

DISCUSSION

In this investigation, we detail new methods of automatic cytopathology that use intrinsic molecular signatures, rather than the distribution of stains and cellular or nuclear morphology for diagnosis. SCP provides a rapid measurement of cellular biochemistry and identifies reproducible variations that distinguish healthy and diseased specimens by analyzing thousands of cells. The most notable advantage SCP offers to conventional cytopathology is an objective and reproducible differentiation, independent of fatigue, experience, and inter-observer variability.

This paper presents, for the first time, findings, which indicate that SCP detects disease in cells from the oral cavity that do not yet show morphological abnormalities. Considerable attention has been placed on abnormal clinical samples that cannot be well diagnosed by conventional cytopathology, such as samples associated with ill-defined terms such as 'atypical squamous cells'. Despite having normal morphology, cells with MACs and squamous atypia, and even normal appearing cells from the majority of cancerous samples, still share biochemical hallmarks with their morphologically cancerous counterparts. SCP provides a compositional analysis, which permits the rapid and unbiased analysis of a cytological sample. These results indicate that SCP is a method of superior sensitivity for detecting disease.

Detection of diseased cells by SCP reported here was made blindly: cells, which clustered away from normal cells in the PCA scores plots were found to have spectral signatures of disease. These cells were then analyzed by conventional cytopathology to confer their diagnosis. In each of the disease case studies presented, the few morphologically abnormal cells in each sample clustered furthest from the normal cells in the PCA scores plot. The spectra from the morphologically abnormal cells presented unique and reproducible spectral patterns contributed by their disease states. Spectral trends and PCA distribution patterns were observed with the comparison of reactive cells to both normal and cancer cells. It was observed throughout a number of clinical case studies that a subtle spectral variation is present in the majority of cells from a diseased sample. These variations are exaggerated in cells in which disease has progressed to the stage of morphological compromise. These observations of spectral trends allow SCP to not only track MACs, but also to detect them at earlier stages, which have not compromised cellular morphology.

For the each of the cytopathological cases presented in this study, SCP has produced spectral patterns and contributions unique to each disease and their progressive states. These spectral contributions can be directly related to the variations in chemical composition of their representative cells, which may be due to anatomical origin (such as the abundance of cytokeratin in tongue cells), drug metabolites, or viral transfection. In general, however, these variations are products of many complex spectral patterns, because of the fact that the spectra sample is a superposition of all compositional changes. In addition, spectra do not just indicate abundance, but detect changes in spectra, which are due to the interactions of the biological components.[23]

Conclusions

SCP is a sensitive technique that can detect discrete biochemical changes present in individual cells that are not reflected in their morphology. Blind, unsupervised detection of dysplasia, viral infection, drug metabolism, and anatomical origin was carried out successfully by SCP. It is concluded that SCP will provide an unsupervised method of differentiated exfoliated cells and offer needed advancements in preventive medical cytopathology to improve patient care.

ACKNOWLEDGEMENT

Support of this research from grant CA 090346 from the National Cancer Institute of the NIH is gratefully acknowledged.

DISCLOSURE/CONFLICT OF INTEREST

The authors declare no conflict of interest.

1. Davies L, Welch HG. Epidemiology of head and neck cancer in the United States. Otolaryngol Head Neck Surg 2006;135:451–457.
2. Bastacky S, Ibrahim S, Wilczynski SP, *et al.* The accuracy of urinary cytology in daily practice. Cancer Cytopathol 1999;87:118–128.
3. Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations. JAMA 2001;285: 1500–1505.
4. Raghu G, Mageto YN, Lockhart D, *et al.* The accuracy of the clinical diagnosis of new-onset idiopathic pulmonary fibrosis and other interstitial lunch disease: a prospective study. Chest 1999;116: 1168–1174.
5. Diem M, Matthaus C, Chernenko T, *et al.* Infrared and Ramam spectroscopy and spectral imaging of individual cells. In: Salzer R, Siesler HW (eds). Infrared and Raman Spectroscopic Imaging. Wiley-VCH: Weinheim, 2009.
6. Diem M. Introduction to Modern Vibrational Spectroscopy. John Wiley & Sons: New York, 1993.
7. Bird B, Romeo MJ, Diem M, *et al.* Cytology by infrared microspectroscopy: automatic distinction of cell types in urinary cytology. Vib Spectrosc 2008;48:101–106.
8. Romeo MJ, Mohlenhoff B, Jennings M, *et al.* Infrared and microspectroscopic studies of epithelial cells. Biochimica et Biophysica Acta 2006;1458:915–922.

9. Diem M, Miljkovic M, Romeo MJ, *et al.* Method of reconstituting cellular spectra from spectral mapping data. USA Patent. 2009.
10. Mohlenhoff B, Romeo M, Diem M, *et al.* Mie-type scattering and non-beer-lambert absorption behavior of human cells in infrared microspectroscopy. Biophys J 2005;88:3635–3640.
11. Bassan P, Byrne HJ, Bonnier F, *et al.* Resonant Mie scattering in infrared spectroscopy of biological materials—understanding the 'dispersion artefact'. Analyst 2009;134:1586–1593.
12. Papanicolaou GN. The cell smear method of diagnosing cancer. Am J Public Health 1948;38:202–205.
13. Adams MJ. Chemometrics in Analytical Spectroscopy, 2nd edn. Royal Society of Chemistry: Cambridge, 2004.
14. Fisher SE, Harris AT, Chalmers JM, *et al.* Head and neck cancer: a clinical overview, and observations from synchrotron-source mid-infrared spectroscopy investigation. In: Diem M, Griffiths PR, Chalmers JM (eds). Vibrational Spectroscopy for Medical Diagnosis. John Wiley & Sons Ltd.: Chichester, 2008.
15. deVeld DCG, Skurichina M, Witjes MJH, *et al.* Autofluorescence and diffuse reflectance spectroscopy for oral oncology. Lasers Surg Med 2005;36:356–364.
16. Schwartz RA, Gao W, Weber CR, *et al.* Noninvasive evaluation of oral lesions using depth-sensitive optical spectroscopy. Cancer 2009;115:1669–1679.
17. Maitland KC, Gillenwater AM, Williams MD, *et al. In vivo* imaging of oral neoplasia using a miniaturized fiber optic confocal reflectance microscope. Oral Oncol 2008;44:1059–1066.
18. Schwartz RA, Gao W, Daye D, *et al.* Autofluorescence and diffuse reflectance spectroscopy of oral epithelial tissue using a depth-sensitive fiber-optic probe. Appl Opt 2008;47:825–834.
19. Rahman M, Chaturvedi P, Gillenwater AM, *et al.* Low-cost, multimodal, portable screening system for early detection of oral cancer. J Biomed Opt 2008;13:1–3.
20. Romeo MJ, Mohlenhoff B, Diem M. Infrared micro-spectroscopy of human cells: causes for the spectral variance of oral mucosa (buccal) cells. Vib Spectrosc 2006;42:9–14.
21. Kim KH, Schwartz F, Fuchs E. Differences in keratin synthesis between normal epithelial cells and squamous cell carcinomas are mediated by vitamin A. Proc Natl Acad Sci 1984;81:4280–4284.
22. Schubert JM, Bird B, Papamarkakis K, *et al.* Spectral cytopathology of cervical samples: detecting cellular abnormalities in cytologically normal cells. Lab Invest (submitted).
23. Diem M, Papamarkakis K, Schubert J, *et al.* The infrared spectral signatures of disease: extracting the distinguishing spectral features between normal and diseased states. Appl Spectrosc 2009;63:307A–318A.
24. Newkirk KA, Holsinger FC. Cancers of the head and neck. In: Feig BW, Berger DH, Fuhrman GM (eds). The M.D. Anderson Surgical Oncology Handbook. Williams and Williams Publishers: Lippincott, 2006.