Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

# *In vivo* Raman spectroscopic identification of premalignant lesions in oral buccal mucosa

S. P. Singh
Atul Deshmukh
Pankaj Chaturvedi
C. Murali Krishna

SPIE

# *In vivo* Raman spectroscopic identification of premalignant lesions in oral buccal mucosa

**S. P. Singh,[a] Atul Deshmukh,[a] Pankaj Chaturvedi,[b] and C. Murali Krishna[a]**
[a]Advanced Center for Treatment Research and Education in Cancer, Chilakapati Laboratory, Kharghar, Navi-Mumbai, 410210, India
[b]Tata Memorial Hospital, Department of Surgical Oncology, Mumbai, 400012, India

**Abstract.** Cancers of oral cavities are one of the most common malignancies in India and other south-Asian countries. Tobacco habits are the main etiological factors for oral cancer. Identification of premalignant lesions is required for improving survival rates related to oral cancer. Optical spectroscopy methods are projected as alternative/adjunct for cancer diagnosis. Earlier studies have demonstrated the feasibility of classifying normal, premalignant, and malignant oral *ex-vivo* tissues. We intend to evaluate potentials of Raman spectroscopy in detecting premalignant conditions. Spectra were recorded from premalignant patches, contralateral normal (opposite to tumor site), and cancerous sites of subjects with oral cancers and also from age-matched healthy subjects with and without tobacco habits. A total of 861 spectra from 104 subjects were recorded using a fiber-optic probe-coupled HE-785 Raman spectrometer. Spectral differences in the 1200- to 1800-$cm^{-1}$ region were subjected to unsupervised principal component analysis and supervised linear discriminant analysis followed by validation with leave-one-out and an independent test data set. Results suggest that premalignant conditions can be objectively discriminated with both normal and cancerous sites as well as from healthy controls with and without tobacco habits. Findings of the study further support efficacy of Raman spectroscopic approaches in oral-cancer applications.  [DOI: 10.1117/1.JBO.17.10.105002]

Keywords: oral cancer; premalignant lesions; optical spectroscopy; *in vivo* Raman spectroscopy; LDA; leave-one-out.

Paper 11685SS received Nov. 24, 2011; revised manuscript received Mar. 3, 2012; accepted for publication Sep. 4, 2012; published online Oct. 1, 2012.

## 1 Introduction

Oral cancer constitutes about 10% of all cancer cases in India and is most common cancer among males. Despite significant advances in treatment modalities, the five-year disease-free survival rate is around 50 percent, which is often attributed to the late detection of disease.[1] Squamous cell carcinoma is the most common representing 90% to 95% of all oral malignancies. Oral squamous cell carcinomas are preceded by clinically visible changes in oral mucosa, which are termed as premalignant lesions and conditions. Common precancerous conditions associated with oral cavity are leukoplakia, erythroplakia, oral submucous fibrosis, erosive lichen planus, and changes associated with sideropenic dysphagia. Surveillance and biopsy of precancers are mammoth tasks, especially in populous countries like India. For example, incidence of leukoplakia itself is up to 1% of general population. Clinical examination followed by microscopic examination of biopsies, studying morphological alterations in the architecture, and arrangement of epithelial strata is the current practice of diagnosis of premalignant pathologies. The premalignant pathosis are reported as varying grades of dysplasia, which ranges from mild to severe depending upon histopathological parameters suggested for denoting the grades of dysplasia. But it lacks accuracy, reproducibility, and requires significant experience on part of the clinician. Marked inter-observer subjectivity among the pathologists while reporting the grades of dysplasia is the greatest obstacle in accurately diagnosing and predicting the treatment outcome of premalignant pathologies. The biopsy procedures are traumatic, painful, and time consuming. It has also been shown that early diagnosis leads to better prognosis, i.e., increased five-year survival rates up to 90%. Hence it would be ideal to have a noninvasive tool for risk characterization, which can be used even in those areas where expert clinicians or pathologist may not be available.

Optical-based diagnostic aids are promising new technologies for improving screening and detection of epithelial malignancies in several organ sites. During oral carcinogenesis structural and biochemical changes occur in both the epithelium and stroma altering the optical and biological properties of dysplastic and cancerous tissue. These changes act as molecular signatures and optical methods like fluorescence, Raman and Fourier-transform infrared spectroscopy can exploit the same for classification of cancerous and normal conditions.[2–7] Raman effect is based on inelastic scattering of photons where energy is exchanged between incident light and the molecule leading to change in wavelength. Only a tiny fraction of photons show this behavior resulting in weak Raman signals. High fluorescence background is an additional problem for biological specimens. With the invention of lasers and sensitive detectors such as charged-coupled-devices (CCDs), Raman spectroscopy of weakly scattering substances such as tissues is now possible. Further, use of excitation photons in near infrared region reduces fluorescence interference from biological tissues. Use of optical fibers for guiding laser light to the desired site and to collect Raman photons facilitates *in vivo* measurements.[8–10] Raman spectroscopic differentiation of normal and cancerous conditions of prostate, esophagus, skin, cervix,

Address all correspondence to: C. Murali Krishna, Chilakapati Laboratory, Advanced Center for Treatment Research and Education in Cancer (ACTREC), Tata Memorial Center, Kharghar, Sector 22, Navi Mumbai, 410210, India. Tel: +91-22-2740 5039; E-mail: mchilakapati@actrec.gov.in

0091-3286/2012/$25.00 © 2012 SPIE

and other forms of cancer have already been reported in the literature.[10–15] As far as oral cavity is concerned, very few studies including ours have shown potential in recording good quality *in vivo* Raman spectra under clinically implementable time.[16,17] In a recent study we have demonstrated that contralateral normal and cancerous sites in oral cancer subjects can be discriminated by *in vivo* Raman spectroscopy.[18] In the present study we have aimed to evaluate potentials of Raman spectroscopy in discriminating premalignant conditions. We have recorded *in vivo* spectra from premalignant patches, contralateral normal, and cancerous site of malignant subjects as well as of healthy controls with and without tobacco habits. Spectra were subjected to multivariate tools unsupervised principal component analysis (PCA) and supervised linear discriminant analysis (LDA) followed by evaluation with leave-one-out cross-validation (LOOCV) and independent test data. Findings of the study have been discussed in the paper.

## 2 Materials and Methods

### 2.1 *Sample Details*

A total of 104 subjects under four categories, subjects with oral cancer, subjects with oral cancer and premalignant patches, age-matched healthy controls without any history of tobacco use, and healthy habitual tobacco users were recruited in the study. Spectra were recorded on tumor and opposite normal site of subjects with buccal mucosa cancerous lesion, which are referred to as "tumor" and "contralateral normal," respectively. We have also recorded spectra from healthy subjects with and without tobacco habits referred to as "habitual tobacco users" and "healthy controls," respectively. Informed written consent of each participant was taken. History of all subjects was documented to ensure cause of cancer as well as type and duration of tobacco habits. Subjects only with tobacco-related cancers were recruited in the study. Median age of subjects with oral cancer, cancer subjects with premalignant patches, habitual tobacco users, and healthy control were 46, 51, 41, and 50 years of age. To ensure uniformity in spectral acquisition, spectra were recorded from opposing buccal surfaces of canine, first premolar, second premolar, first molar, and second molar on both right and left sides. On an average eight spectra (four from contralateral mucosa and four from tumor) from 50 subjects with oral cancer corresponding to 215 spectra from contralateral normal side and 225 from tumor side were recorded. For 30 healthy controls (15 with and 15 without tobacco habits) on an average 10 spectra from right and left buccal mucosa corresponding to 150 each from habitual tobacco users and healthy controls were recorded. Tobacco habits of these subjects were recorded and subjects with smoking habits ≥10 years were recruited. Average time of tobacco habit was ∼14.3 years. A total of 121 spectra from premalignant patches on contralateral side of 24 subjects with oral cancer were also recorded. To avoid any differences because of the mouth environment, subjects were allowed to wash their mouths with water before spectral acquisition. These details are also summarized in Table 1.

The lesions were diagnosed clinically and verified histopathologically by incisional biopsy from the tumor site. Normal mucosa, i.e., healthy controls and habitual tobacco users, and premalignant patches were verified by clinical assessment by a trained senior oral pathologist. No biopsy was taken from normal mucosa owing to ethical limitations.

In view of interlesional and intralesional variability in a tissue, each of the 861 spectra was treated as individual sample. This is due to the fact that clinically homogenous appearing pathologic patches can be heterogeneous at microscopic view. Also often an epithelial tissue specimen can have adjacent normal and anaplastic regions, therefore a pathologist has to examine several sites to decide a sample as normal or abnormal.[19,20] Moreover, for prospective RS applications like surgical demarcation, which requires knowledge on exact boundaries of malignancy, such an approach would be more practical.

### 2.2 *Raman Spectroscopy*

*In vivo* spectra were recorded with a fiber-optic probe coupled HE-785 commercial Raman spectrometer (Jobin-yvon-Horiba, France). Briefly, this system consists of a diode laser (Process Instruments) of 785-nm wavelength as excitation source, a high-efficiency (HE) spectrograph with fixed 950 gr/mm grating coupled with CCD (Synapse). The instrument has no movable parts and spectral resolution as per manufacturer's specification was ∼4 $cm^{-1}$. Commercially available InPhotonics (Inc, Downy St. USA) probe consisting of 105-$\mu$m excitation fiber and 200-$\mu$m collection fiber (NA-0.40) was used to couple excitation source and detection system. As per specifications of manufacturer of the Inphotonics probe, theoretical spot size and depth of field are 105 $\mu$m and 1 mm, respectively. Equal distance during spectral acquisition for all measurements was maintained by adding a spacer at the tip of the probe. To avoid contamination among subjects prior to record spectra from any individual probe was disinfected with CIDEX (Johnson and Johnson, Mumbai, India) solution. Spectral acquisition parameters were $\lambda$ex-785 nm, laser power-80 mW, spectra were integrated

**Table 1** Summary of total number of cases analyzed in the study.

| Sr. no. | Category | Site | Median age | Total no. of spectra/cases | Spectra/cases used for standard model | Spectra/cases used for test prediction |
|---|---|---|---|---|---|---|
| 1. | Contralateral normal site of oral cancer subjects | Buccal mucosa | 46 yrs | 215/50 | 125/27 | 90/23 |
| 2. | Tumor site of oral cancer subjects | Buccal mucosa | 46 yrs | 225/50 | 139/27 | 86/23 |
| 3. | Premalignant patches on oral cancer subjects | Buccal mucosa | 51 yrs | 121/24 | 85/18 | 36/6 |
| 4. | Healthy controls (no tobacco habits) | Buccal mucosa | 50 yrs | 150/15 | 100/10 | 50/5 |
| 5. | Habitual tobacco users | Buccal mucosa | 41 yrs | 150/15 | 100/10 | 50/5 |

**Fig. 1** Photographic representation of instrument set up used in the study.

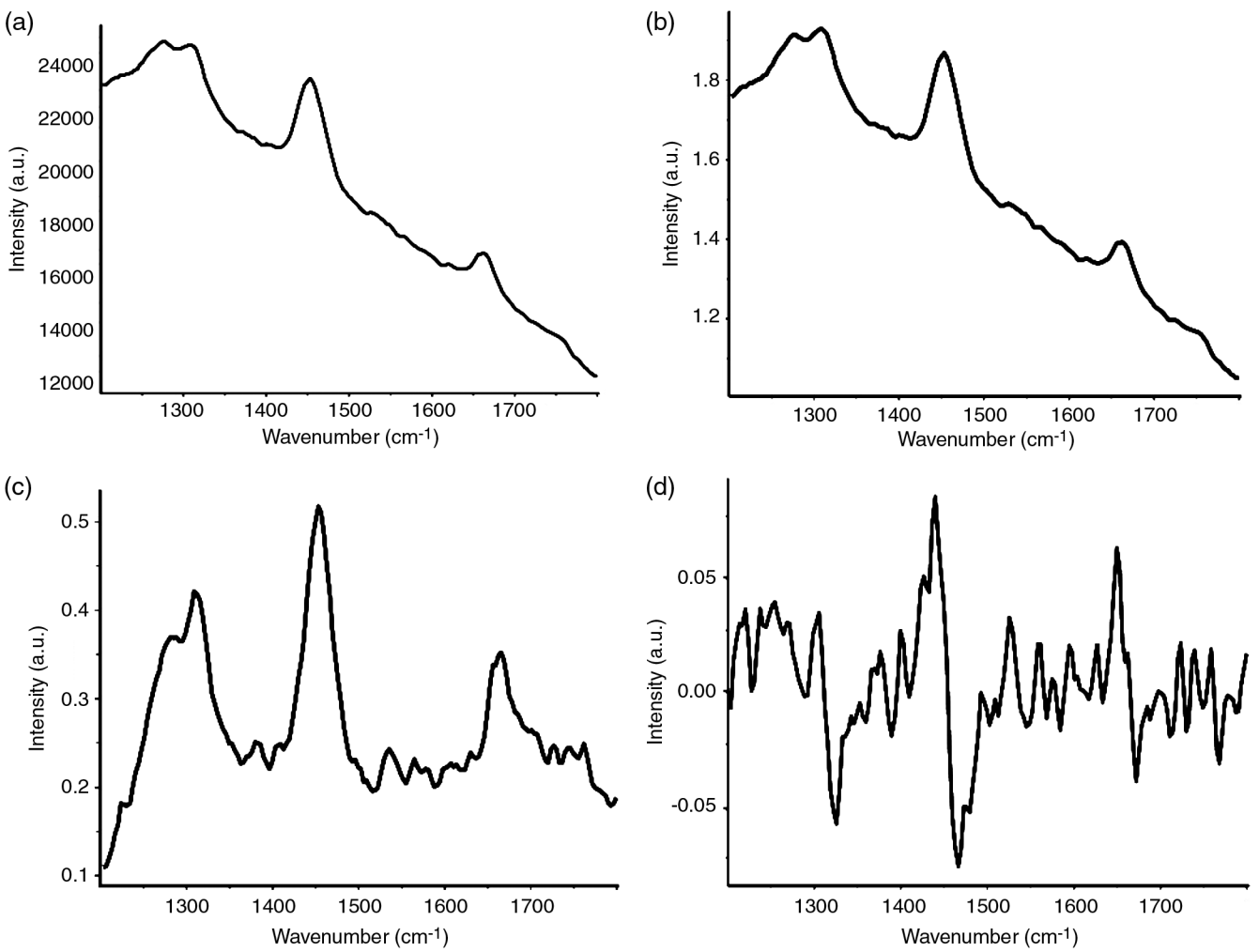


**Fig. 2** Spectral preprocessing for a typical *in vivo* spectrum (1200 to 1800 $cm^{-1}$): (a) raw spectrum; (b) spectrum after CCD response correction; (c) spectrum after background subtraction; (d) first derivative spectrum.

for 3 s and averaged over three accumulations. Photographic representation of the instrument is shown in Fig. 1.

### 2.3 *Data Analysis*

*In vivo* spectra from all five groups were corrected for CCD response with a NIST certified SRM 2241 material followed by subtraction of spectral contribution from optical elements. To minimize influence of slow-moving background first derivative of spectra were computed using Savitzky-Golay method. Typical spectrum at different preprocessing steps is shown in Fig. 2. Our previous studies related to oral cancers have demonstrated the utility of 1200 to 1800 $cm^{-1}$ region in successfully classifying normal and malignant conditions, we have used the

same region for analysis.[17,18] Further, this region is less influenced from fiber interference.

Spectra in the 1200 to 1800 $cm^{-1}$ region were then subjected to unsupervised PCA as well as supervised discrimination analyses (LDA) using algorithms implemented in MATLAB based in-house software.[21] In the first step, unsupervised PCA was performed to bring out classification among different groups. Score of factors were used for classification. These findings were also verified by supervised PC-LDA using significant principal components ($P < 0.05$) as input. In order to avoid overfitting of data factors accounting up to ~95% variance were chosen for LDA.[22] Classification efficiency of standard models was evaluated by LOOCV and independent test dataset. A summary of patient accrual in designing the study is shown in Table 1.

Average spectra were computed from background subtracted underivatized spectra for each class and baseline corrected by fitting a fifth-order polynomial function. Baseline corrected spectra were used for spectral comparisons among different groups.

## 3 Results and Discussion

### 3.1 *Spectral Features*

Mean baseline-corrected spectra from contralateral normal-site, age-matched healthy controls without any tobacco habit, premalignant patches, habitual tobacco users. and tumor sites are shown in Fig. 3. Spectral features from contralateral normal, healthy controls, and habitual tobacco users are dominated by lipid rich features indicated by $C = O$ band of esters, strong $\delta CH_2$ bend, two sharp features in amide III region, and a sharp peak in amide I region. Tumor spectra are dominated by protein bands indicated by broad amide III, broad and shifted $\delta CH_2$, and broad amide I. These findings corroborate earlier reports of *ex vivo* and *in vivo* conditions.[5–7,17,18,23,24] Spectra from habitual tobacco users are also dominated by lipid bands but exhibit minor differences, as compared with healthy controls, such as a minor shift in amide III and $\delta CH_2$ bend as well as broadening of amide I region, which could suggest changes in protein secondary structures. Spectra from premalignant patches show similarities such as broadening of amide III, amide I, and $\delta CH_2$ region with spectra from tumors. Long-term tobacco use can cause an increase in the number of proliferating epithelial cells in the upper-aero digestive tract of tobacco users leading to inflammation and can be considered as a preliminary event for changes culminating in the development of oral SCC.[25] Spectral differences observed between healthy controls, tobacco users, and contralateral normal could be due to varying degree of hypercellularity, which can be considered as indicative of transformation prone mucosa. Under normal conditions, clear-cut stratification between different layers like epithelium (superficial portion), lamina propria (subepithelial connective tissue), and submucosa (deeper portion of connective tissue) can be seen. Biochemically, lamina propria contains reticulin and collagen fibers, while submucosa is rich in adipose tissue. In case of tumors or other pathological conditions, there is loss in architectural arrangement of different layers; therefore, loss of lipid features is expected as content of different layers are mixed. In addition to this, pathological conditions cells have large amounts of surface proteins, receptor proteins, enzymes, antigens, and antibodies, which may give rise to a protein-dominated spectrum.[26] In our recent study we have shown that architectural and morphological organization of tissue components are the hallmark of spectral signatures.[27] Observed variations in Raman spectra of normal and pathological conditions could also be attributed to this.

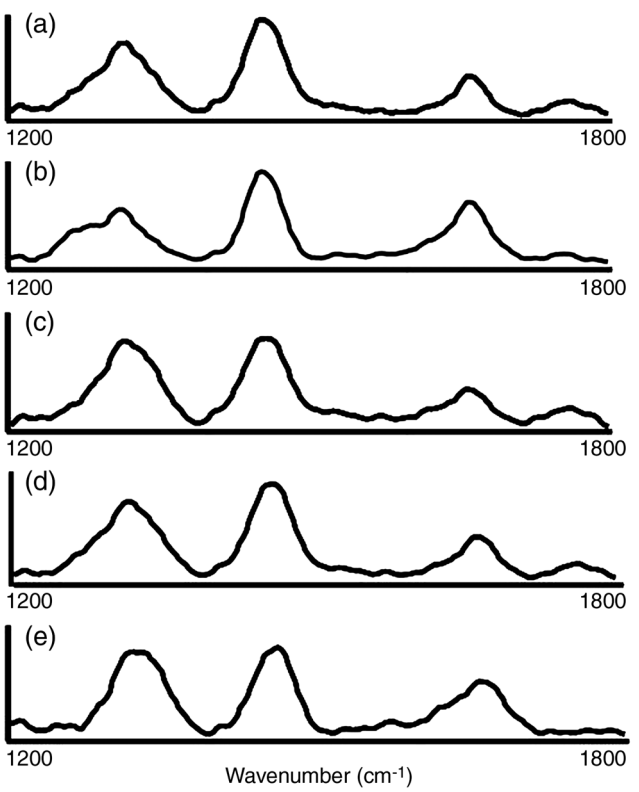


**Fig. 3** Mean *in vivo* Raman spectra from (a) contralateral normal; (b) healthy controls; (c) premalignant patches; (d) habitual tobacco users; (e) tumor sites.

### 3.2 *Statistical Analysis*

In first step 125 spectra from contralateral normal, 139 from tumor site, and 85 from premalignant patches were pooled and analyzed by PCA. It is an unsupervised method that explores patterns in the data set. It decomposes spectra data into a small number of independent variations called factors and contributions of these factors to each spectrum are called score. A total of 10 factors were used and score of factor 1 and 2 provided the best classification. Loading plots of the same and scatter plot are shown in Fig. 4. Minimally overlapping clusters belonging to normal, tumor, and premalignant spectra were obtained. Since PCA is often used as a data overview tool, which helps in identifying trends and patterns in data, supervised classification method LDA was also explored. It is method of classification that maximizes variability between groups and minimizes within group variability by maximizing the ratio between-class variance to the within-class variance in any particular data set. A total 24 factors contributing ~95% variance were chosen for analysis [Fig. 5(a)]. First, three factors contributing maximum variance were visually analyzed to determine extent of partition among groups. Scatter plot shown in Fig. 5(b) suggests that clusters belonging to spectra from premalignant patches makes separate clusters as compared with minimally overlapping clusters of contralateral normal and tumor groups. These results are also summarized in Table 2. As can be seen, 117, 80, and 134 spectra from contralateral normal, premalignant, and tumors, respectively, were correctly

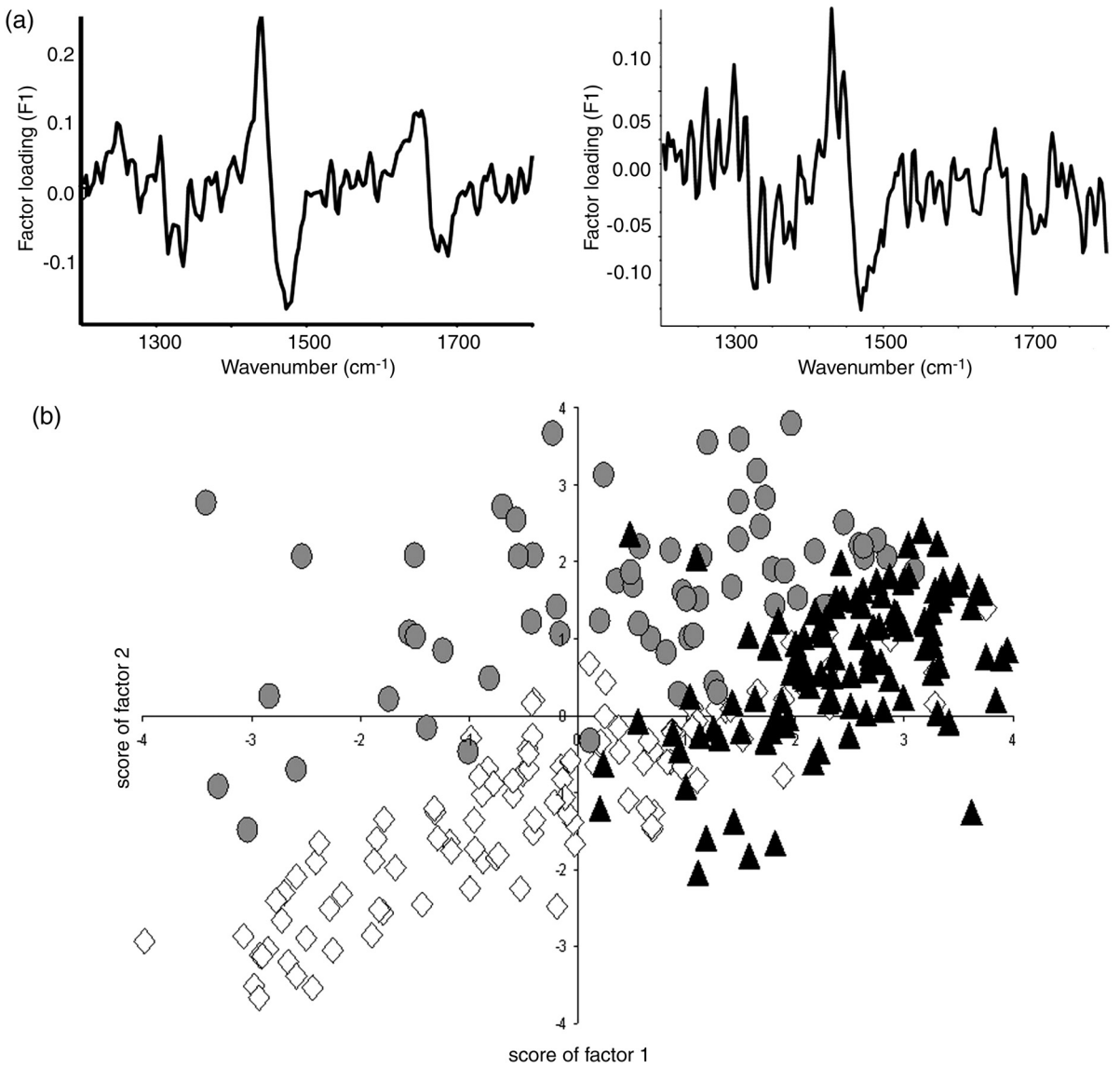


**Fig. 4** PCA of contralateral normal (□), tumor (▲), and premalignant spectra (•): (a) loading plots of factor 1 and factor 2; (b) scatter plot.

classified. Average classification efficiency of ~94.7% was observed. None of the contralateral normal or tumor spectra were misclassified as premalignant. LOOCV was performed to evaluate classification efficiency. Cross-validation, also called as rotation estimation, is a technique for assessing performance of a predictive model with a hypothetical validation set when an explicit validation set is not available. As the name suggests, leave-one-out (LOO) involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data, and results are averaged over the rounds. LOOCV results are shown in Table 2. 109 of 125 spectra from contralateral sites were correctly classified. Of the 16 misclassifications, one was premalignant and 15 were tumors. In case of tumors, 121 out of 139 spectra were correctly classified. Among the 18 misclassifications, only four were premalignant while 14 were contralateral normal. In the premalignant group, 66 out of 85 spectra were correctly classified. Of the 19 misclassifications, 12 were contralateral normal, and seven were tumors. Further, the test prediction efficiency of the classifier model was evaluated using remaining 90, 86, and 36 spectra from contralateral normal, tumor, and premalignant sites, respectively, as independent test data set. Of the 90 test normal spectra, 67 (74%) were correctly predicted, while 21 (23%) as premalignant and two as tumor (2%). In the case of tumor, of the 86 test spectra, 74 (86%) were correctly predicted. Out of 12 wrong predictions, four (5%) were premalignant, and eight (9%) were contralateral normal. Of the 36 premalignant test spectra, 26 (72%) were correctly predicted—seven (19%) as contralateral normal, and the remaining three (8%) as tumor. These results are summarized in Table 2. As a whole prediction efficiency of standard model for tumor spectra was highest as compared with models of contralateral normal and premalignant. Spectra from premalignant patches are specific and can be objectively discriminated with contralateral normal and tumor spectra.

As mentioned earlier, tobacco chewing and smoking are one of the etiological factors for development of oral cancers. Long-term exposure of tobacco-related carcinogens to mucosa of a healthy individual results in morphological and biochemical changes finally leading to development of a premalignant lesion.

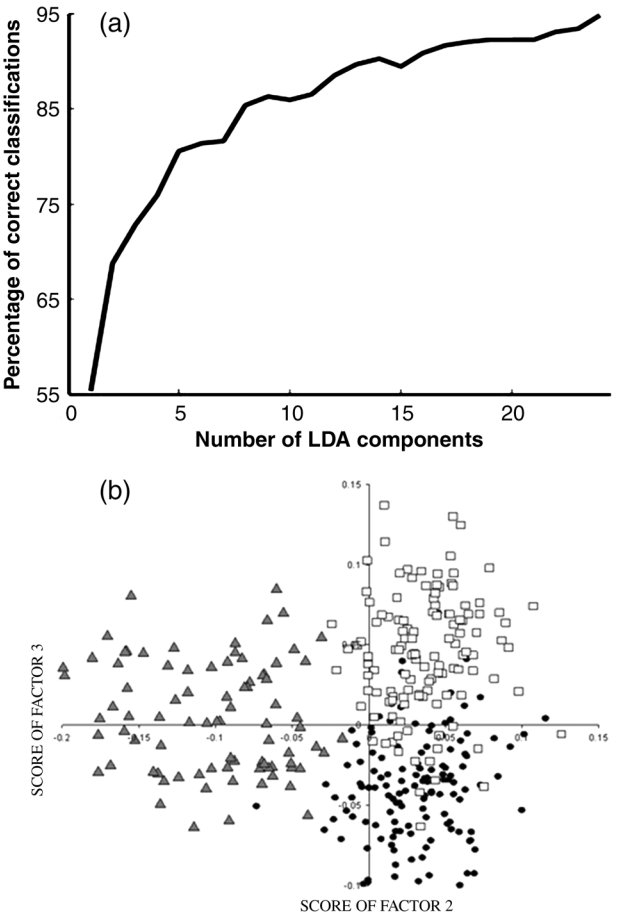


**Fig. 5** LDA of contralateral normal (□), tumor (•), and premalignant spectra (▲): (a) scree plot; (b) scatter plot.

Ideal diagnostic methods should be able to differentiate between premalignant and healthy subjects with and without tobacco habits. Therefore in the next step to evaluate potentials of Raman spectroscopy in discriminating premalignant conditions with closely related habitual tobacco users, we have analyzed spectra from premalignant patches against healthy subjects with and without tobacco habits. 100, 100, and 85 spectra from healthy controls, habitual tobacco users, and premalignant patches, respectively, were pooled and subjected to PCA to identify pattern in data. A total of 10 factors were used and score of factor 1 and 2 provided the best classification. Loading plots of the same and scatter plot are shown in Fig. 6. Separate cluster belonging to premalignant spectra and two minimally overlapping clusters of habitual tobacco user and healthy control spectra, respectively were obtained. In the next step, LDA was performed to explore feasibility of classification. A total of 18 factors corresponding to ~95% variance were used. The first three of the total of 18 factors were visually analyzed to determine degree of separation [Fig. 7(a)]. As shown in Fig. 7(b), three independent clusters belonging to healthy control, premalignant, and habitual tobacco users were obtained. As can be seen from Table 3, 97, 96, and 78 spectra from healthy controls, habitual tobacco users, and premalignant patches, respectively, were correctly classified. Overall classification efficiency of ~92% was observed. LOOCV results are shown in Table 3, 97 of 100 spectra from healthy controls were correctly classified, and the remaining three were misclassified

**Table 2** Confusion matrix for LDA, leave-one-out cross validation, and independent test data prediction analysis of contralateral normal, premalignant, and tumor spectra (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions).

| | Normal | Pre Mal | Tumor | Class. Eff. |
|---|---|---|---|---|
| Normal | 117 | 0 | 8 | 93.6% |
| Pre Mal | 2 | 80 | 3 | 94% |
| Tumor | 5 | 0 | 134 | 96% |
| Leave-one-out cross validation | | | | |
| | Normal | Pre Mal | Tumor | Class. Eff. |
| Normal | 109 | 1 | 15 | 87% |
| Pre Mal | 12 | 66 | 7 | 77.6% |
| Tumor | 14 | 4 | 121 | 87% |
| Independent test data predictions | | | | |
| | Normal | Pre Mal | Tumor | Prediction Eff. |
| Normal | 67/90 | 21 | 2 | 74% |
| Pre Mal | 7 | 26/36 | 3 | 72% |
| Tumor | 8 | 4 | 74/86 | 86% |

as habitual tobacco users. 95 out of 100 spectra from tobacco users were correctly classified. The five misclassifications were as healthy controls. Of the total 85 spectra from premalignant patches 68 were correctly classified. Most of misclassifications (15 out of 17) were with closely related habitual tobacco users and only two as healthy controls. Prediction efficiency of the standard model was evaluated using 50, 50, and 36 spectra from healthy controls, habitual tobacco users, and premalignant patches, respectively, as independent test data set. Of the 50 healthy test spectra, 47 (94%) were correctly predicted, and three (6%) wrong predictions were tobacco users. 43 (86%) out of 50 test tobacco users spectra were correctly predicted, and the remaining seven (14%) wrong predictions were healthy controls. Of the 36 test premalignant spectra, 20 (55%) were correctly predicted as premalignant. Of the 16 wrong predictions, 15 (44%) were as habitual tobacco users and only one was as healthy control. These results are summarized in Table 3. Most of the misclassifications of premalignant test spectra are with habitual tobacco users and are consistent with results shown in Table 2; i.e., misclassifications of premalignant spectra as contralateral normal. Since both the contralteral mucosa of subjects with oral cancer and mucosa of habitual tobacco users are exposed to tobacco-related carcinogen for longer time similar changes in comparison to healthy controls are expected. These findings corroborate our own study on long-term exposure related changes in buccal mucosa.[24]

In this study we have attempted to delineate premalignant spectra from contralateral normal and tumor spectra as well as from healthy subjects with and without tobacco habits. In the first stage, tumor spectra gave best prediction efficiency (86%) as compared to contralateral normal (74%) and premalignant (72%). In this case most of the misclassification seemed to occur between contralateral normal and premalignant spectra.

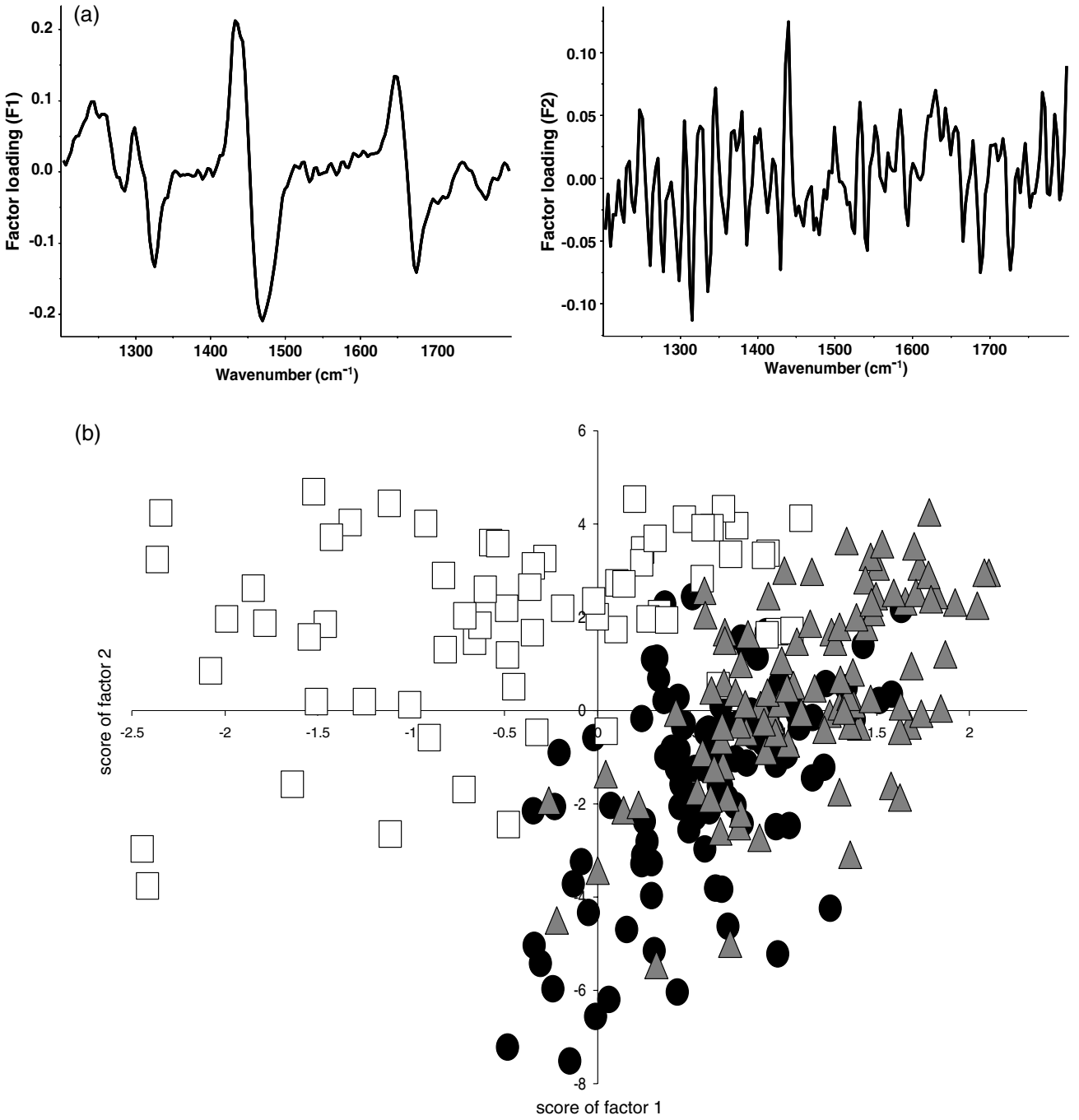


**Fig. 6** PCA of healthy controls (•), habitual tobacco users (▲) and premalignant spectra (□): (a) loading plots of factor 1 and factor 2; (b) scatter plot.

This is probably due to the fact that premalignant patches in the study were from the contralateral side. Further, our probing area is around 100 to 200 $\mu$m, since transformation of a premalignant zone may not be uniform, possibility of acquiring data from a normal site (under tobacco exposure) cannot be completely ruled out. This also explains observed misclassification across premalignant and malignant, as numbers of instances in this case are few as malignant conditions represent a higher degree of transformation as compare to premalignants.

When we consider the second system, i.e., discrimination of premalignants from healthy controls with and without tobacco habits, prediction efficiency of healthy controls without any tobacco habits was the highest (94%). Very few misclassifications with habitual tobacco users was observed, which can be explained on the basis of the fact that extent of tobacco related exposure may not be uniform across whole mucosa. There is always a possibility that mucosa at few places is still healthy. Most of the misclassifications were observed between premalignant and habitual tobacco users spectra. Once again the argument of nonuniformity of mucosa as well as effect of long-term tobacco exposure might explain the observed misclassifications. Overall, our study demonstrated the feasibility of discriminating premalignants from contralateral normal and tumor sites of subjects with oral cancer as well from healthy controls with and without tobacco habits. Prospectively, further studies incorporating pure premalignants,

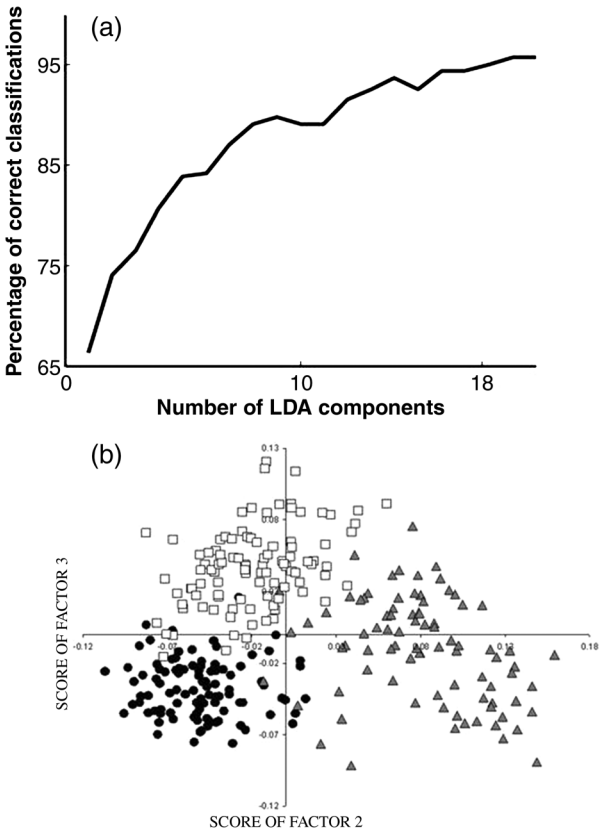


**Fig. 7** LDA of healthy controls (•), habitual tobacco users (□) and premalignant spectra (▲): (a) scree plot; (b) scatter plot.

**Table 3** Confusion matrix for LDA, leave-one-out cross validation, and independent test data prediction analysis of healthy controls without any tobacco habits, habitual tobacco users, and premalignant spectra (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions).

| | Healthy control | Tobacco user | Pre-Mal | Class. Eff. |
|---|---|---|---|---|
| Healthy control | 97 | 3 | 0 | 97% |
| Tobacco user | 4 | 96 | 0 | 96% |
| Pre-Mal | 0 | 7 | 78 | 91.7% |
| Leave-one-out cross validation | | | | |
| | Healthy controls | Tobacco user | Pre-Mal | Class. Eff. |
| Healthy control | 97 | 3 | 0 | 97% |
| Tobacco user | 5 | 95 | 0 | 95% |
| Pre-Mal | 2 | 15 | 68 | 80% |
| Independent test data predictions | | | | |
| | Healthy control | Tobacco user | Pre-Mal | Prediction Eff. |
| Healthy control | 47/50 | 3 | 0 | 94% |
| Tobacco user | 7 | 43/50 | 0 | 86% |
| Pre-Mal | 1 | 15 | 20/36 | 55% |

unlike premalignant patches, should be carried out. Another aspect, site-wise histopathology should also be considered in view of possible nonuniformity of buccal mucosa. Nevertheless, findings of the study demonstrate the applicability of noninvasive, *in vivo* Raman spectroscopic methods in oral cancer applications such as discrimination of normal, tumor, and premalignants.

## Acknowledgments

This work was carried out under project no: BT/PRI11282/MED/32/83/2008, Department of Biotechnology, Government of India. Authors would like to acknowledge all healthy volunteers who participated in the study.

## References

1. D. M. Parkin et al., "Global cancer statistics, 2002," *CA Cancer J. Clin.* **55**(2), 74–108 (2005).
2. D. C. De Veld et al., "The status of in vivo autofluorescence spectroscopy and imaging for oral oncology," *Oral Oncol.* **41**(2), 117–131 (2005).
3. J. G. Wu et al., "Distinguishing malignant from normal oral tissues using FTIR fiber-optic techniques," *Biopolymers* **62**(4), 185–192 (2001).
4. K. Venkatakrishna et al., "Optical pathology of oral tissue: a Raman spectroscopy diagnostic method," *Curr. Sci.* **80**(5), 101–105 (2001).
5. C. M. Krishna et al., "Micro-Raman spectroscopy for optical pathology of oral squamous cell carcinoma," *Appl. Spectrosc.* **58**(9), 1128–1135 (2004).
6. R. Malini et al., "Discrimination of normal, inflammatory premalignant, and malignant oral tissue: a Raman spectroscopy study," *Biopolymers* **81**(3), 179–193 (2006).
7. C. M. Krishna et al., "Raman signatures of tissue biopolymers: diagnosis of oral cancer and inflammatory conditions," in *Biopolymer Research Trends*, T. S. Nemeth, Eds., pp. 189–209, Nova Science Publishers, Happauge, NY (2007).
8. A. Nijssen et al., "Towards oncological application of Raman spectroscopy," *J. Biophoton.* **2**(1–2), 29–36 (2009).
9. C. Kendall et al., "Vibrational spectroscopy: a clinical tool for cancer diagnosis," *Analyst* **134**(6), 1029–1045 (2009).
10. P. Crow et al., "Assessment of fiber optic near-infrared Raman spectroscopy for diagnosis of bladder and prostate cancer," *Urology* **65**(6), 1126–1130 (2005).
11. M. G. Shim et al., "In vivo near-infrared Raman spectroscopy: demonstration of feasibility during clinical gastrointestinal endoscopy," *Photochem. Photobiol.* **72**(1), 146–150 (2000).
12. C. Kendall et al., "Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus," *J. Pathol.* **200**(5), 602–609 (2003).
13. P. J. Caspers et al., "*In vivo* confocal Raman microspectroscopy of the skin: noninvasive determination of molecular concentration profiles," *J. Invest. Dermatol.* **116**(3), 434–442 (2001).
14. A. Mahadevan-Jansen et al., "Development of a fiber optic probe to measure NIR Raman spectra of cervical tissue in vivo," *Photochem. Photobiol.* **68**(3), 23–32 (1998).
15. U. Utzinger et al., "Near-infrared Raman spectroscopy for in vivo detection of cervical precancers," *Appl. Spectrosc.* **55**(8), 955–959 (2001).
16. K. Guze et al., "Parameters defining the potential applicability of Raman spectroscopy as a diagnostic tool for oral disease," *J. Biomed. Opt.* **14**(1), 014016 (2009).
17. S. P. Singh et al., "Raman spectroscopy in head and neck cancers: towards oncological applications," *J. Cancer Res. Ther.* **8**(6), S126–S132 (2012).
18. S. P. Singh et al., "In vivo Raman spectroscopy for oral cancers diagnosis," *Proc. SPIE* **8219**, 82190K (2012).
19. D. L. Kasper et al., "Neoplastic diseases," in *Fauci, Harrison's Principles of Internal Medicine*, 16th ed., pp. 421–430, McGraw-Hill Professional, Canada (2004).
20. S. L. Robbins, R. S. Cotran, and V. Kumar, *Pathologic Basis of Disease*, Saunders/Elsevier, Philadelphia (1994).

21. A. D. Ghanate et al., "Comparative evaluation of spectroscopic models using different multivariate statistical tools in a multicancer scenario," *J. Biomed. Opt.* **16**(2), 025003 (2011).
22. T. J. Harvey et al., "Factors influencing the discrimination and classification of prostrate cancer cell lines by FTIR micro spectroscopy," *Analyst* **134**(6), 1083–1091 (2009).
23. H. Zhiwei et al., "Integrated Raman spectroscopy and trimodal wide-field imaging techniques for real-time in vivo tissue Raman measurements at endoscopy," *Opt. Lett.* **34**(6), 758–760 (2009).
24. S. P. Singh et al., "Tobacco related field effects in oral cancers: a Raman spectroscopic observation," in *AACR International Conference on New Horizons in Cancer Research: Biology to Prevention to Therapy* (2011).
25. D. M. Shin et al., "Sequential increase in proliferating cell nuclear antigen expression in head and neck tumorigenesis, a potential biomarker," *J. Natl. Cancer Inst.* **85**(12), 971–978 (1993).
26. A. T. Harris et al., "Raman spectroscopy in head and neck cancer," *Head Neck Oncol.* **2**(26), 1–6 (2010).
27. A. Deshmukh et al., "Raman spectroscopy of normal oral buccal mucosa tissues: study on intact and incised biopsies," *J. Biomed. Opt.* **16**(12), 127004 (2011).