Development of *in vivo* Raman diagnostic methodologies for oral pre-cancers and cancers

By

Surya Pratap Singh

[LIFE09200804011]

Tata Memorial Centre

Mumbai

A thesis submitted to the Board of Studies in Life Sciences

In partial fulfillment of requirements For the Degree of

DOCTOR OF PHILOSOPHY

Of

HOMI BHABHA NATIONAL INSTITUTE

APRIL, 2014

HOMI BHABHA NATIONAL INSTITUTE

Recommendations of the Viva Voce Board

As members of the Viva Voce Board, we certify that we have read the dissertation prepared by Surya Pratap Singh entitled "Development of *in vivo* Raman diagnostic methodologies for oral pre-cancers and cancers" and recommend that it may be accepted as fulfilling the dissertation requirements for the Degree of Doctor of Philosophy.

----------Date:

Chairperson: -- Dr. S.M. Zingde

----------Date:

Convener: -- Dr. C. Murali Krishna

----------Date:

Member 1:-- Dr. G. Krishnamoorthy

----------Date:

Member 2:-- Dr. G.B. Maru

----------Date:

Member 3:-- Dr. Pankaj Chaturvedi

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to HBNI.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

Date:

Place: Dr. C. Murali Krishna

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfilment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Navi Mumbai,

April 2014 Surya Pratap Singh

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. This work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Navi Mumbai,

April 2014. Surya Pratap Singh

CERTIFICATE

I certify that the thesis titled 'Development of *in vivo* Raman diagnostic methodologies for oral pre-cancers and cancers' submitted for the degree of Doctor of Philosophy by Mr. Surya Pratap Singh is a record of the research carried out by him during the period August 2008 to April 2014 under my supervision. This work has not formed the basis for the award of any degree, diploma, associate ship or fellowship at this or any other institute or university.

Navi Mumbai,

April 2014 Dr. C. Murali Krishna

Dedicated to

My Mom.....

ACKNOWLEDGEMENTS

I take this opportunity to convey my gratefulness to one and all those who have supported and guided me through the entire tenure of PhD work.

First and foremost, a sincere and heart filled gratitude to my mentor Dr. C. Murali Krishna for being such a wonderful and motivating guide. His stimulating suggestions, experience and encouragement accompanied by the freedom of thought that he granted, has not only helped me to discover my potential but also kept me sparked throughout my Ph.D tenure. I will never forget the generosity that he offered on several occasions. Sir, I thank you for everything and will always be proud to have been mentored by you.

I am thankful to Dr. Rajiv Sarin (Director, ACTREC), Dr. Surekha Zingde (Ex. Deputy Director, ACTREC) and Dr. Shubhada Chiplunkar (Deputy Director, ACTREC) for providing me an opportunity to work in this institution and the excellent infrastructure. I also thank DBT for funding the project and ACTREC for my fellowship. I am fortunate to have Dr. Surekha Zingde (ACTREC), Dr. G. Krishnamoorthy (TIFR), Dr. G.B.Maru (ACTREC) and Dr. Pankaj Chaturvedi (TMH) as my doctoral committee members. Their expert comments, critical analysis of the results and helpful suggestions have contributed significantly to the work.

A special thanks to Dr. Pragya (TMH), Dr. Poonam (TMH), Dr. Vikram (TMH), Dr. Atul Deshmukh (ACTREC) and Dr. Lekha (ACTREC) for their extremely precious clinical inputs in the work. I would like extend my deepest regards to Mrs. Arti for helping in tissue collection and histopathological analysis. I would like to acknowledge Dr. Alam, Chrismita, Richa, Gaurav, Shubhangi, Sai, Khushi and Pooja for their timely help at different stages of the work.

I want to express my warmest thanks to all past and current Chilakapati lab members Dr. Kalyan Kumar, Dr. Mahidhar Kodali, Avinash, Sandeep, Deepak, Shridhar, Ashvini, Mahesh, Isha, Rubina, Aditi, Piyush, Tanmoy, Priyanka and Pramod for keeping the lab environment lively and pleasant. I am especially grateful to Aditi, for all the help with the data collection, analysis as well as during manuscript and thesis writing in the last 2 years of my tenure.

I owe a lot to all my friends Piyush, Zahid, Vikrant, Manohar, Lalit, Nikhil, Neetu, Monica, Padma, Lumbini and the entire student's community of ACTREC who kept the tough and tense days of my Ph.D as joyous as possible. I am extremely thankful to some of my oldest friends Ravi, Rajiv, Manish, Vivek and Nidhi for being patient listeners and keeping my days fun filled.

I would like to thank Dr. K. Amin, Mr. Madan (biorepository, ACTREC), all the staff from common instrument facility, photography, library, IT, administration and accounts department of ACTREC for their constant help and support.

I am very much thankful to, my family for their constant support and encouragement. You all always have been my greatest strength and source of inspiration. Whatever I have achieved today is all because of your blessings and sacrifices. I feel so proud and blessed to be your son.

I also thank all others not mentioned here and who have been equally helpful towards me in various ways.

Finally, I thank almighty GOD who has helped and blessed me through the entire voyage and have made this day possible.

Surya Pratap Singh

Contents

Synopsis

SYNOPSIS

1. Name of the Student: Surya Pratap Singh

2. Name of the Constituent Institution: Tata Memorial Centre

3. Enrolment No.: LIFE09200804011

4. Title of Thesis: Development of *in vivo* Raman Diagnostic methodologies for oral pre-cancers and cancers

5. Board of Studies: Life Sciences

INTRODUCTION

Cancer is the second most common cause of morbidity and mortality in the world today, after cardiovascular disorders. Six million people die due to cancer every year. Oral cancer is the 15th most common cancer in the world and two-thirds of it occurs in the developing countries. In comparison to the U.S. population where it represents only ~3%, in India this cancer accounts for >30% of all malignancies [1,2]. India tops in the prevalence of oral cancer in the world and it remain the most common cancer amongst males. Oral cancer is the third most common cancer amongst women in India, after cervical and breast cancer [3,4]. Tobacco (both smoking and chewing) is regarded as the major cause of oral cancer. Alcohol consumption has a strong synergistic effect, as evident from increased risk seen in smokers plus drinkers [5,6]. The prognosis of patients with oral cancer is largely determined by the stage at which the disease is presented, as determined by extent of the tumour, presence of lymph-node metastases and distant metastases. Treatment strategies generally consist of surgery combined with postoperative radiotherapy and have a favourable prognosis for early lesions. In the past decade, the role of organ-preservation protocols, with combined chemo-radiation and surgery for salvage in oral cancer therapy, has increased [7-9]. Non-changing low disease free survival rate for oral cancers can be attributed to the fact that most of the oral squamous cell carcinoma (OSCC) present at a late stages (III or IV) [10-13].

Clinical examination and biopsies followed by histopathological analysis is considered as the gold standard for diagnosis and surveillance of oral cancer. However, the method has several limitations such as: inability in screening and detecting early malignancy associated changes; difficulty in recognizing subtle clinical changes in precancerous lesions or in a normal mucosa that

are indicative of neoplastic transformation; inability in distinguishing premalignant lesions from more common benign or inflammatory conditions; clinical or histological risk stratification lacks accuracy , reproducibility and requires large experience on part of the clinician [14-22]. Moreover, surveillance and biopsy of precancers is a mammoth task especially in populous countries like India. For example incidence of leukoplakia itself is up to 1% of general population. Considering these facts, it is imperative to develop a new rapid and accurate diagnostic method for early oral cancer detection.

Recent research has demonstrated that optical diagnostic methods can be used as alternative or adjunct to existing methods of cancer diagnosis. A variety of optical techniques like fluorescence, Fourier-transform infrared and Raman spectroscopy have been explored in cancer diagnosis [23-33]. These methods are capable of providing biochemical and morphological information in short time, which can be used for online diagnosis. Fluorescence based diagnosis of oral cancer began with use of exogenous fluorophores followed by autofluorescence studies [23-26]. Even though these methods involve simpler instrumentation,limited information and use of multiple excitation wavelengths have rendered its applicability for routine clinical usage. Fourier transform infrared spectroscopy (FT-IR) is absorption based vibrational spectroscopy method. Studies on *ex vivo* tissues have shown that differences due to loss of triglycerides, alterations in protein content and changes in keratin level can be considered as markers of oral malignancy [27,28]. However, these methodologies are less suitable for *in vivo* applications as water, the major component of biological tissues, is highly absorptive in the mid-IR range. ATR based methodologies could be useful in circumventing this difficulty.

Synopsis

Raman effect is based on inelastic scattering of photons. Unlike FT-IR, Raman spectroscopy does not suffer from water interference. However, week signals, as only very small fractions of photons (1 in 108) are inelastically scattered, is a major drawback associated with this technique [29]. Raman spectra of biological tissues are also often swamped by parasitic fluorescence. However, latest developments in light sources (lasers) and detectors (CCDs) have made Raman spectroscopy of biological samples like tissues and cells feasible. Further, use of near infrared excitation photons *e.g.* 785, 830 or 850 nm make this technique less harmful and also minimize the associated fluorescence. Most important attribute of Raman spectroscopy lies in its adaptation for *in vivo* applications. Using optical fibers laser light can be delivered to the desired site and Raman photons can also be collected. In view of above attributes Raman spectroscopy is projected as an ideal tool in pursuing biomedical applications. Raman spectroscopic differentiation of *ex vivo* normal and pathological conditions of oral, breast, cervix, colon, stomach ovarian and other forms of cancers have already been reported in the literature [30-38]. *In vivo* Raman measurements from bladder and prostate, oesophagus, skin, cervix and arteries are also reported [39-46].

RATIONALE AND OBJECTIVES:

Earlier studies on *ex vivo* oral tissues have demonstrated the feasibility of classifying normal, malignant, premalignant and inflammatory conditions by Raman spectroscopy [31,33]. **The work in this thesis aims towards developing and evaluating potential of *in vivo* laser Raman spectroscopy methods for non-invasive and objective diagnosis of oral cancers and precancers under clinical setting.**

Following are the specific objectives of the thesis -

Synopsis

1. Standardization of data acquisition and analysis methods on *ex vivo* oral tissues and correlation with histopathology and biochemical estimation.
2. To demonstrate feasibility of acquiring and classifying *in vivo* Raman spectra from buccal mucosa of normal, cancerous and pre-cancerous subjects and correlation with histopathology.
3. Exploring Raman spectral features of oral cancer cells with definite characters related to oral cancer.

Objective 1: ***Standardization of spectral acquisition and data analysis methods***

Earlier studies on *ex vivo* tissues using a modular instrument have demonstrated potentials of Raman spectroscopic methods in classifying normal, tumor, premalignant and inflammatory conditions. Lipid rich spectral features in normal and predominant protein features in tumor conditions were observed [31,33]. This objective was taken up to evaluate the reproducibility of spectral features with a fiberoptic probe coupled instrument. Spectra from pathologically verified, 36 pairs of oral biopsies (tumor and cut margin) were acquired. Biopsies were collected in liquid nitrogen and stored at -80^{0}C until use, from biorepository, ACTREC.

Instrument details and spectral acquisition

Spectra were acquired using 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, and a HE-785 spectrograph coupled with a CCD (Synapse) as dispersion and detection elements. The spectrograph is equipped with a fixed 950 gr/mm grating and spectral resolution, as specified by manufacturer, is ~4 cm^{-1}. Commercially available InPhotonics (Inc, Downy St. USA) probe consisting of 105 μm excitation fiber and 200 μm 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 μm and 1mm, respectively. A XYZ precision stage along with a probe holder was assembled to record spectra of *ex vivo* tissues. During spectral acquisition, biopsy samples were placed on CaF2 window and mounted on XYZ translational stage. Spectra were recorded with a spacing of ~2 mm in XY direction with parameters: laser power-80 mW, integration-10 seconds and 5-accumulations. These parameters were kept constant during all measurements.

Spectral pre-processing and multivariate analysis

Pre-processing of Raman spectra was performed by a standard protocol which involves correction for CCD response with a NIST certified standard reference material-2241 (SRM- 2241) material followed by subtraction of background signals from optical elements. To remove interference of the slow moving background, first derivatives of spectra (Savitzky-Golay method and window size 3) were computed. First derivative and vector normalized spectra in 1200-1800 cm^{-1} region were used for multivariate analysis by employing PC-LDA. Mean spectra were computed by averaging all variations on Y-axis keeping X-axis constant. Baseline correction of mean un-derivatized spectra was performed by fitting 5^{th} order polynomial function and were used for comparison across different groups.

Feasibility of classification between normal and tumor spectra was explored by multivariate supervised Principal Component-Linear Discriminant Analysis (PC-LDA) method using algorithms implemented in MATLAB (Mathworks Inc.) based in-house software [50]. PCA is the routinely used method for data compression and visualization, while LDA provides data classification based on an optimized criterion which is aimed for more class separability. LDA can be used in

combination with PCA (PC-LDA) to increase the efficiency of classification. For this, PCA scores obtained using a set of few PCs with maximum variance amongst data are used as input data for LDA-based classification. The advantage of doing this is to remove or minimize noise from the data and concentrate on variables important for classification. In our analysis, PC-LDA models were further validated by leave-one-out cross-validation (LOOCV) and independent test data.

Development of standard model and evaluation with independent test data

Standard models of normal and tumor conditions were developed using 63 and 68 spectra from 8 pairs of normal and tumor tissues, respectively. LOOCV yielded sensitivity and specificity of 88 and 79%. Remaining 256 and 296 spectra from 28 pairs of normal and tumor tissues, respectively were used as independent test data. Prediction efficiency of 80 and 95% for normal and tumors, respectively was observed. Corroborating earlier observations, mean spectrum of normal conditions was dominated by lipid bands while proteins were predominant in tumor spectrum [31,33]. Overall findings of the study confirmed the reproducibility of spectral features.

Correlation with band intensity and biochemical estimations

In the next step, lipid and protein rich spectral profiles of normal and tumor tissues were correlated with band intensity and biochemical estimation. Integrated area associated with lipid (1440 cm^{-1}) and protein bands (1450 and 1660 cm^{-1}) were calculated using curve-fitting algorithms of GRAMS/AI software (Thermo Scientific). Statistically significant (p value <0.0001) difference between average intensity of lipid and protein bands for normal (1.42 $\pm$ 0.25 and 0.51$\pm$0.12) and tumor tissues (0.43$\pm$0.18 and 1.46$\pm$0.29) was observed. Intensity plot of another protein band (1660 cm^{-1}) also yielded similar information *i.e.* high for tumors (1.12$\pm$0.19) and low for normal (0.89$\pm$0.28). These spectral features were then correlated with biochemical estimation of total lipid,

total protein and phopsholipids from same tissues. Total Protein, total lipid and phospholipids were estimated with Folin-Lowry, Floch and Rouser method, respectively [47-49]. The statistical comparisons were performed by unpaired Student's t–test and p value <0.05 was considered statistically significant. Corroborating spectral features, high protein to lipid and phospholipid ratio for tumor tissues (2.15±0.41 and 24.13±2.12) with respect to normal tissues (0.72±0.22 and 16.12±2.28) was observed.

Study on origin of spectral features in normal oral tissues

Spectral features of normal conditions show an abundance of lipids while tumors are rich in proteins. This study was undertaken to understand the origin of spectra in normal tissues and its influence on classification with tumor. Raman spectra from superior (epithelium) and inferior (connective tissue) surfaces of 10 *ex vivo* normal intact and incised oral tissues were acquired. Spectra obtained from upper and lower surfaces of intact oral tissue showed lipid and protein signatures due to histological arrangement of lipid and collagen molecules in the epithelium, lamina propria and connective tissue. Spectra from the superior and inferior surfaces of intact biopsy showed overlapping cluster after PCA, probably due to spectral contribution from entire length of tissue. On the other hand spectra from same surfaces after epithelium separation are different. However, spectra of all four groups of normal tissues also gave exclusive clusters when tested against tumor spectra. Overall findings of this study demonstrate that spectra recorded from the superior or inferior surface of an intact tissue may have contributions from deeper layers and has no bearing on classification with tumors. ***[J Biomed Opt, 16 (11), 2011]***

Objective 2: ***In vivo Raman spectroscopy for diagnosis of normal, cancer and precancerous conditions.***

Synopsis

The above standardized spectral acquisition and data analysis protocols were used for the *in vivo* studies employing the same instrumental set-up. Uniformity during spectral acquisition across all measurements was ensured by recording spectra as per the teeth positions *i.e.* buccal surfaces opposite of canine, first premolar, second premolar, first molar and second molar on both right and left side was considered as reference point and spectra were acquired. To avoid any differences because of mouth environment, subjects were allowed to wash their mouth with water before spectral acquisition. In order to maintain constant focal length during all measurements, a detachable, metallic spacer of length 5 mm was added at the tip of the fiberoptic probe. This provided flexibility during spectral acquisition, and since it can be disinfected, patient to patient contamination was also avoided. Prior to obtaining spectra from any individual, probe was disinfected with CIDEX solution (Johnson and Johnson, Mumbai, India). Spectral acquisition parameters were: laser power-80mW, integration time-3 seconds and 3-accumulations. *In vivo* Raman spectra from contralateral normal and cancerous lesion of 113 subjects were acquired. Spectra from 40 individuals were used for developing standard models and the remaining as test. *In vivo* spectra were also acquired from 50 subjects with only premalignant patches. Of these, spectra from 24 subjects were used for developing standard model and remaining as test. Spectra were corrected for CCD response followed by subtraction of background signals as per the previously described procedure.

Spectra from contralateral normal were dominated by lipid features indicated by C=O band of esters, strong δCH_2 bend, two sharp bands in amide III region, and a sharp peak in amide I region. Predominant protein bands indicated by broad amide III, broad and shifted δCH_2, and broad amide I were observed in mean tumor spectra. These findings corroborate earlier reports of *ex-vivo* and *in*

vivo conditions [31,33,51,52]. Spectra from premalignant patches show features associated with tobacco induced hypercellularity and profile similar to tumor spectra in amide III, amide I, and δCH_2 regions.

Development of standard models and evaluation with independent test data

Standard models were developed using 170 spectra from contralateral normal areas (40 subjects), 192 spectra from tumor sites (40 subjects) and 113 spectra premalignant patches (24 subjects). Mean ages of subjects with cancerous and premalignant lesions were 48.66 years and 51.33 years, respectively. First derivative and vector normalized spectra in 1200-1800 cm^{-1} region were used as input for PC-LDA. LOOCV yielded efficiency of 86, 91 and 91% for normal, premalignant and tumor spectra, respectively. Remaining 274 spectra from contralateral normal areas (73 subjects), 181 spectra from tumor sites (73 subjects) and 93 spectra from premalignant patches (26 subjects) were used as independent test data set and prediction efficiencies of 79, 60 and 86% for contralateral normal, premalignant, and tumor, respectively was observed.

Influence of variability in tumor grade and differentiation status on classification was also explored. Findings suggest that it has no influence on classification with normal or precancerous conditions.

Misclassifications between different groups can be primarily attributed to mucosal heterogeneity. Spectra from tumors gave best prediction efficiency (86%) followed by contralateral normal (79%) and premalignant (60%). Misclassification of tumor spectra as contralateral normal can be explained on the basis of fact that spectra were recorded at different points therefore, possibility of acquiring spectra from normal or inflammatory patches in a tumor cannot be completely ruled out. Maximum misclassification was observed between contralateral normal and premalignant spectra. This is probably due to the fact that premalignant patches in the study were from contralateral side.

Further, our probing area is around 100-200 μm, since transformation of a premalignant zone may not be uniform, possibility of acquiring data from a normal site cannot be completely ruled out. This also explains observed misclassification across premalignant and malignant, as numbers of instances in this case are very few as malignant conditions represent higher degree of transformation as compare to premalignant. ***[J Biomed Opt, 17 (105002), 2012; J Cancer Res Ther,8,2012; Proc SPIE 8219,2012]***

Study on classification of different premalignant lesions in oral cavity

A wide array of precancerous conditions like leukoplakia, erythroplakia, oral lichen planus, oral submucous fibrosis (OSMF), erythematosus etc. has been implicated in the development of oral cancer. Leukoplakia and OSMF are two of most common pre-cancerous conditions found in Indian population. However, clinical manifestations of both conditions are very different. Leukoplakia is described as a white patch or plaque that cannot be characterized clinically or pathologically as any other disease. OSMF is a chronic progressive condition where fibroelastic changes of oral mucosa along with epithelial atrophy leads to stiffness of mucosa resulting in trismus and inability to eat. Despite the general accessibility of the oral cavity during physical examination, many malignancies are not diagnosed until late stages of disease. In order to explore potentials of Raman spectroscopy in classifying these two conditions 62 OSMF spectra from 14 subjects and 53 leukoplakia spectra from 12 subjects were analyzed against contralateral normal and tumor spectra. PC-LDA followed by LOOCV yielded efficiency of 49 and 57% for leukoplakia and OSMF, respectively. Misclassifications can be explained on the basis of varying grade of thickness of a patch and the fact that often oral cancer subjects are presented with multiple premalignant conditions.

Study on cancer field effects or malignancy associated changes

Subtle changes in the oral mucosa because of tobacco abuse/unknown etiological factors may serve as prognostic markers. These changes also referred as Cancer field effects or Malignancy associated changes (CFE / MACs), are shown to be primarily associated with development of secondary tumors. In order to explore potential of Raman spectroscopy in detecting these changes, a separate study using 722 *in vivo* Raman spectra from 84 subjects was carried out under following five categories-

- Cancer and Contralateral normal (cancer and tobacco habit)
- Healthy controls (no tobacco habit, no cancer)
- Habitués healthy controls (no cancer, tobacco habit)
- Non-habitués contralateral normal (cancer, no tobacco habits)

Mean and difference spectra suggested that loss of lipid, and features representing proteins and DNA are characteristics of all pathological conditions, with respect to healthy controls. PC-LDA results suggest that Raman characteristics of mucosa of healthy controls are exclusive, while that of habitués healthy controls are similar to the contralateral normal mucosa, suggesting carcinogen induced field changes can be identified. It was also found that cluster of non-habitués contralateral normal mucosa is different from habitués healthy controls, indicating malignancy associated changes are different from carcinogen induced changes and can be identified with Raman spectroscopy. The non-invasiveness and use of harmless excitation wavelength impart several advantages to this method, and thus prospectively has potential to become an ideal mass screening tool in public health programs. ***[Analyst 138, 2013]***

Objective 3: ***Exploring Raman spectral features of oral cancer cells with definite characters related to oral cancer.***

Loss of keratin is shown to be a prospective diagnostic marker for oral cancers. The present objective was taken up to explore potential of Raman spectroscopy in identifying spectral markers related to loss of keratin in oral cancer-derived cell lines.

Keratins belong to the intermediate filament (IF) family of proteins and are one of the most widely used markers for oral cancers. These are not expressed in normal oral tissues, but are expressed in oral cancers. Aberrant expression of keratins 8 and 18 is most common change in human oral cancer. Epithelial tissues express different pairs of keratins depending upon the epithelial cell type and stage of differentiation (*e.g.* all stratified squamous epithelia express K5 and K14, whereas K8 and K18 are seen in simple epithelia) [53-55]. Recently, it has been shown that knockdown of K8 in the OSCC-derived cell line AW13516 leads to a substantial reduction in tumorigenicity, cell-motility, and cell invasion, indicating role of keratin 8/18 in invasion and metastasis as well as in promoting malignant transformation [56,57]. We hypothesized that identification of spectral contribution from keratin (K8/18) protein in squamous cell carcinoma derived cells could serve as additional marker for oral cancer diagnosis.

Cell line and spectral acquisition

We have chosen tongue cancer derived AW13516 cell line [56]. Cells expressing keratin 8/18 are called as vector-controls and cells with reduced expression are called as knockdown for K8/18 [57]. Cells were grown up to 80% confluence and synchronized by growing under serum free conditions. Cells were collected using a cell scraper and pelleted after washing with PBS and centrifugation at 2000 rpm for 10 minutes. Cell pellets of three independent experiments in duplicate were used for

recording Raman spectra. A total of 123 and 96 spectra from knockdown clones and vector control, respectively, were acquired using already described fiberoptic probe-coupled Raman system. Spectral acquisition parameters were laser power-80 mW, integration-8 seconds and 6-accumulations.

Multivariate analysis

Spectral pre-processing was performed by as per the previously described procedure. Pre-processed spectra in 1200-1800 cm^{-1} region were utilized for PC-LDA and LOOCV yielded ~63% classification efficiency. In order to identify specific spectral contribution from keratin, spectra from purified keratin using same set-up was also acquired. However, no specific Raman bands associated with keratin presence or absence was observed. The differences between knockdown and vector control cells could be attributed to the morphological changes induced due to loss of keratin 8/18. Morphological differences among both groups were established using confocal microscopy and live cell imaging [57]. It was observed that due to loss of K8, knockdown cells have symmetric contracted epithelial appearance as compared to vector controls. ***[Proc. SPIE 8225,2012]***

Raman micro-spectroscopic studies

Morphological differences due to keratin loss between both groups were further established by Raman imaging. K8 knock-down and vector control cells were grown on a cover slip and mounted in water on a glass slide and placed under the microscope. Spectra were acquired using WITec Raman alpha300 R (WITec GmbH) imaging system. Briefly, this system consists of a 532nm laser as excitation source and spectrograph with 600 gr/mm grating. The laser light is focused on the sample using an oil immersion Zeiss 63X objective (NA-0.55) and Raman scattering was detected by CCD coupled with the spectrograph.

Raman maps for knock-down and vector control cells were generated using K-means clustering method. Different clusters corresponding to cell membrane and nucleus were obtained. Features from vector control cells suggest protruding prominent actin based microfilaments and elongated shape, while knockdown cells show loss of filaments and round epithelial shape. Mean spectra from different clusters corresponding to cell membrane, microfilaments, sub cellular components, and nucleus were extracted and compared with spectra recorded from purified K8/18 protein. Findings suggest that spectra from different cellular compartments (nucleus, cell membrane and cytoplasm) can be obtained. Raman signals due to loss of keratin could not be observed; however, morphological differences between both groups were established by Raman mapping.

Summary and conclusions

Work reported in this thesis supports application of Raman spectroscopy in oral cancer diagnosis. Reproducibility of strong lipid features of normal, and protein rich spectral features of tumor tissues was established and correlated with band intensity calculations and biochemical estimation. Origin of these signals in normal tissues was understood and contribution of deeper layers on spectral profile was demonstrated. Potential of Raman spectroscopy in identifying subtle changes induced by loss of keratin was explored in tongue cancer cell line and established through Raman mapping. To the best of our knowledge, for the first time, we have demonstrated the feasibility of acquiring good quality *in vivo* Raman spectra under clinically implementable time, and classifying normal, cancer, and precancerous conditions, in Indian population. Potential of Raman spectroscopy in identifying earliest pre-neoplastic changes associated with carcinogen exposure or unknown etiological factors in uninvolved normal mucosa were also evaluated. Future studies involving pure

premalignant subjects and rigorous evaluation of the standard models may help in realizing translation of these technologies for routine clinical usage.

Referecnes:

1. Ferlay, J., H. R. Shin, F. Bray, D. Forman, C. Mathers and D. M. Parkin (2010). "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008." Int J Cancer **127**(12): 2893-2917.
2. Parkin, D. M., F. Bray, J. Ferlay and P. Pisani (2005). "Global cancer statistics, 2002." CA Cancer J Clin **55**(2): 74-108.
3. Saman, D. M. (2012). "A review of the epidemiology of oral and pharyngeal carcinoma: update." Head Neck Oncol **4**: 1.
4. Dikshit, R., P. C. Gupta, C. Ramasundarahettige, V. Gajalakshmi, L. Aleksandrowicz, R. Badwe, R. Kumar, S. Roy, W. Suraweera, F. Bray, M. Mallath, P. K. Singh, D. N. Sinha, A. S. Shet, H. Gelband, P. Jha and C. Million Death Study (2012). "Cancer mortality in India: a nationally representative survey." Lancet **379**(9828): 1807-1816.
5. Boffetta, P., S. Hecht, N. Gray, P. Gupta and K. Straif (2008). "Smokeless tobacco and cancer." Lancet Oncol **9**(7): 667-675.
6. Jeng, J. H., M. C. Chang and L. J. Hahn (2001). "Role of areca nut in betel quid-associated chemical carcinogenesis: current awareness and future perspectives." Oral Oncology **37**(6): 477-492.
7. Smith, R. B., J. C. Sniezek, D. T. Weed and M. K. Wax (2007). "Utilization of Free Tissue Transfer in Head and Neck Surgery." Otolaryngology -- Head and Neck Surgery **137**(2): 182-191.
8. Vergeer, M. R., P. A. H. Doornaert, D. H. F. Rietveld, C. R. Leemans, B. J. Slotman and J. A. Langendijk (2009). "Intensity-Modulated Radiotherapy Reduces Radiation-Induced Morbidity and Improves Health-Related Quality of Life: Results of a Nonrandomized Prospective Study Using a Standardized Follow-Up Program." International Journal of Radiation Oncology*Biology*Physics **74**(1): 1-8.
9. Boscolo–Rizzo, P., F. Maronato, C. Marchiori, A. Gava and M. C. Da Mosto (2008). "Long-Term Quality of Life After Total Laryngectomy and Postoperative Radiotherapy Versus Concurrent Chemoradiotherapy for Laryngeal Preservation." The Laryngoscope **118**(2): 300-306.
10. Mignogna, M. D., S. Fedele and L. L. Russo (2004). "The World Cancer Report and the burden of oral cancer." European Journal of Cancer Prevention **13**(2): 139-142.

11. Mignogna, M. D., S. Fedele, L. Lo Russo, E. Ruoppo and L. Lo Muzio (2002). "Costs and effectiveness in the care of patients with oral and pharyngeal cancer: analysis of a paradox." European Journal of Cancer Prevention **11**(3): 205-208.
12. Mignogna, M. D., S. Fedele, L. L. Russo, E. Ruoppo and L. L. Muzio (2001). "Oral and pharyngeal cancer: lack of prevention and early detection by health care providers." European Journal of Cancer Prevention **10**(4): 381-383.
13. Silverman,S., Jr. (1988). "Early diagnosis of oral cancer." Cancer **62**: 1796-1799.
14. Abbey, L. M., G. E. Kaugars, J. C. Gunsolley, J. C. Burns, D. G. Page, J. A. Svirsky, E. Eisenberg, D. J. Krutchkoff and M. Cushing (1995). "Intraexaminer and interexaminer reliability in the diagnosis of oral epithelial dysplasia." Oral Surg Oral Med Oral Pathol Oral Radiol Endod **80**(2): 188-191.
15. Sandler, H. C. (1962). "Cytological screening for early mouth cancer. Interim report of the veterans administration co-operative study of oral exfoliative cytology." Cancer **15**(6): 1119-1124.
16. Sankaranarayanan, R., K. Ramadas, G. Thomas, R. Muwonge, S. Thara, B. Mathew, B. Rajan and G. Trivandrum Oral Cancer Screening Study (2005). "Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial." Lancet **365**(9475): 1927-1933.
17. Mignogna, M. D. and S. Fedele (2005). "Oral cancer screening: 5 minutes to save a life." Lancet **365**(9475): 1905-1906.
18. Thomson, P. J. (2002). "Field change and oral cancer: new evidence for widespread carcinogenesis?" Int J Oral Maxillofac Surg **31**(3): 262-266.
19. Frist, S. (2003). "The oral brush biopsy: separating fact from fiction." Oral Surg Oral Med Oral Pathol Oral Radiol Endod **96**(6): 654-655.
20. Eisen, D. and S. Frist (2005). "The relevance of the high positive predictive value of the oral brush biopsy." Oral Oncol **41**(7): 753-755; author reply 756.
21. Potter, T. J., D.-J. Summerlin and J. H. Campbell (2003). "Oral malignancies associated with negative transepithelial brush biopsy." Journal of Oral and Maxillofacial Surgery **61**(6): 674-677.
22. Bhoopathi, V., S. Kabani and A. K. Mascarenhas (2009). "Low positive predictive value of the oral brush biopsy in detecting dysplastic oral lesions." Cancer **115**(5): 1036-1040.
23. Poh, C. F., L. Zhang, D. W. Anderson, J. S. Durham, P. M. Williams, R. W. Priddy, K. W. Berean, S. Ng, O. L. Tseng, C. MacAulay and M. P. Rosin (2006). "Fluorescence Visualization Detection of Field Alterations in Tumor Margins of Oral Cancer Patients." Clinical Cancer Research **12**(22): 6716-6722.
24. De Veld, D. C. G., M. J. H. Witjes, H. J. C. M. Sterenborg and J. L. N. Roodenburg (2005). "The

status of in vivo autofluorescence spectroscopy and imaging for oral oncology." Oral Oncology **41**(2): 117-131.

25. Lane, P. M., T. Gilhuly, H. Zeng, C. F. Poh, P. M. Williams, L. Zhang, M. P. Rosin, C. E. MacAulay, S. Ng and P. Whitehead (2006). "Simple device for the direct visualization of oral-cavity tissue fluorescence." Journal of Biomedical Optics **11**(2): 024006-024006.
26. Betz, C. S., H. Stepp, P. Janda, S. Arbogast, G. Grevers, R. Baumgartner and A. Leunig (2002). "A comparative study of normal inspection, autofluorescence and 5-ALA-induced PPIX fluorescence for oral cancer diagnosis." Int J Cancer **97**(2): 245-252.
27. Wu, J. G., Y. Z. Xu, C. W. Sun, R. D. Soloway, D. F. Xu, Q. G. Wu, K. H. Sun, S. F. Weng and G. X. Xu (2001). "Distinguishing malignant from normal oral tissues using FTIR fiber-optic techniques." Biopolymers **62**(4): 185-192.
28. Schultz, C. P., K. Z. Liu, P. D. Kerr and H. H. Mantsch (1998). "In situ infrared histopathology of keratinization in human oral/oropharyngeal squamous cell carcinoma." Oncol Res **10**(5): 277-286.
29. Raman C.V. and Krishnan K.S. (1928). "A change of wavelength in light scattering." Nature **121,**501.
30. Mahadevan-Jansen, A. and R. R. Richards-Kortum (1996). "Raman spectroscopy for the detection of cancers and precancers." Journal of Biomedical Optics **1**(1): 31-70.
31. Venkatakrishna, K., Jacob Kurien, Keerthilata M. Pai, Manna Valiathan, Nirmala Nagesh Kumar, C. Murali Krishna, G. Ullas and V.B. Kartha (2001). "Optical pathology of oral tissue: A Raman Spectroscopy diagnostic method." Current Science **80**:101-105.
32. V. B. Kartha, R. Jyothi Lakshmi, K.K. Mahato, C. Murali Krishna**,** C. Santosh, and K.Venkatakrishna (2003). "Raman Spectroscopy in Clinical Investigations." Indian Journal Physics **77**B, 113-123.
33. Malini, R., K. Venkatakrishna, J. Kurien, K. M. Pai, L. Rao, V. B. Kartha and C. M. Krishna (2006). "Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: a Raman spectroscopy study." Biopolymers **81**(3): 179-193.
34. Krishna, C. M., N. B. Prathima, R. Malini, B. M. Vadhiraja, R. A. Bhatt, D. J. Fernandes, P. Kushtagi, M. S. Vidyasagar and V. B. Kartha (2006). "Raman spectroscopy studies for diagnosis of cancers in human uterine cervix." Vibrational Spectroscopy **41**(1): 136-141
35. Chowdary, M. V. P., K. K. Kumar, J. Kurien, S. Mathew and C. M. Krishna (2006). "Discrimination of normal, benign, and malignant breast tissues by Raman spectroscopy." Biopolymers **83**(5): 556-569.

36. Chowdary, M. V., K. K. Kumar, K. Thakur, A. Anand, J. Kurien, C. M. Krishna and S. Mathew (2007). "Discrimination of normal and malignant mucosal tissues of the colon by Raman spectroscopy." Photomed Laser Surg **25**(4): 269-274

37. Kalyan Kumar, K., A. Anand, M. V. P. Chowdary, Keerthi, J. Kurien, C. Murali Krishna and S. Mathew (2007). "Discrimination of normal and malignant stomach mucosal tissues by Raman spectroscopy: A pilot study." Vibrational Spectroscopy **44**(2): 382-387.

38. Krishna, C. M., G. D. Sockalingum, R. A. Bhat, L. Venteo, P. Kushtagi, M. Pluot and M. Manfait (2007). "FTIR and Raman microspectroscopy of normal, benign, and malignant formalin-fixed ovarian tissues." Anal Bioanal Chem **387**(5): 1649-1656.

39. Crow, P., A. Molckovsky, N. Stone, J. Uff, B. Wilson and L. M. WongKeeSong (2005). "Assessment of fiberoptic near-infrared raman spectroscopy for diagnosis of bladder and prostate cancer." Urology **65**(6): 1126-1130.

40. Shim, M. G., B. C. Wilson, E. Marple and M. Wach (1999). "Study of Fiber-Optic Probes for in Vivo Medical Raman Spectroscopy." Appl. Spectrosc. **53**(6): 619-627.

41. Shim, M. G., L.-M. Wong Kee Song, N. E. Marcon and B. C. Wilson (2000). "In vivo Near-infrared Raman Spectroscopy: Demonstration of feasibility during clinical gastrointestinal endoscopy." Photochemistry and Photobiology **72**(1): 146-150.

42. Nijssen, A., T. C. Bakker Schut, F. Heule, P. J. Caspers, D. P. Hayes, M. H. Neumann and G. J. Puppels (2002). "Discriminating basal cell carcinoma from its surrounding tissue by Raman spectroscopy." J Invest Dermatol **119**(1): 64-69.

43. Chrit, L., C. Hadjur, S. Morel, G. Sockalingum, G. Lebourdon, F. Leroy and M. Manfait (2005). "In vivo chemical investigation of human skin using a confocal Raman fiber optic microprobe." Journal of Biomedical Optics **10**(4): 044007-044007.

44. Mahadevan-Jansen, A., M. F. Mitchell, N. Ramanujam, A. Malpica, S. Thomsen, U. Utzinger and R. Richards-Kortum (1998). "Near-infrared Raman spectroscopy for in vitro detection of cervical precancers." Photochem Photobiol **68**(1): 123-132.

45. Buschman, H. P., E. T. Marple, M. L. Wach, B. Bennett, T. C. Bakker Schut, H. A. Bruining, A. V. Bruschke, A. van der Laarse and G. J. Puppels (2000). "In Vivo Determination of the Molecular Composition of Artery Wall by Intravascular Raman Spectroscopy." Analytical Chemistry **72**(16): 3771-3775.

46. Nijssen, A., S. Koljenović, T. C. B. Schut, P. J. Caspers and G. J. Puppels (2009). "Towards oncological application of Raman spectroscopy." Journal of Biophotonics **2**(1-2): 29-36.

47. Lowry, O. H., N. J. Rosebrough, A. L. Farr and R. J. Randall (1951). "Protein measurement with the folin phenol reagent." Journal of Biological Chemistry **193**(1): 265-275.

48. Folch, J., M. Lees and G. H. S. Stanley (1957). "A Simple method for the isolation and purification of total lipids from animal tissues." Journal of Biological Chemistry **226**(1): 497-509.

49. Rouser, G., S. Fleischer and A. Yamamoto (1970). "Two dimensional thin layer chromatographic separation of polar lipids and determination of phospholipids by phosphorus analysis of spots." Lipids **5**(5): 494-496.

50. Ghanate, A. D., S. Kothiwale, S. P. Singh, D. Bertrand and C. M. Krishna (2011). "Comparative evaluation of spectroscopic models using different multivariate statistical tools in a multicancer scenario." J Biomed Opt **16**(2): 025003.

51. Huang, Z., S. K. Teh, W. Zheng, J. Mo, K. Lin, X. Shao, K. Y. Ho, M. Teh and K. G. Yeoh (2009). "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.

52. Guze, K., M. Short, S. Sonis, N. Karimbux, J. Chan and H. Zeng (2009). "Parameters defining the potential applicability of Raman spectroscopy as a diagnostic tool for oral disease." J Biomed Opt **14**(1): 014016.

53. Moll, R., W. W. Franke, D. L. Schiller, B. Geiger and R. Krepler (1982). "The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells." Cell **31**(1): 11-24.

54. Coulombe, P. A. and M. B. Omary (2002). "'Hard' and 'soft' principles defining the structure, function and regulation of keratin intermediate filaments." Curr Opin Cell Biol **14**(1): 110-122.

55. Vaidya, M. M., A. M. Borges, S. A. Pradhan and A. N. Bhisey (1996). "Cytokeratin expression in squamous cell carcinomas of the tongue and alveolar mucosa." Eur J Cancer B Oral Oncol **32B**(5): 333-336.

56. Tatake R. J., N. Rajaram, R. N. Damle, B. Balsara, A. N. Bhisey and S. G. Gangal (1990) “Establishment and characterization of four new squamous cell carcinoma cell lines derived from oral tumors”, Journal of Cancer Research and Clinical Oncology, **116**(2),179-186.

57. Alam H, T. K. Samrat, S. N. Dalal and M. M. Vaidya (2011). “Loss of keratins 8 and 18 leads to alterations in 6 4-integrin-mediated signalling and decreased neoplastic progression in an oral-tumour-derived cell line.” Journal of Cell Science, **124**, 2096-2111.

List of publications

1 Atul Deshmukh*, **S.P.Singh***, Pankaj Chaturvedi, C. Murali Krishna, "Raman spectroscopy of normal oral buccal mucosa tissues: A study on intact and incised biopsies" J Biomed Opt16 (12), 2011 (*Equal contribution). (DOI: 10.1117/1.3659680.) [Atul performed maxillofacial surgical procedure on biopsies, S.P.Singh recorded Raman spectra and analyzed the data]

2 **S. P. Singh**, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna,"Raman spectroscopy in head and neck cancers: towards oncological applications" J Cancer Res Ther. S2, 8, 2012. (DOI: 10.4103/0973-1482.92227)

3 **S. P. Singh**, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna, "In vivo Raman spectroscopy methods for oral cancers diagnosis" Proc. of SPIEVol. 8219, 82190K, 2012. (DOI: 10.1117/12.905453)

4 **S. P. Singh**, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna, "In vivo Raman spectroscopic identification of premalignant lesions in oral buccal mucosa" J Biomed Opt, 17(10), 105002,2012. (DOI: 10.1117/1.JBO.17.10.105002)

5 **S. P. Singh**, Hunain Alam, Milind M. Vaidya, C. Murali Krishna,"Raman spectroscopic study of keratin 8 knockdown oral squamous cell carcinoma cells" Proc. of SPIEVol. 8225, 82251C, 2012. (DOI:10.1117/12.905450)

6 **S. P. Singh**, Aditi Sahu, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna, "In vivo Raman spectroscopy of oral buccal mucosa: A study on malignancy associated changes (MAC)/cancer field effects (CFE)" Analyst,138, 4175-4182, 2013 (DOI: 10.1039/C3AN36761D)

Surya Pratap Singh

Doctoral Committee

S. No.	Name	Designation	Signature	Date
1.	Dr. S. M. Zingde	Chairman		6/5/13
2.	Dr. C. Murali Krishna	Convener/Guide		6/5/13
3.	Dr. G. Krishnamoorthy	Member		7/5/13
4.	Dr. G.B.Maru	Member		06.05.2013
5.	Dr. Pankaj Chaturvedi	Member		6/5/13

Dr. S. V. Chiplunkar
(Chairperson, Academic
and training programmes)

Dr. S. V. Chiplunkar
(Deputy Director)
CRI, ACTREC

Dr. R. Sarin
(Director)
ACTREC

Dr. K.S. Sharma
Director Academics,
Tata Memorial Centre

List of Figures

List of Figures

List of Figures

List of Tables

List of Tables

Chapter 1

Introduction

1.I Oral Cancers: An overview

Cancer is a group of disease characterized by uncontrolled growth, and caused by external (tobacco, chemicals, radiation, and infectious organisms) as well as internal factors (inherited mutations, hormones, immune conditions etc.) [1,2]. It is one of the most common causes of morbidity and mortality today, with more than 10 million new cases and more than 6 million deaths each year worldwide. More than 20 million people around the world live with a diagnosis of cancer, and more than half of all cancer cases occur in the developing countries. Cancer is responsible for about 20 and 10 % of all deaths in high and low income countries, respectively. It is projected that by 2020 there will be 15 million new cancer cases and 10 million cancer deaths each year [3]. As per the population based registries in India, the age adjusted incidence rates vary from 44 to 122 per 100,000 males and 52 to 128 per 100,000 females. The Mumbai cancer registry has reported the age-adjusted mortality rate (AAMR) of 62 per 100,000 males and 58 per 100,000 females [4].

1.I.1 Epidemiology of oral cancers: Oral cancer is a major cause of cancer morbidity and mortality among Indian males. Squamous cell carcinoma of the oral cavity ranks as the 15th most common cancer in the world and 10th most frequent in males [1].The Indian subcontinent accounts for one-third of the world burden. Compared to U.S. population, where oral cavity cancer represents only about 3% of all malignancies, it accounts for over 30% of all cancers in India. The age-adjusted rates of oral cancer varies from over 20 in India and 10 in United States to less than 2 per 100,000 in the Middle East [1,5]. India tops in the prevalence of oral cancer in the world and remains the most common cancer among males and the third most common cancer after cervical and breast cancer amongst females [5,6].

1.I.2 <u>Etiology of oral cancers:</u> The major risk factors for oral cancer include tobacco, alcohol consumption and viral infections. The association of tobacco chewing and smoking as well as alcohol consumption has been very well established. A strong association of cigarette smoking and alcohol consumption and oral cancer has been reported in the western countries, whereas smoking and chewing tobacco has been associated with oral cancers in South Asia, including India. It is estimated that in India, about 65% of men and 33% of women use some form of tobacco [7].

1.I.2.a *Smokeless Tobacco:* Smokeless tobacco is consumed in two forms: chewing tobacco which is available in the form of loose-leaf-cut or shredded and snuff which is either moist or dry. Moist snuff is used for applying, dipping or sucking and may be available as commercial packed products. Smokeless tobacco products are used either alone (chewed or snuff) or in various combinations with areca nut, betel leaves or lime. High incidence of oral cancer in the Indian sub-continent has been attributed to heavy tobacco-chewing habits. In India, tobacco is mostly consumed with areca nut, lime or in betel quid; practice of using tobacco as dentifrice is also prevalent. The estimated risk for developing oral cancer in tobacco chewers is about two to four times higher as compared to non-chewers [8,9]. The lesions reported due to frequent use of various forms of smokeless tobacco are: leukoplakia, erythroplakia, snuff dippers lesion, tobacco and lime dippers lesions, verrucose hyperplasia and submucosal deposits. These lesions are considered to be precursors of oral squamous cell carcinoma and are generally seen to develop at the site of product application. About 80% cases of oral cancers reported from the Indian subcontinent involve buccal mucosa, particularly the lower buccal sulcus or posterior buccal mucosa where tobacco containing quid is placed [10]. It is estimated that about 50% of the oral cancer cases in South East Asia are attributed to the use of smokeless tobacco resulting in 50,000 new cases from this region; out of

which 36,000 are from India alone. The use of smokeless tobacco in the form of 'nass' (powdered tobacco with lime and oil) is also prevalent in Iran and former Soviet Central Asian Republics [11].

1.I.2.b *Smoking:* Smoking is practiced by the use of cigarettes or bidis. The principle impact of tobacco smoking is seen in higher incidence of cancers of the lungs, larynx, oesophagus, pancreas and bladder. Bidi smoking is associated with cancer of oropharynx as well as larynx. Parts of the Indian subcontinent and Latin America have incidence of palate cancer because of reverse smoking. It was estimated that habit of smoking is prevalent in 25% of the Indian population out of which about 47-51% of males and 52-95% of females were bidi smokers [12,13]. In ten case control studies done by Rahman *et al.* in 2005, it was estimated that about 24.1 to 31.4 % of oral cancers can be attributed to the bidi smoking habit [14]. Tobacco-related cancers account for nearly 50% of all cancers among men and 25% of all cancers among women. There are predictions that there will be 7-fold increase in tobacco-related cancer morbidity by 2025. Further there may be an overall increase by 20% of cancer deaths simply related to tobacco use by the year 2020 [4].

1.I.2.c *Alcohol:* There is a strong association between high alcohol consumption and oral cancer. Epidemiological studies carried out in India and abroad have shown that increased alcohol consumption is causally associated with cancers at various sites, mainly oral cavity, pharynx, larynx, and oesophagus [4]. Many prospective and case–control studies show a 2–3-fold increased risk for cancer of the oral cavity, pharynx, larynx and oesophagus in people who consume 50 g of alcohol a day (equal to approximately half a bottle of wine), compared to non-drinkers [15]. These effects were dose-dependent and smoking might have a synergistic effect. A carefully designed French study demonstrated that alcohol consumption of more than 80 g a day (approximately 0.7 liter of wine) is associated with a relative risk (RR) of 18 for development of oesophageal carcinoma, which translates into an 18-fold higher cancer risk in those exposed to this amount of

alcohol when compared to non-drinkers, whereas smoking more than 20 cigarettes a day resulted in an increased RR of only five. However, both factors act synergistically, resulting in an increased RR of 44 [15].

1.I.2.d *Viral infection:* Oral squamous cell carcinoma has been associated with Epstein Bar Virus (EBV) and Human Papilloma Viruses (HPV) although their putative role is controversial. EBV has been detected in oral hairy leukoplakia (OHL) in HIV infected immunocompromised as well as non immunocompromised patients [16,17]. However, OHL is benign in nature and has been rarely observed to progress to malignancy. A study using PCR-Southern analysis by D'Costa *et al.* has shown that 25% of oral squamous cell carcinoma biopsies were positive for EBV as compared to the normal mucosa [18]. Several studies done on various ethnic groups have also shown the presence of EBV in oral cancer lesions and tumour biopsy samples albeit with varying proportions, which might be a reflection of variations associated with differences in clinico-pathology of tumors of patients with differing lifestyles [19,20]. High risk HPV-16 has been seen to be predominantly present in OSCC, however etiological role of other viruses like HPV cannot be ruled out. The differences in the detection rate of HPVs in oral cancers are due to high amount of heterogeneity among different studies, attributed to variations in prevalence and multiple detection techniques used. The precise role of HPV in oral carcinogenesis is not very well known but oncogenic early proteins E5, E6 and E7 of high risk HPV are reported to promote mitotic defects, aneuploidy and chromosomal instability by their propensity to integrate in the host DNA at fragile sites [21,22].

1.I.3 <u>Premalignant lesions of the oral cavity:</u> Oral carcinoma is a multi-step disease and it is accepted that most of the oral tumors are often preceded by visible changes in the oral mucosa accompanied with certain high risk lesions termed as 'precancerous lesions'. A precancerous lesion is defined as any morphological alteration in a tissue which makes it more susceptible to malignant transformation as compared to its adjacent apparently normal tissue [23]. The predominant types of oral pre-cancers include: leukoplakia, erythroplakia, submucous fibrosis and lichen planus, Figure 1.1.

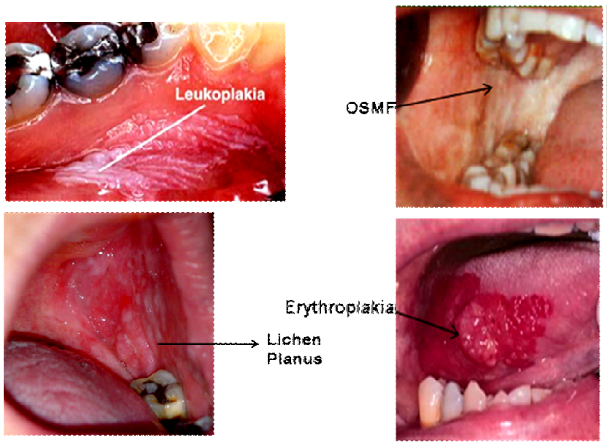


Figure 1.1: Different premalignant lesions of the oral cavity

1.I.3.1 *Leukoplakia*: Leukoplakia, first termed by Schwimmer in 1877 is defined as any white oral lesion that cannot be characterized clinically or pathologically [24]. Some leukoplakias develop from papillary surface and are called verrucous or verruciform leukoplakia, while others develop from surface irregularities and are called granular or nodular leukoplakia. Typical transformation rate of leukoplakia ranges from 0.16 % to 6% which rises to 14% when dysplasia is present [25]. It is estimated that patients with leukoplakia carry five-fold higher risk as compared to controls; it is therefore advisable to rule out any suspicion by a conventional biopsy. The clinico-pathological behavior of leukoplakia varies across geographic and ethnic populations due to differences in the dietary intake and tobacco habits. In Western population, leukoplakia occurs

predominantly on the tongue whereas in India buccal mucosa is the more common site in tobacco chewers (26). More often the leukoplakia lesion is seen to develop at the site of placement of smokeless tobacco product and the severity of lesion is also seen to correlate with duration and amount of tobacco used [10]. Clinically, leukoplakias are treated with anti-inflammatory and antimycotic agents along with complete withdrawal of etiological agents for a few weeks. If the disease does not subside, then surgical excision, preferably with laser surgery is performed to remove the persistent leukoplakia.

1.I.3.2 ***Erythroplakia:*** Erythroplakia is an oral precancerous stage which is defined as a red fiery patch that cannot be characterized either clinically or pathologically [27]. This definition however excludes inflammation induced red patches in the oral cavity. The sites affected by erythroplakia are majorly soft palate, floor of the mouth and buccal mucosa. The potential of malignant transformation of erythroplakia is very high varying from 14 to 50%, which is higher than other precancerous lesions [28]. Some erythroplakias are smooth whereas others are nodular or glanular in nature. Some lesions comprise of interspersed red and white patches which are called as erythroleukoplakia. It is estimated to be prevalent in about 0.02 to 0.83% populations of South and South East Asia. Tobacco and alcohol are considered to be the major risk factors associated with the development of erythroplakia. However, the possible role of *Candida albicans* and HPV as an etiologic cofactor in erythroplakia is currently being investigated. Surgical removal of the lesion either with cold knife or laser is the usual recommended treatment modality.

1.I.3.3 ***Submucous fibrosis:*** Oral Submucous Fibrosis (SMF) is a precancerous condition characterized by burning sensation, blanching and stiffening of oral mucosa, rigidity of lip, palate and tongue leading to difficulty in mastication and phonation [29]. This condition is prevalent mostly in the South and South East Asian countries and among the Asian immigrants to UK and

Africa. The etiological agents associated with the incidence of SMF are chewing areca nut and betel quid, and micronutrient deficiency of iron, zinc and essential vitamins. The rate of malignant transformation of submucous fibrosis was found to be 7.6% in a long term study carried out for 17 years [30].

1.I.3.4 ***Oral Lichen Planus:*** It presents as small white raised lesion or plaques that might resemble leukoplakia. The common sites affected by the lesions are buccal mucosa, dorsal tongue, gingival, labial mucosa and vermillion of the lip. Studies have shown that the risk of lichen planus to undergo malignant transformation varies from 0.4 to 5% over a period of observation from 0.5 to 20 years. It is caused when autotoxic T cells induce apoptosis of epithelial cells leading to chronic inflammation. The treatment mainly comprises of anti-inflammatory agents, usually topical corticosteroids [31].

1.II Current Screening/Diagnosis methods

Screening for a disease can be defined as: 'the application of a test or tests to people who are apparently free from the disease in question in order to sort out those who probably have the disease from those who probably do not'[32]. Well-known examples of screening methods are Pap test for cervical cancer and mammography for breast cancers. In contrast to screening, diagnostic test is applied to a patient who has abnormal signs or symptoms in order to establish the presence of a disease and subject the patient to treatment. Among the screening tests or diagnostic aids presently available for oral cancer, some have been used and studied for many years while others have recently become commercially available. These mainly include Conventional Oral Examination (COE), oral cytology, staining with toluidine blue or tolonium chloride and light based detection system.

1.II.1 ***Conventional Oral Examination (COE):*** A conventional oral examination (COE), using normal (incandescent) light, has long been the standard method for oral cancer screening. A numbers of reports have suggested that COE may have limited value as a method for detecting pre-cancerous or early cancerous lesions. Conversely, other studies have reported a relatively high degree of sensitivity, specificity and positive predictive value of COE. Study carried out by Sankaranarayanan *et al.* over nine years involving over 130,000 individuals for the first time reported significant increase in survival rates among males with high-risk habits, such as tobacco use [33]. Although no increase in survival was observed for the overall population, this was the first clear evidence to support the efficacy of an oral cancer screening program, as measured by reduced mortality. It prompted others to call for the broader use of oral screening measures throughout the world [34]. Although COE may be effective as a screening test, there are still many problems with this approach. First, approximately 5-15% of the general population has oral mucosal abnormalities and vast majority of these lesions are benign in nature [35-37]. Second, only a small percentage of leukoplakias are progressive or become malignant and COE cannot discriminate between these lesions and their non-progressive counterparts. Furthermore, while COE may detect a number of clinical lesions and a small percentage of those may exhibit histological features of premalignancy, some precancerous lesions may be lurking within mucosa that appears clinically normal, cannot be detected by COE alone. Recent report from Thomson *et al.* has shown that 9/26 consecutive patients (36%) with a newly diagnosed HNSCC had histologic evidence of dysplasia or microinvasive cancer in a biopsy from clinically normal mucosa from the corresponding, contralateral anatomic site [38]. Therefore, while COE may be useful in discoverring some oral lesions, it does not identify all potentially premalignant lesions, nor does it accurately detect the small proportion of biologically relevant lesions that are likely to progress to cancer.

1.II.2 ***Oral cytology:*** Oral exfoliative cytology is another commonly used adjunct to oral visual inspection. The mechanism of cytology, regardless of its application to cervical, bladder or oral mucosal lining, is based upon the fact that dysplastic and cancerous cells tend to have fewer and weaker connections to each other and to their neighboring normal cells in the surrounding tissue. Dysplastic and cancerous cells therefore, tend to "slough off" or exfoliate preferentially and can easily be collected from the surface of the lesion. These abnormalities can be identified under a microscope by a trained pathologist. However, high number of false negative results ranging from 30% for cancerous to 60% for dysplastic lesions has limited applicability of this approach for routine usage. The poor results of manual cytology are primarily attributed to the fact that cytology instruments do not sample the deepest layers of the oral lesion. This is essential, since unlike cervical cancer, the deepest layer of the lesion, the basal cell layer, is often the only layer that contains abnormal cells. Furthermore, since the sensitivity of cytology is often dependent on a tedious and manual visual search for a potentially rare abnormality on the microscope slide, the precancerous or cancerous cells collected on the slide may not be detected by the laboratory pathologist.

To overcome the above mentioned limitations of traditional oral cytology, oral brush biopsy was introduced in the year 1999. This method utilizes an improved brush to obtain a complete trans-epithelial biopsy specimen with cellular representation from each of the three layers (basal, intermediate and superficial) of the lesion. The oral brush biopsy does not require topical or local anesthetic and causes minimal bleeding (pin-point bleeding) and pain. Examples of well-known applications of brush biopsies include fiberoptic bronchoscopy (bronchial), ureteral retrograde brush biopsy (renal or ureter tissue), cholangiography (bile duct structure), pancreatic ductal brush biopsies and others, including endometrial, nasopharynx, and GI tract applications (rectal, gastric,

esophageal, colon). Use of brush biopsy technology for oral cancer diagnosis was introduced in 2000 under the commercial name OralCDx (CDx Laboratories, Suffren, NY). The improved accuracy of the OralCDx brush biopsy is also due to advanced, highly specialized, computer-assisted neural network-based analysis. Any abnormality in cell morphology, altered cell size, degree of keratinization, nuclear staining intensity, and size are characterized and analyzed by the software. After verification from a pathologist, specimens are classified as negative (no epithelial abnormality), atypical (abnormal epithelial cells, but of uncertain diagnostic significance), and positive. A standard incisional/excisional scalpel biopsy is performed on cases with atypical or positive cells. The oral brush biopsy has been criticized for adding time and cost to the diagnosis of oral lesions without additional benefit to the patient [39-41]. Because the brush biopsy detects only cellular atypia, positive oral brush biopsy results must be confirmed with a scalpel biopsy for definitive diagnosis. This results in requirement for two procedures, rather than one to establish a diagnosis. The need to perform two procedures may significantly delay diagnosis. Overall, it is a method of identifying unsuspected oral cancers found during a visual examination, at early and curable stages.

1.II.3 ***Toluidine blue staining:*** Toluidine blue (TB) also called Tolonium chloride has been used for more than 40 years as an aid in detection of mucosal abnormalities of the cervix and the oral cavity. TB is a metachromatic, acidophilic dye that binds preferentially to tissues undergoing rapid cell division (inflammatory, regenerative and neoplastic tissue) or to sites of DNA change. The binding results in the staining of abnormal tissue in contrast to unstained adjacent normal mucosa. Over the years, mixed results have been a persistent feature of studies with TB and thus several authors have expressed reservations with the technique. A recent report has shown that TB staining might be useful in indentifying clinically evident lesions having potential for

transformation [42]. Results suggest that TB can preferentially stain lesions with higher degrees of dysplasia exhibiting high-risk clinical features. Most importantly it can predict risk and outcome of visible oral lesions with little to no microscopic evidence of dysplasia. However, these studies have not been extended to determine whether TB screening can help identify and predict the risk of progression of lesions that cannot be visualized with COE. Overall, TB appears to be useful in detecting carcinomas but is positive in only ~50% of lesions with dysplasia. In addition, it frequently stains common, benign conditions such as non-specific ulcers. The high rate of false positive stains and the low specificity in staining dysplasia are some of the well known limitations of the technique. Additional point of debate is about the intensity of staining and whether or not pale blue staining should be regarded as positive result. A recent study suggests that only dark royal blue staining should be regarded as positive [43-47]. All carcinomas stained dark royal blue and histology showed nuclear staining. Benign lesions had no nuclear staining and were more often pale blue in color. These limitations do not however preclude its usefulness as an adjunct to clinical examination and case-finding, even in primary care. In the experienced hands, TB staining may be useful in the evaluation of oral lesions and as an adjunct in the surveillance of high-risk individuals, such as patients at risk for a second primary lesion.

1.II.4 ***Comment on the limitations of the current diagnostic methods:*** Histological risk stratification, currently the gold standard of oral cancer diagnosis requires biopsy and microscopic examination by a pathologist. However, following are the major limitations of the method:

- Removal of tissue or biopsy is an inherently invasive procedure and carries certain risk of complications in some cases when involvement of vital anatomy is encountered. It also

limits the amount of tissue available for study. Furthermore, once removed, the tissue can undergo biochemical changes which can lead to artifacts.

- Sampling errors in collecting or interpreting biopsy can be significant. In many diseases, tissue involvement is not uniform and diseased tissue is 'invisible' to the physician taking the biopsies. In such cases only a tiny fraction of the suspected area is sampled, and the tissue obtained may not be a true representative of the diseased area. Additionally, histological findings are inherently prone to interpretation bias as expert pathologists evaluating the same tissue sample do not always come to the same conclusion
- Some of the early oral cancers are clinically indistinguishable from benign lesions. It is difficult to recognize subtle clinical changes in precancerous lesions or in a normal mucosa that are indicative of early neoplastic transformation. Additionally, difficulty in discriminating premalignant lesions from more common benign inflammatory conditions further limits its applicability.

Overall, there is no satisfactory mechanism to adequately screen and detect oral premalignant changes and early malignancies; clinical risk stratification lacks accuracy, reproducibility and requires large experience on part of the clinician. Surveillance and biopsy of pre-cancers is a mammoth task especially in populous countries like India. For example incidence of leukoplakia itself is up to 1% of general population. As mentioned earlier, oral cancers in the Indian subcontinent are majorly associated with tobacco chewing and smoking habits. It is therefore conceivable that the primary prevention of the disease would involve activities to avoid the use of tobacco and alcohol. Secondary prevention includes activities that are aimed to detect the disease in an early stage which would lead to better prognosis and lesser morbidity. In view of all aforementioned limitations of the conventional diagnostic methods, need of an alternate method of

diagnosis is warranted. Optical spectroscopic methods, due to their ability in providing real-time, objective, non-intrusive and online information are being projected as alternative or adjunct method of cancer diagnosis.

1.III Optical spectroscopy in oral cancer diagnosis

Spectroscopy is the study of interaction of electromagnetic radiation with matter. There are three aspects to a spectroscopic measurement: absorption, spontaneous emission and scattering. Diagnostic techniques based on spectroscopy have the ability to link the biochemical and morphological properties of tissues to individual patient care. If applied successfully, optical spectroscopy can represent an important step forward towards improving diagnostic and therapeutic medical applications. Studies have demonstrated that optical spectroscopy methods have the potential to fulfill the need for improved screening and diagnosis of cancers. Optical measurements provide quantitative information that can be analyzed instantaneously for an objective diagnosis even in the hands of a non-expert operator. Devices to make these measurements have become inexpensive, robust, and portable because of advances in computing, fiber optics, and semiconductor technology. Approaches based on reflectance, fluorescence and Fourier-transform infrared spectroscopy have shown potential for improved detection of oral cancers. In the following sections a brief introduction on these techniques and their potential applications in oral cancer diagnosis is presented.

1.III.1 ***Tissue Reflectance:*** Reflectance spectroscopy is a spectroscopic technique which measures the difference in reflectance of two beams of light incident on a surface with different linear polarizations. The term 'Reflectance' generally refers to the fraction of incident electromagnetic power reflected at the interface. Diagnosis based on tissue reflectance had been used for long time to examine cervix abnormalities. Recently, this technology has been adapted for

use in the oral cavity and is being currently marketed under commercial names ViziLite Plus and MicroLux DL [48-50]. These products are intended to enhance the identification of oral mucosal abnormalities. With both systems, the patient must first rinse with 1% acetic acid solution prior to visual examination using a blue-white light source. ViziLite Plus uses a disposable light packet, while the MicroLux unit offers a reusable, battery-powered light source. The 1% acetic acid wash helps remove surface debris and causes mild dehydration which in turn may increase the visibility of epithelial cell nuclei. Under blue-white illumination, normal epithelium appears bluish while abnormal epithelium appears distinctly white (aceto-white). ViziLite Plus also provides a tolonium chloride solution which is intended to aid in marking of any acetowhite lesion for subsequent biopsy, once the light source is removed. Evidences in support of reflective tissue systems as adjunct in detecting premalignant lesions are quite sparse and contradictory. Recently a multi-centre study by Epstein *et al.* involving 134 patients has shown that ViziLite examination did not significantly improve lesion detection when compared to COE [51]. Another clinical survey by Huber *et al.* on 150 subjects demonstrated that diagnosis based on ViziLite has better prognosis than COE [52]. Although one lesion was detected with ViziLite (that was not observed using incandescent light alone,) the false positive finding of the benign suggests that while the sensitivity of the technique may be relatively high, its specificity and PPV are probably low. In summary, published studies to date suffer from numerous experimental design issues, especially the critical comparison to the diagnostic gold standard (scalpel biopsy) in all cases. Furthermore, based upon the current suggested usage of these devices, it is unclear what added benefit would they provide to a practicing clinician. If a clinician is able to clinically identify a lesion, they are obligated to obtain a definitive diagnosis in order to direct the treatment of the patient's lesion [53,54]. Thus, subjective improvement of one's ability to see a lesion would provide minimal diagnostic advantage to the

practicing dentist or the patient, unless the test can also discriminate indolent lesions from those that are more biologically worrisome. On the other hand, some reports hint that this technique may help identify lesions that cannot be seen with incandescent light. Well-controlled clinical trials are needed that specifically investigate the ability of these devices to detect precancerous lesions that are not identified by COE alone. If such discrimination can be confirmed, it would support the use of this technology as a true screening device.

1.III.2 ***Fluorescence spectroscopy:*** Optical spectroscopic methods probes the energy levels of a molecule. The energy level of a molecule is defined as its characteristic state, which is related to the molecular structure of a molecule and to the energetics and dynamics of any chemical processes that the molecule may undergo. The ground state of a molecule is defined as the state of lowest energy while the states of higher energy are called excited states (Figure 1.2). When a molecule is illuminated at an excitation wavelength lying within the absorption spectrum of that molecule, it absorbs the energy and gets activated from its ground state (S_0) to an excited singlet state (S_1), with an electron in the same spin as its ground state. The molecule can then relax back from the excited state to the ground state by generating energy either non-radiatively or radiatively, depending upon the local environment. In a non-radiative transition, relaxation occurs by thermal generation (dashed arrows). In a radiative transition, relaxation occurs *via* fluorescence at specific emission wavelengths (solid arrow). Fluorescence generation occurs in three steps: thermal equilibrium is achieved rapidly as the electron makes a non-radiative transition to the lowest vibrational level of the first excited state; the electron then makes a radiative transition to a vibrational level of the ground state; and finally, a non-radiative transition to the lowest vibrational level of the ground state [55]. A fluorescence emission spectrum represents the fluorescence

intensity measured over a range of emission wavelengths at a fixed excitation wavelength and can provide biochemical information.

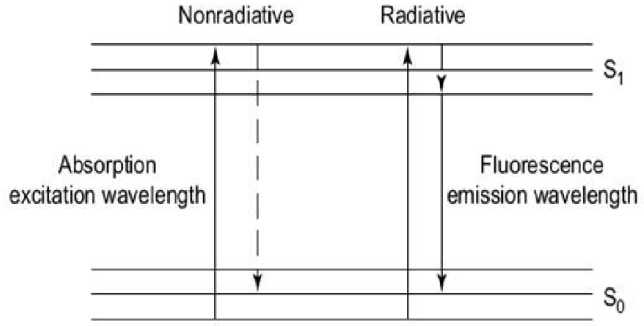


Figure 1.2: Energy level diagram illustrating the phenomena of absorption and fluorescence

Application of these methods to oral cancer diagnosis started with the use of exogenous light-sensitive drugs ('photo sensitizer'). In this method, first the photo sensitizers are introduced in tumors, either through systemic or topical application or by administration of a precursor, such as protoporphyrin IX (PpIX). The photo sensitizer produces singlet oxygen upon excitation with light of a certain wavelength, which damages vital cell organelles, inducing death of cells in the direct environment. Since some of the sensitizers were believed to accumulate in malignant tissues, they could possibly serve as markers as well. Some promising results have been obtained in oral cancer diagnosis using these methods [56-58]. However, certain limitations in term of specificity of the photo sensitizers, time lag between and after application have rendered applicability of this approach for use in regular screenings of high-risk patient groups. Furthermore, the application of photo sensitizers leaves the patients with a temporary sensitivity to light, which may negatively affects his daily life.

In the late 1970s, oral cancer diagnosis based on auto-fluorescence (also called natural or endogenous fluorescence) was started. Naturally occurring fluorophores such as collagen, elastin, keratin and NADH after excitation with a suitable wavelength produces auto-fluorescence. The

fluorophores can be located in the tissue matrix or in cells. Presence of diseases can lead to changes in blood concentration, nuclear size distribution, collagen content or epithelial thickness, which in turn can alter the concentration, scattering and absorption efficiencies of the natural fluorophores. For example, the epithelial layer shields the strongly fluoresceing collagen layer and therefore the recorded fluorescence signal will be lower in case of hyperplasia. Conversely, excessive keratin production by lesions may result in an increase in auto-fluorescence intensity. Cell metabolism may increase with malignant changes, which changes the balance between the fluorescent NADH (increase) and non-fluorescent NAD+ (decrease) [59]. In case of oral cavity, the first report came from Harris and Werkhaven *et al.,* when they reported difference in auto-fluorescence based on porphyrin band between healthy and tumor mucosa [60]. These findings were established later and it was proposed that fluorescence was probably produced by microorganisms living on ulcerating or necrotic surfaces [61,62]. In a study with 130 patients, autofluorescence patterns were acquired with a similar set-up and 91.1% sensitivity and 84.3% specificity in distinguishing malignant from benign lesions was reported [63]. Autofluorescence spectroscopy is the other facet of the same technique where, light source, usually in the near-UV to visible wavelength range is used to excite the tissue through a fiber and the fluorescence produced is recorded by a spectrograph, while the reflected light is filtered out. The recorded spectra can be analyzed by multivariate techniques like Principal Components Analysis (PCA), wavelength ratios and artificial neural networks. Studies have been carried out to distinguish oral lesions (cancerous, dysplastic or benign) from healthy oral mucosa [64,65]. Chen *et al.* using 300 nm for excitation have reported positive predictive values (PPV) of 94% and 93%, for abnormal and normal tissues, respectively [66,67]. Gillenwater *et al.* have recorded *in vivo* autofluorescence spectra from oral mucosa of 8 healthy volunteers and 15 patients with premalignant or malignant lesions using 337, 365 and 410 nm excitations. Decreased

intensity in the blue spectral regions, and increased fluorescence around 635 nm (porphyrin) was observed. Based on the ratio between these values, sensitivity of 82% and a specificity of 100% were reported [68]. A very recent study by Shaizu *et al.* involving habitual tobacco users and premalignant lesions have shown that autofluorescence spectroscopy in combination with linear discriminant analysis can be used for identifying oral cavity disorders caused by long-term tobacco habits. Their findings suggest that lower collagen level and increased level of redox value can be considered as prognostic marker for oral cancer risk [69]. Large scale validations of these findings are required before translation for routine usage.

1.III.3 ***Fourier-transform infrared spectroscopy:*** Fourier-transform infrared spectroscopy (FTIR) is an absorption based optical spectroscopy technique and is being projected as a powerful tool to study the structure and functions of cellular components in tissues. In 1990, Rigas *et al.,* had shown spectral differences between normal and malignant tissues with FTIR and suggested its potential application to a variety of biological and medical problems [70]. In 1995, Jackson *et al.,* presenting infrared spectra of human central nervous tissue and human breast carcinoma, discussed implications of the presence of collagen in tissues for diagnosis of the disease state, with particular reference to cancer [71]. In a recent review, Bellisola *et al.* (2012) have summarized its biomedical applications with a focus on cancer diagnosis [72]. Both fiber-optic and FTIR micro-spectroscopic studies in oral cancers are reported. Wu *et al.* have shown that on the basis of lipid and protein content, normal and tumor oral tissues can be discriminated. The spectra of normal tissues were rich in C=O and C-H bands while those of tumor tissues were rich in N-H and O-H stretching bands with low C=O band. Authors also suggested that the most common difference between normal and tumor sections was C=O band at 1745 cm^{-1}. They validated the results by comparing spectral findings with stained histological sections [73]. In another study

Fukuyama *et al.* have shown on 10 normal sub-gingival tissues (NST) and 15 oral squamous cell carcinoma (OSCC) tissues that normal spectra is strongly influenced by the presence of collagen. Normal and tumor spectra show maximum differences in bands at 1431 1482 cm^{-1}, 1183 and 1274 cm^{-1}. They also suggested that spectral influence by keratin, which exists in the ectodermal cells may be present, indicated by disappearance of 1368 cm^{-1} band in OSCC [74]. Study by Schultz *et al.* to assess changes in biochemistry of well and poorly differentiated oral/oropharyngeal squamous cell carcinoma (SCC) by infrared microspectroscopy has demonstrated that DNA and keratin can provide distinct differences between normal and SCC biopsies. Bivariate histogram analysis of cell components (*e.g.* DNA and keratin) indicated that cancer cells produce relatively homogeneous and clearly abnormal cell biochemistry, whereas differentiated epithelial cells present a very heterogeneous distribution of cellular components. Using these features, tissues containing abnormal or cancer cells can easily be distinguished from normal epithelial structures. The abnormal distribution of keratin in poorly differentiated SCC and in keratin pearls (present only in well-differentiated SCC) can also help in understanding the process of malignant tissue transformation [75]. Although there are reports regarding use of FTIR for the analysis of human tissues and cells, the application to the field of clinical medicine is still poor. This could be due the fact that these methodologies are less suitable for *in vivo* and *in situ* studies as water, the major component of biological tissues, is highly absorptive in the mid-IR range. This problem can be minimized to some extent, by subjecting samples to mild vacuum conditions for dehydration before spectral acquisition. Limitations in terms of sample thickness is another hindrance which can affect applicability of IR-spectroscopy. Attenuated total reflectance (ATR) based methodologies could be useful in circumventing this difficulty. In this approach, the spectra are recorded in reflectance mode and with minimum or no sample preparation. Further, with the advent of flexible and

inexpensive infrared fibers, few studies have demonstrated the efficacy of fiberoptic evanescent wave spectroscopy based infrared spectroscopy (FTIR-FEWS) in diagnosis of skin cancers and melanomas [76]. These systems are based on ATR elements, where samples are placed in contact IR transmitting prisms or flat waveguides that serve as ATR elements. The evanescent waves that escape from ATR elements are absorbed in samples of wavelength that is specific to the sample. Therefore by measuring the transmission of an ATR element in contact with a sample, IR absorptions can be recorded. The advantage of this method is that sample thickness has no influence over spectral measurements and can be used for *in vivo* or *in situ* measurements.

1.IV Raman Spectroscopy: An overview

Raman Effect, named after its discoverer, the Indian physicist Sir C. V. Raman, who first observed it in 1928, is based on inelastic scattering of light. Raman won the Nobel Prize in Physics in 1930 for this discovery, accomplished using filtered sunlight as a monochromatic source of photons, a colored filter as a monochromator, and the human eye as detector [77,78]. The Raman effect arises when a photon, incident on a molecule, interacts with the electric dipole of the molecule. In classical terms, this interaction can be viewed as a perturbation of the molecule's electric field. Most of the photons are elastically scattered *i.e.* with same energy as the incident photons, the phenomenon called 'Rayleigh scattering'. However, a small fraction of photons (approximately 1 in 10^7) are scattered at optical frequencies different from, and usually lower than, the frequency of the incident photons. The process leading to this inelastic scatter is termed as Raman Effect. The virtual state description of scattering is shown in Figure 1.4. Numerically, energy difference between the initial and final vibrational levels, v , or Raman shift in wavenumber (cm^{-1}) can be calculated using the following equation.

$$\nu = 1/\lambda_{\text{incident}} - 1/\lambda_{\text{scattered}}$$

Where, $\lambda_{incident}$ and $\lambda_{scattered}$ are the wavelengths (in cm) of the incident and Raman scattered photons, respectively. In a typical vibrational Raman spectroscopy, the differences in energy between the incident photon and the Raman scattered photon is equal to the energy of a vibration of the scattering molecule. A plot of intensity of scattered light against energy difference is called as Raman spectrum. The vibrational energy is ultimately dissipated as heat. Because of the low intensity of Raman scattering, the heat dissipation does not lead to a measurable temperature rise in the system. At room temperature, as per the Boltzmann distribution law, the thermal population of excited vibrational states is low, although not zero. Therefore, the initial state is the ground state, and the scattered photon will have lower energy (longer wavelength) than the exciting photon. This is called Stokes scattering and is often measured in Raman spectroscopy, Figure 1.3. However, small fractions of the molecules remain in vibrationally excited states. Raman scattering from these molecules leaves them in the ground state and scattered photons appear at higher energy (shorter wavelength), Figure 1.3. This is called anti-Stokes-shifted Raman spectrum and it is weaker than the Stokes-shifted spectrum. The ratio of anti-Stokes to Stokes intensity at any vibrational frequency is a measure of temperature. The anti-Stokes spectrum is used when the Stokes spectrum is not directly observable, for example because of poor detector response or spectrograph efficiency.

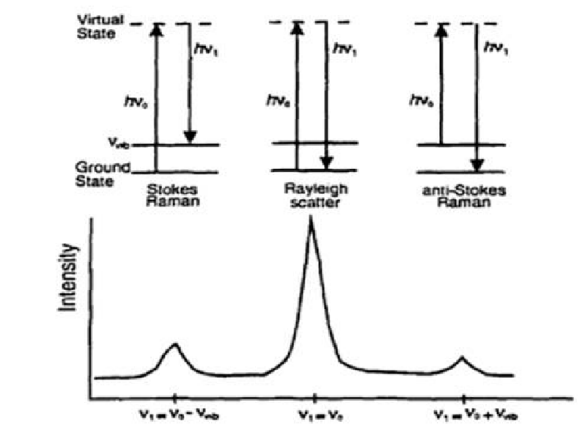


Figure 1.3: Energy level diagrams for Stokes and Anti-Stokes Raman scattering

Although Raman spectroscopy and conventional Infra Red (IR) spectroscopy measure the vibrational energies of molecules, these methods involve different photo-physical processes. For a vibrational mode to be IR active, dipole moment of the molecule must change. Therefore symmetric stretching is normally inactive in IR spectroscopy. In Raman Spectroscopy, a molecule is considered as Raman active, if there is a change in polarizability of the molecule. It is a measure of the ease with which an electron cloud around the molecule can be disturbed by the external electric field. Raman scattering intensity is proportional to the square of the induced dipole moment*, i.e.,* to the square of polarizability derivative. Molecules with little or no symmetry modes are likely to be active in both infrared and Raman though, in general, bands that appear strong in Raman are weak in IR, and *vice-versa* [79].

As mentioned earlier, Raman scattered photon is generated when incident photon cause changes in the vibrational state of a molecule. Since this frequency shift is unique to specific molecular vibrations of the molecule, qualitative analysis became feasible; either by direct comparison of the spectra of known and unknown materials recorded consecutively or by comparison of the spectra of unknown compounds with catalogs of reference spectra. Thus

identification of 'chemical moieties' provides 'molecular fingerprint' of the sample. One of the major drawbacks of Raman spectroscopy is its inherent low sensitivity as only 1 among 10 million of incident photons is Raman photon. Hence, this technique demands sophisticated instrumentation *i.e.* powerful excitation source, high throughput spectrograph and sensitive detection systems. The advent of compact high power laser excitation sources, high throughput spectrographs and high-quantum efficiency, low noise CCD detectors resulted in a flurry of research activities in Raman spectroscopy. Apart from advances in instrumentation, there have been innovative approaches for enhancement of Raman signals - surface enhanced Raman spectroscopy (SERS): Attachment of gold or silver colloid lead to thousand to 10 million fold increase in Raman signal intensity, enabling single molecule level detection [80-82]. Resonance Raman spectroscopy (RRS): Marked enhancement of Raman signal when excitation wavelength matches or is close to transition state [83]. Confocal Raman microspectroscopy: Microscope coupled approach with high spatial resolution ~1 μm [84-86]. Coherent anti-Stokes Raman scattering (CARS): a pump and probe process and orders of magnitude stronger than spontaneous Raman emission [87,88].

1.IV.1 ***Instrumentation:*** A typical Raman system consists of four major components: an excitation source, filters, spectrograph and detector. A sample is illuminated with a laser beam in the ultraviolet (UV), visible or near-infrared (NIR) region. Scattered light is collected, filtered and sent through spectrograph to generate Raman spectrum. Schematic representation of a typical Raman instrument is shown in Figure 1.4. A brief introduction about individual components is provided in the following sections.

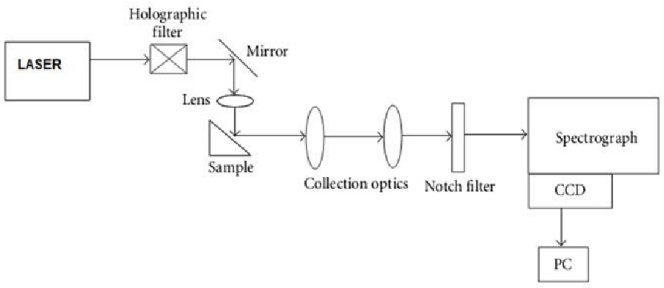


Figure 1.4: Schematic representation of a typical Raman spectrometer

1.IV.1.a *Excitation Source (LASER):* Typically a laser system is made up of lasing medium (atom, molecule or ion), a resonant cavity and an excitation source (electrical, radiation or others), Figure 1.5. The excitation source takes the atoms or ions of the lasing medium to a higher energy state. Transition from this state to a lower state gives the laser radiation, which is amplified by stimulated emission due to multiple (or single) passes through the resonant cavity.

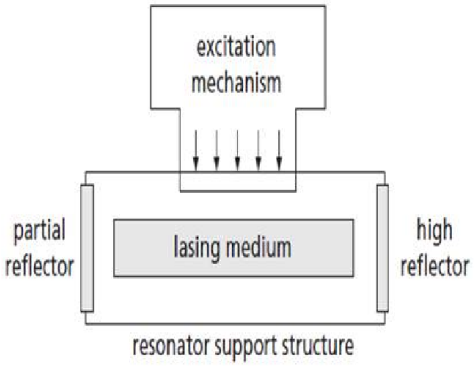


Figure 1.5: Schematic diagram of a typical laser system

Broadly, lasers can be classified into three categories- solid, liquid and gas lasers. Generally, active medium of a solid-state laser consists of a glass or crystalline "host" material to which "dopants" are added. The widely used semiconductor lasers are also included in this category. Solid state lasing media are typically optically pumped, using either a flash lamp or arc lamp, or by laser diodes. Diode-pumped solid-state lasers tend to be much more efficient, and have become much more common as the cost of high power semiconductor lasers has decreased. In the present study,

work was carried out using a continuous wave (CW) diode laser of 785 nm. Additional information about this laser has been provided in Chapter 2 under the methodology section. Liquid lasers are based on strongly absorbing organic dye molecules such as rhodamines or coumarins in an organic solvent. The dye is pumped optically by a flash lamp to produce population inversion followed by stimulated emission to produce a laser gain. Lasing begins when light is absorbed by a dye, exciting it from lowest singlet to high energy level, now the dye falls to slightly lower state within the same singlet band which serves as upper lasing level. A typical dye laser can operate over a wavelength range 30-40 nm. Gas lasers, on basis of lasing transitions can be categorized as neutral atom and ionic. Neutral atom lasers emit throughout UV, visible and IR ranges, while ion lasers emits mostly in UV through the near IR region.

1.IV.1.b ***Filters:*** These are used to filter out the intense Rayleigh scatter or background signals. The advent of holographic notch filters have eliminated the need for cumbersome, low throughput, multistage scanning spectrometers, and enabled rapid spectral data acquisition. Holographic notch filters can have a very steep-edged blocking bandwidth at the excitation wavelength, with contrast ratios greater than one million, while transmitting 90% of the light with a relatively flat baseline above and below the excitation wavelength. Holographic filters used in delivery path can block unwanted emission from the laser that would otherwise swamp the Raman spectrum, transmitting 85–95% of the excitation line. Various categories of holographic notch filters include:

1.IV.1.b.i ***Long-pass filters:*** A long-pass (LP) filter is an optical interference or coloured glass filter that attenuates shorter wavelengths and transmits (allows to pass) longer wavelengths over the active range of the target spectrum (ultraviolet, visible, or infrared). Longpass filters can

have a very sharp slope (referred to as edge filters) and are described by the cut-off wavelength at 50 percent of peak transmission.

1.IV.1.b.ii *Short-pass filters:* A short-pass (SP) filter is an optical interference or coloured glass filter that attenuates longer wavelengths and transmits (allows to pass) shorter wavelengths over the active range of the target spectrum (usually the ultraviolet and visible region).

1.IV.1.b.iii *Band-pass filters:* Band-pass filters only transmit a range of wavelength band, and block others. The width of such a filter is expressed in the wavelength range which it lets through and can be anything from less than Angstroms to a few hundred nanometers. Such a filter can be made by combining an LP- and an SP filter.

1.IV.1.c *Spectrograph:* The main function of the spectrograph is to disperse the light from an object into its component wavelengths. There are four essential components of the spectrograph: (1) an entrance slit (2) a collimating element which may be a lens or a mirror to obtain parallel rays when pass through one point of the entrance slit (3) a dispersing element, usually a grating which spreads the light intensity in space as a function of wavelength and (4) a focusing element to form an image of the entrance slit at the detector focal plane. As shown in Figure 1.6, once light enters from the entrance slit, it is redirected by the first mirror towards the grating. Its function is to disperse the incoming parallel light from the entrance slit to its component wavelengths such that they are no longer parallel but instead leave the grating at slightly different angles depending on the wavelength. The surfaces of these mirrors are polished with aluminum, silver or gold and are reflective in the wavelength region of the light involved. A typical grating can be characterized by equally spaced grooves, called facets. Classically ruled gratings may be plano or concave and possess grooves parallel to each other. Holographic grating grooves may be either parallel or of unequal distribution in order that system performance may be optimized. Holographic gratings are

generated on plano, spherical, and many other surfaces. When monochromatic light is incident on a grating surface, it is diffracted into discrete directions. The light diffracted by each groove combines to form set of diffracted wave fronts. The utility of a grating depends on the fact that there exists a unique set of discrete angles along which, for a given spacing between grooves, the diffracted light from each facet is in phase with the light diffracted from any other facet, leading to constructive interference.

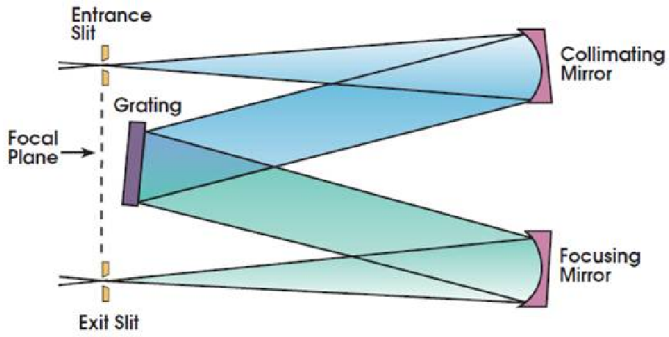


Figure 1.6: Basic components of a typical spectrograph

1.IV.1.d *Detector:* This helps in recording the intensity of the Raman signal at each wavelength. A CCD detector can be visualized as a rectangular array of photosensitive elements or pixels arranged in horizontal rows and vertical columns, configuration known as a focal plane array (FPA). In spectroscopic applications, the spectral or wavelength direction corresponds to the horizontal rows and the column pixels are usually summed, *i.e.* 'binned', vertically, providing intensity at each wavelength. CCD detectors are fabricated on a monolithic silicon chip. Typical dimensions of CCDs of 1024 × 256 pixels, each of which is 25 μm on a side, covering an area of about 25 mm × 6 mm. CCD technology has improved to such an extent that quantum efficiencies of 90% can be achieved from the visible to the near-IR wavelength range. Different types of CCDs, front illuminated, thinned back-illuminated and front- or back-illuminated deep depletion, are used for different applications. For the near-IR region, deep-depletion CCDs are the best. The depletion

region penetrates at least 30 μm into the silicon, providing a larger volume to generate photoelectrons, which results in increased quantum efficiency. Etaloning effect is a common problem for back-thinned CCDs operating in the near-IR region. Due to the larger transparency of silicon at these wavelengths, photons arising from the broad background fluorescence of the tissue are not fully absorbed in the shortened path length of the thinned silicon wafer. The result is that the thinned silicon wafer acts as an etalon for the back-reflected fluorescence, producing sharp peaks in the signal from this otherwise featureless background. Since these peaks are of similar line width to the tissue Raman features, they are exceedingly difficult to correct. Special surface treatment is used that randomizes the silicon wafer thickness on a wavelength scale, is being used by various manufacturers to overcome this problem.

1.V Computational Techniques

As mentioned earlier histopathological diagnosis is considered as gold standard for oral cancer diagnosis. However, this approach suffers from subjectivity, as it involves thorough visual inspection of the suspected section of the tissue under the microscope by an experienced pathologist. The fatigue factors due to examination of large number of samples and inexperience have been reported to exacerbate the error rate in the conventional approach of cancer diagnosis [89-91]. These problems can be surmounted in spectroscopic diagnosis. An important aspect of optical spectroscopy is objectivity. This is because spectral data are amenable to statistical tools. The application of these tools facilitates the computation of mathematical parameters derived from spectral data for classification. The distinct feature of this approach is that it is devoid of visual decision making and the system (computer) is completely blind to the sample that is being analyzed. Analysis of the data generated from a spectroscopy experiment can be performed in two different ways: univariate and multivariate. Univariate analysis using optical density values is generally

performed in colorimetric estimations of different biomolecules [92]. In this case knowing either of dependent and independent variables, a solution for the second variable can be calculated. In contrast to univariate approach, multivariate analysis involves observation and analysis of more than one statistical variable at a time. In the following section, discussion regarding multivariate analysis is presented.

1.V.1 ***Multivariate analysis:*** Data generated from infrared or Raman experiments consist of results of observations of multiple variables (wave-numbers) for a number of individuals (diseased or healthy). Each variable may be regarded as constituting a different dimension, such that if there are 'n' variables (IR or Raman bands) each object may be said to reside at a unique position in an abstract entity referred to as n-dimensional hyperspace. This hyperspace is necessarily difficult to visualize. The underlying theme of multivariate analysis (MVA) is simplification or dimensionality reduction. This can occur in one of two ways; either using an unsupervised or a supervised learning algorithms. In general, unsupervised methods such as principal component analysis (PCA) and hierarchical cluster analysis (HCA) are used to assess the 'natural' differences and similarities between spectra. These methods are employed to discover structure in the data and can be used to 'cluster' samples into groups by producing scatter plots (PCA) and dendrograms (tree-like figures; HCA). By contrast, supervised methods like linear discriminant analysis (LDA) and artificial neural networks (ANNs) are 'calibrated' with some known existing parameters about the sample. A *priori* knowledge is used in the construction of the LDA or ANN model followed by validation of model with test data or cross validation [92-94]. In following sections, a brief description about multivariate data analysis methods employed in the thesis is presented.

1.V.1.a *Principal Component Analysis (PCA):* PCA is one of the most widely used unsupervised multivariate methods for data compression and over-viewing. Main function of PCA is to identify trends, pattern and outliers in the data set [95, 96]. It decomposes a set of spectra into their most common variations (factors) and produces small set of well defined numbers (scores) for each sample that represent the amount of variation present in the spectrum. Presumably, these "factors" can also be used for reconstructing the spectrum by multiplying each one by a different constant scaling factor and adding the results together until the new spectrum closely matches the unknown spectrum. The scaling constants used to reconstruct the spectra are generally known as scores. A pictorial representation of PCA methodology is shown in Figure 1.7**.**

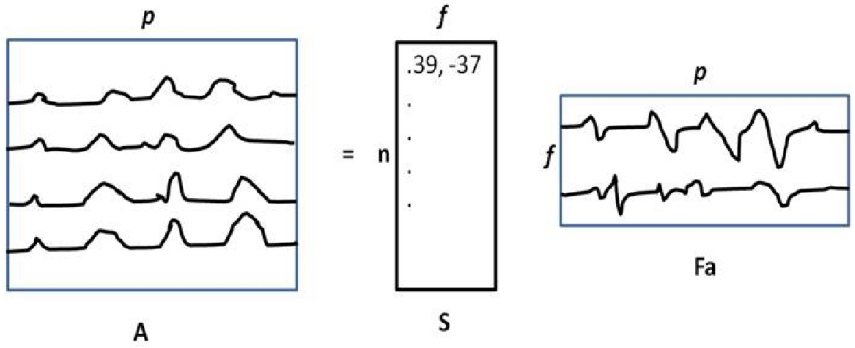


Figure 1.7: Schematic depiction of working principle of PCA, A= Original spectral data; n = number of spectra; S = PCA scores, p = number of data points; F = PCA factors (Eigenvectors, Loadings); f = number of principal components.

Spectra from a large number of samples, which belong to a given class, are subjected to an eigenvalue-eigenvector analysis. The eigenvectors can be assumed as the spectral equivalents for the principal components of these samples, while the scores correspond to the contribution of each PC to a given sample spectrum. Each sample spectrum can be regenerated by linear summation of the products of each eigenvector with its corresponding scores. The spectra of given set of samples may have contribution from a limited number of initial factors and also from the later eigenvectors accounting for day to day variations and unwanted noise. Only few initial Eigenvectors represent

the true variations in the given set of samples (often called as primary factors), whereas the remaining may be close to zero with no practical contribution to the spectra (secondary factors). Typically, the Eigenvalues of the primary factors are much larger than those of the secondary factors. The number of significant factors can be decided by a comparative statistical method called total percent variance analysis, which is actually a measure of the importance of each factor for reconstructing the spectra. Since the PCA factors represent the variations in the data, and the Eigenvalues are the relative weights of each of the factors, the Eigenvalues can also be thought of as the amount of variance in the data that is represented by that factor [97]. By summing the Eigenvalues, an estimate can be made of how much variance is accounted for by the PCA factors.

Before PCA is applied to a training set, the data is mean centered *i.e.* the mean spectrum (average spectrum) calculated from all of the spectra is subtracted from every calibration spectrum. Mean centering has the effect of enhancing the subtle differences between the spectra. Since the eigenvectors or principal components represent the changes in the spectral data that are common to all the calibration spectra, removing the mean simply removes the first most common variation before the data is even processed by the PCA algorithm. By iteratively eliminating each independent variation from the calibration spectra in series, it is possible to create a set of eigenvectors (principal components) that represent the changes that are common to all. There are two different methods used to calculate the Principal Components of a set of data; the NIPALS algorithm, and Decomposition of covariance [98]. After data has been fully processed by the PCA algorithm, it is reduced to two main matrices: the eigenvectors (spectra) and the scores (the eigenvector weighting values for all the calibration spectra). The matrix of the model is expressed by equation $\boldsymbol{A = S\ F + EA}$ (where $\boldsymbol{A}$ is an $\boldsymbol{n}$ by $\boldsymbol{p}$ matrix of spectral absorbances, $\boldsymbol{S}$ is an $\boldsymbol{n}$ by $\boldsymbol{f}$ matrix of score values for all of the spectra, and $\boldsymbol{F}$ is an $\boldsymbol{f}$ by $\boldsymbol{p}$ matrix of eigenvectors. The $\boldsymbol{EA}$

matrix called the matrix of residual spectra is the errors in the model's ability to predict the calibration data and has the same dimensionality as the ***A*** matrix. ***n*** is the number of samples (spectra), ***p*** is the number of data points (wavelengths) used for calibration, and ***f*** is the number PCA eigenvectors) [98]. PCA aims to summarize the overall variability, which includes both the divergence between groups, and the variation within groups. However, to assess the relationship between the different clusters, an adequate method should focus on between-group variability, while neglecting within-group variation.

1.V.1.b *Linear Discriminant Analysis* *(LDA):* LDA is a classification method which provides linear transformation of n-dimensional feature vectors (or samples) into an m-dimensional space ($m < n$), so that samples belonging to the same class are close together but samples from different classes are far apart from each other [99,100]. It is a supervised classification method, as the categories to which objects are to be classified is known before the model is created. The objective of LDA is to determine the best fit parameters for classification of samples as well prediction of unknown samples [101]. LDA is based on Bayes' theory where a classification model is developed assuming the probability distribution within all groups is known, and that the prior probabilities for groups are given, and sum to 100% over all groups. It is based on the normal distribution assumption and the hypothesis that the covariance matrices of the two (or more) groups are identical. This means that the variability within each group has the same structure. The only difference between groups is that they have different centers. This method also allows for a probabilistic assignment of individuals to each group, as in Bayesian clustering methods. The estimated covariance matrix for LDA is obtained by pooling covariance matrices across groups. When the variability of each group does not have the same structure (unequal covariance matrix), and the shape of the curve separating the groups is not linear, Quadratic Discriminant Analysis

(QDA), another variable of LDA is used. QDA may perform better in situations where the different groups being classified have their main variability in different directions, but only when the training sets used are large. The Mahalanobis distance is a way of measuring the distance of an observation to the centers of the groups, and uses ellipses to define the distances. Discriminant analysis is a type of qualitative calibration, where category group variable is used for the classification and not a continuous measurement as would be the case for a quantitative calibration [102,103]. LDA will not perform well on data sets where the discriminatory information is not in the mean, but is in the variance of the data. The results of the LDA classification are the predicted class for each sample, presented in the form of Confusion matrix. The confusion matrix is used for visualization of classification results from supervised methods. It carries information about the predicted and actual classifications of samples, with each row showing the instances in a predicted class, and each column representing the instances in an actual class.

1.V.1.c *Principal Component-Linear Discriminant Analysis* *(PC-LDA):* The prime difference between LDA and PCA is that former relies more on feature selection while LDA provides data classification [104]. In PCA, the shape and location of the original data sets changes when transformed to a different space whereas LDA does not change the location but only tries to provide better class separability and draw a decision line between the given classes. In recent times the use of PCA for data reduction, followed by linear discrimination algorithms (LDA) for classification, has become a well established method for spectroscopic data analysis with a number of research groups employing the method specifically for discrimination of Raman spectra. This technique is actually the cascade application of LDA on the factors resulting from PCA. PC-LDA relies on data transformation using PCA a step prior to LDA, which ensures that variables submitted to LDA are perfectly uncorrelated, and that their number is less than that of samples under study.

An important consideration when applying LDA following PCA is the number of PCs to be included; too few may result in lack of enough information, while too many increases the amount of noise in the data, and may lead to overfitting [100-105].

1.V.1.d ***Leave-one-out cross-validation (LOOCV):*** The ideal method for developing and validating efficacy of any diagnostic techniques is to validate results with an independent test data set [106-108]. An algorithm or series of algorithms is fit to the data in the training set using the empirical or statistical method of choice, and the criteria for classification into specific categories is determined. Classification of the spectra in the test set determines the unbiased accuracy of the algorithm. However, in cases of small data sets (as is often the case in pilot studies), division of the data into training and test sets is not feasible. The leave-one-out cross-validation method is a popular alternative to independent test sets. In the leave-one-out method, one spectrum is removed from the data set and the algorithm is driven using the remaining spectra. The algorithm is then tested using the removed spectrum. This process is repeated for every spectrum in the data set, such that an estimate of the potential accuracy of future algorithms developed using the method in question can be calculated [108].

1.V.1.e ***Curve fitting/Spectral deconvolution:*** The goal of spectral curve fitting is to mathematically create individual peaks from a spectrum that when added together, match the original data. Using an optimally fitting function, spectra are resolved into individual components and peak parameters like location (frequency), height (intensity), FWHM and area under band are calculated. By computing scatter intensity plots, the relative contribution of biomolecules associated with these bands can be obtained. The Levenberg-Marquardt algorithm is the most commonly used non-linear least squares fitting method [109-111]. In this method, variables in each peak are iteratively adjusted in an attempt to minimize the x^2 value (called as reduced chi-squared). It can be

defined as a weighted difference measure between the actual and measured data and calculated by following equation.

$$x^2 = \sum_{i=0} (actual\ i - calculated\ i\ /\ RMS\ noise)^2\ /\ (n\text{-}f)$$

The actual and calculated values are the measured and calculated data, respectively. The RMS noise is the estimated Root Mean Squared noise in the actual data over the fitted region. 'n' is the number of data points in the fitted region and 'f' is the total number of variables from all the peak and baseline functions. Thus, n-f is the number of degrees of freedom.

One of the biggest problems in interpreting spectroscopy results is determining how many peaks actually exist in the data. If the peaks were all nicely resolved, it would be easy and the need to use Peak Fitting to fit the data would be moot. Usually, non-linear peak fitting methods are applied because the peaks of interest are badly overlapped and therefore not directly measurable [98]. There are methods that give good approximations of how many peaks exist in a data set. One of these is calculating the Second Derivative. A second derivative of a data set indicates how slope of the data is changing. Presumably, at the top of a peak (or the bottom of a valley) the slope changes sign (usually from positive to negative, going left to right). In the second derivative this will appear as minima in the curve. The number of distinct minima and their positions give a good indication of number of peaks present in the data and their locations.

As with the number of peaks, peak shapes that have a real (physical) meaning for the set of data being analyzed are also important. Peak functions such as Gaussian, Lorentzian or mixed are used to generate an optimum function for fitting and calculations, using the following formulae [98].

1. **Gaussian :** $f(x) = He\text{-}(x\text{-}x0/w)2(4ln(2))$ [x0 = position; H = height; W = FWHH]
2. **Lorentzian:** $f(x) = H / 4\ (x\text{-}x0 / w)2 + l$
3. **Mixed (Gauss + Lorz):** $f(x) = (1\text{-}M)\ (Gauss) + (M)\ (Lorentz)$ [M = mixture (% Lorz)]

Choosing a line shape closely modeling the data is imperative to generate a good fit. For example, Lorentzian line shapes are as obvious the choice for NMR data, Vogit line shapes are for FT-IR data. However in many cases, trial and error is the only method that works.

1.VI An overview of Biomedical applications of Raman Spectroscopy

Raman spectroscopy has been extensively applied in diverse fields such as: analytical chemistry for process monitoring; detection of pollutants in environmental science; forensics applications include detection of gunpowder residues, blood, semen; identification of hazardous chemicals in security; industries like pharmaceuticals, petrochemical, semiconductor and in research areas -geology, nuclear science, material science, art and archaeology [112-117]. In recent years there has been a remarkable increase in the application of Raman spectroscopy to the field of medicine. This has come out of the awareness that Raman, like IR spectroscopy, is a vibrational spectroscopic technique capable of providing details on the chemical composition, molecular structure, and molecular interactions in cells and tissues. As a disease leads to changes in the molecular composition of affected tissues, these changes should be reflected in the spectra. Furthermore, if the spectral changes are specific enough for a particular disease state, they can, in principle, be used as phenotypic markers of the disease. The wealth of information provided by Raman spectra provides wide applicability, ranging from quantitative determination of the chemical composition of tissues or analyte concentrations in blood to real-time *in vivo* tissue classification [118-123]. In the following sections, a short note on applications of *ex vivo*, *in vivo* and micro Raman spectroscopy in cancer diagnosis is presented.

1.VI.1 ***Ex-vivo Raman spectroscopy:*** Spectral patterns present specific quantitative biochemical and morphological information about the examined tissues, depending on the cellular metabolic rate, vascularity, intravascular oxygenation and alterations in tissue morphology. Raman

spectroscopic studies pertaining to almost every part of human body is reported in literature. Alfano *et al.* in the year 1991 reported the Raman spectrum of breast tissue, using a FT- Raman spectrometer. They demonstrated that relative intensity of bands at 1651 and 1445 cm^{-1} could be used as diagnostic marker for differentiating normal and pathological tissues [124]. This was supported by studies carried out by Redd *et al*. (1993), Frank *et al.* (1994 and 1995), Hanlon *et al.* (2000), Haka *et al.* (2005), Chowdhary *et al*. (2006), K. Kumar *et al.* (2008) at different excitation wavelengths [125-131]. It was found that normal tissue spectra were dominated by Raman bands of fatty acids, while the Raman spectra of benign and malignant tissues were dominated by protein bands and these conditions can be classified using multivariate analytical tools. Gniadecka *et al.* in the year 2004 has developed a Raman system for acquiring Raman spectra of skin tissues. Findings have demonstrated the feasibility of classifying malignant skin lesions from normal skin tissue with a sensitivity of 84% and specificity of 97% [132]. Along similar lines, study by Nijssen *et al.* have shown applicability of high-wavenumber region (2800–3125 cm^{-1}) for classification of basal cell carcinoma (BCC) and normal skin tissue. In this study over 500 Raman spectra from 28 tissue samples of BCC and normal skin tissue were analyzed, and a discriminative accuracy of 100% sensitivity and 99% specificity was achieved [133]. Extensive work has been carried out by Mahadevan-Jansen *et al.* towards non-invasive and objective diagnosis of the cervix cancers using Raman spectroscopy. Classification using empirical peak intensities and unbiased multivariate algorithms based on principal component and linear discriminant analysis provided an average sensitivity and specificity of 88% and 92%, respectively [134]. Krishna *et al.* utilizing 785 nm excitation wavelength and PCA based limit test approach have demonstrated that normal and pathological cervix tissues can classified with sensitivity and specificity of ~99% [135,136]. Differences between both groups can be ascribed to non-collagenous proteins and DNA.

Vidyasagar *et al.* have demonstrated that Raman spectroscopy can also be used for predicting radiation response in cervical tumors [137]. Studies of human lung tissues have primarily focussed on combination of optical spectroscopy with endoscopic procedures and the discrimination of normal bronchial surface from premalignant lesions. Yamazaki *et al.* have demonstrated Raman spectroscopic applications for lung tissue analysis. Over 200 cancerous and noncancerous lung tissue samples were analyzed after formalin fixation. Discrimination with 91% sensitivity and 97%specificity was achieved [138]. Studies carried out by Stone *et al.* have demonstrated efficacy of Raman spectroscopy in classifying normal and abnormal esophageal tissues against normal and abnormal epithelial tissues from the larynx, tonsil, stomach, bladder and prostate. Comparison of prediction models with pathology provided sensitivities of 73–100% and specificities of 90–100% [139-142]. Studies carried out by Koljenovic and Krafft *et al.* on brain tissues have demonstrated potentials of Raman spectroscopy in identifying *ex vivo* normal and abnormal brain tissues as well as intracranial tumors [143,144]. Extensive work carried out by Morris *et al.* on human and animal bone tissues have shown that Raman spectroscopy can be used for identifying osteoporosis, bone fractures and deformities [145-149]. In addition to this Studies related to classification of normal and pathological tissues in ovary, colon, stomach and prostrate cancers are also reported in literature [150-155]. Overall these studies have demonstrated potentials of Raman spectroscopic methods in objective diagnosis of normal and pathological conditions.

1.VI.2 ***In vivo Raman spectroscopy:*** Even though Raman based optical diagnosis of tissues has shown potential in detection of potentially pre-cancerous tissues and real-time tissue evaluation with a high degree of sensitivity and specificity, removal of tissue or biopsy is an inherently invasive procedure. In most cases, this procedure has some associated morbidity and in some cases poses a high risk when involvement of vital anatomy is encountered. The actual promise

of these technologies lies in non-invasive or *in vivo* diagnosis of cancers. The most important advantage of Raman spectroscopy is that spectral acquisition can be carried out under *in vivo* conditions without disruption of native tissue conformation. Following studies have demonstrated potentials of *in vivo* Raman spectroscopy in cancer diagnosis:

After successful classification of normal, benign and cancer breast tissues, in the year 2006 *in vivo* Raman studies were undertaken for intraoperative tumor margin assessment in 9 patients undergoing partial mastectomy procedure [156]. Even though data set were small, 100% sensitivity and specificity was observed. Interestingly, Raman spectra from one margin correlated to a cancerous lesion which was grossly invisible, and upon postoperative pathological findings the margin was deemed positive which then required a second operation for excision. These findings provided evidence of extreme sensitivity of Raman spectroscopy in identification of pathological conditions. Bitar *et al.* in the year 2010 along similar lines have demonstrated feasibility of transcutaneous spectral acquisition from the DMBA-induced tumors and their margins in Sprague-Dawley rats [157]. Due to the obvious accessibility to skin, *in vivo* Raman spectroscopic analysis is a very practical and rapid, convenient approach for real-time cancer diagnosis. In the year 2008, Lieber *et al* developed a portable confocal Raman device along with a hand-held probe and demonstrated feasibility of *in vivo* spectra acquisition on 19 patients. Successful discrimination between normal, inflamed, basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) with efficiency of 95% was achieved [158]. Application of Raman spectroscopy in brain cancer diagnosis is widely reported and it could provide a means for *in vivo* evaluation of brain tissue, thereby eliminating unnecessary tumor resection during brain tumor surgeries, where preservation of uninvolved normal tissue is extremely crucial. *In vivo* studies on an animal model conducted by Kirsch *et al* in the year 2010, have demonstrated the efficacy of Raman spectroscopy in delineating

tumor margins in a cortical and subcortical melanotic tumor model [159]. This was further supported by another study by Beljebbar *et al.* in the same year on *in vivo* identification of C6 glioblastoma implanted in rat brains [160]. Prospectively, brain tumor surgeries could greatly benefit from Raman spectroscopic methods for histological evaluation *in situ* and by improvement of tumor margin clearance. Huang *et al.* have developed a rapid-acquisition image-guided Raman endoscopy system with 785-nm excitation to acquire *in vivo* spectra from gastric tissues within 0.5 s during clinical gastroscopic examinations. Using this setup, first the feasibility of classifying *in vivo* spectra from normal and neoplastic sites, followed by differentiation between benign and malignant ulcers in the stomach with 90 and 94% accuracy, respectively was demonstrated. Additional recruitments enabled independent test data evaluation using ant colony-linear discriminant technique, and prediction sensitivity of 89.3% and specificity of 97.8% was achieved. In recent studies, authors have integrated this Raman endoscopic imaging system with white-light reflectance, auto fluorescence imaging and narrow-band imaging modalities to enable both Raman spectra and endoscopy to be visualized simultaneously [161-166]. In the year 2011 same authors have demonstrated possibility of a fully automated on-line Raman spectral diagnostics framework integrated with a multimodal image-guided Raman technique for real-time *in vivo* cancer detection with endoscopy using 305 patients, followed by a study to detect pre-neoplastic stages like intestinal metaplasia, dysplasia [167]. This multi-modal image guided Raman endoscopy system was also employed for esophageal cancer detection by implementing a highly accurate biomolecular modeling algorithm and an accuracy of 96% was achieved [168]. Studies for developing a miniature confocal fiber-optic Raman probe useful during endoscopy-guided biopsies for better lesion identification have also been undertaken [169]. To enable *in vivo* applications in lung cancer diagnosis several groups are focusing on developing an optimized endoscopic probe for

bronchoscopic diagnosis. Recent study by Short *et al.* have demonstrated that including Raman spectroscope as part of available white-light/auto-fluorescence bronchoscopy set-up improves sensitivity for detection of preneoplastic lesions. *In vivo* study involving 26 patients using PC-LDA and fitting of combination of biomolecular models yielded 96% sensitivity and 91% specificity [170]. Several *in vivo* studies have been carried out in the last decade with a portable fiber-probe based Raman spectroscopic system for diagnosis of cervical abnormalities. Mahadevan-Jansen *et al*. in the year 1998 have demonstrated that cervical pre-cancers can be distinguished from benign tissues with *in vivo* Raman spectroscopy [171]. Same group in the year 2001, 2007 and 2009 demonstrated the feasibility of acquiring *in vivo* Raman spectra of more than 200 cervical cancer subjects in less than 5s acquisition time. Spectral findings were correlated against histopathology, methylene blue staining and coloposcopy [172-174]. Classification between normal, cancerous and different pathlogical grades has been explored. Kanter *et al.* in the year 2009 performed another study to demonstrate the potential of Raman spectroscopy in detecting subtle variations due to hormonal effects caused by menopause and menstruation cycle on both normal and dysplastic cervical tissues [175,176]. Duraipandian *et al.* recently demonstrated that Raman spectroscopy in conjunction with biomolecular modeling can be a powerful diagnostic tool for identifying hormone/menopause-related variations in the native squamous epithelium of normal cervix, as well as for assessing the effect of *Vagifem* treatment on postmenopausal atrophic cervix *in vivo* during clinical colposcopic inspections [177]. Overall these studies indicate that Raman spectroscopy has the potential to become a highly sensitive tool for detecting and accurately classifying normal and diseased conditions while greatly reducing the need for biopsies and the associated morbidity.

1.VI.3 ***Raman microspectroscopy:*** The process of spontaneous Raman scattering, however, produces a very weak signal, often orders of magnitude weaker than its fluorescence

counterpart. In order to address disadvantage of weak Raman signals from biological macromolecules, Puppels *et al.,* in 1990, developed novel confocal Raman micro-spectrometer for non-resonant Raman spectroscopy. With this instrument, it was possible to study single cells and chromosomes with high spatial resolution ($1\mu m^3$) [178]. Same group, in the year 1991, have reported typical spectra of the cell nucleus, chromosome and cytoplasm in human white blood cells and different granulocytes [179]. It was followed by exploring possibility of combining fluorescence activated cell sorting and Raman microspectroscopy. These techniques were applied to investigate the presence and sub-cellular location of carotenoids in cells isolated from human peripheral blood of healthy individuals. Carotenoids bands around 1157 cm^{-1} and 1525 cm^{-1} were high in T-lymphocytes and NK-cells. Findings of the study have provided possibilities of suggested protective role of carotenoids against cancer development [180,181]. Hawi *et al.* in the year 1996 have demonstrated its efficacy in characterizing normal and malignant human hepatocytes in cultured cells and liver tissues. Differences in the intensity of bands at 1040 and 1083 and 1182 cm^{-1} were observed in normal, cirrhotic and cancerous liver tissues [182]. Uzunbajakava *et al.* in 2003 showed potential of nonresonant confocal Raman imaging in mapping distributions of DNA and protein in individual single human cells [183]. Krafft *et al.* in the same year, utilizing PCA, generated single cell map of embryonic lung epithelial fibroblasts, human osteogenic sarcoma cells and human astrocytoma cells. Authors concluded that the origin of spectra strongly depends on the hydration state and the conformation of the surrounding molecules [184]. The same group, in the year 2005, has reported identification of sub-cellular structures by spectral signatures using non-resonant mapping on single human lung fibroblast cells [185]. Stone *et al.* in the same year demonstrated identification of prostate cancer cells using Raman microspectroscopy in combination with multivariate tools [186]. Matthäus *et al.* in year 2006 have reported Raman and infrared micro-

spectroscopic images of human cells at different stages of mitosis. They have monitored the distribution of condensed nuclear chromatin and other biochemical components during mitosis. Findings suggest that spectral intensities depend on the overall chromatin density variation in the individual sub-phases of mitosis [187]. Another study by the same group in the year 2007 first time reported label-free detection of mitochondria in human cervical cancer cell lines (HeLa). In this study, authors have compared results of fluorescence and Raman microscopy to visualize mitochondrial distribution. In the Raman maps, location of nuclei, nucleoli as well as sub-cellular organelles in the cytoplasm were observed. This study opened the door for non-invasive, *in vitro* studies of biological aspects of cell such as the dynamics of mitochondrial movement, drug uptake, apoptosis, and other effects [188]. In the same year Yu *et al.* used confocal Raman microspectroscopy to identify differences among normal and transformed human breast epithelial cell lines. Findings suggest that DNA duplication activities in tumorigenic cell nuclei are significantly higher than in normal cells [189]. Happ *et al.* in the year 2007 have analyzed retinoic acid induced differentiation of neuroblastoma cells by confocal Raman microscope. Authors were able to trace the presence of nor-adrenaline neurotransmitter as a marker for the differentiation process [190]. Schulze *et al.* in the year 2010 reported identification of murine and human embryonic stem cells on basis of Raman spectra. Findings suggested that the spectra reflected different molecular events connected with stem cell differentiation, including an increase in cell size, increase of tissue-specific proteins, a decrease of the proliferative activity and drop of the nuclear RNA content [191]. Zoladek *et al.* in the year 2011 using confocal Raman microspectroscopy performed time-course spectral imaging of live breast cancer cells undergoing apoptosis. The maps of DNA bands indicated an increase in signal intensity in apoptotic cells, which was attributed to chromatin condensation. The Raman maps of lipids indicated a high

accumulation of membrane phospholipids and highly unsaturated non-membrane lipids in apoptotic cells [192].

Nijssen *et.al* in the year 2002, demonstrated applicability of Raman microspectroscopy in distinguishing basal cell carcinoma from surrounding noncancerous tissue with pseudo-color Raman images of tissue sections. Efficacy of Raman spectroscopy in identifying biochemical changes accompanying malignancy was demonstrated. In this study molecular composition of different layers and structure of skin were studied and classification models with 100% sensitivity and 93% specificity were developed [193]. Koljenovic *et al.* in the year 2002 have reported discrimination between spectra of vital, tumor and necrotic region in glioblastoma with 100% accuracy [194]. Same group in the year 2005 have explored feasibility of understanding infiltration of meningioma within dura matter, the known cause of meningioma recurrence. Spectral assignment was performed on cluster average spectra of pseudo-color maps after correlation with histopathology [195]. Along similar lines, same authors, in the year 2006 have used it for characterizing biochemical composition of normal bronchial tissue in order to understand lung cancer progression. Raman maps generated at different stages of carcinogenesis suggest DNA rich features in basal layers while lipids are predominant in the superficial layers of the epithelium [196]. Krafft *et al*. in the year 2008 demonstrated that biochemical changes associated with gliomas and pediatric lung pathologies can be identified with Raman and FTIR imaging. They reported different sensitivity for both modalities and supported combined use of Raman and FTIR for in-depth analysis [197,198]. Beljebbar *et al.* in the year 2009 have demonstrated efficacy of Raman imaging as diagnostic tool for identification of normal and adenocarcinoma colonic tissues. Using unsupervised hierarchical cluster analysis (HCA), authors have identified the molecular composition and distribution of biomolecules within tissues [199]. In the same year, reports from

Tan *et al.* and Shetty *et al*. supported its applicability in discrimination of normal, premalignant and cancerous cervix and oesophageal tissues [200,201]. Kamemoto *et al.* in the year 2010 have shown differences in the normal and cancerous regions of cervix epithelium on the basis of collagen content [202]. Krafft *et al.,* in the year 2011 analyzed Raman and FTIR images of liver tissue using different statistical methods *e.g.* vertex component analysis (VCA), K-means cluster analysis (KCA), Fuzzy C-means cluster analysis and PCA. They concluded that VCA maps provides high molecular contrast as compared to other cluster analysis methodologies and can be interpreted in a similar way like H&E stained sections [203]. Abramczyk *et al.* recently reported Raman imaging of breast tissues from 150 patients. They observed that the composition of carotenoids and lipids in cancerous tissue is different from normal tissues. Raman spectra of normal tissue are dominated by monounsaturated fatty acid (oleic acid) while that of cancerous tissues are dominated by arachidonic acid derived cyclic eisanoids catalyzed by cyclooxygenase [204]. Froukje *et al.* have recently analyzed the spectral contributions of individual histopathological structures in oral mucosa. They concluded that origin of pre-dominating lipids bands can be attributed to the presence of adipose tissues in normal conditions and spectra have contributions from deeper layers [205].

1.VII Raman spectroscopy in oral cancers diagnosis

Raman spectroscopic applications in oral cancers started with analysis of normal and dysplastic tissue in a rat model by Schut *et al.* in the year 2000. Dysplasia in the palate was induced by topical application of the carcinogen 4-nitroquinoline 1-oxide. Raman spectra of normal and dysplastic tissues were obtained with 100 s integration time and classification was explored. Sensitivity and specificity of 100% in normal, low/high-grade dysplasia and carcinoma *in situ* were observed [206]. This was followed by study on human oral cancers biopsies by Venkatakrishna *et al.* in the year 2001. They have recorded 140 spectra of 49 biopsies using 785 nm excitation and SpeX Triax 320

spectrometer. Using PCA based multivariate analysis average classification efficiency of 88% was observed [207]. In the year 2004, study carried out by Krishna *et al.* has shown applicability of formalin fixed oral tissues for optical pathology. Using micro-Raman spectroscopy they have acquired spectra of formalin-fixed normal and cancerous oral tissues. Findings suggest significant differences in the epithelial region of normal and malignant samples however, no such differences were observed in sub-epithelial regions. Major differences between normal and malignant spectra seem to arise from the protein composition, conformational/structural changes, and possible increase in protein content in malignant epithelia [208]. In the year 2006, Malini *et al.* using 785 nm excitation and 30 s integration time have demonstrated efficacy of Raman spectroscopic methods in discriminating normal, cancerous, precancerous and inflammatory conditions. A total of 50 normal, 50 malignant, 10 inflammatory, and 5 premalignant biopsy specimens were utilized and 216 spectra (79 normal, 90 malignant, 37 inflammatory, and 10 premalignant) were acquired. Lipid rich features in normal conditions and predominant protein features were observed in tumors and other pathological conditions. Classification between different groups was explored using PCA coupled with multiparameter 'limit test' and 100% sensitivity and specificity was observed [209]. In the same year Oliveira *et al.* using FT-Raman spectroscopy and 1064-nm Nd:YAG laser line demonstrated that DMBA (7,12-dimethybenzanthracene) induced oral cancer in the buccal pouch of hamsters can be identified. Major differences between normal and malignant spectra seem to arise from the composition, conformational, and structural changes of proteins, and possible increase of its content in malignant epithelia. A total of 123 spectra were acquired and analyzed statistically by PCA and 91 and 69% sensitivity and specificity, respectively were observed [210]. Hu *et al.* in the year 2008, using confocal Raman microscpectroscopy acquired spectra of 66 human oral mucosa tissues (43 normal and 23 malignant). PCA along with calculation of areas under 1004, 1156, 1360

1587 and 1660 cm^{-1} bands were used as a classification method [211]. Recent study by Sunder *et al.* has demonstrated that oral carcinomas of different pathological grades can also be identified with Raman spectroscopy. On the basis of protein and lipid conformations authors have noted changes in the relative intensities of bands at 1656 cm^{-1}, 1440 cm^{-1} and 1450 cm^{-1} [212].

In vivo Raman spectroscopic study for identifying site wise variations in the oral cavity was reported by Guze *et al.* in the year 2009. A total of 51 subjects (25 Caucasian and 26 Asian) were recruited and *in vivo* spectra of buccal mucosa, tongue, floor of mouth, lip and hard palate with 1 s acquisition time were acquired. Fiberoptic probe of 1.8 mm diameter and 0.75 m length consisting of 200 μm excitation fiber surrounded by 27 ultralow 100 μm collection fiber along with filters at proximal and distal ends was used to acquire spectra. By analyzing high wavenumber region (2800-3100 cm^{-1}) authors have concluded that subject ethnicity does not have any influence on spectra; however different sites can be discriminated on basis of level of keratinization [213]. Along the similar lines Bergholt *et al.* in the year 2011, have characterized *in vivo* Raman spectroscopic properties of different anatomical regions (inner lip, attached gingiva, floor, dorsal tongue, ventral tongue, hard palate, soft palate, and buccal) in the oral cavity in the fingerprint region (800-1800 cm^{-1}). A total of 402 high-quality in vivo oral Raman spectra were acquired from 20 subjects. Fitting of reference biochemicals (hydroxyapatite, keratin, collagen, DNA, and oleic acid) and partial least squares-discriminant analysis (PLS-DA) were employed to assess the inter-anatomical variability. Findings suggest that histological characteristics of different sites have influence on the *in vivo* Raman spectra. Different sites can be classified with PLS-DA algorithms with overall sensitivity and specificity of 85% [214].

Aims and Objectives

These two, proof of principle studies have demonstrated that Raman spectra can be measured *in vivo* from the oral cavity. However, to the best of our knowledge no study is reported on use of *in vivo* Raman spectroscopic methods for discriminating normal and pathological conditions. **The studies reported in the thesis aims towards developing *in vivo* Raman spectroscopic method for the diagnosis of oral cancers and pre-cancers.**

Towards this end the following **Objectives** were undertaken:

1. Standardization of data acquisition and analysis methods on *ex vivo* oral tissues and correlation with histopathology and biochemical estimations.
2. To demonstrate feasibility of acquiring and classifying *in vivo* Raman spectra from buccal mucosa of normal, cancerous and pre-cancerous subjects and correlation with histopathology.
3. Exploring Raman spectral features of oral cancer cells with definite characters related to oral cancer.

Chapter 2

Raman spectroscopy of ex vivo tissues

Introduction

Five-year disease free survival-rates for oral cancer is around 50%, which is majorly attributed to the fact that patients present at late stages of the disease [2-6]. Optical spectroscopic methods are being actively pursued as possible alternatives to the existing methods and also as a modality for early detection [48-76]. *Ex vivo* Raman spectroscopic studies for classifying normal and pathological conditions of oral cancers are reported in literature [206-212]. Malini *et al* using a diode laser of 785 nm shown that spectra of normal tissues are rich in lipids while protein features are predominant in tumors. By utilizing PCA and 'limit test' based multivariate approaches they achieved 90% sensitivity and specificity in discriminating normal, premalignant, inflammatory and tumor conditions [209].

The present study was undertaken as a standardization step before initiating *in vivo* studies. In the following sections (1) adaptation of the prospective *in vivo* set-up for *ex vivo* and microscopic applications (2) standardization of spectral acquisition and analysis protocols on *ex vivo* tissues (3) histopathological and biochemical correlation of spectral data and (4) study on origin of spectra in normal tissues is presented.

2.I Adaptation of the Raman spectroscope for *ex vivo* and microscopic studies

A fiberoptic probe coupled HE-785 (Jobin-Yvon-Horiba, France) commercial instrumental set-up was procured and assembled. This system consists of a diode laser of 785 nm wavelength (PI-ECL-785-300-FC). The laser has dimensions of 9.25" x 2.5" x 4.25", is thermoelectrically cooled, made of AlGaAs, requires FC fitting for fiber connection and maximum output is 300 mW. Spectrograph is equipped with a fixed 950 gr/mm grating and CCD based (1024x256-BIDD-SYN) detection system. The CCD is thermoelectrically cooled, back illuminated deep-depleted type and consists of 1024 X 256 pixels of 26 μm X 26 μm size. Optical filtering of unwanted noise including Rayleigh

signals is accomplished through 'Superhead', the other component of the system. Optical fibers are used to couple excitation source and detection system with the 'Superhead'. Ball probe is the other component of the system which can be attached to 'Superhead' for remote applications including *in vivo* measurements. The ball probe is made of a hollow aluminum core with a lens at the tip. Since all the filtering and collection mechanisms are inside the 'Superhead', the ball probe merely functions as carrier for laser light delivery and collection of photons. Figure 2.1 shows the ball probe attached to the Raman system.

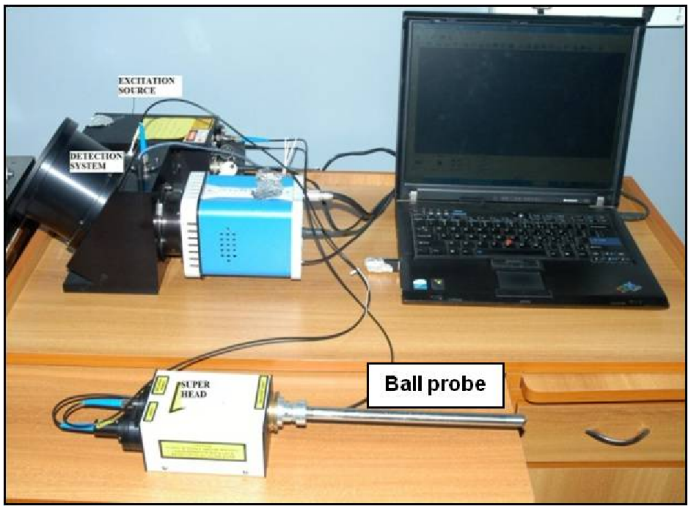


Figure 2.1: Ball probe attached to the HE-785 Raman spectrometer

This probe is large in size and has dimensions of 10" x 2.5", which makes it highly inconvenient and impractical for *in vivo* applications. Additionally, spectra acquired using this probe are marred by high background and interfering signals from fiber. A typical spectrum acquired using ball probe at 52 mW, 10 s acquisition time, over 5 cycles shown in Figure 2.2.

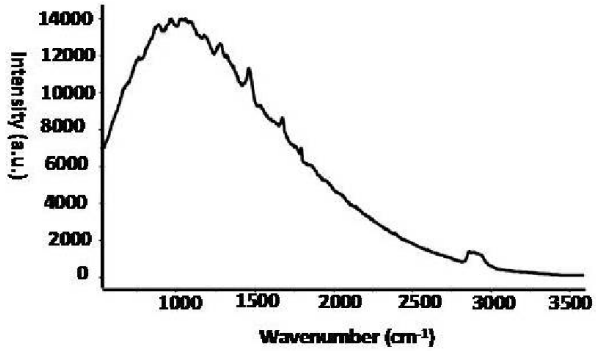


Figure 2.2: Typical *in vivo* spectrum acquired using ball probe

Because of poor spectral quality, high noise and patient's discomfort, studies using this probe were discontinued. Another commercially available fiberoptic probe (InPhotonics Inc., Downy St, USA) was procured and tested. Pictorial comparison of the ball probe and InPhotonics probe is presented in Figure 2.3.

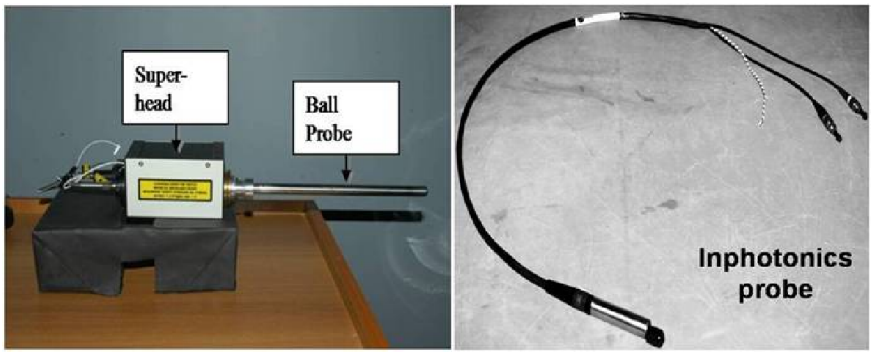


Figure 2.3: Comparison of Ball and Inphotonics probe

Figure 2.3 shows that the InPhotonics probe is much smaller in length (10 cm) and diameter (1.25 cm) in comparison to the ball probe. In contrast to the ball probe, InPhotonics probe is a coaxial, two-fiber probe. One fiber is used for excitation (105 μm, NA-0.40) and another for collection (200 μm, NA-0.40). The overlap between the two fiber ends is optimized by using a lens to focus the laser as well as to collect the scattered radiation. As per specifications of the manufacturer of the Inphotonics probe, the theoretical spot size and depth of field are 105 μm and 1 mm, respectively.

All of the filtering mechanisms are contained in a 0.5" (25.4 mm) diameter stainless-steel jacket which makes this probe ideal for remote hand-held applications. The beam path inside the probe is shown in the Figure 2.4.

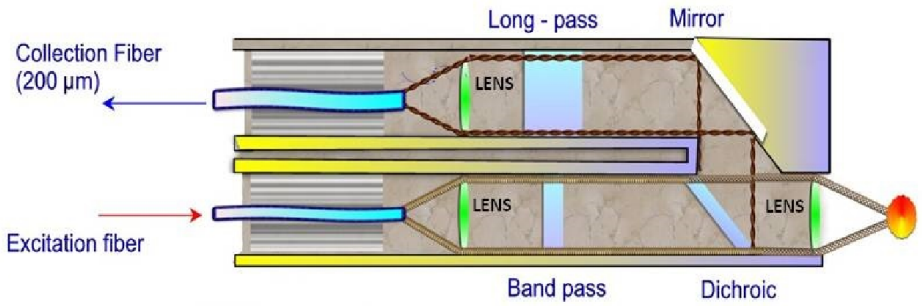


Figure 2.4: Focusing and filtering optics inside the Inphotonics Raman probe

It is to be noted that at the end of the excitation fiber, a lens is used to collimate the laser light. A band-pass filter removes the silica Raman bands and transmits only the pure laser light. The dichroic filter transmits the laser line to be focused by another lens onto the sample. The same lens gathers the light which is scattered at 180^{o} from the laser direction (backscattering geometry). The collected signal is then reflected by the dichroic filter through a long-pass filter assembly that transmits only the Stokes scattered light. This last filter set attenuates the Rayleigh signals and also prevents the silica Raman bands that arises in the collection fiber. Finally, another lens is used to focus the Raman scattered light from the sample, onto the outgoing fiber. Small dimensions of this probe provide required flexibility to the clinician as well as comfort to the patient while recording *in vivo* spectra. Besides flexibility and comfort this probe also gives better spectra in shorter acquisition time. Additional information regarding adaptation of InPhotonics probe for *in vivo* measurements is presented in Chapter 3, under the methodology section.

In order to utilize InPhotonics probe for *ex vivo* measurements a probe holder along with XYZ sample stage was assembled. This helps in acquiring spectra at different points with fixed spacing

which might lead to a better correlation with histopathology. Photographic representation of the set-up used for acquiring spectra of *ex vivo* tissues is shown in Figure 2.5.

Figure 2.5: Raman spectroscope employed for *ex vivo* applications

The 'Superhead' component of the system can also be coupled to a microscopic objective. Raman microscopic set-up was assembled by attaching a suitable microscopic objective to the 'Superhead'. This set up has been successfully utilized for recording spectra of cell smears and serum samples. Photographic representation of the assembled Raman microprobe set-up is shown in Figure 2.6.

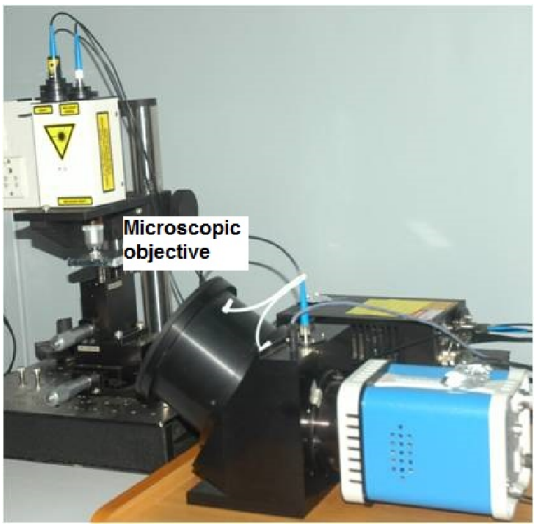


Figure 2.6: Assembled Raman microprobe set up for microscopic applications

Ex vivo tissues were used for standardization of spectral acquisition and data analysis protocols. In the following section the results obtained and their evaluations are discussed.

2.II Raman spectroscopy of *ex vivo* tissues: Reproducibility of spectral features

Earlier studies have demonstrated that spectra of normal tissues are rich in lipids while proteins are predominant in tumor. This study was undertaken to evaluate the reproducibility of the spectral features of normal and tumor tissues. Using fiberoptic probe coupled Raman spectroscope spectra of normal and cancer oral biopsies were acquired and correlated with histopathology.

2.II.1 Methodology

2.II.1.a *Clinical samples:* The study was approved by institutional ethics committee, Tata Memorial Center (Project No. 605). Pathologically verified 36 pairs of buccal mucosa biopsies (tumor and cut-margin) samples were collected in liquid nitrogen and stored at -80^{0}C until use. Information about clinicopathological parameters such as age, sex, tobacco habit and tumor grade were obtained from electronic medical record (EMR) of Tata Memorial Hospital.

2.II.1.b *Spectral acquisition:* Spectra were acquired using above described fiberoptic probe coupled HE-785 Raman system (Figure 2.5). Tissues were passively thawed and kept on a calcium fluoride (CaF_2) window. The window was placed under illumination zone of the fiberoptic probe on XYZ precision stage. Spectra were acquired at different points with an average spacing of ~2 mm. Spectral acquisition parameters were: laser power–80 mW, acquisition time–10 seconds and averaged over 5 accumulations. Identical conditions were maintained during all measurements.

2.II.1.c *Spectral pre-processing:* Spectral pre-processing using Labspec 5.0 software (HORIBA Jobin Yvon) was performed as per the already existing standard protocol [104,133]. In first step, the wavelength dependency of the detector and the polarization dependence of the optical elements were measured using a calibration standard (standard reference material number- 2241; NIST, Gaithersburg, MD, USA). The spectrum associated with the instrument response, was

divided to the measured Raman spectra. The spectral contribution of optical elements *i.e.* the background signals, were obtained by acquiring spectra of CaF_2 window without the sample under similar conditions. The response corrected background spectrum was subtracted. In order to remove influence of slow moving fluorescence background, first derivative spectrum was computed using Savitzky–Golay filter mechanism (window size-3). Correction for spectral differences due to relative intensity changes or variation in sample thickness was performed by vector normalization. First derivative and vector normalized spectra were interpolated in to desirable spectral regions and were used as input for multivariate analysis. Typical *ex vivo* spectrum at different pre-processing steps is shown in Figure 2.7.

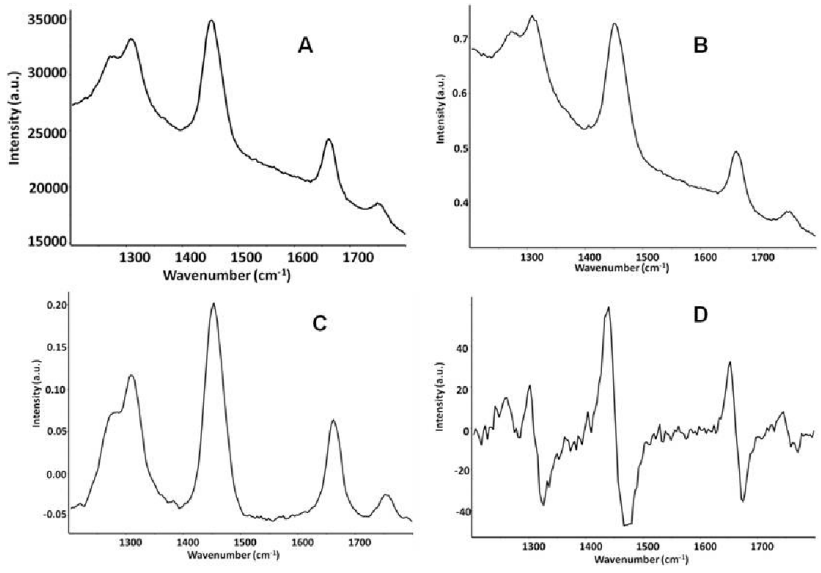


Figure 2.7: Typical *ex vivo* spectrum at different pre-processing steps A. Raw spectrum B. CCD response corrected C. Background corrected D. First Derivative

2.II.1.d *Spectral-comparisons:* Average spectra were computed from the background subtracted spectra (prior to derivatization) for each class and baseline corrected by fitting a 5th order polynomial function. These baseline corrected spectra were used for spectral comparisons

across all groups. Difference spectra were computed by subtracting mean normal spectra from tumor. In this case all positive bands belong to tumor and negative are from normal tissues.

2.II.1.e *Multivariate analysis:* Principal Component-Linear Discriminant Analysis (PC-LDA) was employed for data analysis. As already described in the Chapter 1 under computational techniques section, this method is combination of PCA and LDA. In first step, PCA is performed for data reduction and principal components are generated. Of this significant principal components ($p<0.05$) are used as input for LDA based discrimination. This helps in removing or minimizing the noise from the data and concentrating on variables important for classification. In order to avoid over-fitting of the data, as a thumb rule, total number of factors selected for analysis were less than half the number of the spectra in the smallest group. This method of analysis has been utilized by various groups [103,151]. PC-LDA models were validated by leave-one-out cross-validation (LOOCV) and independent test data. Algorithms for these analyses were implemented in MATLAB (Mathworks Inc.) based in-house software [215]. The results of PC-LDA can be depicted as scatter plots, generated by plotting various combinations of scores of factors used for discrimination. These results can also be presented in form of a confusion matrix, where all diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions.

2.II.2. Results and Discussion

2.II.2.a *Spectral features:* Mean spectra along with standard deviation of normal and tumor tissues are shown in Figure 2.8. Spectra of normal conditions were dominated by C=O (ester) band at 1750 cm^{-1}, strong CH_2 bend at 1450 cm^{-1} and two sharp features around 1300 cm^{-1} which are attributable to lipids (Figure 2.8A). Predominant protein bands indicated by strong and broad amide I (1660 cm^{-1}), broad δCH_2 (1450 cm^{-1}) and broad features in the amide III were seen in mean tumor spectrum (Figure 2.8B). These features were further confirmed by computing difference

spectrum. Mean normal spectrum was subtracted from mean tumor spectrum, thus positive bands belongs to tumor while negative bands are from normal. As shown in Figure 2.8C**,** strong negative ester band (1750 cm^{-1}), δCH_2 (1450 cm^{-1}) and two sharp features around 1300 cm^{-1}) indicating predominant lipid composition were observed in case of normal. Positive band indicating presence of DNA (1340 cm^{-1}), high protein content (amide I and III) were seen in case of tumor.

Origin of lipid features in normal conditions had been primarily attributed to biochemical composition and architectural arrangement of different layers. In case of malignancy, demarcation between different layers is lost, leading to mixing of content from different layers, which might give rise to protein rich spectrum. Further, fact that cancerous cells secrete large amount of antigen, antibodies and receptor proteins could also contribute to origin of protein rich spectrum. Origin of negative DNA band observed in case of tumors can be attributed to increase in number of cells due to hypercellualrity or inflammation, preliminary events in cancer development. Overall, spectral features corroborating with already reported studies were observed for normal and tumor tissues and thus reproducibility of spectral features was established.

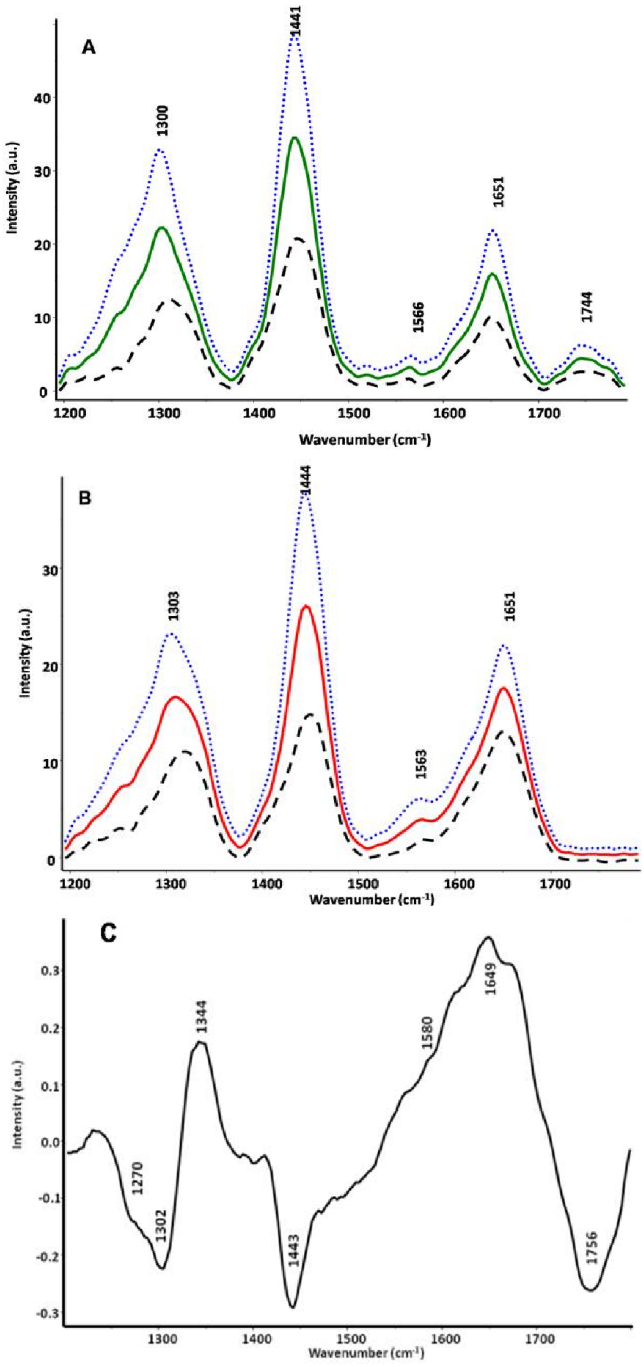


Figure 2.8: Spectral features of normal and tumor tissues A. normal B. tumor C. difference spectra (Tumor - Normal). (Solid line: Mean spectrum, Dotted line: Mean + standard deviation, broken line:mean – standard deviation)

2.II.2.b *Classification of normal and tumor spectra:* PC-LDA was performed to explore the feasibility of classification between normal and tumor spectra. Standard models for *ex vivo* normal and tumor conditions were developed using 63 and 68 spectra from 8 normal and 8 tumor tissues, respectively. PC-LDA was performed using 7 factors contributing ~88% correct classification efficiency. Scatter plot using score of factor 2 and 3 was generated and shown in Figure 2.9. Two independent clusters belonging to normal and tumor spectra respectively were obtained.

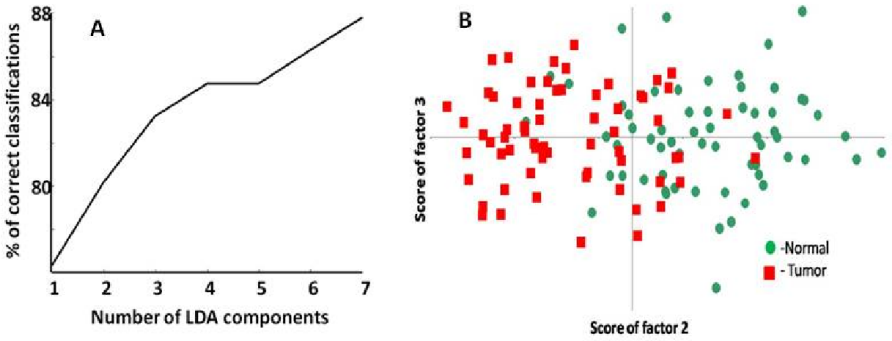


Figure 2.9: PC-LDA of normal and tumor spectra: A. Scree plot B. Scatter plot

Similar results are presented in form of confusion matrix shown in Table 2.I. Here diagonal elements are true positive and ex-diagonal elements are false positive predictions. As can be seen 56 of 63 normal spectra (89%) and 56 of 68 tumor spectra (82%) were correctly classified (Table 2.1A). These findings were validated by leave-one-out cross validation (LOOOCV) and results are summarized in Table 2.IB. Classification efficiencies of 89 and 79% were observed for normal and tumor conditions, respectively. Independent test data set comprising of 256 normal and 296 tumor spectra of 28 tissues each was used for estimating prediction efficiencies of the standard models. As shown in Table 2.IC, 205 of 256 normal and 284 of 296 tumor spectra were correctly identified.

Prediction efficiencies of 80 and 96% were observed for standard models of normal and tumor conditions, respectively.

A	Normal	Tumor	% classification eff.
Normal	55	8	87.30
Tumor	10	58	85.29
B	Leave-one-out cross validation		
	Normal	Tumor	% classification eff.
Normal	52	11	82.53
Tumor	15	53	77.94
C	Independent test data prediction		
	Normal	Tumor	% prediction eff.
Normal	205/256	51	80.07
Tumor	12	284/296	95.94

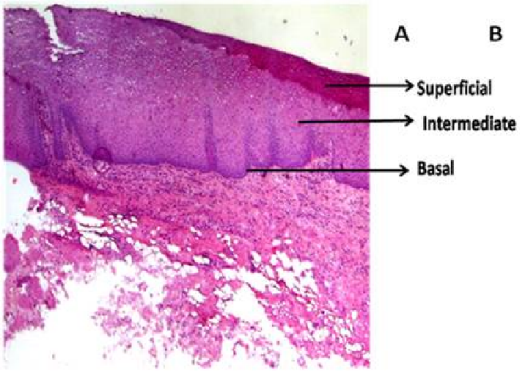
Table 2.I: Summary of classification between normal and tumor spectra A. Standard model; B. LOOCV; C. Independent test data (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions)

Minor misclassification among normal and tumor spectra can be attributed to the sample heterogeneity *i.e.* presence of normal regions in a tumor biopsy or presence of analogous inflammatory or hyperplasic patches in a normal specimens. Normal tissues used in the study were collected from tumor adjacent areas, therefore possibility of acquiring spectra from an inflammatory region cannot be completely ruled out. This is further supported by the fact that majority of oral cancer patients have possibilities of developing second primary tumors in normal appearing mucosa because of long-term tobacco exposure. Overall, findings of the study demonstrated that spectra of normal and tumor biopsies are reproducible and can be objectively classified.

Influence of sample heterogeneity on prediction efficiency of standard models was further explored by correlating spectral predictions of normal and tumor biopsies against their respective histopathological reports and the findings are discussed in the following section.

2.III Correlation of the spectral data with histopathology

Five (5) µm sections of 8 normal and 8 tumor tissues were collected, stained and correlated with their corresponding LOOCV spectral predictions. Two of the eight normal tissues show no misclassification. A representative true normal section along with its spectral prediction is shown in Figure 2.10A. As per the histopathological report this section appears to be completely normal, indicated by clearly visible stratification of epithelium into basal, intermediate and superficial layers, with prominent evagination of epithelial rete pegs and invaginations of connective tissue papilla. The connective tissue is also normal as no inflammatory infiltration was observed. LOOCV spectral prediction further corroborates with histopathological findings, as all 7 spectra acquired from this tissue were correctly identified (Figure 2.10B).



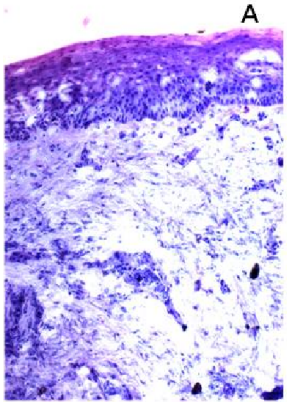
Spectra	Original input	Prediction after LOOCV
Tissue X.1	Normal	Normal
Tissue X.2	Normal	Normal
Tissue X.3	Normal	Normal
Tissue X.4	Normal	Normal
Tissue X.5	Normal	Normal
Tissue X.6	Normal	Normal
Tissue X.7	Normal	Normal

Figure 2.10: A: Histopatholgocial section of a true normal specimen B: LOOCV spectral predictions

Another representative section along with its spectral prediction of normal tissue is shown in Figure 2.11. Analysis of histopathological section reveals that there is loss of stratification and presence of dysplastic features in the lower 1/3rd of the epithelium (Figure 2.11A). Additionally, there is increase in thickness of the intermediate and the superficial layers of the epithelium and para-keratinization can also be seen. Also some areas in the basal and parabasal layer show abnormality indicated by altered staining properties and increased nuclear cytoplasmic ratio (increase in size of

the nucleus). LOOCV predictions are shown in Figure 2.11B. Corroborating with histopathological report, 3 out of 11 spectra acquired from this tissue were wrongly predicted. This can be attributed to above discussed abnormality in the tissues specimen. Overall, misclassifications of normal spectra as tumor could be primarily attributed to increased keratotis in stratum superfacialis and hyperplasic or inflammatory changes in the epithelium.

A



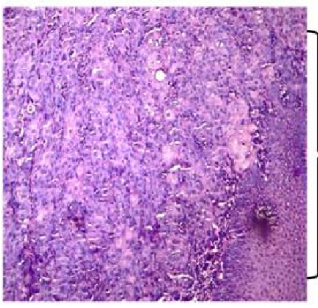
B

Spectra	Original Input	Prediction after LOOCV
Tissue Y.1	Normal	Tumor
Tissue Y.2	Normal	Normal
Tissue Y.3	Normal	Tumor
Tissue Y.4	Normal	Normal
Tissue Y.5	Normal	Tumor
Tissue Y.6	Normal	Normal
Tissue Y.7	Normal	Normal
Tissue Y.8	Normal	Normal
Tissue Y.9	Normal	Normal
Tissue Y.10	Normal	Normal
Tissue Y.11	Normal	Normal

Figure 2.11: A. Histopathological section of normal specimen B. LOOCV Spectral predictions

In case of tumors three (3) out of 8 tissues show no misclassification. A representative true tumor tissue section is shown in Figure 2.12A**.** The tissue show moderately differentiated squamous cell carcinoma (MDSCC), with complete loss of epithelial stratification and areas of tumor islands with neoplastic cells and inflammatory patches. LOOCV spectral predictions were found to be corroborating with the histopathological findings, none of the 7 spectra acquired from this tissue were misclassified (Figure 2.12B).

A

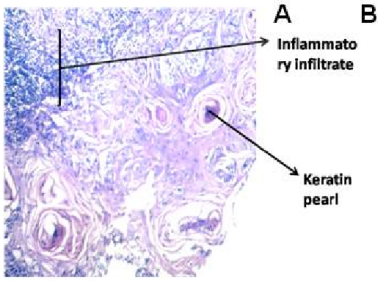


B

Spectra	Original Input	Prediction after LOOCV
Tumor Tissue X.1	Tumor	Tumor
Tumor Tissue X.2	Tumor	Tumor
Tumor Tissue X.3	Tumor	Tumor
Tumor Tissue X.4	Tumor	Tumor
Tumor Tissue X.5	Tumor	Tumor
Tumor Tissue X.6	Tumor	Tumor
Tumor Tissue X.7	Tumor	Tumor

Figure 2.12: A. Histopathological section of true tumor specimen B. LOOCV Spectral predictions

Another representative section of the tumor tissue is shown in Figure 2.13. Histopathological analysis of this tissue revealed that the tissue is of well differentiated squamous cell carcinoma type (WDSCC), indicated by presence of epithelial islands and keratin pearls (Figure 2.13A). In case of WDSCC tumors, level of anaplasia is less therefore cells are more differentiated and closer to the tissue of origin *i.e.* normal. Additionally increased stromal width and inflammatory regions were also seen. LOOCV spectral predictions are shown in Figure 2.13B. Corroborating with histopathological findings, 4 of the total 11 spectra acquired from this tissue were misclassified as normal, which can be primarily attributed to presence of large amount of stroma (normal connective tissue).



Spectra	Original Input	Prediction after LOOCV
Tumor Tissue Y .1	Tumor	Normal
Tumor Tissue Y.2	Tumor	Tumor
Tumor Tissue Y.3	Tumor	Normal
Tumor Tissue Y.4	Tumor	Normal
Tumor Tissue Y.5	Tumor	Tumor
Tumor Tissue Y.6	Tumor	Normal
Tumor Tissue Y.7	Tumor	Tumor
Tumor Tissue Y.8	Tumor	Tumor
Tumor Tissue Y.9	Tumor	Tumor
Tumor Tissue Y.10	Tumor	Tumor
Tumor Tissue Y.11	Tumor	Tumor

Figure 2.13: A. Histopathological section of tumor specimen B. LOOCV Spectral predictions

Most of the tumor tissues used in the study belong to well differentiated category as tissue sections show numerous keratin pearls along with adjacent areas of dysplastic or hyperplasic or relatively normal mucosal epithelial areas. All these reasons could be attributed to the misclassification of tumor spectra as normal.

Overall findings of the study suggest that Raman spectral profiles are signature of the architectural arrangement of different layers in a tissue and misclassifications between normal and tumor spectra can be correlated with their respective histopathology.

2.IV Raman signals in normal tissues: Study on intact and incised biopsies

Lipid rich Raman signals in normal tissues had been attributed to membrane lipids. It is believed that Raman scattering from upper layers can reach more efficiently to detectors as compared to deeper layers owing to losses due to multiple scattering [207-209]. However, in view of recent developments in deep tissue Raman spectroscopy, this fact needs revisiting [216]. This study was undertaken to explore origin of Raman signals in normal conditions by acquiring spectra from intact and incised biopsies and its influence on classification with tumors.

2. IV.1 Methodology

2.IV.1.a *Intact and incised biopsy components:* Ten scalpel biopsies of approximately 8 x 6 x 5 mm dimensions from contralateral buccal mucosa of 10 oral cancer patients were collected. Tissue samples were collected in phosphate buffer saline (PBS) and transferred to liquid nitrogen. Two frozen sections, each of 5μm thickness were obtained longitudinally by orienting epithelium and connective tissue in order and used for histopathological examination. In the first step, Raman spectra of epithelium (superior surface of biopsy) referred as *'intact epithelium'* and the connective tissue (inferior surface of biopsy) termed as *'intact connective tissue'*, were acquired. In the second step, epithelium was separated from connective tissues using surgical blade no.11 attached to Bard-Parker (BP) handle. The procedure of separation of epithelial and connective tissue component was as per the routine technique used in maxillofacial surgery practice. A schematic representation of the experimental protocol is shown in Figure 2.14. Raman spectra of the same superior and inferior surfaces of incised biopsies were acquired and termed as *'separated epithelium-upper'* and *'separated connective tissue-lower'*.

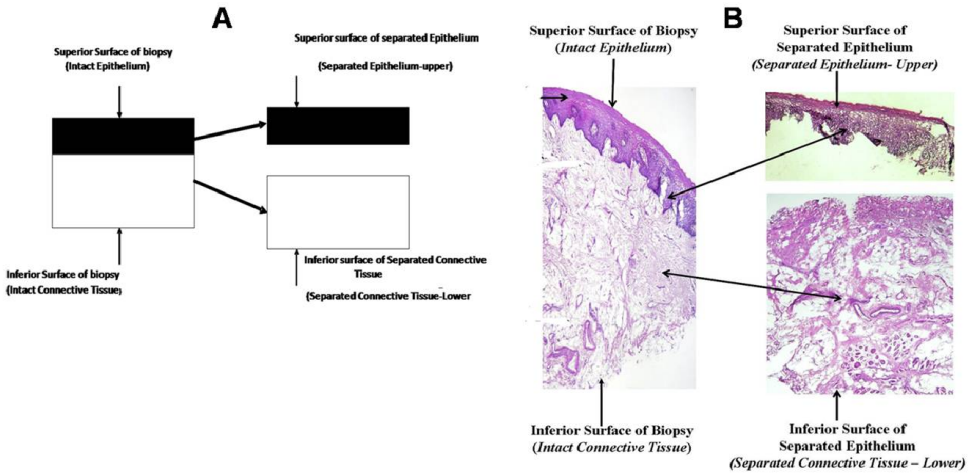


Figure 2.14: A. Schematic presentation of protocol of spectral acquisition: B. Histological Sections of Buccal mucosa; Intact and Separated Components

2.IV.1.b *Spectral acquisition and pre-processing*: On an average 5-6 spectra from all four surfaces of 10 normal tissues were recorded using previously described fiberoptic probe coupled HE-785 commercial Raman spectrometer (Figure 2.5). A total of 234 spectra (68-*intact epithelium,* 53-*intact connective tissue*, 54-*separated epithelium-upper* and 59-*separated connective tissue-lower*) were acquired. We have also recorded 128 spectra from 15 histopathologically verified tumor biopsies. Spectra were acquired as per the already described protocol (Section 2.II.1.b) at different points with an average spacing of ~2 mm. Spectral acquisition parameters were: laser power–80 mW, acquisition time–10 seconds and averaged over 5 accumulations. Spectral pre-processing was performed by as per the previously described protocol (Section: 2.II.1.c). First derivative and vector normalized spectra in 1200-1800 cm^{-1} region were used as input for multivariate analysis.

2.IV.1.c *Spectral comparisons:* Average spectra were computed from the background subtracted spectra (prior to derivatization) for each class and baseline corrected by fitting a 5^{th} order

polynomial function. These baseline corrected spectra were used for spectral comparisons across all groups.

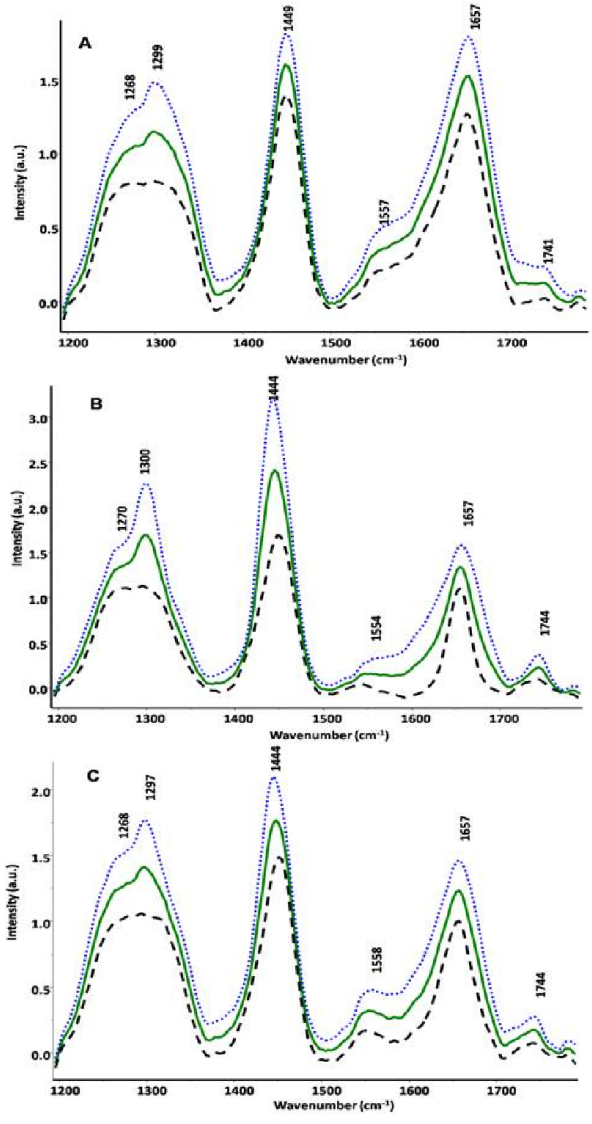
2.IV.1.d *Statistical analysis:* Both unsupervised PCA and supervised LDA methods were used to explore feasibility of classification between different groups using algorithm implemented in MATLAB based in-house software [216].

2.IV.2. Results and Discussion

2.IV.2.a *Spectral features:* Mean Raman spectra of *'intact epithelium', 'intact connective tissue'*, *'separated epithelium-upper'*, *'separated connective tissue-lower'* and tumors in 1200-1800 cm^{-1} region along with standard deviations are shown in Figure 2.15. Spectral features of all four groups (Figure 2.15A-D) of normal tissues are suggestive of predominant lipid features indicated by ester bands (1745 cm^{-1}), strong δCH_2 bend (1450 cm^{-1}), sharp features around amide III region and sharp amide I (1660 cm^{-1}). Dominating protein features indicated by broad amide III, broad and shifted δCH_2 as well as broad amide I can be seen in mean tumor spectrum (Figure 2.15E). Spectra of intact and separated connective tissue show strong lipid features in comparison to spectra from epithelial surfaces. Likewise, spectra of intact and separated epithelium are similar.

Biochemically, oral buccal mucosa can be divided in to three parts, epithelium (rich in protein), lamina propria (rich in laminin, reticulin proteins and collagenous material) and submucosa (rich in lipids) [217]. In the first step when spectra were acquired from superior surface of intact biopsies, entire buccal epithelium, lamina propria and portion of submucosal adipose tissue could have contributed to the spectrum. It may be the reason for Raman signals of lipids and proteins observed in the spectra (Figure 2.15A). Raman spectra obtained from inferior surface, Figure 2.15B show strong lipid features along with proteins due to the fact that submucosal adipose tissue zone falls within the depth of field of the probe and also Raman cross-section of lipids is

larger [129,218]. Origin of protein features in this case could be due to contribution of the epithelial surfaces. In contrast to intact biopsies, spectra from superior and inferior surfaces of incised biopsies show distinct features. Spectra from *'separated epithelium-upper'*, is rich in protein features while spectra of *'separated connective tissue-lower'* show predominant lipid features (Figure 2.15 C,D).



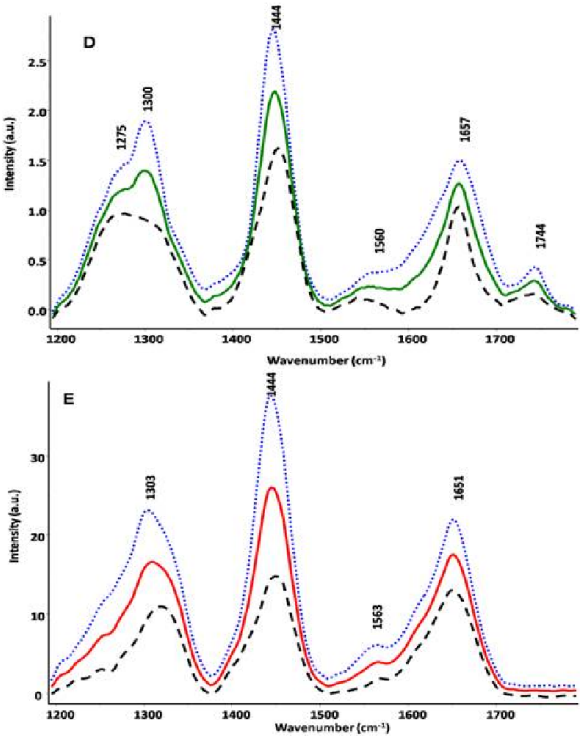


Figure 2.15 : Mean and standard deviation spectra: (A) intact epithelium, (B) intact connective tissue, (C) separated epithelium-upper, (D) separated connective tissue-lower, and (E) tumor (Solid line: Mean spectra; dotted line: Mean + Standard deviation; broken line : Mean - Standard deviation)

2.IV.2.b *Multivariate analysis:* In order to understand above mentioned variations and similarities across the spectra of different classes and to explore the feasibility of classification, multivariate analysis using both unsupervised (PCA) and supervised (LDA) methods was carried out. Loading plots of factors used for PCA are shown in Figure 2.16. In first step, spectra of all four groups of normal tissues (*'intact epithelium'*, *'intact connective tissue', 'separated epithelium-upper'* and *'separated connective tissue-lower'*) were pooled and PCA was carried out. Best result obtained in our analysis is shown in Figure 2.17A, wherein spectra from intact tissues (*'intact epithelium'* and *'intact connective tissue')* gave highly overlapping clusters, while spectra from *'separated epithelium-upper'* and *'separated connective tissue-lower*' showed a tendency of classification, although they overlapped with clusters of intact tissues. In the next step, spectral data from all four

groups of normal tissues were subjected to PCA in two stages. First data from intact tissues (*'intact epithelium'* and *'intact connective tissue')* was used for PCA, no classification could be achieved (Figure 2.17B). Conversely, PCA of second group *i.e.* separated sections gave better classification, Figure 2.17C.

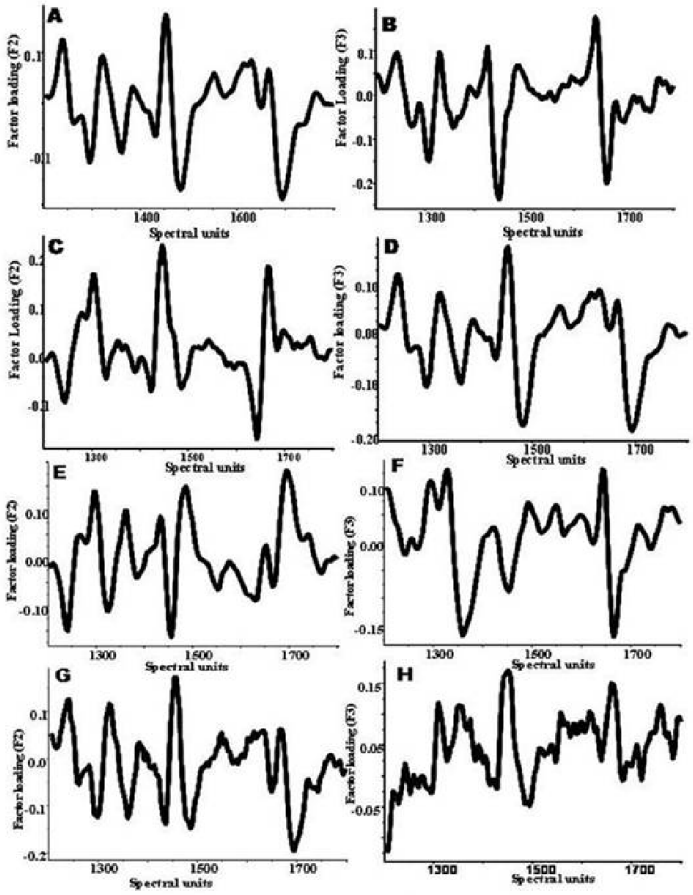


Figure 2.16: Loadings of factor 2 and 3 used for PCA of intact tissues and separated sections [(A) and (B)]; intact tissues [(C) and (D); separated sections [(E) and (F)]; PCA along with tumor tissues and intact and incised biopsies [(G) and (H)].

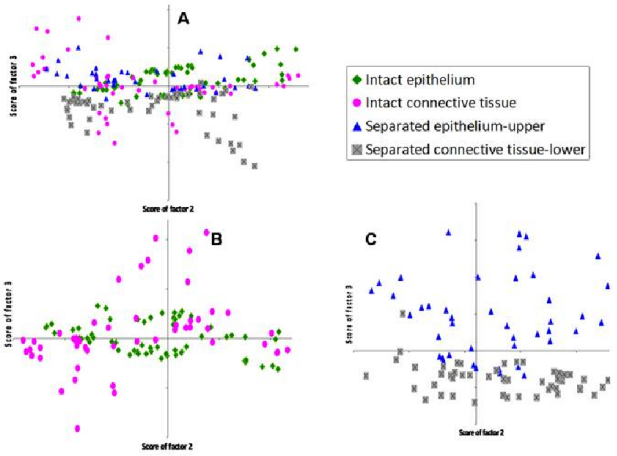


Figure 2.17: PCA of spectra acquired from four groups of normal oral tissues: (A) intact and incised biopsies; (B) intact tissues; (C) incised tissue

These results probably suggest that, spectral differences across intact tissues *('intact epithelium'* and *'intact connective tissue')* are minimal as compared to that of separated sections (*'separated epithelium-upper'* and *'separated connective tissue-lower'*). Further influence of surface orientation on classification with tumors was explored by analyzing 128 spectra acquired from 15 tumor biopsies. As can be seen from Figure 2.19, tumor spectra gave exclusive cluster suggesting surface orientation does not have any influence on classification.

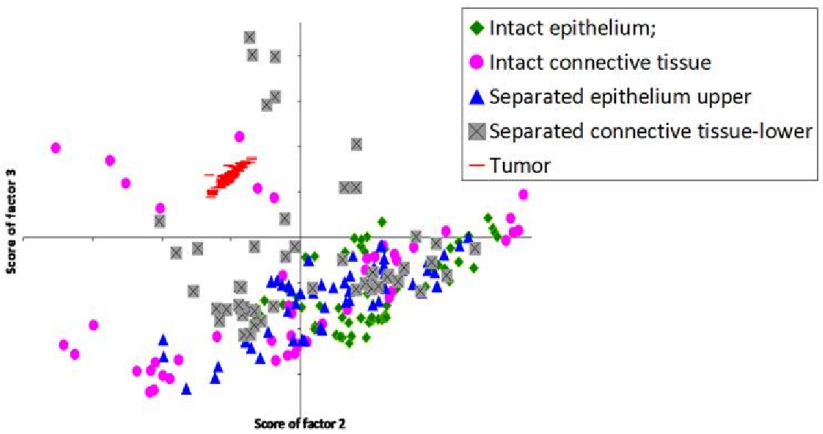


Figure 2.18: PCA of tumor spectra along with four groups of normal tissues

In the next step, supervised analysis using LDA was performed. Forty (40) factors accounting up to 95% correct classification were used (Figure 2.20A). As shown in Figure 2.20B, four clusters corresponding to *'intact epithelium',' intact connective tissue'*, *'separated epithelium-upper'* and *'separated connective tissue-lower*', respectively were obtained. Further analysis of scatter plot reveals that clusters corresponding to *'intact epithelium'* and *'intact connective tissue'* are closer while clusters of *'separated epithelium-upper'* and *'separated connective tissue-lower*' are placed far apart from each other.

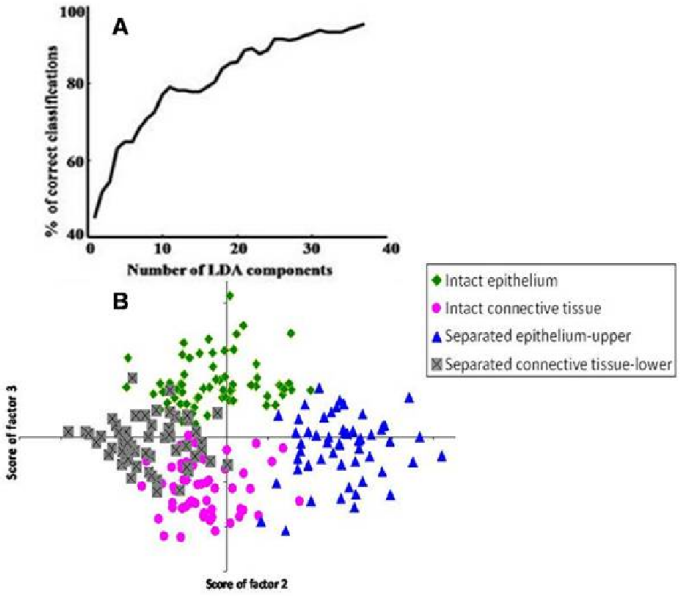


Figure 2.19: LDA of spectra from intact and incised biopsies: A. Scree plot B. Scatter plot

Summary of classification for all groups is summarized in confusion matrix shown in Table 2.IIA. In case of intact tissue, 65 out of 68 spectra of *'intact epithelium'* and 50 out of 53 spectra of *'intact connective tissue'* were correctly classified. In case of incised biopsies, 52 out of 54 spectra of '*separated epithelium-upper'* and 55 out of 59 spectra of '*separated connective tissue-lower'* were correctly classified. Overall, classification efficiency of 94% was observed. These findings were validated by LOOCV and results are shown in Table 2.IIB. As can be seen maximum misclassifications were observed for intact tissues. Of the 16 misclassifications of *'intact*

epithelium' spectra, 10 are with '*separated epithelium-upper'* 2 with '*intact connective tissue'* and 4 with *'separated connective tissue-lower*'. Among the 21 misclassifications of the *'intact connective tissue'* spectra, 9, 4 and 8 spectra classified as *'intact epithelium', 'separated epithelium-upper'* and *'separated connective tissue-lower*', respectively. But the situation is different when the spectra of separated sections were considered. In this case, most of spectra overlap with intact tissues *e.g.* of the 25 misclassifications of *'separated epithelium-upper'* only 4 are misclassified as '*separated connective tissue-lower*' group. Similarly, only 2 out of 16 spectra of '*separated connective tissue-lower*' group were misclassified as *'separated epithelium-upper'*. 21 and 14 misclassifications of *'separated epithelium-upper'* and '*separated connective tissue-lower*' were with *'intact epithelium'* and *'intact connective tissue',* respectively. This could be attributed to the fact that spectral features of intact tissues might have signals from the entire length of the tissue *i.e.* both the epithelial and connective tissue components, supporting the findings of PCA (Figure 2.18).

A	Intact epithelium	Intact connective tissue	Separated epithelium –upper	Separated connective tissue-lower
Intact epithelium	65	1	2	0
Intact connective tissue	2	50	1	0
Separated epithelium – upper	0	2	52	0
Separated connective tissue-lower	1	3	0	55
Leave-one-out cross validation				
B	Intact epithelium	Intact connective tissue	Separated epithelium –upper	Separated connective tissue-lower
Intact epithelium	52	2	10	4
Intact connective tissue	9	32	4	8
Separated epithelium – upper	10	11	29	4
Separated connective tissue-lower	7	7	2	43

Table 2.II: Summary of classification between intact and incised oral normal buccal mucosa spectra A: LDA, B: Leave-one-outcross-validation (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions)

Influence of surface orientations on classification with tumors was also evaluated by LDA. Similar to earlier observation (Figure 2.19), an exclusive cluster of tumor spectra was obtained (Figure 2.21). These results are presented in form of a confusion matrix shown in Table 2.III. As can be seen only 12 out of 128 (9%) spectra of tumor are wrongly classified, which can be attributed to previously described sample heterogeneity. Overall, findings of LDA corroborate with PCA results and suggest surface orientation of normal tissues does not have any bearing on classification with tumors.

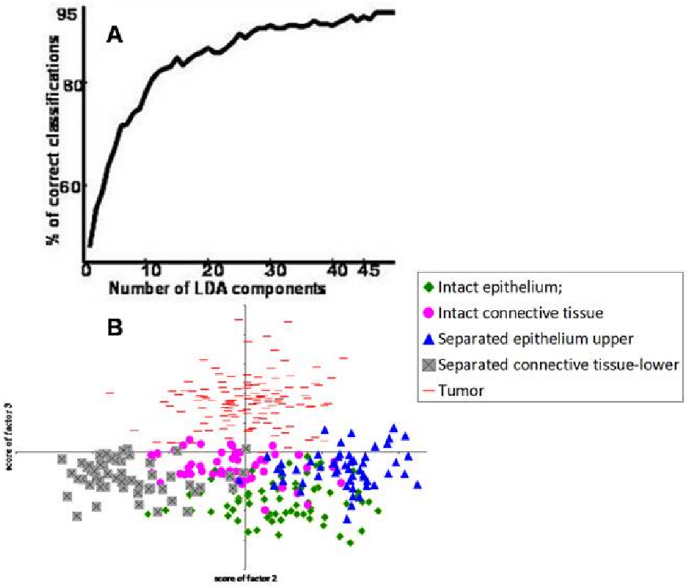


Figure 2.20: LDA of tumor spectra along with intact and incised biopsies: A. Scree plot B. Scatter plot

A	Tumor	Intact Epithelium	Intact connective tissue	Separated epithelium-upper	Separated connective tissue-lower
Tumor	127	0	1	0	0
Intact Epithelium	0	65	1	2	0
Intact connective tissue	0	2	50	1	0
Separated epithelium-upper	0	0	2	52	0
Separated connective tissue-lower	0	1	3	0	55
Leave-one-out cross validation					
B	Tumor	Intact Epithelium	Intact connective tissue	Separated epithelium-upper	Separated connective tissue-lower
Tumor	116	3	5	2	2
Intact Epithelium	0	51	3	9	5
Intact connective tissue	0	13	28	6	6
Separated epithelium-upper	0	12	12	28	2
Separated connective tissue-lower	0	4	3	2	50

Table 2.III: Summary of classification of intact and incised oral normal buccal mucosa against tumor spectra A:LDA, B:Leave-one-out cross-validation (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions)

Findings of the study suggest that architectural and morphological organization of tissue components are the hallmark of spectral signatures. Spectra obtained from upper and lower surface of intact oral buccal mucosal biopsy showed lipid and protein signatures due to histological arrangement of lipid and collagen molecules in the connective tissue. On the other hand, spectra from same surfaces after epithelium separation seem to be different but they overlap with intact tissue spectra. Clusters of intact tissues overlap among themselves while clusters from separated sections remain exclusive. Therefore, it can be assumed that spectra recorded from either surface will have features from entire volume of probing area. This is probably due to collection of signals even from deeper layers of tissue. These findings are in tune with existing literature [219,220]. However, surface orientations of normal tissue seem to have no bearing on classification with tumors.

2.V Correlation of the spectral data with biochemical composition of the tissue

This study was undertaken to correlate spectral features of normal and tumor oral tissues with the underlying biochemical composition. Spectral parameters derived from curve-deconvolution were correlated with biochemical estimations of total protein, total lipid and phospholipids content of the same tissues.

2.V.1 Methodology

2.V.1.a *Clinical samples:* Pathologically verified 20 pairs of buccal mucosa biopsies (tumor and normal) samples were collected in liquid nitrogen and stored at -800C until use, from biorepository, ACTREC.

2.V.1.b *Spectral acquisition:* Spectra were acquired using previously described fiberoptic probe coupled HE-785 Raman system (Figure 2.5). Tissues were passively thawed and kept on

CaF_2 window. The window was placed under illumination zone of the fiberoptic probe on a XYZ precision stage. Spectra were acquired at different points with an average spacing of ~2 mm. Spectra were acquired at 80 mW for 10 seconds and averaged over 5 accumulations. These conditions were kept constant during all measurements.

2.V.1.c *Spectral pre-processing*: Spectral pre-processing using Labspec 5.0 software was performed as per the previously described protocol (Section 2.II.1.c). Briefly, it involves correction for CCD response followed by subtraction of background signals and vector-normalization. Spectra interpolated in 1200-1800 cm^{-1} region were baseline corrected by fitting 5^{th} order polynomial function. These spectra were used as input for curve- deconvolution studies.

2.V.1.d *Total protein estimation:* Total protein content of tissues was estimated as per the procedure of Lowry *et al.*[247]. This method is based on the biuret reaction of aromatic amino acid tyrosine and tryptophan in proteins with Folin Ciocalteau's phenol reagent (FC reagent) leading to the formation of a blue color. Tumor and adjacent normal tissues were homogenized in appropriate buffer. One hundred (100) µl of either blank/ standard (5-40 µg BSA solution)/ sample were taken in test tubes. To this, 1 ml of freshly prepared CTC solution (0.8 N NaOH, 10% sodium dodecyl sulphate (SDS) in D/W) was added and the tubes were vortexed. After incubating for 10 min at room temperature, 500 µl of FC reagent (1:6 dilution) was added, the tubes were vortexed and incubated in the dark for 30 minutes at RT. All the samples and standards were run in duplicates. Absorbance at 750 nm was measured using a spectrophotometer (U-2001, Hitachi, Japan). The concentration of protein (expressed in per mg of tissue weight) in the samples was determined from the standard curve.

2.V.1.e *Total lipid estimation:* The total lipid content of the tissues was estimated using the protocol of Floch *et al*. [248]. In this method, firstly the tissue was homogenized in

chloroform/methanol (2/1) mixture in a final volume 20 times to the volume of the tissue sample (*e.g.*1g in 20 ml of solvent mixture). After dispersion, the whole mixture was agitated for 15-20 minutes in an orbital shaker at room temperature. The liquid phase of the homogenate was recovered after centrifugation at 1000 rpm for 15 minutes. This was followed by washing with 0.2 volumes (4 ml for 20 ml) of 0.9% sodium chloride (NaCl) solution. After vortexing the mixture was centrifuged at low speed (2000 rpm) to separate the two phases. The upper phase was removed by siphoning. The remaining solvent along with the interface was washed with methanol/water (1/1) solution without mixing the whole preparation. The lower chloroform phase containing lipids was evaporated under a nitrogen stream and lipid content was expressed per mg tissues weight.

2.V.1.f *Phospholipid estimation:* Estimation of phospholipids was performed by the method of Rouser *et al.* [249]. This method is based on the principle of converting the organic phosphorus (P) to inorganic phosphorus (P), which on reaction with ammonium molybdate, forms phosphomolybdic acid. This in turn, on reduction with ascorbic acid, forms a stable color which can be read optically at 800 nm. Dried total lipid samples and standards (1 to 5 µg P/tube) were dissolved in 0.65 ml perchloric acid and placed in heated block for about 30 minutes or until the yellow color disappears. After cooling, 3.3 ml water, 0.5 ml of molybdate solution and 0.5 ml of ascorbic acid solution were added to the tubes and tubes were agitated on a vortex after each addition. The tubes were placed in a boiling water bath for 5 minutes and the absorbances of cooled samples (including the standards) were read at 800 nm. Amount of phospholipids were calculated using standards curve and expressed in per mg of tissue weight.

2.V.1.g *Curve deconvolution*: As already mentioned in Chapter 1 under the computational technique section, curve deconvolution or peak fitting is a method of computing the intensity associated with a band by resolving spectra in to individual components. Nonlinear peak

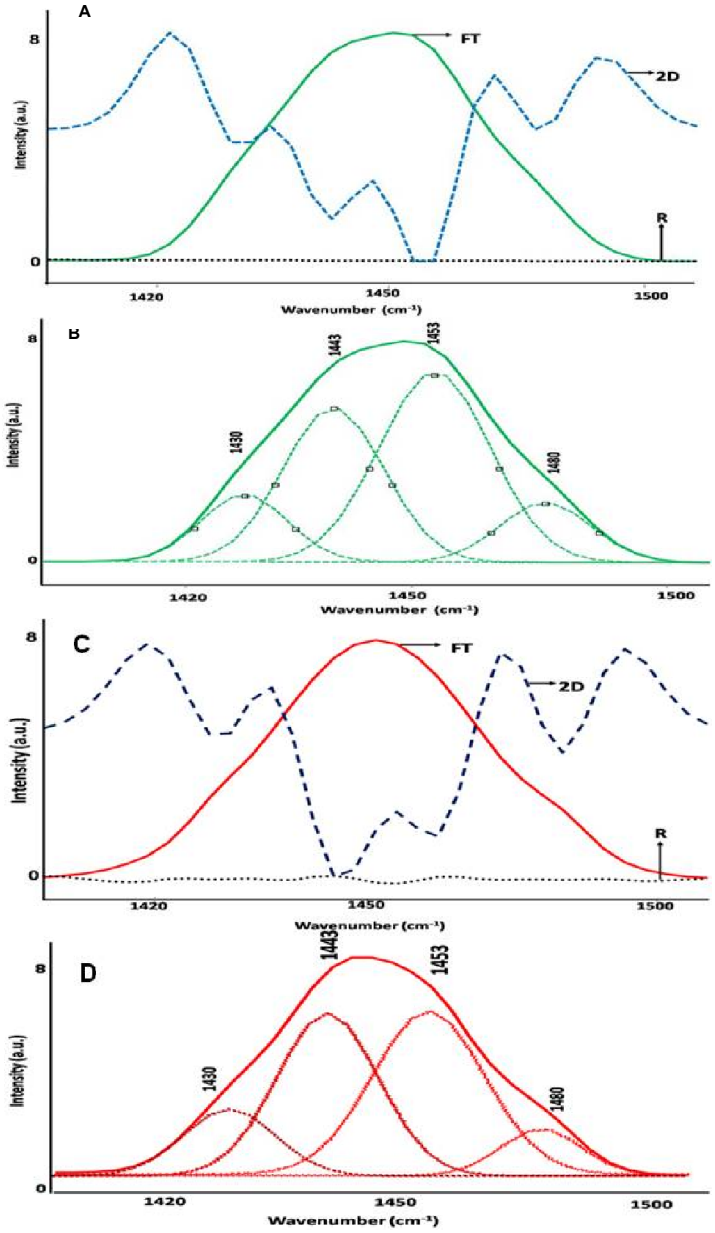
fitting method based on Lavenberg-Marquardt algorithm was employed [109-111]. In this method, the sum of the squared differences between observed and computed spectra are minimized to get the best fit. Spectra in the 1200 to 1800 cm^{-1} region were used for analysis. Normal and tumor spectra were resolved into 8 and 9 bands, respectively. Each band was characterized by four parameters: shape factor, peak position, peak intensity and full width at half maximum (FWHM). Scatter intensity plots of lipid (1440 cm^{-1}) and protein bands (1450 and 1660 cm^{-1}) were generated to understand the relative contributions of these biomolecules. In our curve analysis, the Gaussian band led to a good fit. The fitted spectra (FT), residual (R) depicting differences between the original and fitted spectra, and 2nd derivative spectra (2D) depicting exact peak positions were also generated.

2.V.1.h *Statistical analysis*: The data were expressed as mean $\pm$ SD and statistical comparisons were performed by unpaired student 't' test. p value less than 0.05 was considered as statistically significant.

2.V.2 Results and Discussion

Spectra of normal and tumor tissues were interpolated to specific regions and subjected to curve-deconvolution analysis. Four bands at 1430 cm^{-1}, 1443 cm^{-1}, 1453 cm^{-1} and 1480 cm^{-1} in δCH_2 region (1400-1500 cm^{-1}) were observed in case of normal and tumor spectra (Figure 2.21A-D). Origin of band at 1440 cm^{-1} has been attributed to CH_2 and CH_3 bending modes of lipids. Area associated with this band was computed, averaged and subjected to unpaired student 't' test analysis to determine the difference between normal and tumor tissues. Statistically significant ($p < 0.0001$) difference between the average intensity of the lipid band (1440 cm^{-1}) in normal (1.42 ± 0.25) and tumor (0.43 ± 0.18) tissues was observed (Figure 2.21E).

Origin of band at 1450 cm^{-1} has been assigned to C-H deformation modes of amino acid side chains of the proteins. To correlate protein rich features of tumor tissues, area associated with this band was computed, averaged and subjected to unpaired student 't' analysis. Similar to earlier observation, significant ($p < 0.0001$) difference between average values of area associated with 1450 cm^{-1} band of normal (0.51 ± 0.12) and tumor (1.46 ± 0.29) tissues was observed (Figure 2.21F).



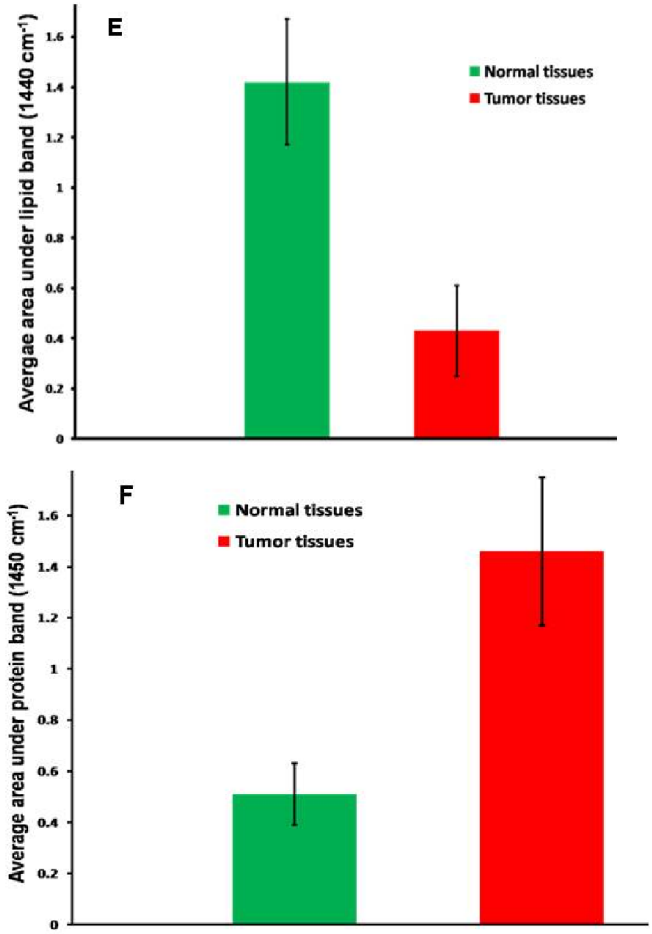
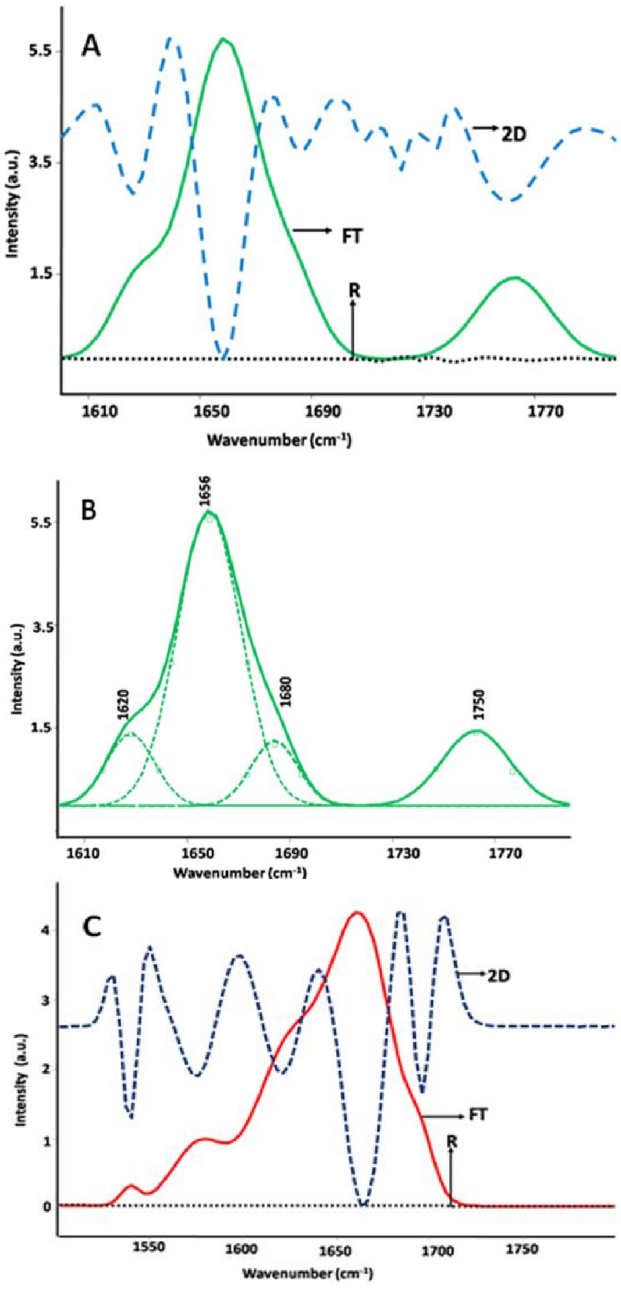


Figure 2.21: Curve deconvolution of mean normal (A-B) and tumor (C-D) spectra in δCH_2 region (2D- 2nd derivative; FT- Fitted spectrum; R- Residual) E-average area under 1440 cm^{-1} band; F- average area under 1450 cm^{-1} band

In next step, curve-fitting in the amide I region of normal and tumor spectra was performed. Four bands at 1620 cm^{-1}, 1656 cm^{-1}, 1680 cm^{-1} and 1750 cm^{-1} were fitted in the amide I region (1600-1800 cm^{-1}) of normal spectrum (Figure 2.22A-B). Five bands at 1540 cm^{-1}, 1590 cm^{-1}, 1628 cm^{-1}, 1655 cm^{-1} and 1690 cm^{-1} were fitted in the mean tumor spectrum in 1500-1800 cm^{-1} range (Figure 2.22C-D). Vibrational modes around 1655 cm^{-1} has been attributed to amide I (C=O stretch) of proteins. Intensity associated with this band was calculated, averaged and subjected to unpaired student 't-test' analysis. Statistically significant difference (p value <0.0001) was observed between

the average intensity associated with 1660 cm^{-1} bands in normal (0.89±0.28) and tumor (1.12±0.19) tissues (Figure 2.22E).



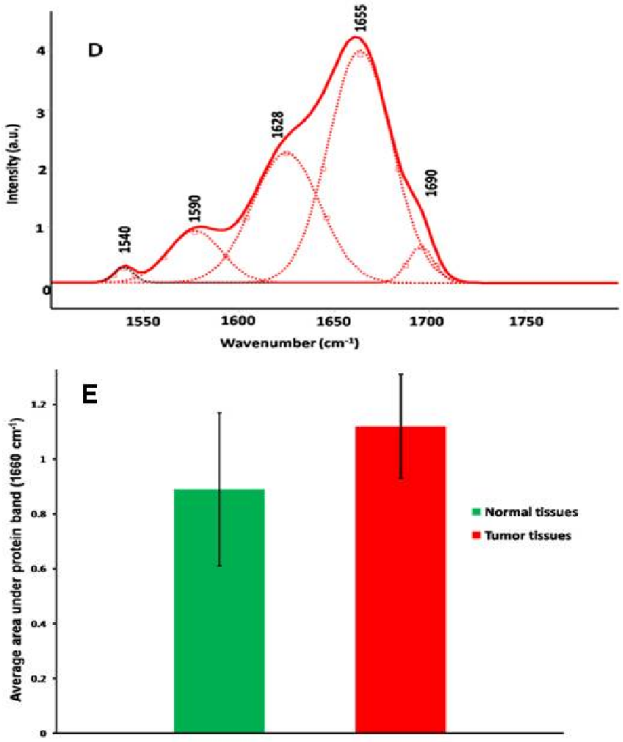
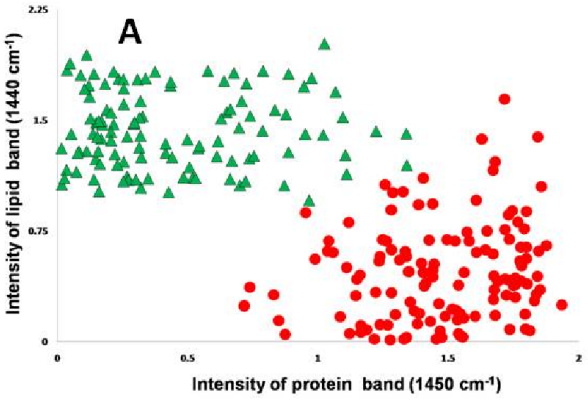


Figure 2.22: Curve deconvolution of mean normal (A-B) and tumor (C-D) spectra in amide I region (2D-2nd derivative; FT- Fitted spectrum; R- Residual) E-average area under 1660 cm^{-1} band

Scatter intensity plots were generated to explore the feasibility of classification between normal and tumor tissues. Plots of lipid band (1440 cm^{-1}) against protein bands (1450 cm^{-1}&1660 cm^{-1}) are shown in Figure 2.23A,B. Corroborating earlier observations, two separate clusters indicating higher intensity of lipid band (1440 cm^{-1}) in normal tissues in comparison to tumors were obtained. Similarly, the intensity of the protein band (1450 cm^{-1} and 1660 cm^{-1}) was high in case of tumors as compare to normal tissues.



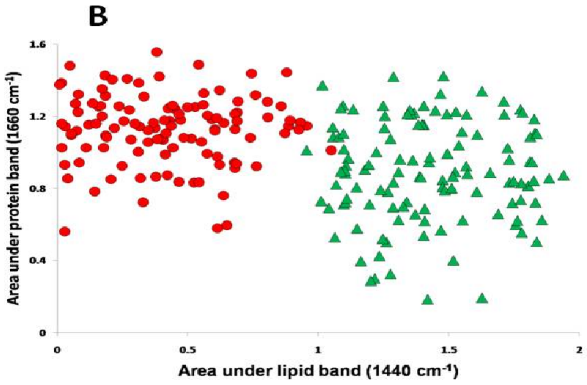


Figure 2.23: Scatter intensity plots of lipid (1440 cm^{-1}) and protein (1450 cm^{-1} and 1660 cm^{-1}) bands of normal and tumor tissues

Further band intensity of lipid and protein bands were correlated with biochemical composition of tissues by estimating amount of total lipid, total protein and phospholipid content from same tissues. The protein to lipid and phospholipid ratio was calculated and plotted. As shown in Figure 2.24A, in accordance with the spectral features, protein to lipid ratio was high in case of tumors (2.15±0.41) in comparison to normal (0.72±0.22). The protein to phospholipid ratio yielded similar results *i.e.* high in tumors (24.13±2.12) and low in normal (16.12±2.28) (Figure 2.24B). The differences with respect to protein-lipid and protein-phospholipid ratio between both groups were found to be statistically significant ($p < 0.05$).

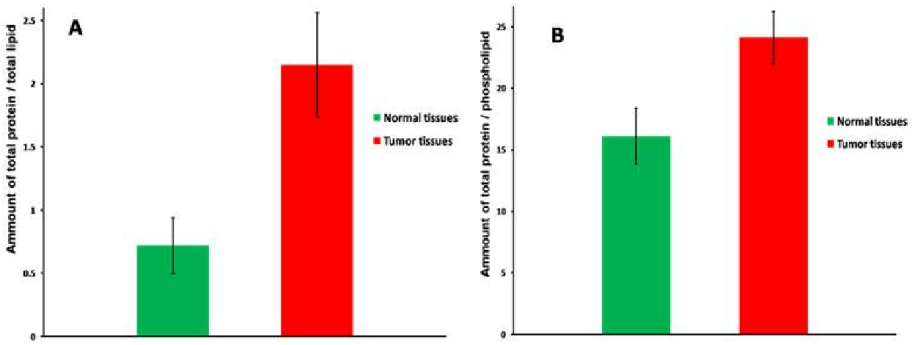


Figure 2.24: Plot of protein to lipid (A) and phospholipid (B) ratio of normal and tumor tissues

Lipids are the main constituents of cell membrane which are required for maintaining the cell shape and regulation of various cell processes. The alterations in membrane lipids are important aspects of malignant transformation. Lipids**,** particularly polyunsaturated fatty acids, are the major class of biomolecules susceptible to oxidative damage by reactive oxygen species generated by tobacco exposure. This could be one of the reasons for the decreased amounts of lipids in tumors, as suggested by Raman spectra and confirmed by biochemical estimation. A decrease in the phospholipid concentration can be attributed to dysfunction of the cell membrane leading to an increase in its degradation. The findings of this study indicate that spectral features are hallmark of the biochemical composition of a tissue. The presence of tumor in the body produces a number of deleterious effects such as anorexia, nausea, impaired digestion and cachexia. Malignant cells are known to differ biochemically in many ways from normal cells. Cancer cells show a variety of alterations on the cell surface and also display disturbed membrane, as compared to their normal counterparts. Spectral features as well as biochemical estimation suggest that the lipid to protein ratio is high in normal tissue and low in case of tumors. Overall, findings of curve-deconvolution and biochemical estimation correlate very well and corroborate spectral profile noted in earlier studies.

Summary: Studies discussed in this Chapter were carried out as standardization steps before initiating the *in vivo* measurements. Following is the brief summary of the work presented in this Chapter:

1. A fiberoptic probe coupled *in vivo* Raman spectroscope was procured and assembled. This set up was utilized for *ex vivo* measurements by attaching probe holder and XYZ precision

stage. Same instrument was utilized as a Raman microprobe by attaching a microscopic objective.

2. Data acquisition and analysis protocols were standardized by acquiring spectra of *ex vivo* normal and tumor tissues. Reproducibility of spectral features was established. Lipid rich features were observed in normal spectra while proteins were predominant in tumor spectra. Standard models of normal and tumor conditions with predictions efficiency of 80 and 96 %, respectively were developed.
3. Misclassifications between normal and tumor tissues were analyzed by correlating spectral predictions with their histopathology. Findings suggest that misclassification between both groups can be primarily attributed to the tissue heterogeneity *i.e.* presence of normal regions in a tumor biopsy and *vice-versa.*
4. Raman signals of normal tissues were analyzed by acquiring spectra of intact and incised oral biopsies. Findings demonstrated that morphological and architectural arrangements of different layers in a tissue contribute to the spectral signatures. Influence of surface orientation of normal tissues on classification with tumors was also assessed and it was found that orientation does not have any bearing on classification with tumor.
5. Lipid and protein rich spectral features of normal and tumor tissues were correlated with biochemical estimations. Spectral parameters derived from curve-deconvolution analysis were found to be highly correlating with biochemical measurements.

Chapter 3

In vivo Raman spectroscopy of oral cancers

Introduction

Standardization of data acquisition and analysis methodologies on *ex vivo* tissues were described in Chapter 2. In the present Chapter, findings of *in vivo* Raman spectroscopic methods in oral cancer diagnosis are presented. In the first section, studies pertaining to classification of normal, cancerous and premalignant conditions are discussed. This is followed by discrimination of healthy controls with and without tobacco habits and classification of oral sub-mucous fibrosis (OSMF) and leukoplakia, two of the most common precancerous conditions in Indian population. In the second section data on Raman spectroscopic identification of "cancer field effects or malignancy associated changes" the earliest event in oral carcinogenesis are presented.

3.I. Identification of normal, cancer and premalignant conditions

Visual examination followed by biopsy of the potential cancerous and precancerous lesions is the standard diagnosis procedure for oral cancers. However, the invasiveness of the procedure and fact that often most of the practitioners and patients are reluctant to perform a confirmatory biopsy limits its applicability for diagnosis of early lesions [90]. Non-invasive and objective diagnosis using optical spectroscopic techniques is being actively pursued as novel alternative for biopsy. This study was undertaken to demonstrate the feasibility of acquiring good quality *in vivo* Raman spectrum under clinically implementable time and to explore classification between normal, cancerous and precancerous conditions in the oral buccal mucosa.

3.I.1 Methodology

3.I.1.a *Clinical samples:* The study was approved by Institutional ethics committee, Tata Memorial Hospital, Mumbai, India (Project number: 605). Subjects with proper mouth opening and referred for surgery at ACTREC were recruited for the study, after obtaining an informed written

consent. Clinico-pathological details such as tobacco usage, age, sex and tumor grade of all subjects was obtained from electronic medical record (EMR) of Tata Memorial Hospital, Mumbai, India. History of all subjects was documented to ensure cause of cancer along with type and duration of tobacco habits. Subjects with only tobacco related cancers were recruited in the study. *In vivo* spectra of contralateral normal (opposite side of tumor) and cancerous lesion (tumor) were acquired from 113 subjects (101-male, 12-female) with tobacco associated, pathologically verified oral squamous cell carcinoma (OSCC). Spectra were also recorded from premalignant patches on the contralateral side of the 50 (male) oral cancer subjects. Median age of subjects with cancerous lesion and premalignant lesion were 51 and 49 years, respectively. Fifteen (15) age-matched healthy controls (13-male, 2-female) without any history of tobacco habit were recruited as negative controls. Average age of these subjects was 51 years. Fifteen (15) subjects (14-male, 1-female) with similar and long-term tobacco habit were also recruited. Screening criterion of tobacco habit for more than 10 years was applied for recruitment of these subjects and average time of tobacco consumption was ~14.5 years. Median age of these subjects was 50 years. These details are summarized in Table 3.I. The cancerous lesions were diagnosed clinically and verified histopathologically by incisional biopsy from the tumor site. Normal mucosa *i.e.* healthy controls, contralateral normal, habitual tobacco users, and premalignant patches were verified by clinical assessment conducted by a trained senior oral pathologist. No biopsy was taken from normal mucosa owing to ethical limitations.

Sr. No.	Category	Site of acquisition	Total no. subjects (M&F)	Median age (yrs)	Cancer	Tobacco habit	Total no. of spectra
1.	Contralateral Normal	Buccal mucosa	113 (101M, 12-F)	51	YES	YES	444
2.	Tumor	Buccal mucosa	113 (101M, 12-F)	51	YES	YES	337
3.	Premalignant patches	Buccal mucosa	50 (50 M)	49	YES	YES	206
4.	Healthy controls	Buccal mucosa	15 (13M, 2F)	51	NO	NO	150
5.	Habitual tobacco users	Buccal mucosa	15 (14M, 1F)	50	NO	YES	150

Table 3.I: Subject accrual details

3.I.1.b *Instrumentation:* *In vivo* spectra were acquired using the Inphotonics probe coupled HE-785 instrument described in Chapter 2 (Section: 2.I). Raman set-up used for acquiring *in vivo* spectrum is shown in Figure 3.1.

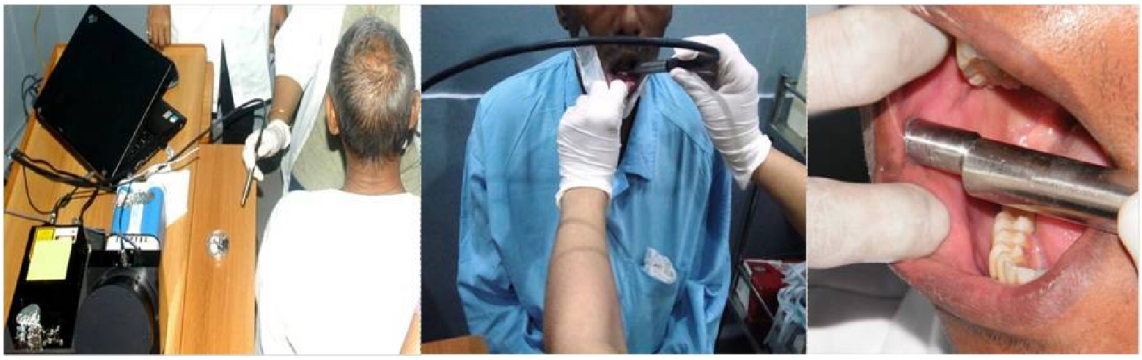


Figure 3.1: Raman spectroscope utilized for *in vivo* measurements

Previously described commercially available Inphotonics fiberoptic probe (Figure 2.4) was adapted for *in vivo* measurements. The working distance of the probe is 5 mm. To maintain a constant focus during all measurements a metallic, detachable spacer of length 5 mm was attached at the tip of the probe. The photographic representation of the original and modified fiberoptic probe is shown in Figure 3.2**.** The attachment of spacer provided flexibility during *in vivo* measurements and since it can be disinfected, patient to patient contamination was also avoided.

Figure 3.2 : Fiberoptic probe used for *in vivo* measurements (A-original, B-with spacers)

3.I.1.c *Spectral acquisition:* Previous *in vivo* studies have demonstrated efficacy of Raman spectroscopy in identifying site-wise differences in the oral cavity [213,214]. To exclude spectral differences associated with different sites, spectra in the present study were acquired only

from buccal mucosa, both on right and left sides. In order to ensure uniformity, spectra were collected as per the teeth positions. Buccal surfaces opposing canine, first premolar, second premolar, first molar and second molar both on right and left side were considered as reference point and spectra were acquired. Pictorial representation of sites of spectral acquisition is presented in Figure 3.3**.**

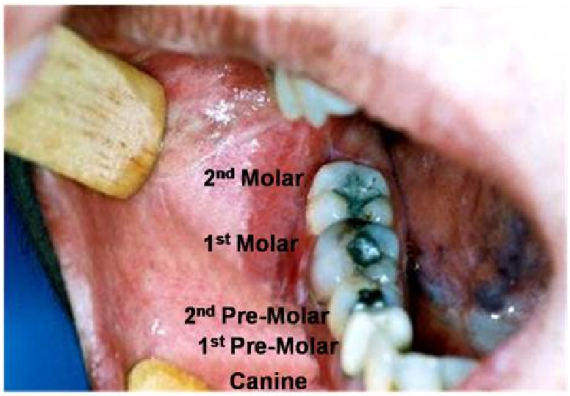


Figure 3.3: Pictorial representation of spectral acquisition sites on buccal mucosa

To avoid any differences because of mouth environment, subjects were requested to wash their mouth with water before spectral acquisition. To avoid contamination among subjects, spacers were disinfected with CIDEX (Johnson and Johnson, Mumbai, India) solution after each acquisition. Spectral acquisition parameters were: laser power-80 mW, integration- 3 seconds and averaged over 3-accumulations. These parameters were kept constant during all measurements. On an average ~8 spectra (4 from contralateral mucosa and 4 from tumor) from 113 subjects with oral cancer corresponding to 444 contralateral normal and 337 tumor spectra were obtained. A total of 206 spectra from premalignant patches on contralateral side of 50 subjects with OSCC were also acquired. For 30 healthy controls (15 with and without tobacco habits each), 10 spectra from right and left buccal mucosa corresponding to 150 each from habitual tobacco users and healthy controls were acquired. These details are presented in Table 3.I.

3.I.1.d *Spectral pre-processing:* Spectral pre-processing was performed as per the already described protocol including correction for CCD response followed by subtraction of background signals acquired under similar acquisition conditions (Section:2.II.1.c**)**. To minimize the influence of slow moving background first derivative of spectra were computed using Savitzky-Golay method (window size-3). Relative intensity associated differences across different groups were corrected by vector normalization. First derivative, vector normalized spectra were used as input for multivariate analysis. Typical *in vivo* spectrum at different pre-processing stages is shown in Figure 3.4.

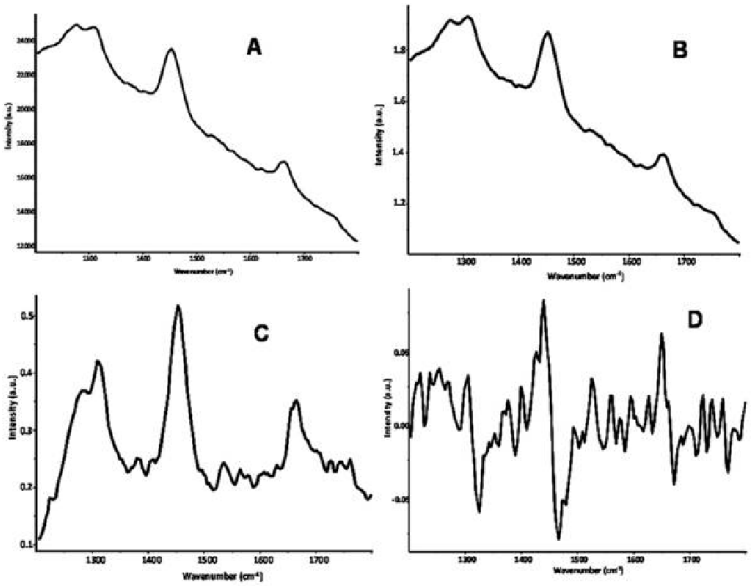


Figure 3.4: Typical *in vivo* spectrum at different pre-processing steps A. Raw spectrum B. CCD response corrected C. Background corrected D. First Derivative

3.I.1.e *Spectral comparisons:* Average or mean spectra were computed by averaging variations on Y-axis keeping X-axis constant from background subtracted underivatized spectra for each class. These spectra were baseline corrected by fitting a 5^{th} order polynomial function and were used for computing difference spectrum as well as comparisons across different groups.

3.I.1.f *Statistical analysis:* PC-LDA method using algorithms implemented in MATLAB based in-house software was utilized to explore feasibility of classification among different groups [216]. Different spectral regions like full-range, fingerprint and high-wave-number were explored for classification. Best classification was achieved in 1200-1800 cm^{-1} region. Further, this region is least influenced by fiber interference. Findings of PC-LDA were evaluated using LOOCV and independent test data prediction.

3.I.2 Results and discussion

3.I.2.a *Spectral features:* Mean baseline corrected spectra of contralateral normal, tumor and premalignant patches sites along with standard deviations are shown in Figure 3.5 A-C. Spectra of contralateral normal show strong lipid features indicated by C=O band of esters (1744 cm^{-1}), strong δCH_2 bend (1442 cm^{-1}), sharp features in amide III region and sharp peak in amide I region (1660 cm^{-1}), Figure 3.5A. Tumor spectra are dominated by protein bands indicated by broad amide III, broad and shifted δCH_2 and broad amide I and loss of ester band, Figure 3.5B. Spectra of premalignant patches show features of both normal and tumor conditions. Minor differences such as broadening of amide III, red shifted amide I and δCH_2 were observed in comparison to contralateral normal. In contrast to tumor spectra, prominent ester (1740 cm^{-1}) and sharp amide I bands were observed in spectra of premalignant conditions, Figure 3.5C.

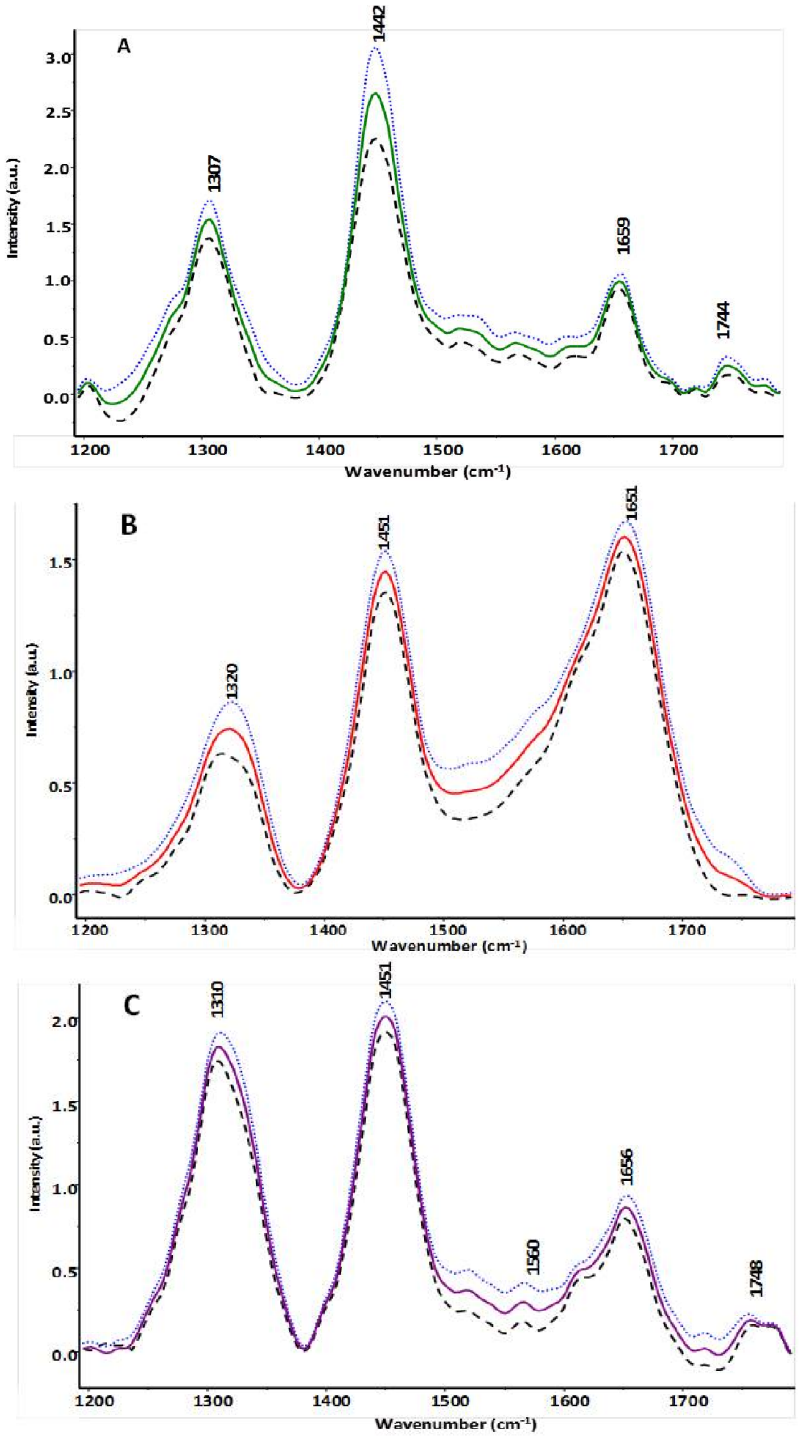


Figure 3.5: Mean spectra along with standard deviations A. Contralateral normal B. Tumor C. Premalignant. [Solid line: mean spectra, Dotted line: mean+ standard deviation, Broken line: mean – standard deviation

Difference spectra were computed to understand similarities and dissimilarities across normal, cancer and premalignant groups. Tumor and premalignant spectra were subtracted from contralateral normal. In this case all positive bands belongs to contralateral normal while negative bands are from pathological conditions *i.e.* tumor and premalignant. As shown in Figure 3.6A, strong lipid features indicated by positive bands at 1741 cm^{-1}, 1441 cm^{-1}, amide III were observed in difference spectrum of contralateral normal. Negative bands around 1340 cm^{-1} and 1658 cm^{-1} which are suggestive of predominant DNA and protein features were observed in tumors. It is a well known fact that in pathological conditions there is loss in architectural arrangement of different layers therefore; loss of lipid features is expected as contents of different layers are mixed. Additionally, pathological conditions cells have large amounts of surface proteins, receptor proteins, enzymes, antigens, and anti-bodies which may give rise to a protein-dominated spectrum. Difference spectra generated by subtracting premalignant spectra from contralateral normal is shown in Figure 3.6B**,** strong negative DNA band (1340 cm^{-1}) along with minor differences around δCH_2(1442 cm^{-1}), amide I (1660 cm^{-1}) region were observed. Increase of DNA in pathological conditions can be associated with higher hypercellularity or inflammatory index. Spectra of tumor was also subtracted from premalignant, in this case positive bands belongs to premalignant while negative bands are from tumors, Figure 3.6C. Similar to contralateral normal difference spectra, lipid rich features indicated by positive bands around 1740 cm^{-1}, 1305 cm^{-1} and 1440 cm^{-1} were observed in case of premalignant. Negative band around 1660 cm^{-1} indicate high protein content in tumors as compare to premalignants.

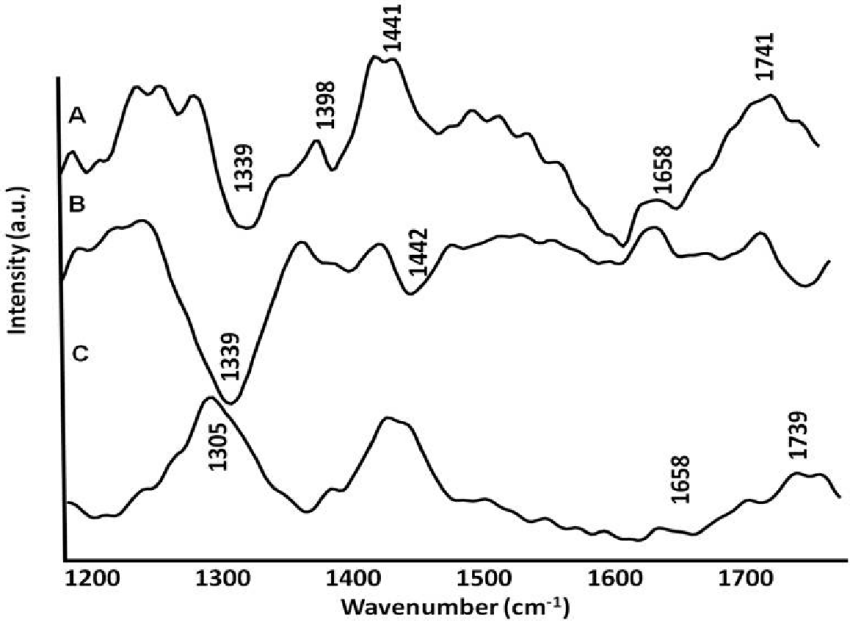


Figure 3.6: Comparison of difference spectra across different groups: A. contralateral normal – Tumor; B. contralateral normal – Premalignant; C. Premalignant – Tumor.

All the vibrational modes were assigned based on available literature [221,222]. These findings corroborate with earlier reports and are in tune with basic understanding of neoplastic processes [213,214]. As it is well known, precancerous and cancerous conditions are fundamentally more cellular and hence biochemically have higher level of proteins. In contrast to cancers premalignant lesions contain more stromal components, which might explain relatively high lipid content observed in premalignant tissues. The observed spectral features can provide vital clues in understanding the differences in the biochemical composition of normal, precancerous and cancerous oral tissues, which may be used for optical diagnosis.

3.I.2.b *Classification of normal, premalignant and tumors:* Above mentioned spectral variations were further utilized for classification by employing PC-LDA. Standard models were developed using 170 contralateral normal (40 subjects), 192 tumor (40 subjects) and 113 premalignant spectra (24 subjects). First derivative and vector normalized spectra were used as input for multivariate analysis. Fifteen (15) factors accounting up to ~92% correct classification

were used for PC-LDA, Figure 3.7A. Scatter plot generated after plotting score of factor 2 and 3 is shown in, Figure 3.7B**.** Three clusters belonging to normal, premalignant and tumor, respectively were obtained.

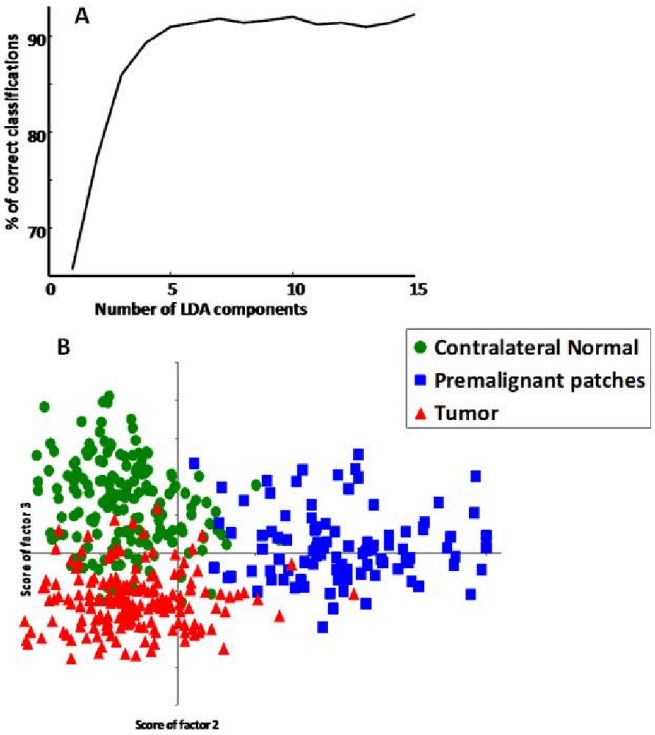


Figure 3.7: PC-LDA of contralateral normal, premalignant and tumor conditions A. Scree plot B.

These results are summarized in form of a confusion matrix presented in Table 3.IIA. Here diagonal elements are true-positive while ex-diagonals indicate false positive predictions. As can be seen that 151/170 (89%), 106/113 (94%) and 181/192 (94%) spectra of contralateral normal, premalignant and tumors, respectively were correctly classified. None of the premalignant spectra misclassified with tumor. LOOCV was performed to evaluate efficiency of these models. As shown in Table 3.IIB, 146 of 170 (86%) spectra from contralateral normal site were correctly classified. Of the 24 misclassifications 22 were with tumors and remaining 2 with premalignants. In case of tumors, 176 out of 192 (92%) spectra were correctly classified. Of the 16 misclassifications, 12 were with

contralateral normal and 4 with premalignant. In case of premalignant, 103 out of 113 (91%) spectra were correctly classified. Of the 10 misclassifications, 8 were with contralateral normal and remaining 2 with tumor.

A.	Contralateral Normal	Premalignant patches	Tumor
Contralateral Normal	151/170	1	18
Premalignant patches	7	106/113	0
Tumor	9	2	181/192

Leave-one-out cross-validation

B.	Contralateral Normal	Premalignant patches	Tumor
Contralateral Normal	146/170	2	22
Premalignant patches	8	103/113	2
Tumor	12	4	176/192

Independent test data prediction

C.	Contralateral Normal	Premalignant patches	Tumor
Contralateral Normal	219/274	18	37
Premalignant patches	25	56/93	12
Tumor	8	17	156/181

Table 3.II: Summary of classification between *in vivo* Raman spectra of normal, premalignant and tumor conditions: A. Standard model, B. LOOCV, C. Independent test data prediction. (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions)

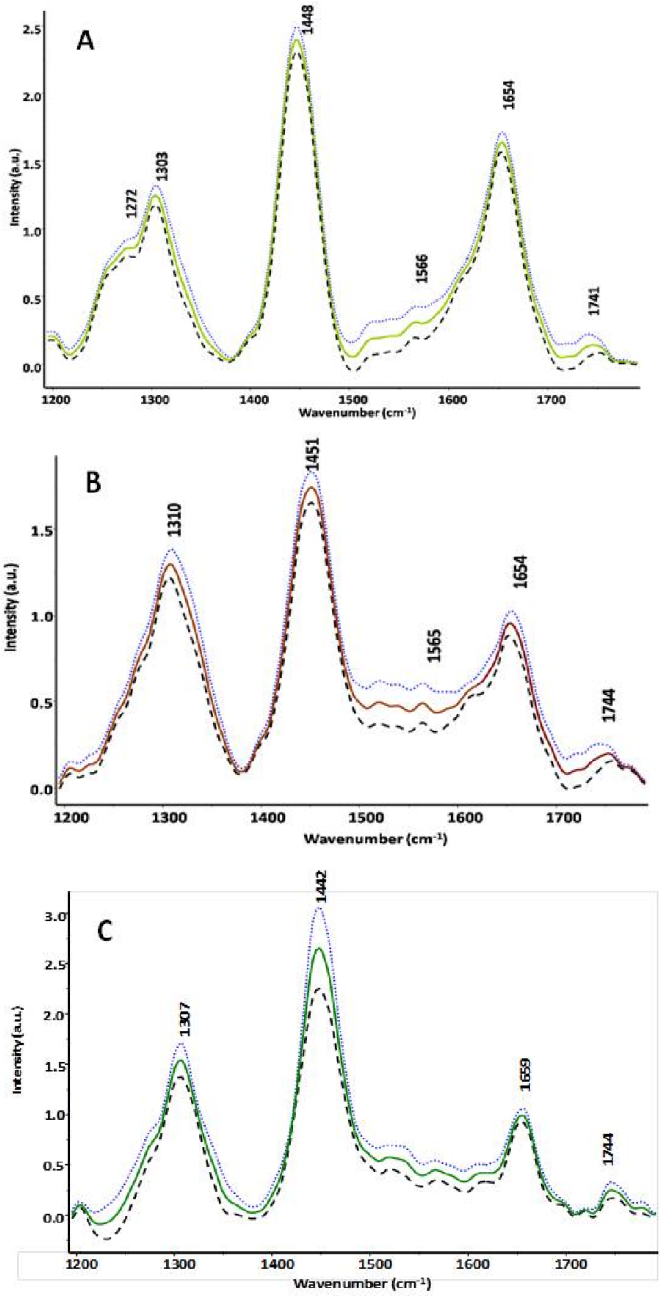
Further, test prediction efficiency of the classifier models was evaluated by an independent test data set comprising of 274 contralateral normal (73 subjects), 181 tumor (73 subjects) and 93 premalignant (26 subjects) spectra. As shown in Table 3.IIC, of the 274 test contralateral normal spectra, 219 were correctly predicted, and of the 55 wrong predictions, 18 were with premalignant and remaining 37 with tumor. Similarly, 156 out of 181 test tumor spectra were correctly identified. Among the wrong predictions, 17 were premalignant and 8 were contralateral normal. In case of premalignants, 56 out of 93 spectra were correctly predicted. Among the 37 misclassifications, 25 were contralateral normal and remaining 12 were tumor. Prediction efficiency of 79, 60 and 86% for standard models of contralateral normal, premalignant and tumor, respectively was observed.

Misclassifications between different groups can be primarily attributed to mucosal heterogeneity. Highest prediction efficiency was observed in case of tumors (86%) followed by contralateral normal (79%) and premalignant (60%). Misclassification of tumor spectra as contralateral normal can be explained on the basis of tumor heterogeneity. Spectra in the present study were recorded at different points therefore, possibility of acquiring spectra from normal or inflammatory patches in a tumor cannot be completely ruled out. Maximum misclassification was observed between contralateral normal and premalignant spectra. This is probably due to the fact that premalignant patches in the study were from contralateral side. Further, our probing area is around 100-200 μm, since transformation of a premalignant zone may not be uniform, possibility of acquiring data from a normal site cannot be completely ruled out. This also explains observed misclassification across premalignant and malignant, as numbers of instances are very few as malignant conditions represent higher degree of transformation as compared to premalignant.

3.I.2.c *Classification of healthy controls with and without tobacco habits:* Tobacco chewing and smoking are the main etiological factors associated with oral cancer. Long-term exposure of tobacco related carcinogens to the mucosa of healthy individual results in morphological and biochemical changes leading to the development of premalignant lesions, which ultimately transform into cancer. An ideal diagnostic/screening method should be able to discriminate between premalignant lesions and closely associated tobacco induced invisible changes in the mucosa. Therefore, in the next step, spectra of premalignant patches were analyzed against normal contralateral mucosa exposed to tobacco, healthy controls with and without tobacco habits.

Mean baseline corrected spectra of healthy controls, habitual tobacco users, contralateral normal and premalignant patches are shown in Figure 3.8A-D. Spectra of healthy controls suggest strong lipid features indicated by ester band (1741 cm^{-1}),sharp amide I (1654 cm^{-1}), δCH_2 (1448 cm^{-}

[1]), bands around 1272 cm^{-1} and 1303 cm^{-1}, Figure 3.8A. Almost similar spectral profile indicating predominant lipid features was observed for contralateral normal and habitual tobacco users. Minor differences such as shift in amide III and δCH_2 band and broadening of the amide I region suggesting changes in protein secondary structures were observed with respect to healthy controls, Figure 3.8B,C.



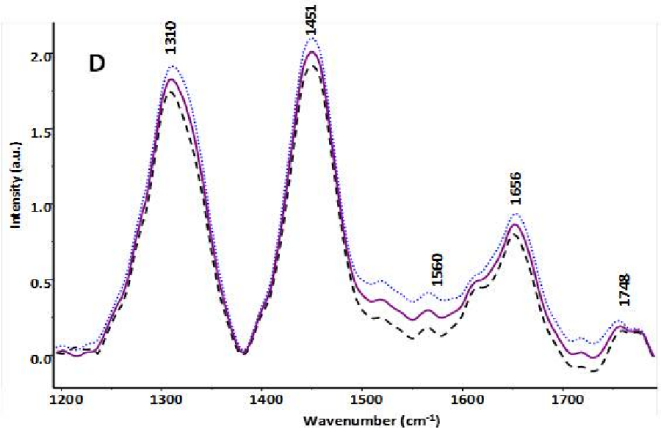


Figure 3.8: Mean spectra along with standard deviations A. Healthy controls without tobacco habit, B. Habitual tobacco users, C. Contralateral normal, D. Premalignant. (Solid line: mean spectra, Dotted line: mean+ standard deviation, Broken line: mean – standard deviation)

Difference spectra were computed to explore spectral similarities and dissimilarities across all groups. Spectra of contralateral normal, habitual tobacco users and premalignant patches were subtracted from healthy controls**,** Figure 3.9A-C**.** In this case positive bands belong to healthy controls and negative are from pathological conditions *i.e.* premalignant, contralateral normal and habitual tobacco users. Negative band around 1340 cm^{-1} indicating DNA was observed in all pathological conditions. Origin of this band can be attributed to long term tobacco associated hypercellularity or inflammation induced changes in the mucosa [223]. Intensity of this band was highest for premalignants in comparison to other groups, as it represents higher degree of cellularity (Figure 3.9A). Almost similar intensity of the DNA band suggesting similar level of hypercellularity was seen in contralateral normal and habitual tobacco user difference spectrum, Figure 3.9B,C. Positive δCH_2 (1440 cm^{-1}) , amide I and ester (1740 cm^{-1}) of healthy control suggests strong lipid features in comparison to other groups.

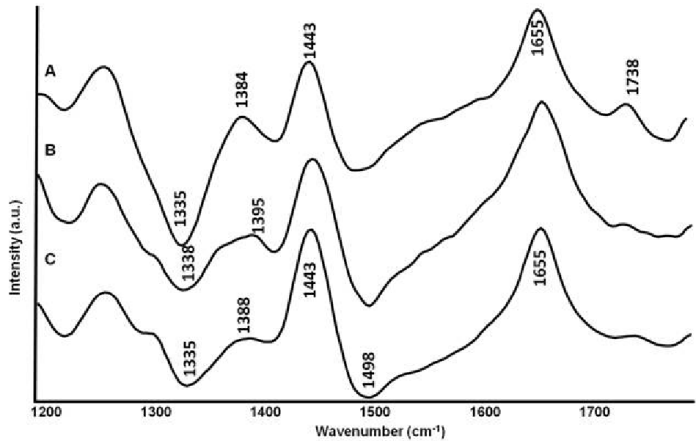


Figure 3.9: Comparison of difference spectra across different groups: A. healthy control – premalignant, B. healthy control – contralateral normal, C. healthy control – habitual tobacco users

Mean spectrum of premalignant patches was subtracted from habitual tobacco users. In this case positive bands belong to habitual tobacco users and negative bands are from premalignant, Figure 3.10A. Strong negative band around 1320 cm^{-1} indicating DNA and δCH_2 (1440 cm^{-1}) band suggesting higher degree of hypercellularity was observed in case of premalignant. Difference spectrum of contralateral normal and habitual tobacco users, shown in Figure 3.10B, suggest almost similar spectral profile with minor differences in amide III, amide I and δCH_2 bands.

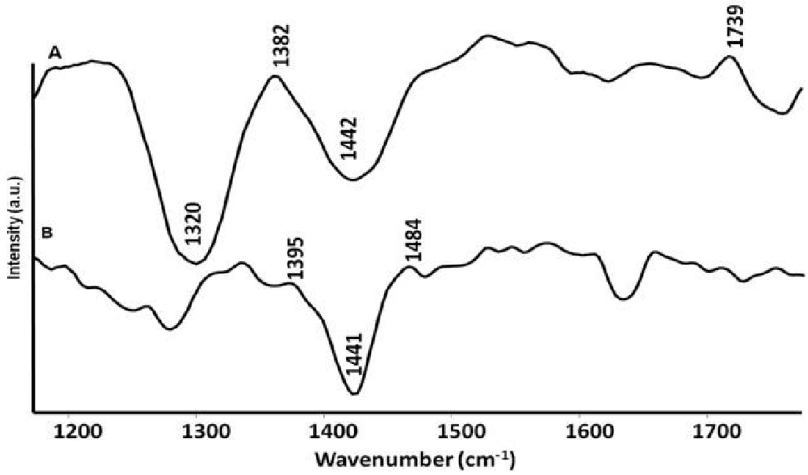


Figure 3.10: Comparison of difference spectra across different groups: A. habitual tobacco users– premalignant, B. habitual tobacco users – contralateral normal

These differences were further explored for classification using PC-LDA followed by LOOCV. A total of 583 spectra (150-healthy controls without tobacco habit; 170-contralateral normal under tobacco influence; 113-tobacco associated premalignant patches; 150-healthy controls with long-term tobacco habits) were pooled and analyzed. Seventeen (17) factors contributing up to 86% correct classification were used, Figure 3.11A. Scatter plot generated by plotting score of factor 2 and 3 is shown in, Figure 3.11B**.** Separate clusters were obtained for spectra of healthy controls without tobacco habit and premalignant patches. Spectra of contralateral normal and habitual tobacco users show overlap, which can be attributed to similar and long-term tobacco exposure associated changes in the oral mucosa.

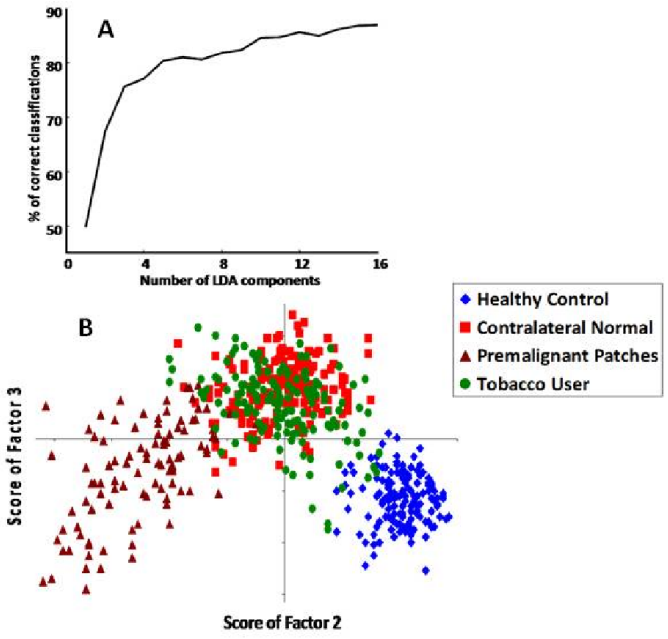


Figure 3.11: PC-LDA of healthy control, contralateral normal, premalignant conditions, habitual tobacco user A. Scree plot B. Scatter plot

Confusion matrix of PC-LDA findings is shown in Table 3.IIIA**.** All 150 spectra of healthy controls without any history of tobacco habits were correctly classified. In case of premalignant, 96/113 spectra were correctly classified. Among the 17 misclassifications, 15 were with habitual tobacco

user and 2 with contralateral normal. None of the premalignant spectra misclassified with tumor or healthy controls, suggesting exclusive features in these groups. In case of contralateral normal, among the 30 misclassifications 27 were with closely related habitual tobacco users and of the remaining 3, 2 and 1 were with premalignant and healthy controls, respectively. Similar trend was observed in case of habitual tobacco users, in this case of the 25 misclassifications, 20 were with contralateral normal.

A	Healthy Control	Contralateral Normal	Premalignant patches	Habitual tobacco user
Healthy Control	150/150	0	0	0
Contralateral Normal	1	140/170	2	27
Premalignant patches	0	2	96/113	15
Habitual tobacco user	5	20	0	125/150

Leave-one-out cross validation

B	Healthy Control	Contralateral Normal	Premalignant patches	Habitual tobacco user
Healthy Control	147/150	2	0	1
Contralateral Normal	1	133/170	3	33
Premalignant patches	0	6	86/113	21
Habitual tobacco user	6	25	0	119/150

Table 3.III: Summary of classification between *in vivo* Raman spectra of healthy control, contralateral normal, premalignant and habitual tobacco user A. Standard model, B. LOOCV. (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions)

In the next step, LOOCV was performed to evaluate classification efficiency and results are shown in Table 3.IIIB. Maximum misclassification was observed between contralateral normal and habitual tobacco users, which can be attributed to similar changes in buccal mucosa due to long-term tobacco exposure. Among the 37 misclassifications of contralateral normal, 33(89%) were with habitual tobacco users, 3(8%) with premalignant and remaining 1(3%) was with healthy control. Similarly in case of habitual tobacco users, of the 31 misclassifications, 25(81%) were with contralateral normal and remaining 6 (19%) with healthy controls. Twenty-seven (27) spectra from premalignant patches were misclassified, of these 21 (78%) were with habitual tobacco users and 6

(12%) were with contralateral normal. Overall classification efficiencies of 98, 78, 76 and 79% were observed for healthy controls, contralateral normal, premalignant and habitual tobacco users, respectively.

Highest classification efficiency was observed in case of healthy controls without any tobacco habits (98%). In this case very few misclassifications with habitual tobacco users were 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 is still healthy at few places. Misclassifications of premalignant spectra with habitual tobacco users are consistent with results shown in Table 3.II, *i.e.* misclassifications of premalignant spectra as contralateral normal. Since both the contralateral 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. This also explains major overlap/misclassifications observed between mucosa of contralateral normal and habitual tobacco users. Overall, findings of the study demonstrated the feasibility of discriminating healthy controls with and without tobacco habits against premalignants.

3.I.2.d *Classification of different premalignant lesions:* A wide array of precancerous conditions like leukoplakia, erythroplakia, oral lichen planus, oral sub-mucous fibrosis (OSMF), erythematosus etc. have been implicated in the development of oral cancer. Leukoplakia and OSMF are two of most common precancerous conditions found in Indian population. Clinical manifestations of OSMF and leukoplakia are very different. Leukoplakia is described as a white patch or plaque that cannot be characterized clinically or pathologically as any other disease. OSMF is a chronic progressive condition where fibroelastic changes of oral mucosa along with epithelial atrophy leads to stiffness of mucosa resulting in trismus and inability to eat. This study was taken

up to evaluate potentials of Raman spectroscopy in discriminating leukoplakia and OSMF, two of the most widely occurring premalignant conditions in Indian population.

Mean baseline corrected spectrum of leukoplakia and OSMF along with standard deviation is shown in Figure 3.12. Major differences in the C=O stretching vibrations of the peptide backbone were seen between both groups. Broadening of amide I region in case of OSMF can be attributed to differences in protein/collagen content of both conditions. In addition to amide I, minor differences in amide III and δCH_2 region were also seen.

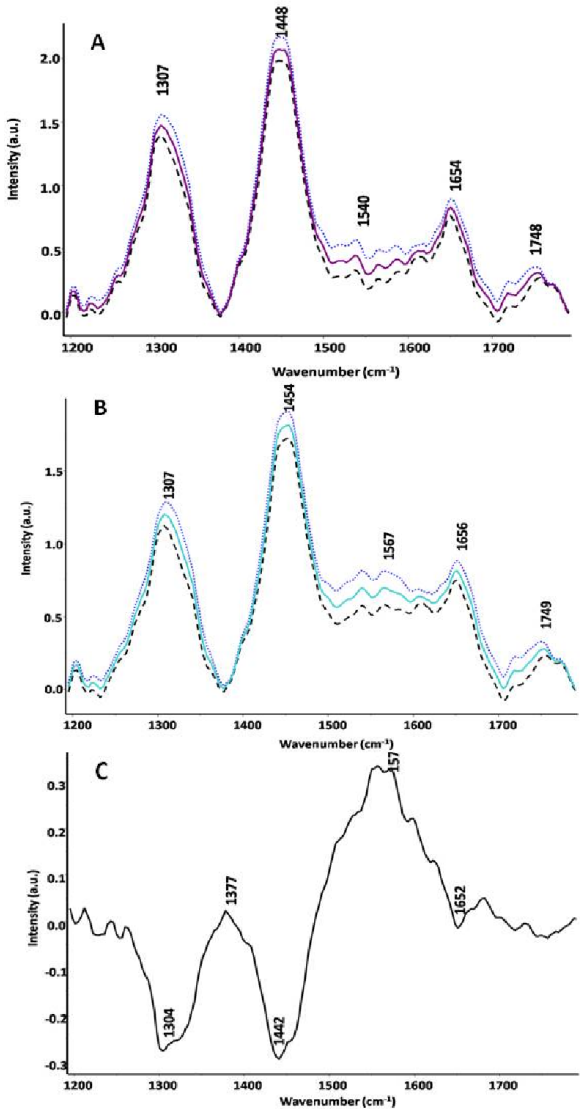


Figure 3.12: Mean spectra along with standard deviations A. Leukoplakia B. OSMF C. Difference spectrum (leukoplakia-OSMF) [Solid line: mean spectra, Dotted line: mean+ standard deviation, Broken line: mean – standard deviation]

Difference spectra obtained by subtracting spectra of OSMF from leukoplakia is shown in Figure 3.12C**.** In this case positive bands belong to leukoplakia while negative bands are from OSMF. Positive bands around 1558 $cm^{-1}$and 1571 cm^{-1} has been assigned to C=C vibrations from porphyrin ring and guanine/adenine rings, respectively [221,222]. Presence of these bands in leukoplakia spectra can be attributed to bacterial infections associated with poor mouth hygiene and increase in cellularity or inflammatory responses associated with the disease. Negative bands in amide III, amide I and δCH2 regions of the subtracted OSMF spectra suggests differences in protein or collagen content in both conditions.

These differences were utilized for exploring feasibility of classification between leukoplakia and OSMF conditions. A total of 170 and 192 spectra from contralateral normal and tumor sites of 40 subjects were analyzed against 62 OSMF (14-subjects) and 51 leukoplakia (12-subjects) spectra. Twelve (12) factors contributing up to 85% correct classification were used for analysis, Figure 3.13A. As can be seen from scatter plot shown in Figure 3.13B, two minimally overlapping clusters of spectra from OSMF and leukoplakia, respectively were obtained.

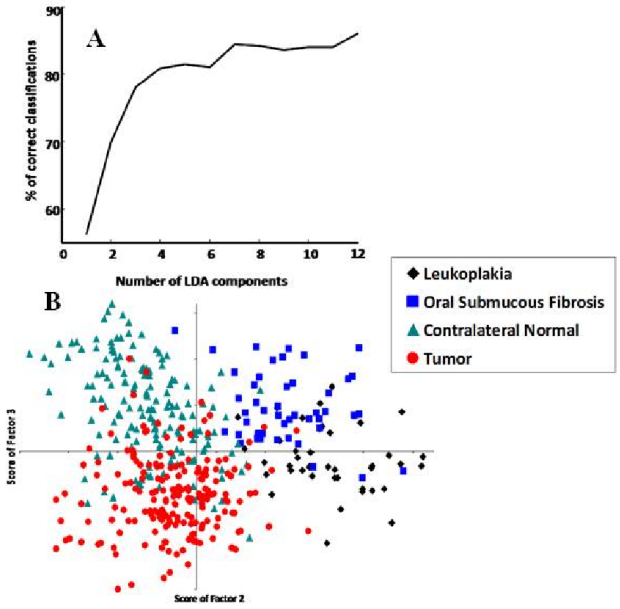


Figure 3.13: PC-LDA of leukoplakia, OSMF, contralateral normal and tumor conditions, A. Scree plot B. Scatter plot A. Scree plot B. Scatter plot

Confusion matrix of PC-LDA and LOOCV analysis is shown in Table 3.IV, as can be seen 147/170 (86%) contralateral normal and 179/192 (93%) tumor spectra were correctly classified. In case of leukoplakia 37/51 (72%) spectra were correctly classified, most of the misclassifications (13 out of 14) were with OSMF spectra. Only 1 spectrum was wrongly classified as normal. None of the leukoplakia spectra were misclassified as tumor. Similar trend was observed for OSMF; in this case 46/62 (74%) spectra were correctly classified. Of the 16 misclassifications, 13 were with leukoplakia, 2 with contralateral normal and 1 with tumor. Findings of PC-LDA were evaluated by LOOCV and results are shown in Table 3.IVB. In case of leukoplakia, of the 26 misclassifications 23 were with OSMF. Similarly in case of OSMF of the 27 misclassifications, 25 were with leukoplakia. Similar to earlier observation contralateral normal and tumor spectra were classified with an average classification efficiency of ~88%.

A	Leukoplakia	OSMF	Contralateral Normal	Tumor
Leukoplakia	37/51	13	1	0
OSMF	13	46/62	2	1
Contralateral Normal	3	1	147/170	19
Tumor	1	3	9	179/192

Leave-one-out cross validation

B	Leukoplakia	OSMF	Contralateral Normal	Tumor
Leukoplakia	25/51	23	1	2
OSMF	25	35/62	2	0
Contralateral Normal	3	3	144/170	20
Tumor	2	5	11	174/192

Table 3.IV: Summary of classification between *in vivo* Raman spectra of leukoplakia, OSMF, contralateral normal, and tumor A. Standard model, B. LOOCV. (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions)

Overall, a tendancy of classification was observed between leukoplakia andOSMF. Very few spectra from both groups are wrongly classified as normal or tumor, which further supports already described findings regarding objective identification of precancerous conditions. Misclassifications among premalignant conditions, leukoplakia and OSMF can be explained on the basis of varying

grade of thickness of a precancerous patch and the fact that often oral cancer subjects are presented with multiple premalignant conditions. Overall, findings of the study support applicability of Raman spectroscopic methods in identifying premalignant conditions in oral cavity.

3.I.2.e *Correlation with histopathology:* Malignancies arising from the oral cavity are epithelial in nature and therefore classified as squamous cell carcinomas (SCC). More than 90% of the tumors are of SCC type [224-226]. As per the degree of differentiation it can be divided in to three groups (i) well differentiated squamous cell carcinoma (WDSCC)-more than 75% keratinization (ii) moderately differentiated squamous cell carcinoma (MDSCC)-approximately 25-75% keratinization (iii) poorly differentiated squamous cell carcinoma (PDSCC) - less than 25% keratinization. The majority of oral cancer cases are MDSCC type [225]. A clear relationship between histopathological differentiation and prognosis has not been established, although lack of differentiation has been associated with rapid growth and spread of the disease [225,226]. The present study was undertaken to assess the influence of variability in tumor pathological grades on classification with normal conditions.

On the basis of tumor grade and differentiation status 86 out of 113 subjects enrolled in the study were categorized in to three groups (47-MDSCC, 13-WDSCC and 26-PDSCC). A total of 255 tumor spectra (130-MDSCC, 48-WDSCC and 77-PDSCC) were obtained from these subjects. These spectra were analyzed against 170 contralateral normal spectra of 40 subjects. PC-LDA followed by LOOCV was performed to explore the feasibility of classification. Twelve (12) factors accounting up to 65% correct classification were used, Figure 3.14A. Scatter plot obtained by plotting score of factor 2 and factor 3 is shown in Figure 3.14B. Two separate clusters of contralateral normal and tumor spectra were obtained. No clear classification between different pathological tumor grades was observed.

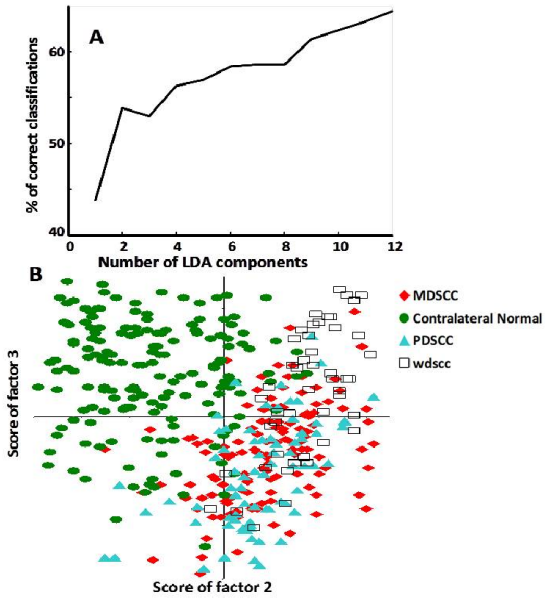


Figure 3.14: PC-LDA of contralateral normal, MDSCC, PDSCC and WDSCC tumors A. Scree plot B. Scatter plot A. Scree plot B. Scatter plot

	MDSCC	Contralateral normal	PDSCC	WDSCC
MDSCC	37	3	55	35
Contralateral Normal	12	129	13	16
PDSCC	38	0	31	8
WDSCC	16	0	7	25

Leave-one-out cross validation

	MDSCC	Contralateral normal	PDSCC	WDSCC
MDSCC	60	2	40	28
Contralateral Normal	12	135	8	15
PDSCC	19	1	45	12
WDSCC	8	0	6	34

Table 3.V: Summary of classification between contralateral normal, MDSCC, PDSCC and WDSCC tumors A. Standard model, B. LOOCV. (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions)

PC-LDA and LOOCV results are summarized in form of a confusion matrix shown in Table 3.V. Contralateral normal spectra were classified with an efficiency of 77%. Major overlap was observed in case of tumors of different pathological grades. Only 37% spectra from MDSCC, 29% spectra from PDSCC and 50% spectra from WDSCC were correctly classified. Most of the misclassifications are among the different pathological grades. This can be attributed to tumor

heterogeneity *i.e.* in a tumor there is always a possibility of presences of mixed pathologies. Overall, findings of the study suggest that spectra of normal and tumor conditions can be objectively classified. However, variability in tumor grade seems to have no bearing on classification with normal conditions.

3.II Study on cancer field effects or malignancy associated changes

Appearance of a clinically visibly precancerous lesion or condition in oral cancers is often preceded by microarchitectural changes in the oral cavity. Occurrences of these clinically or histologically unrecognizable subtle perturbations are attributed to 'malignancy-associated-changes' (MACs) or cancer-field-effects (CFEs)', terms often used interchangeably. These effects have been primarily associated with the development of second primary tumors (SPTs) in the oral cavity. Subjects suffering from SPT pose serious challenges in clinical evaluation, planning of treatment regimen and post-treatment quality of patient's life in terms of associated morbidity. Identification of CFEs or MACs may serve as a novel screening tool which can reduce the morbidity and mortality associated with multiple potentially malignant transforming fields. In the present study, feasibility of *in vivo* Raman spectroscopic identification of changes that are indicative of neoplastic transformation has been explored.

3.II.1 Methodology

3.II.1.a *Clinical samples:* Fourteen (14) subjects (8-male,6-female) with buccal mucosa SCC lesion and without any history of tobacco usage were included (non-habitués contralateral) in addition to the subjects accrued for the earlier study (Section: 3.I.1.a). These details are presented in Table 3.V. Briefly, it includes 40 subjects (male) with tobacco associated pathologically verified buccal mucosa lesion, 15 subjects (13-male, 2-female) without any history of tobacco habits

(healthy controls), 15 subjects (14-male,1-female) with long-term tobacco habit (habitués healthy controls). Median age of tumor, habitués healthy control, non-habitués contralateral and healthy control subjects were 51, 50, 45 and 51 years, respectively.

Sr. No.	Category	Site of spectral acquisition	Total no. of subjects (M&F)	Median age (yrs)	Cancer	Tobacco habit	Total no. of spectra
1	Healthy controls	Buccal mucosa	15 (13M,2F)	51	NO	NO	150
2	Habitues healthy controls	Buccal mucosa	15 (14M,1F)	50	NO	YES	150
3	Non-habitues contralateral	Buccal mucosa	14 (8M,6F)	45	YES	NO	60
4	Contralateral Normal	Buccal mucosa	40 (40 M)	51	YES	YES	170
5	Tumor	Buccal mucosa	40 (40M)	51	YES	YES	192

Table 3.VI: Subject accrual details

3.II.1.b *Instrumentation:* *In vivo* spectra were acquired using previously described fiberoptic probe coupled HE-785 instrument (Figure 3.1), as per teeth positions on buccal mucosa as shown in Figure 3.3. Subjects were requested to wash their mouth with distilled water before spectral acquisitions to avoid any differences because of the mouth environment.

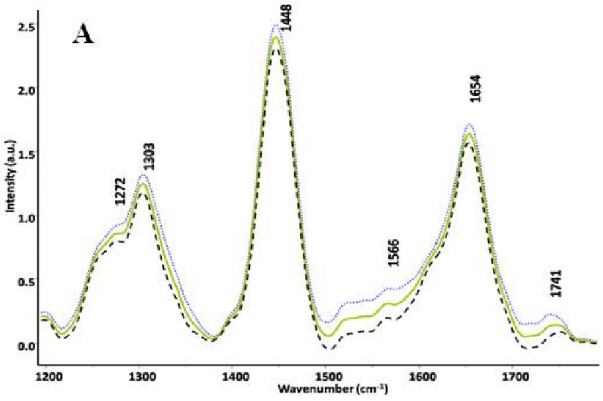
3.II.1.c *Spectral-acquisition*: In cancer patients, spectra were acquired from both contralateral normal (opposite side of tumor) and tumor site to obtain a total of 170 contralateral and 192 tumor spectra. For healthy controls with and without tobacco habits, 10 spectra corresponding to a total of 150 spectra were acquired from the left and right buccal mucosa. Sixty (60) spectra were acquired from contralateral normal side of 14 non-habitués oral cancer subjects. Spectral acquisition parameters were: λ_{ex}-785 nm, laser power-80 mW, integration- 3 seconds and averaged over 3 accumulations.

3.II.1.d *Spectral preprocessing:* Spectral preprocessing involving correction for CCD response and background signals was performed as per the previously described procedure (Section:2.II.1.c**)**. First derivative, vector normalized spectra in 1200-1800 cm^{-1} region were used

as input for multivariate analysis for PC-LDA followed by LOOCV. Mean spectrum for each group was generated by averaging Y axis keeping X axis constant and baseline corrected by fitting a 5th order polynomial function. These were used for spectral comparisons and computing difference spectra.

3.II.2 Results and discussion

3.II.2.a *Spectral features:* Average spectra along with standard deviation of (a) healthy control, (b) contralateral normal, (c) non-habitués contralateral, (d) habitués healthy controls and (e) tumor were computed and are shown in Figure 3.15. Similar to earlier observations, spectral features of healthy controls are dominated by strong lipid features indicated by δCH_2 bend (1440 cm^{-1}), two sharp features in amide III region, sharp peak around amide I (1660 cm^{-1}) and ester band around 1744 cm^{-1}. Broad features in amide III region, broad and shifted δCH_2 and broad amide I in mean tumor spectrum indicate protein dominance. Spectra of contralateral normal and habitués healthy controls are also dominated by lipid bands but exhibit minor differences, with respect to healthy controls, such as minor shift in amide III and δCH_2 bend (1440 cm^{-1}) as well as broadening of amide I region suggesting changes in protein secondary structures. Mean spectra of non-habitués contralateral subjects show spectral profile similar to that of healthy controls with minor differences like sharp amide I and additional bands in ester region (1745 cm^{-1}).



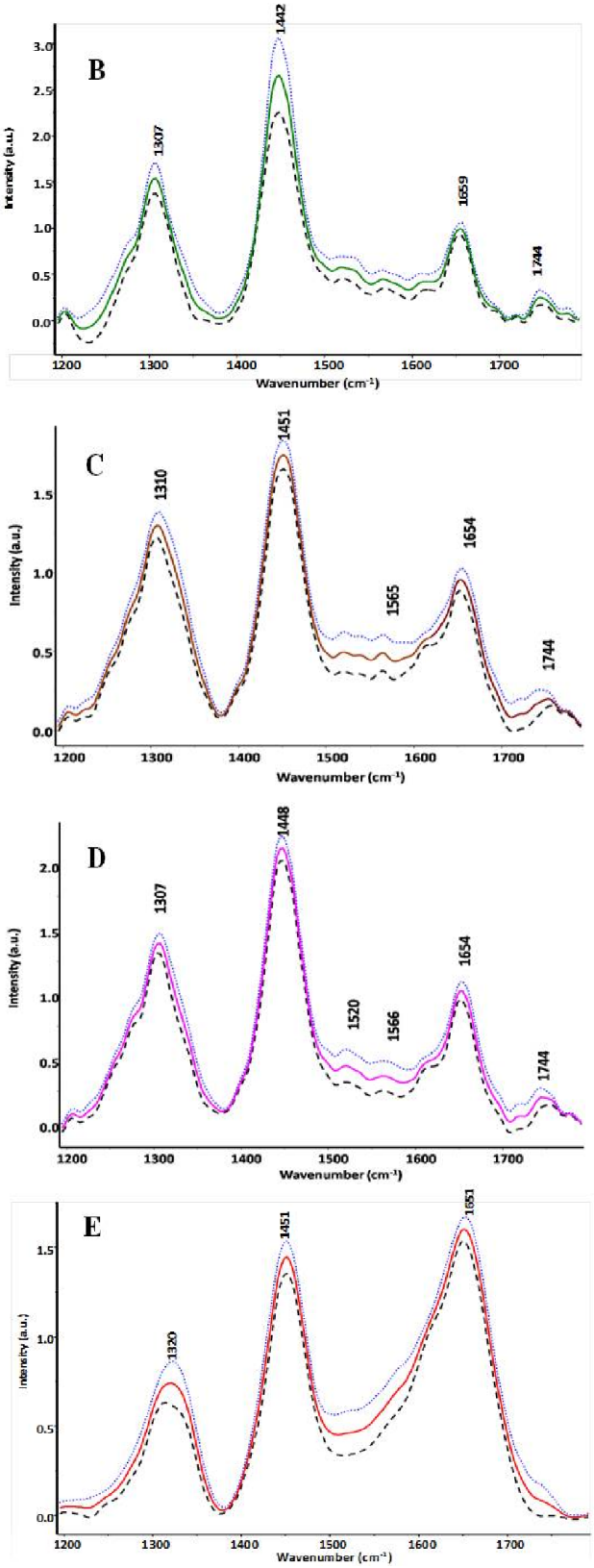


Figure 3.15: Mean spectra along with standard deviations A- Healthy controls B- Habitués healthy controls C- Contralateral normal D- Nonhabitués contralateral normal E-Tumor (Solid line: mean spectra, Dotted line: mean+ standard deviation, Broken line: mean – standard deviation)

Spectral similarities and dissimilarities across various groups were further analyzed by computing difference spectra. In first step, difference spectra of pathological conditions *i.e.* tumor (i), contralateral normal (ii), non-habitués contralateral normal (iii) and habitués healthy controls (iv) were computed by subtracting them from healthy controls. As can be seen from Figure 3.15A-D, all positive peaks correspond to healthy controls and negative bands are from various pathological conditions. Difference tumor spectrum (Figure 3.16A) exhibit loss of lipids (1440, 1300, 1743 cm^{-1}) and presence of DNA (1340, 1480 cm^{-1}) and haemoglobin (~1560 cm^{-1}), which could be attributed to increase in number of dividing nuclei and angiogenesis, respectively, which are known hallmarks of tumorigenesis. Negative bands suggesting presence of DNA (1340, 1480 cm^{-1}) and loss of lipids (1300, 1440, 1743 cm^{-1}) were also seen in difference spectrum of other pathological conditions (Figure 3.16B-D). Long term tobacco exposure is known to cause increase in proliferation of epithelial cells in the upper aero-digestive tract of tobacco users and can be considered as preliminary event for genetic changes culminating in the development of oral SCC.

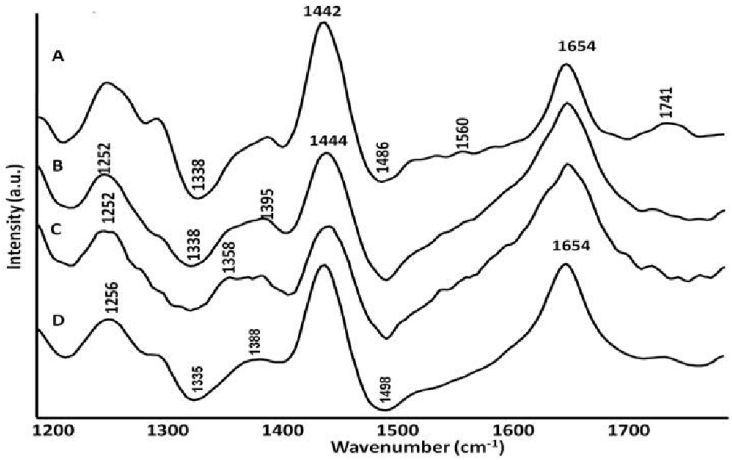


Figure 3.16: Comparison of difference spectra across different groups. A. Healthy Control – Tumor ; B. Healthy Control – contralateral normal; C. Healthy Control - Non-habitués contralateral normal; D. Healthy Control - Habitués healthy control.

To understand influence of long-term tobacco exposure, spectra of tumor, contralateral normal and non-habitués contralateral were subtracted from mean habitués healthy controls spectrum, Figure

3.17A-C**.** In this case positive bands correspond to habitués healthy controls and negative bands to pathological conditions. Once again, tumor spectra show loss of lipids (1300, 1440, 1743 cm^{-1}), predominant protein features (broad 1660 cm^{-1}) and prominent DNA bands (1340, 1480 cm^{-1}), Figure 3.17A. Positive DNA bands along with loss of lipid features (1300,1440,1650, 1743 cm^{-1}) were observed in difference spectrum of non-habitués contralateral, which could be attributed to tobacco induced hypercellularity (Figure 3.17C). These features were further corroborated by computing difference spectra of contralateral normal as shown in Figure 3.16B. No major differences except minor shift in δCH_2 band were observed.

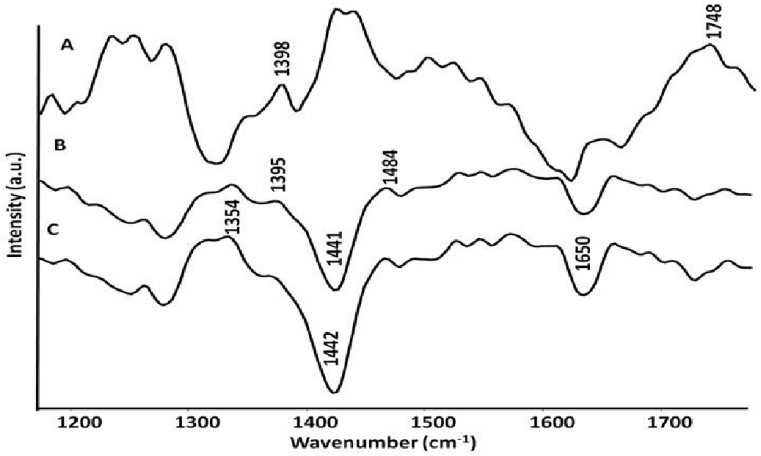


Figure 3.17:Comparison of difference spectra across different groups. A. Habitués healthy control – Tumor; B. Habitués healthy control – Contralateral Normal C. Habitúes healthy control– Non-habitués contralateral normal.

Difference spectrum of tumor and non-habitués contralateral were computed by subtracting from contralateral normal spectra (Figure 3.18A,B**).** Here, positive peaks corresponds to contralateral normal and negative peaks to tumor and non-habitués contralateral normal. Similar to earlier observations loss of lipids and predominant protein features suggested by negative amide I (~1658 cm^{-1}), amide III were seen in difference spectrum of tumors, Figure 3.18A. Difference spectra of

non-habitués contralateral normal suggest differences in lipid content (negative bands 1440, 1743 cm^{-1}). In addition positive DNA bands (1340, 1480 cm^{-1}) were also observed which could be attributed to tobacco induced hypercellularity, Figure 3.18B**.**

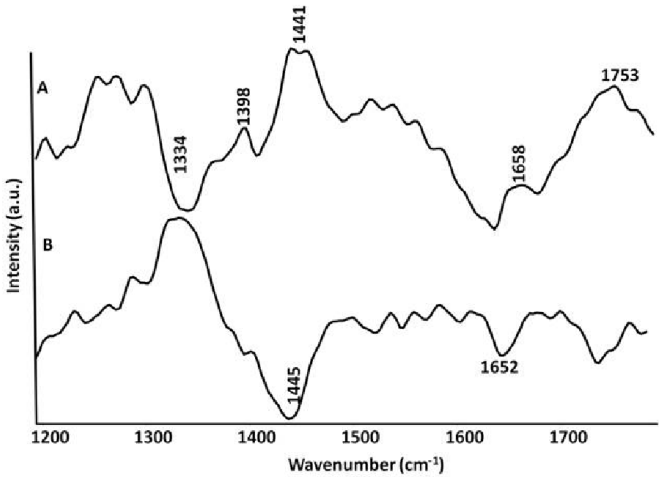


Figure 3.18: Comparison of difference spectra across different groups. A. Contralateral Normal – Tumor; B. Contralateral Normal – Non-Habitúes contralateral normal.

Difference tumor spectrum was computed by subtracting from mean non-habitués contralateral normal spectrum. Observed spectral features in this case corroborate earlier observations, *i.e.* loss of lipids (1300,1440,1743 cm^{-1}) and presence of DNA bands (1340, 1480 cm^{-1}), as shown in Figure 3.19. Overall, major spectral variability was observed in bands associated with DNA (1340, 1480 cm^{-1}), proteins (amide I, III, δCH_2) and lipids (δCH_2 deformation, twisting and ester). Loss of lipids seems to be a common feature for most of the pathological conditions in comparison to healthy controls.

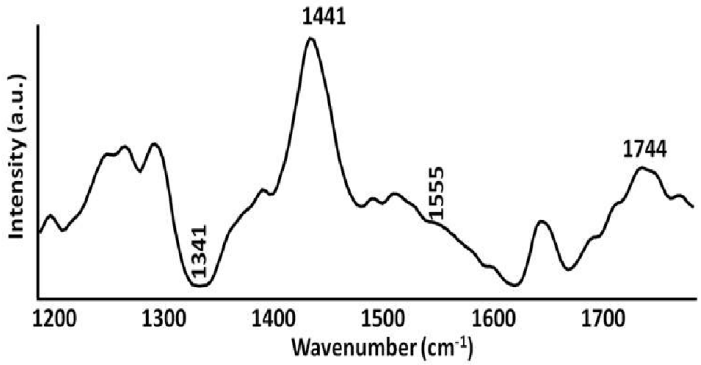


Figure 3.19: Difference spectrum (Non-habitúes contralateral normal – Tumor)

3.II.2.b Classification of healthy controls: In order to utilize above mentioned spectral differences for classification PC-LDA followed by LOOCV was performed. In first step, the efficacy of Raman spectroscopy in correctly identifying negative controls *i.e.* healthy controls without tobacco habit was demonstrated. Spectral data of healthy controls, contralateral normal and tumor sites were subjected to PC-LDA using 5 factors accounting for ~92% correct classification (Figure 3.20A). As shown in Figure 3.20B, cluster belonging to healthy control is exclusive, while that of contralateral normal and tumor show minimal overlap.

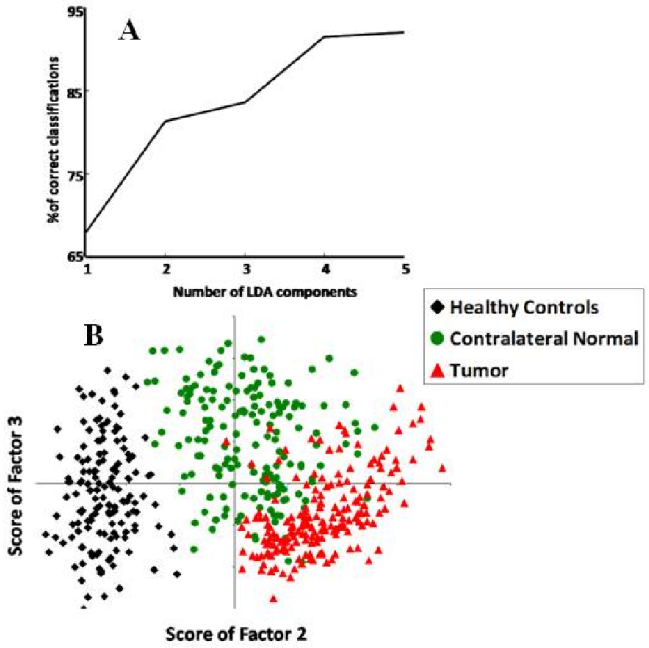


Figure 3.20: PC-LDA of healthy controls, contralateral normal and tumors, A.Scree plot B. Scatter plot

Similar results are presented in form of a confusion matrix shown in Table 3.VII. In case of healthy controls, 148/150 spectra (98%) were correctly classified. Two (2) spectra were wrongly classified as contralateral normal. In case of tumors, 180/192 spectra (94%) were correctly classified. In accordance with already discussed results, most of the misclassifications were with contralateral normal. After LOOCV (Table 3.VIIB) only 2 out of 150 (1%) spectra from healthy control were wrongly classified as contralateral normal. Corroborating the earlier findings, 14 tumor spectra were

misclassified as contralateral normal, while 23 contralateral normal spectra were wrongly classified as tumor. Overlap among contralateral and tumor could be attributed to the presence of transformation zones in visibly normal mucosa and heterogeneous nature of tumors.

A	Healthy control	Contralateral Normal	Tumor
Healthy control	148/150	2	0
Contralateral Normal	4	143/170	23
Tumor	0	12	180/192

Leave-One-Out cross-validation

B	Healthy control	Contralateral Normal	Tumor
Healthy control	148/150	2	0
Contralateral Normal	4	143/170	23
Tumor	0	14	178/192

Table 3.VII: Summary of classification between in vivo Raman spectra of Healthy controls, contralateral normal and tumor A. Standard model, B. LOOCV (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions)

3.II.2.c *Classification of habitués healthy controls:* In order to explore the feasibility of identifying changes due to tobacco exposure, spectral data of habitués healthy controls along with healthy controls, contralateral normal and tumor sites were analyzed by PC-LDA. In this case, 8 factors accounting for ~82 % correct classification were used for analysis, Figure 3.21A. Similar to earlier observation, an exclusive cluster of healthy controls and two minimally overlapping clusters of contralateral normal and tumor sites of habitués oral cancer subjects were obtained (Figure 3.21B). Major overlap was observed between contralateral normal mucosa and habitués healthy controls.

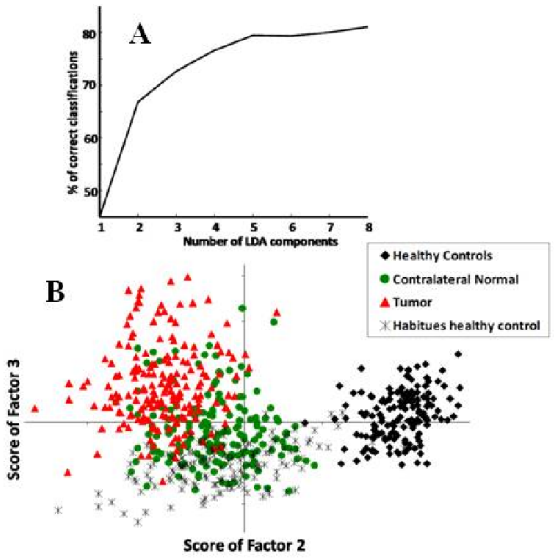


Figure 3.21: PC-LDA of healthy controls, contralateral normal, tumor and habitués healthy controls A. Scree plot B. Scatter plot.

Similar results are shown in Table 3.VIIIA**,** 146/150 spectra (97%) of healthy controls and 168/192 spectra (87%) of tumor were correctly classified. Major overlap was observed between spectra obtained from contralateral normal site of tobacco using oral cancer subjects and habitués healthy controls. 46 spectra of habitués healthy controls were misclassified, of these 41 (89%) were with closely related contralateral normal and remaining 5 with healthy controls. Similarly 30 of the total 51 misclassification (59%) of contralateral normal were with habitués healthy controls. The findings of LOOCV, shown in Table 3.VIIIB, indicate that only 6 spectra from healthy controls were misclassified (4 as contralateral normal and 2 as habitués healthy control) and 52 out of 170 spectra of contralateral normal were wrongly classified (21 as tumor and 31 as habitués healthy controls). In case of habitués healthy controls of 47 misclassifications (42 with contralateral and 5 with healthy controls) were noted. This could be attributed to tobacco induced cancer field effects suggesting that early changes due to carcinogen exposure can be identified. In case of tumor, 26 spectra (19 as habitués healthy and 7 as contralateral normal) were wrongly classified, which could

be due to the presence of analogous regions of inflammation as a result of tobacco exposure within tumor (tumor- heterogeneity) and early transformation zones in visibly normal mucosa.

A	Healthy control	Contralateral Normal	Tumor	Habitués healthy controls
Healthy control	146/150	3	0	1
Contralateral Normal	0	119/170	21	30
Tumor	0	7	168/192	17
Habitués healthy controls	5	41	0	104/150

Leave-one-out cross validation

B	Healthy control	Contralateral Normal	Tumor	Habitués healthy controls
Healthy control	144/150	4	0	2
Contralateral Normal	0	118/170	21	31
Tumor	0	7	166/192	19
Habitués healthy controls	5	42	0	103/150

Table 3.VIII: Summary of classification between *in vivo* Raman spectra of Healthy controls, contralateral normal, tumor and habitués healthy control A. Standard model, B. LOOCV. (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions)

3.II.2.d *Classification of non- habitués cancer subjects* *:* Although tobacco is known major etiological factor for oral cancers, there is a sizeable occurrence of oral cancers in subjects without tobacco habits. To study the transformation in uninvolved mucosa of such cases, analysis was carried out on non-habitués contralateral subjects along with habitués contralateral normal, habitués and non-habitués healthy controls by PC-LDA using 8 factors, accounting up to ~85% correct classification, Figure 3.22A. Similar to earlier observations, an exclusive cluster of healthy controls and overlapping clusters of contralateral normal and habitués healthy controls were obtained. Contralateral normal side of non-habitués cancer subjects also yielded an exclusive cluster (Figure 3.22B).

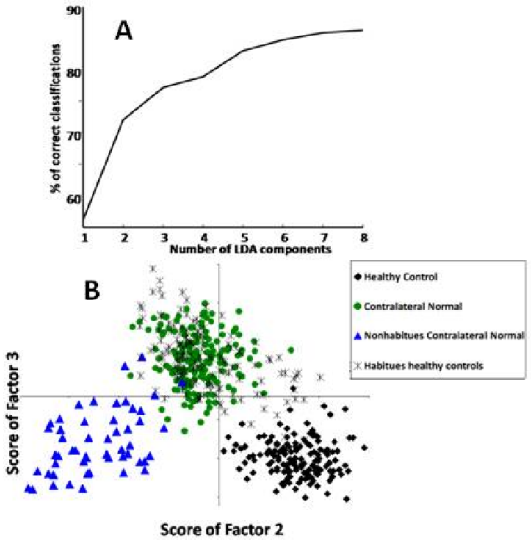


Figure 3.22: PC-LDA of healthy controls, contralateral normal, habitués healthy control and nonhabitués contralateral normal A. Scree plot B. Scatter plot.

PC-LDA and LOOCV results are shown in form of a confusion matrix in Table 3.VIII. As can be seen 148/150 spectra of healthy controls and 134/170 spectra of contralateral normal were correctly classified (Table 3.IXA). None of the spectra of healthy controls and contralateral normal misclassified with non habitués contralateral normal. Corroborating earlier observations, most of the misclassifications (97%) of contralateral normal were with habitués healthy controls. Similarly in case of habitués healthy controls, 23 of the total 28 misclassifications were with contralateral normal. As can be seen from LOOCV results shown in Table 3.IXB, 51 out of 60 spectra of non-habitués contralateral normal were correctly classified. The minor misclassifications were with contralateral (5 spectra) and habitués healthy controls (4 spectra). Overlap between contralateral normal and habitués healthy controls were observed. Of the total 37 misclassifications of contralateral normal, 36 were with habitués healthy controls and 1 with healthy control. In case of habitués healthy controls, of the 29 misclassifications, 24 were with contralateral normal, 4 with healthy controls and 1 with non-habitués contralateral. In this case, both contralateral normal mucosa and non-habitués contralateral mucosa showed no overlap with healthy controls, suggesting

occurrence of CFE/MAC. No overlap between contralateral normal mucosa and non-habitués contralateral mucosa was also observed which suggest that early transformation changes in both groups may be different.

A	Healthy Controls	Contralateral normal	Nonhabitués Contralateral normal	Habitués healthy controls
Healthy Controls	148/150	0	0	2
Contralateral normal	1	134/170	0	35
Nonhabitués Contralateral normal	0	3	53/60	4
Habitués healthy controls	4	23	1	122/150

Leave-one-out cross-validation

B	Healthy Controls	Contralateral normal	Nonhabitués Contralateral normal	Habitués healthy controls
Healthy Controls	148/150	0	0	2
Contralateral normal	1	133/170	0	36
Nonhabitués Contralateral normal	0	5	51/60	4
Habitués healthy controls	4	24	1	121/150

Table 3.IX: Summary of classification between *in vivo* Raman spectra of Healthy controls, contralateral normal, habitués healthy control and nonhabitués contralateral normal A. Standard model, B. LOOCV (diagonal elements are true positive predictions and elements are false positive predictions)

The Findings of the study suggest that subtle changes because of tobacco abuse/unknown etiological factors, which may be indicative of early neoplastic transformation in clinically normal appearing contralateral mucosa *i.e.* CFE/MAC, may be detected by Raman spectroscopy. The non-invasiveness and use of harmless excitation wavelength impart this method several advantages, and thus prospectively Raman spectroscopy has potential to become an ideal mass screening tool in public health programs.

Summary: Studies reported in this Chapter were carried out using a fiberoptic probe coupled Raman spectroscope. To the best of our knowledge, for the first time, we have demonstrated the feasibility of acquiring good quality *in vivo* Raman spectra under clinically implementable time in Indian population. Following is the brief summary of the work presented in this Chapter.

1. Feasibility of objectively classifying premalignant lesions against contralateral normal and tumors was demonstrated. Standard models for all three conditions were developed and evaluated with independent test data.
2. Discrimination of premalignant conditions against closely related habitual tobacco user was explored. Findings suggest that premalignant lesions can be objectively discriminated.
3. Classification of OSMF and leukoplakia, two of the most commonly occurring precancerous conditions in Indian population was explored and feasibility of classification between both conditions was demonstrated.
4. Influence of variability in tumor pathological grades on classification with normal tissues was assessed. Results suggest that variation in the pathological grade of tumor does not have any bearing on classification.
5. Potential of Raman spectroscopy in identifying invisible changes that are indicative of neoplastic transformation was also evaluated. Findings suggest that 'cancer field effects or malignancy associated changes', earliest events in oral carcinogenesis can be identified and supports applicability of Raman spectroscopic methods as a non-invasive mass screening tool.

Chapter 4

Raman microspectroscopy of oral cancer cells

Introduction

Ex vivo and *in vivo* Raman spectroscopic studies have demonstrated that normal, cancerous and precancerous conditions in oral buccal mucosa can be objectively classified. The present study aims at evaluating potentials of Raman spectroscopy in identifying minor changes associated with loss of 'keratin' in oral cancer cells. In the first section, a brief introduction about keratin protein and its applicability as a marker in oral cancers is presented. It is followed by Raman spectroscopic study on cell pellets of keratin expressing and knockdown oral cancer cells is presented. The last section of the Chapter deals with Raman microspectroscopic studies of single vector control and knockdown oral cancer cells.

4.I Keratins: An Overview

The cytoskeleton of all metazoan cells consists of three major filamentous networks namely: microfilaments (MFs), microtubules (MTs) and intermediate filaments (IFs). The integrated network formed by these filament systems is responsible for the mechanical integrity of the cell and is critical for cellular processes like cell division, motility and plasticity [227]. The name "intermediate filaments" comes from their size of diameter (10–12 nm) being intermediate between that of MTs (20-25 nm) and MFs (4-6 nm). They are mostly involved in the maintenance of cell shape, cell organelles, locomotion, intracellular organization, cell division and transport. Intermediate filament proteins, based on the amino-acid sequence identity and polymerization properties and can be subdivided into 6 subgroups. They are preferentially expressed in a tissue specific and differentiation dependent manner. Due to their cell type specific expression, IFs serve as cell type markers and are being explored as diagnostic markers. Pattern of IF proteins expression in different tissues is shown in the Table 4.I.

Class	*IF*	*Molecular weight(KD)*	*Cell type*
Type I	Acidic cytokeratins	40-64	Epithelial cell
Type II	Basic cytokeratins	52-68	Epithelial cell
Type III	Desmin.	53	Muscle
	GFAP	51	Astroglia
	Peripherin	54	Neuronal
	Vimentin	55	Mesenchymal
Type IV	Neurofilament-L	68	Neuronal
	Neurofilament-M	110	Neuronal
	Neurofilaments- H	130	Neuronal
	Nestin	240	CNS stem cell
Type IV	Lamin A	70	Most cell types
	Lamin B	64	Nuclear lamina
	Lamin C	58	Nuclear lamina
Orphan	Phakinin	46	Lens
Orphan	Filensin	83	Lens

Table 4.I: Classification of IF proteins based on their tissue specificity and polymerization

Cytokeratins or keratins (K) are largest subgroup of intermediate filament (IF) proteins preferentially expressed in epithelial tissues. On the basis of molecular weight and isoelectric point these can be subdivided into type I acidic (K9-K28) and type II basic (K1-K8 and K71-K74) keratins. They are obligatory heteropolymers and are assembled in 1:1 molar ratio, consisting of one type I and one type II keratins [227-232]. Epithelial tissues express different pairs of keratins depending on the cell type *e.g.* all stratified squamous epithelia express K5/14 while K8/18 are seen in all simple epithelia [233,234]. All keratin filaments have a highly conserved central coil-coil α-helical 'rod' domain that is flanked by a non-a-helical N-terminal 'head' and C-terminal 'tail' domains of various lengths [232,235]. In addition to cytoprotective functions, they also perform important regulatory functions by modulating certain signaling pathways associated with variety of cellular processes such as protein synthesis, cell growth, cell differentiation, apoptosis and osmoregulation [236-240].

In squamous epithelia like oral epithelia, keratins exhibit a complex expression pattern which is tightly regulated with differentiation program of the tissue. Keratins 8 and 18 (K8 and K18) are predominantly expressed in simple epithelial tissues and perform protective and regulatory

functions like modulation of protein localization, protein targeting/trafficking and protein synthesis. K8 and K18 expression is not observed in stratified adult epithelial tissues like oral buccal mucosa. However, they are often aberrantly expressed in carcinomas including oral SCC and their expression is correlated with invasion and poor prognosis [241,242]. K8/18 along with vimentin has also been associated with drug resistance, metastatic and invasive properties of some carcinomas and melanomas [243].

Gene knockdown refers to an experimental technique by which the expression of one or more genes is reduced. The reduction can occur either through genetic modification or by treatment with a reagent such as a short DNA or RNA oligonucleotide having a complementary sequence to either gene or an mRNA transcript of interest. Recently, Alam *et al.* have generated stable clones of K8 knockdown cells using short hairpin (shRNA) constructs and shown that depletion of K8 expression level leads to decreased tumorigenicity and cell migration accompanied with changes in cell shape and actin cytoskeleton in OSCC derived cell line AW13516 [244]. Furthermore, loss of K8 phosphorylation also resulted in increase in tumorigenic potential of these cells [245].

Studies utilizing Raman spectroscopic methods as biosensors for identifying morphological changes associated with drug treatment, apoptosis or pathological conditions are reported in literature [250-252]. The greatest benefits of this technique are its high sensitivity and capability for non-invasive sensing. Unlike conventional biological assays, biochemical analysis of cells and tissues with Raman spectroscopy does not require the use of fixatives, markers or stains. The present study was undertaken to assess the potentials of Raman spectroscopy in identifying minor changes associated with depletion of keratin expression in oral cancer cells.

4.II Raman spectroscopic study on cell pellets

In the first step cell pellets of vector control and knockdown cells were generated and efficacy of Raman spectroscopy in identifying changes associated with differential level of keratin expression was evaluated.

4.II.1 Methodology

4.II.1.a *Cell line:* AW-13516 cells, derived from poorly differentiated SCC of tongue were used. Cells have morphology of typical epithelial cells, numerous mitotic features individual cell diskeratosis with occasional nuclear and nucleolar abnormalities [246]. The cells were cultured in IMDM media (Gibco), supplemented with 10% fetal calf serum (FCS; Hyclone) and 1% antibiotics solution (Amphotericin B -20 μg/ml, Penicillin -2500 Units/ml, Streptomycin -800 μg/ml of 1XPBS), at 37^0C and under a 5% CO_2 atmosphere [244-246]. Stable clones of already established K8 knockdown AW-13516 cells *i.e.* with reduced K8 protein expression, were chosen and cultured under similar conditions [244]. Cells harboring only the vector were termed as 'vector control cells' and had normal expression of keratin. Cells in which the keratin gene was 'knocked' down with a vector harboring shRNA against Keratin 8 had reduced expression of kerain 8 and were referred to as 'knockdown cells'.

4.II.1.b *Generation of cell pellets:* Approximately $1x10^6$ cells of both vector control and knockdown groups were grown up to 80% confluence. In order to bring all the cells in the same phase of cell-cycle, they were synchronized by growing under serum-free conditions for 24 hrs. Cells were collected by scrapping from plates using a cell scrapper and centrifuged at 2000 rpm for 10 minutes to obtain pellets followed by washing with PBS buffer (150 mM NaCl, 2 mM KCl, 8

mM Na_2HPO_4 and 1 mM KH_2PO_4) twice. The experiments were repeated four times with two pellets each time.

4.II.1.c *Fiberoptic Raman spectroscopy:* Cell pellets of both knockdown and vector control cells were placed on a CaF_2 window and spectra were acquired using already described fiber-optic probe coupled Raman set up along with a XYZ precision stage (Figure 2.5). A total of 123 and 96 spectra were recorded from vector controls and knockdown cells, respectively from four independent experiments.

4.II.1.d *Preprocessing and Multivariate analysis:* Raman spectra obtained from both vector control cells and knock down cell pellets were preprocessed as per the already described protocol including correction for CCD response followed by subtraction of contaminating fibers signals (Section: 2.II.1.c). First derivative and vector normalized spectra in 1200-1800 cm-1 region were used as input for multivariate analysis by PC-LDA followed by LOOCV. Average spectra were computed from the background subtracted spectra prior to derivatization for each class and baseline corrected by fitting a fifth order polynomial function. These baseline corrected, smoothened (Savitzky-Golay method, window size-3) and vector normalized spectra were used for spectral comparisons and computing difference spectrum.

4.II.2 Results and Discussion

4.II.2.a *Spectral features:* Mean baseline corrected spectra along with standard deviation of vector control and k8 knockdown cell pellets are shown in Figure 4.1. Spectra of vector control cells (Figure 4.1A) have bands associated with vibrations of glycine backbone and proline side chains (1204 cm^{-1}), amide III, C-N asymmetric stretching mode of aromatic amino acids (1307 cm^{-1}),δCH_2 deformation (1440 cm^{-1}) and amide I (1660 cm^{-1}). Spectra of knockdown cells show similar

profile to that of vector controls cells along with differences in amide III, δCH_2 and amide I regions (Figure 4.1B).

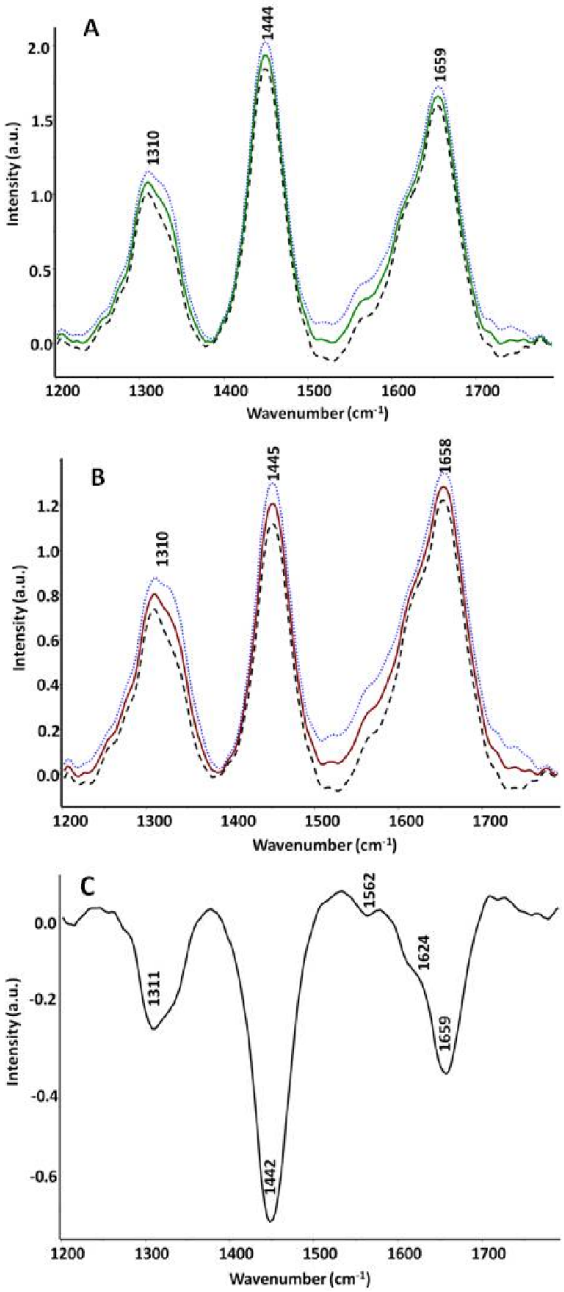


Figure 4.1: Mean spectra along with standard deviations of A. vector control cell pellets, B. K8 knockdown cell pellets, C. Difference spectrum (K8 knockdown – vector control)

Differences in protein content can be primarily attributed to the fact that knockdown and vector controls express differential amount of surface proteins. Difference spectra were computed by subtracting spectra of vector control cells from knockdown cells. In this case positive bands are of knockdown cells and negative bands are of vector controls. Negative bands around amide III, δCH_2 and amide I indicating higher protein content in vector control cells in comparison to knockdown cells were observed, Figure 4.1C.

4.II.2.b *Classification of K8 knock-down and vector control cells:* In order to utilize above mentioned spectral differences for classification, PC-LDA followed by LOOCV was performed using first derivative vector normalized spectra in the 1200-1800 cm-1 region as input. Five (5) factors contributing up to 72% correct classification were used, Figure 4.2A. Score of factor 2 and 3 were explored for classification. Two minimally overlapping clusters of vector control and knockdown cells were obtained, Figure 4.2B.

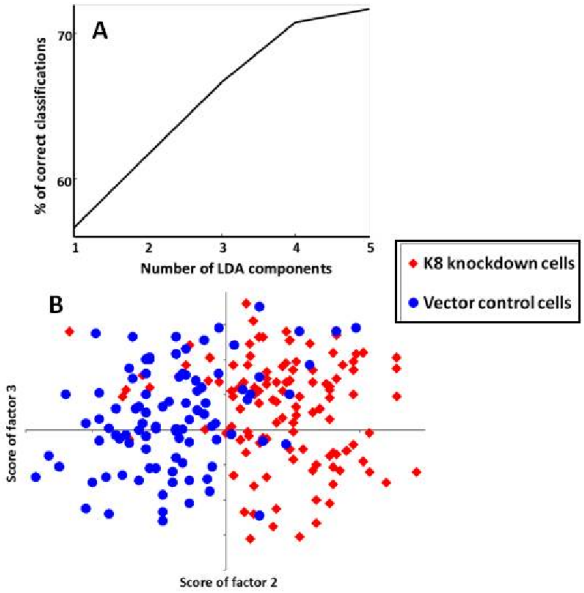


Figure 4.2: PC-LDA of spectra of K8 knock down and vector control cell pellets spectra: A. Scree plot B. Scatter plot

Similar results are presented in form of a confusion matrix in Table 4.II. Here all diagonal elements are true positive predictions while ex-diagonals indicate false positive predictions. Eighty-seven (87) of 123 spectra of vector controls and 70 of 96 spectra of knockdown cells with an efficiency of 71 and 73% respectively were correctly classified. These findings were validated by LOOCV and results are presented in Table 4.IIIB. In this case 79/123 and 56/96 spectra of knockdown and vector control cells, respectively with an average classification efficiency of 63% were correctly classified.

A	K8 Knockdown	Vector control
K8 Knockdown	87	36
Vector control	26	70
Leave-one-out cross validation		
B	K8 Knockdown	Vector control
K8 Knockdown	79	44
Vector control	37	59

Table 4.II: Summary of classification between K8 knockdown and vector control cell pellet spectra. A. Standard model B. LOOCV (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions)

Earlier studies have shown that K8 knockdown leads to changes in the cell morphology [244]. Contrasting morphological differences between both groups could be the main reason behind classification of vector and knockdown cells. Morphological difference between both groups of cells was further analyzed by phase contrast microscopy. As shown in Figure 4.3A**,** vector control cells have more actin based filamentous protrusions and they are elongated in shape. In contrast to vector controls, knockdown cells show very few actin based protrusions and have symmetric contracted epithelial appearance, Figure 4.3B. This pattern was compared with the already reported profile of actin in vector control and knockdown cells stained with FITC-conjugated phalloidin, followed by analysis with confocal microscopy. A reported image of vector control and knockdown

cells depicting differences in cell morphology and loss of actin filaments, adapted from reference 244 is presented in Figure 4.4.

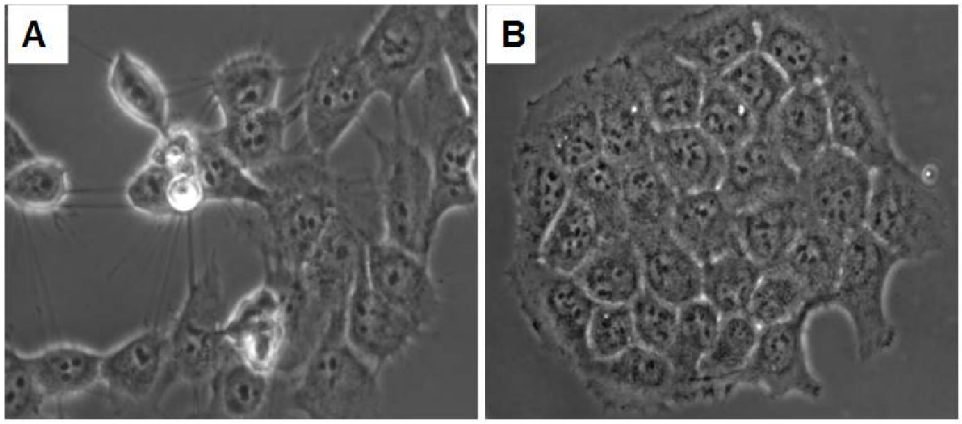


Figure 4.3: Photomicrograph depicting cell morphology of vector control (A) and knockdown cells (B) (Phase contrast microscope, 40X)

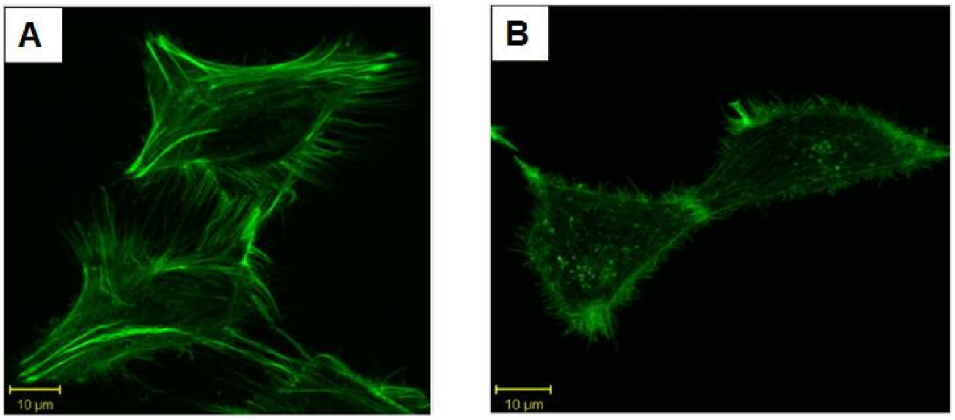


Figure 4.4: Confocal analysis of filament networks in K8 knockdown and vector control clones. (Adapted from Reference 244)

The high rate of misclassifications between knockdown and vector control cells can be primarily attributed to the fact that spectra were acquired from cell pellets rather than single cells. Further, knockdown efficiency of the system was around 80-90%, indicates probability of colocalization of both vector controls and knockdown cells in one pellet which might lead to spectral contamination or overlap. This is further supported by larger spot size of the set up (~105µm), suggesting possibility of spectral acquisition from more than one cell.

4.III Raman microspectroscopic study of single cells

To overcome the above mentioned limitations and to establish morphological difference between knockdown and vector controls cells, Raman mapping studies of single vector control and knockdown cells were taken up.

4.III.1 Methodology

4.III.1.a *Preparation of cells for Raman microspectroscoy:* K8 knockdown and vector control cells were grown on glass cover slips treated with Poly-L-Lysine (0.01% Poly-L-Lysine in milli Q water) for 48 hours. Cells were washed with pre-warmed PBS buffer (150 mM NaCl, 2mM KCl, 8 mM Na2HPO4 and 1 mM KH2PO4) thrice. After washing, the cover slip was mounted on glass slide in PBS and used for acquiring Raman spectra under Raman microscope.

4.III.1.b *Data acquisition*: Raman spectra of single vector control and knock-down cells were acquired using WITec Raman alpha300R (WITec, GmbH) confocal Raman imaging system. Photographic representation of the instrument is shown in Figure 4.5. The microscope is equipped with XYZ piezo-scan and a computer controlled sample stage with a maximum scan area of 200 x 200 x 20 μm, which enables automatic scanning of the specimen. Laser light of wavelength 532nm (Nd:YAG) at 8 mW power was used as excitation source. Laser was focused through 60X water immersion objective (Zeiss, 0.55 N.A.) on single cell. Raman signal was collected with a 100 μm fiber and directed to a 300 mm spectrograph equipped with 600 gr/mm grating and thermo-cooled CCD. Raman spectra were acquired over 70 to 3600 cm^{-1} range. For conducting experiments, a grid consisting of 100x100 pixels and covering an area of 75 x 75 μm was selected over the visible light image and 10,000 spectra/ map with integration time of 50 ms/spectrum and resolution of 0.75 μm were acquired.

Figure 4.5: Raman microspectroscopic set up

4.III.1.c *Raman mapping:* Each data point was first calibrated and corrected for the wavelength-dependent signal detection efficiency of the Raman set up. Interfering background Raman signals originating in the optical elements in the laser light delivery pathway and the glass slide were subtracted from the Raman spectra. The first derivative of the spectra was calculated using the Savitzky- Golay method with a window size of three points followed by normalization. In this way, the influence of non-informative slowly varying fluorescence or background scatter in the spectra was minimized.

Pseudo color Raman maps were constructed using K-means cluster analysis (KCA). KCA works on principle of identifying groups of spectra having resembling spectral characteristics and can easily handle the large amounts of data. Algorithms of KCA were implemented in the WITec project plus software. In KCA, first N spectra are chosen at random from the dataset. These spectra are taken as initial cluster centers. Then the distance of all spectra to these cluster centers is calculated, and the spectra are assigned to the nearest cluster center. Then for each cluster, a new center is calculated, being the average of all spectra assigned to that cluster. This procedure is

repeated until a stable solution is reached. Spectra that are highly similar and part of one cluster are believed to be obtained from areas of very similar molecular/biochemical composition. The cluster-membership information can be plotted as a pseudocolor map by assigning a color to each cluster. This procedure have been used and described in several Raman mapping studies [194-203].

4.III.2 Results and Discussion

4.III.2.a *Pseudo color Raman maps of vector control and knockdown cells:* In contrast to microscopic imaging, spectral imaging can provide both spatial and molecular information, which can help in understanding the distribution of different biochemical components. Using KCA, four and six cluster maps were generated for vector control and knockdown cells, respectively (Figure 4.6 and 4.7). Raman maps of vector control cell depict clear distinction between the membranous (Cluster I), nuclear (Cluster III) and cytoplasmic (Cluster IV) compartments of the cell. Cytoplasmic and nuclear components are separated by a perinuclear zone (Cluster II), which is known to contain extended and dense membrane network (endoplasmic reticulum, golgi apparatus etc.). Corroborating with the confocal image (Figure 4.4A) prominent actin based membrane protrusions and elongated cell shape features of vector control cells were also observed in Raman map. Similarly membranous (Cluster II), cytoplasmic (Cluster III), perinuclear (Cluster IV) and nuclear (Cluster V and VI) regions were seen in Raman map of knockdown cell, shown in Figure 4.7**.** Correlating with the confocal image (Figure 4.4B) loss of actin protrusions and change in cell morphology were observed.

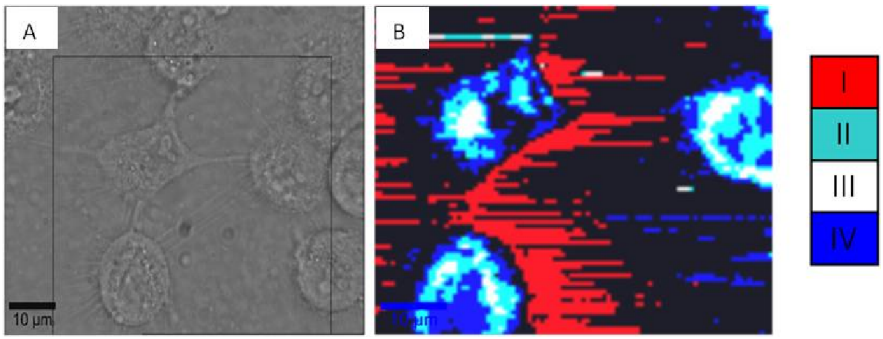


Figure 4.6: A. Photomicrograph of vector control cell B. 4-cluster pseudo color Raman map of vector control cell.

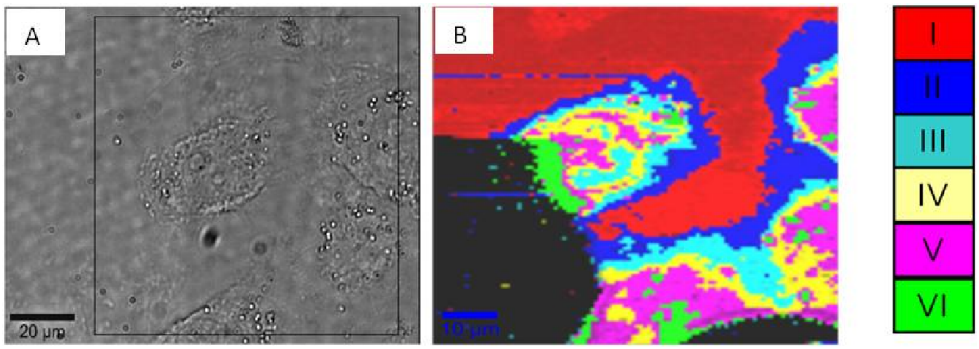


Figure 4.7: A. Photomicrograph of K8 knockdown control cell B. 6-cluster pseudo color Raman map of K8 knockdown cell.

4.III.2.b *Analysis of cluster averaged Raman spectra:* Cluster average spectrum from each cluster annotated to a cellular compartment (Figure 4.8 and 4.9) were extracted for spectral analysis. Clusters ascribed to membranous region in both vector and K8 knockdown cells have protein rich features indicated by Raman bands around 1246 cm^{-1} (amide III), 1455 cm^{-1} (CH_2 bending) and 1667 cm^{-1} (C=C stretching of proteins and α-helical structure), Figure 4.8A and Figure 4.9B. Similarly, clusters ascribed to perinuclear and nuclear compartments of the both cells have predominant protein and nucleic acid features indicated by 1321 cm^{-1} (amide III), 1450 cm^{-1}, 1667 cm^{-1}, 1340 cm^{-1}, 1373 cm^{-1} and 1591 cm^{-1} (C=N and C=C stretching), Figure 4.8B and Figure 4.9D. Clusters ascribed to the nuclear region show higher intensity of 1340 cm^{-1} and 1590 cm^{-1} bands, which can be attributed to large amount of genetic material inside the nucleus, Figure 4.8C and Figure 4.9E, F. Strong protein features indicated by bands at 1320 cm^{-1} (CH deformations of

proteins), 1450 cm^{-1} (δCH_2) and broad 1660 cm^{-1} (amide I) were seen in spectrum extracted from clusters assigned to the cytoplasmic region. Minor bands around 1340 cm^{-1} and 1591 cm^{-1} in this region suggest possibility of free nucleic acid material inside the cytoplasm, Figure 4.8D and Figure 4.9C.

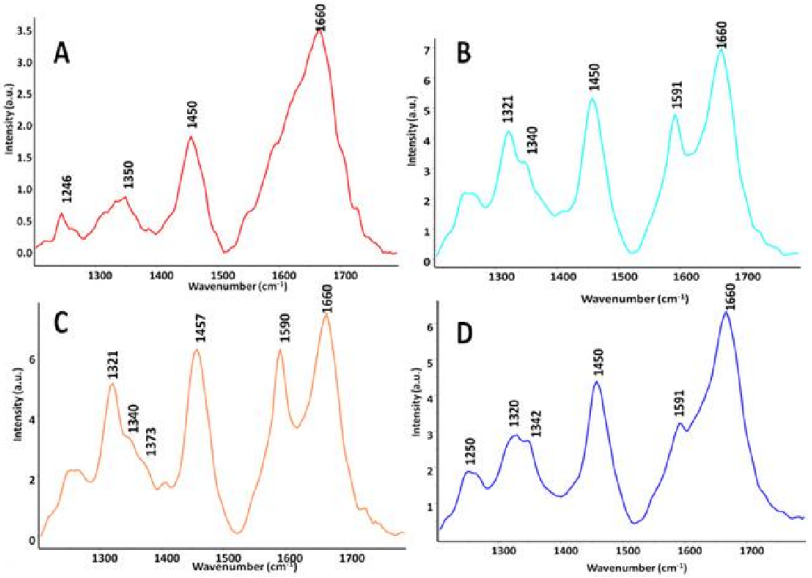


Figure 4.8: Cluster average spectrum of vector control cells A. Cluster I, B. Cluster II, C. Cluster III and D. Cluster IV.

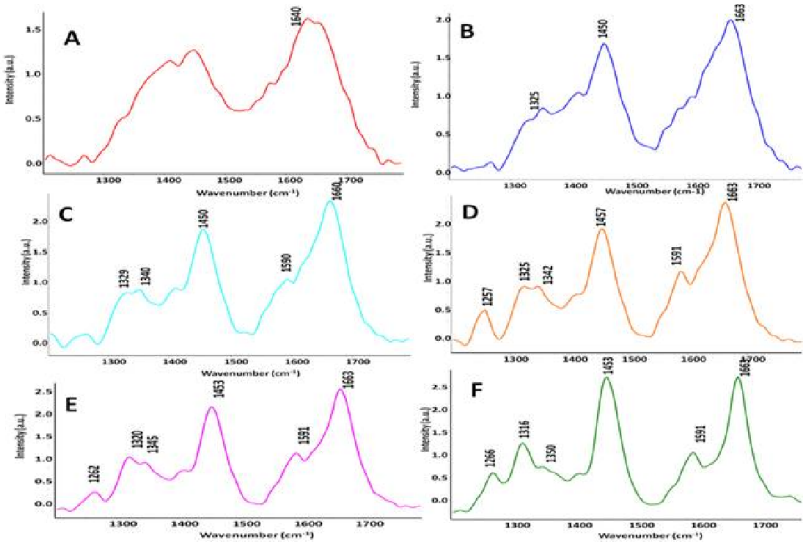


Figure 4.9: Cluster average spectrum of K8 knockdown cells A. Cluster I B. Cluster II, C. Cluster III, D. Cluster IV E. Cluster V and F. Cluster VI

Overall, findings of the study indicate potentials of Raman spectroscopy in identifying minor differences associated with loss of one protein. Morphological differences due to loss of 'keratin' protein were successfully established using Raman maps. Non-invasive phenotyping with Raman microspectroscopy could have far-reaching applications, including identification of cancerous cell phenotypes to aid in disease detection and cancer research, as a biosensor to monitor cell response to drugs for pharmaceutical testing, and as a basic cytology tool to verify cell phenotype in cultures.

Summary: Studies carried out in this Chapter aims at evaluating potentials of Raman spectroscopy in identifying minor changes associated with differential level of 'keratin' expression in oral cancer cells. Following is the brief summary of the work presented in this Chapter:

1. Tongue cancer derived AW13516 cells with reduced (knockdown) and normal (vector control) expression of keratin protein were used. Spectra were acquired from cell pellets and differences in the protein content/secondary structures were observed. PC-LDA followed by LOOCV was explored for classification.
2. Contrasting morphological differences between vector control and knockdown cells could be the main reason behind classification. It was observed that vector control cells are elongated in shape and have more actin based filamentous protrusions. In contrast to vector controls, knockdown cells show very few actin-based protrusions and have symmetric contracted epithelial appearance.
3. Morphological differences between K8 knockdown and vector control cells were further established by Raman microspectroscopic study of single cells. Pseudo color Raman maps were generated by K-means cluster analysis. Different clusters corresponding to membranous, cytoplasmic, perinuclear and nuclear regions of the cell were obtained. Cluster

averaged spectra of the perinuclear and nuclear region are dominated by nucleic acid bands while that of cytoplasmic regions are rich in proteins. Corroborating earlier observations, K8 knockdown cells show very few actin based protrusions and have symmetrical contracted epithelial appearance in contrast to elongated shape of vector control cells.

Chapter 5

Conclusions and future perspectives

Oral cancer is a major health problem in India and other South-Asian countries. India tops the prevalence of oral cancer list in the world. Although the oral cavity is easily accessible to inspection, patients with oral cancer most often present at an advanced stage when treatment is less successful there-by leading to high morbidity and mortality. Early detection remains the best way to ensure patient survival and quality of life. The current gold standard for clinical diagnosis of oral lesions is biopsy and subsequent histopathological confirmation. The process is invasive, time-consuming and prone to inter-observer variability. An alternate method of diagnosis is therefore warranted, that will enable non-invasive diagnosis of oral cavity in individuals with suspicious oral lesions. It is now well recognized that techniques based on optical spectroscopy can play a very important role towards this end. Raman spectroscopic methods are ideal for *in vivo* diagnosis because, is nondestructive, does not require external dyes, and photons can be delivered / collected via fiber-based instrumentation. These methods can provide rapid, *in situ*, objective and near real-time evaluation of a disease with high degree of accuracy. The work reported in the thesis aims at developing and evaluating potentials of *in vivo* Raman spectroscopic methods for early and non-invasive oral cancer diagnosis. The major highlights of the work are as follows:

1. ***Raman spectroscopy of ex vivo tissues***

Fiberoptic probe coupled Raman spectroscope for *in vivo* applications was procured and assembled. This set up was adapted for *ex vivo* measurements by attaching probe holder and XYZ precision stage.

In order to standardize data acquisition and analysis protocols as well as to assess the reproducibility of spectral features, spectra of *ex vivo* normal and tumor tissues were acquired. A total of 683 spectra from 36 pairs of biopsies were obtained. Lipid rich features were observed in

normal spectra while tumor showed predominanty protein bands. Classification with PC-LDA was explored and findings were validated by LOOCV and independent test data. Reproducibility of spectral features was established and objective classification between both groups was obtained. Misclassifications between both groups were analyzed by correlating spectral predictions of normal and tumor biopsies against their respective histopathology. Findings suggest that misclassification between both groups can be primarily attributed to the tissue heterogeneity *i.e.* presence of normal regions in a tumor biopsy and *vice-versa.*

Origin of Raman signals in normal tissues was explored by acquiring spectra of intact and incised oral biopsies. Findings demonstrated that morphological and architectural arrangements of different layers in a tissue contribute to the spectral signatures. Influence of surface orientation of normal tissues on classification with tumors was also assessed and it was found that orientation does not have any bearing on classification with tumor.

Spectral features of normal and tumor tissues were correlated with underlying biochemical composition. Area associated with protein (1450 cm^{-1} and 1660 cm^{-1}) and lipid 1440 cm^{-1}) bands were computed by curve fitting / deconvolution methods. These were correlated with biochemical composition of the tissue by estimating amount of total lipids, proteins and phospholipids. Spectral features as well as biochemical estimation suggest that the lipid to protein ratio is high in normal tissues in comparison to tumors. Spectral parameters derived from curve resolved protein and lipid Raman bands were found to be highly correlating with biochemical measurements.

2. ***In vivo Raman spectroscopy of oral cancers***

To the best of our knowledge, for the first time, we have demonstrated the feasibility of acquiring good quality *in vivo* Raman spectra under clinically implementable time in Indian population. The fiberoptic probe was adapted for *in vivo* measurements by attaching a detachable, metallic spacer of

length 5 mm was attached at the tip of the probe to maintain constant focus during all measurements and to avoid inter-subject contaminations. In order to ensure similar acquisition sites in all subjects, spectra were acquired from buccal mucosa as per the teeth positions. A total of 444 contralateral normal, 337 tumor and 206 premalignant spectra from 163 subjects were obtained. In addition to this 300 spectra were also acquired from 30 healthy controls (15 with and 15 without tobacco habits). Mean and difference spectra suggest predominant lipid features in normal conditions while proteins are rich in tumors. Standard models for contralateral, premalignant and tumor conditions were developed and evaluated with independent test data. Discrimination of premalignant conditions against closely associated habitual tobacco users was also demonstrated. Findings suggest that premalignant conditions in the oral cavity can be objectively classified against normal, tumor as well as closely associated habitual tobacco users. Classification of OSMF and leukoplakia, two of the most commonly occurring precancerous conditions in Indian population was explored and feasibility of classification between both conditions was demonstrated.

Various clinically or histologically unrecognizable micro-architectural changes generally precede the development of a clinically visible precancerous lesions and are attributed to 'malignancy-associated-changes' (MACs) or cancer-field-effects (CFEs)', terms often used interchangeably. Identification of CFEs or MACs may serve as a novel screening tool to reduce the morbidity and mortality associated with multiple potentially malignant transforming fields. In order to evaluate the feasibility of *in vivo* Raman spectroscopic identification of early changes which may be an indicative of neoplastic transformation, 722 spectra of 84 subjects under five categories namely healthy control (no tobacco habit, no cancer), contralateral normal (cancer and tobacco habit), non-habitués contralateral (cancer and no tobacco habit), habitués healthy controls (no cancer, tobacco habit) and tumor (cancer and tobacco habit) were analyzed. Mean and difference

spectra are suggestive of changes in protein, lipid content as well as tobacco induced hypercellularity. PC-LDA results suggest that Raman characteristics of mucosa of healthy controls are exclusive, while that of habitués healthy controls are similar to the contralateral normal mucosa, suggesting carcinogen induced field changes can be identified. It was also found that cluster of non-habitués contralateral normal mucosa is different from habitués healthy controls, indicating malignancy associated changes may be different from carcinogen induced changes and can be identified with Raman spectroscopy. The non-invasiveness and use of harmless excitation wavelength impart several advantages to this method, and thus prospectively has potential to become an ideal mass screening tool in public health programs.

3. *Raman microspectroscopy of oral cancer cells*

Keratins are one of the most widely used markers for oral cancers. Keratin 8 and 18 are expressed in simple epithelia and perform mechanical and regulatory functions in cell. Their expression is not seen in normal oral tissues but is often expressed in oral squamous cell carcinoma. Aberrant expression of keratins 8 and 18 is the most common change in human oral cancer. Study on tongue cancer derived AW13516 cell-line was taken up to evaluate potentials of Raman spectroscopy in identifying minor changes associated with differential level of keratin expression. Cells with reduced expression of keratin 8 protein were termed as 'K8 knockdown' and with normal expression termed as vector controls. In the first step spectra of K8 knockdown and vector control cell pellets were acquired using fiberoptic probe set up. Spectral features of both groups are suggestive of differences in the protein content and secondary structures. These differences were utilized for classification using PC-LDA followed by LOOCV.

Contrasting morphological differences between both groups could be the main reason behind classification. Individual cell morphology was analyzed using live cell imaging and confocal

microscopy. Findings suggest that vector control cells have more actin based filamentous protrusions and they are elongated in shape. In contrast to vector controls, knockdown cells show very few actin-based protrusions and have symmetric contracted epithelial appearance.

Morphological differences between K8 knockdown and vector control cells were further established by generating Raman maps of single cells. Spectra were acquired at 532 nm excitation with a Raman microspectrometer and maps were generated by K-means cluster analysis method. Different clusters corresponding to membranous, cytoplasmic, perinuclear and nuclear regions of the cell were obtained. Spectra of the perinuclear and nuclear region were dominated by nucleic acid bands while that of cytoplasmic regions were found to be rich in proteins. Corroborating earlier observations, K8 knockdown cells show very few actin based filaments and have symmetric contracted epithelial appearance in contrast to elongated appearance along with multiple membrane protrusions of vector control cells.

Overall findings of our study demonstrate the efficacy of Raman spectroscopic methods in conjunction with multivariate analysis tools for unambiguous and non-invasive identification of normal and pathological conditions as well as the early invisible changes which may be an indicative of neoplastic transformation in oral cancers.

Future Directions

The ultimate goal of optical spectroscopic methods is to provide an objective, real-time adjunct/alternative to cancers diagnosis. Studies carried out in the present thesis have successfully demonstrated the feasibility of classifying normal and pathological conditions in oral cancers using Raman spectroscopy in a laboratory/hospital set up. In the coming years, large scale clinical trials

must be conducted to gain the amount of data necessary for developing adequate size training and test set for robust algorithm development and analysis. These models should be tested very vigorously, preferable double-blinded, as multicentric studies, before they are contemplated for routine use. Incorporating a marking system which could be intrinsic to the probe itself should also be investigated. This will help in realizing the surgical boundary demarcation and site-wise histopathology applications. Further improvements in data analysis algorithms is also required for developing less cumbersome, rapid, unambiguous, objective and user friendly interfaces from the point of view of routine clinical use where a clinician or a technician can analyze a given spectrum against all available models to diagnose a case.

The prospective adaptation of Raman spectroscopy for routine clinical diagnosis would decrease the number of follow-up clinic visits and patient anxiety as long wait for histopathological diagnosis would be minimized to a great extent. The technology poses no known risks to the patients, and therefore could be a safe alternative/adjunct to the current diagnostic methods.

References

References

1. Ferlay, J., H. R. Shin, F. Bray, D. Forman, C. Mathers and D. M. Parkin (2010). "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008." Int J Cancer **127**(12): 2893-2917.
2. Parkin, D. M., F. Bray, J. Ferlay and P. Pisani (2005). "Global cancer statistics, 2002." CA Cancer J Clin **55**(2): 74-108.
3. Petersen, P. E. (2009). "Oral cancer prevention and control--the approach of the World Health Organization." Oral Oncol **45**(4-5): 454-460.
4. Murthy, N. S. and Mathew A. (2004). "Cancer epidemiology, prevention and control" Current Science, **86**:4:518-527
5. Dikshit, R., P. C. Gupta, C. Ramasundarahettige, V. Gajalakshmi, L. Aleksandrowicz, R. Badwe, R. Kumar, S. Roy, W. Suraweera, F. Bray, M. Mallath, P. K. Singh, D. N. Sinha, A. S. Shet, H. Gelband, P. Jha and C. Million Death Study (2012). "Cancer mortality in India: a nationally representative survey." Lancet **379**(9828): 1807-1816.
6. Saman, D. M. (2012). "A review of the epidemiology of oral and pharyngeal carcinoma: update." Head Neck Oncol **4**: 1.
7. Shimkhada, R. and J. W. Peabody (2003). "Tobacco control in India." Bull World Health Organ **81**(1): 48-52.
8. Notani, P. N. and K. Jayant (1987). "Role of diet in upper aerodigestive tract cancers." Nutr Cancer **10**(1-2): 103-113.
9. Sankaranarayanan, R., S. W. Duffy, G. Padmakumary, N. E. Day and T. K. Padmanabhan (1989). "Tobacco chewing, alcohol and nasal snuff in cancer of the gingiva in Kerala, India." Br J Cancer **60**(4): 638-643.
10. Warnakulasuriya, K. A. and R. Ralhan (2007). "Clinical, pathological, cellular and molecular lesions caused by oral smokeless tobacco-a review." J Oral Pathol Med **36**(2): 63-77.
11. Brunnmann, KD, Hoffmann D. 'Chemical composition of smokeless tobacco products' in Smokeless tobacco and health, An International perspective, NIH Publ. No.92-3461,US Department of Health and Human Service, National Institute of Health. 1992.96-108.
12. Narayan, K. M., S. L. Chadha, R. L. Hanson, R. Tandon, S. Shekhawat, R. J. Fernandes and N. Gopinath (1996). "Prevalence and patterns of smoking in Delhi: cross sectional study." BMJ **312**(7046): 1576-1579.
13. Jha, P., M. K. Ranson, S. N. Nguyen and D. Yach (2002). "Estimates of global and regional smoking prevalence in 1995, by age and sex." Am J Public Health **92**(6): 1002-1006.
14. Rahman, M., J. Sakamoto and T. Fukui (2005). "Calculation of population attributable risk for bidi smoking and oral cancer in south Asia." Prev Med **40**(5): 510-514.

15. Seitz, H. K. and F. Stickel (2007). "Molecular mechanisms of alcohol-mediated carcinogenesis." Nat Rev Cancer **7**(8): 599-612.
16. Adler-Storthz, K., G. Ficarra, K. V. Woods, D. Gaglioti, M. DiPietro and E. J. Shillitoe (1992). "Prevalence of Epstein-Barr virus and human papillomavirus in oral mucosa of HIV-infected patients." J Oral Pathol Med **21**(4): 164-170.
17. Lozada-Nur, F., J. Robinson and J. A. Regezi (1994). "Oral hairy leukoplakia in nonimmunosuppressed patients. Report of four cases." Oral Surg Oral Med Oral Pathol **78**(5): 599-602.
18. D'Costa, J., D. Saranath, V. Sanghvi and A. R. Mehta (1998). "Epstein-Barr virus in tobacco-induced oral cancers and oral lesions in patients from India." J Oral Pathol Med **27**(2): 78-82.
19. Horiuchi, K., K. Mishima, K. Ichijima, M. Sugimura, T. Ishida and T. Kirita (1995). "Epstein-Barr virus in the proliferative diseases of squamous epithelium in the oral cavity." Oral Surg Oral Med Oral Pathol Oral Radiol Endod **79**(1): 57-63.
20. Mao, E. J. and C. J. Smith (1993). "Detection of Epstein-Barr virus (EBV) DNA by the polymerase chain reaction (PCR) in oral smears from healthy individuals and patients with squamous cell carcinoma." J Oral Pathol Med **22**(1): 12-17.
21. Duensing, S., L. Y. Lee, A. Duensing, J. Basile, S. Piboonniyom, S. Gonzalez, C. P. Crum and K. Munger (2000). "The human papillomavirus type 16 E6 and E7 oncoproteins cooperate to induce mitotic defects and genomic instability by uncoupling centrosome duplication from the cell division cycle." Proc Natl Acad Sci USA **97**(18): 10002-10007.
22. Duensing, S. and K. Munger (2002). "The human papillomavirus type 16 E6 and E7 oncoproteins independently induce numerical and structural chromosome instability." Cancer Res **62**(23): 7075-7082.
23. World Health Organization. Report from a meeting of investigators on the histological definition of precancerous lesions, 1973. CAN/731, Geneva
24. Neville, B. W. and T. A. Day (2002). "Oral cancer and precancerous lesions." CA Cancer J Clin **52**(4): 195-215.
25. Tradati, N., R. Grigolat, L. Calabrese, L. Costa, G. Giugliano, F. Morelli, C. Scully, P. Boyle and F. Chiesa (1997). "Oral leukoplakias: to treat or not?" Oral Oncol **33**(5): 317-321.
26. Daftary DK, Murli PR, Bhosle RB, Gupta PC, Mehta FS, Pindborg JJ. Oral precancerous lesions and conditions of tropical interest. In: SR Prabhu, DF Wilson, DK Daftary, Johnson NW .ed Oral diseases in the tropics. Oxford:OUP,1992:402-28.

References

27. Bouquot, J. E. and H. Ephros (1995). "Erythroplakia: the dangerous red mucosa." Pract Periodontics Aesthet Dent **7**(6): 59-67; quiz 68.
28. Pindborg JJ, Reichart PA, Smith CJ, van der Waal I. Histological typing of cancer and precancer of the oral mucosa. International histological classification of tumours.World Health Organization; 1997.
29. Jeng, J. H., M. C. Chang and L. J. Hahn (2001). "Role of areca nut in betel quid-associated chemical carcinogenesis: current awareness and future perspectives." Oral Oncol **37**(6): 477-492.
30. Murti, P. R., R. B. Bhonsle, J. J. Pindborg, D. K. Daftary, P. C. Gupta and F. S. Mehta (1985). "Malignant transformation rate in oral submucous fibrosis over a 17-year period." Community Dent Oral Epidemiol **13**(6): 340-341.
31. Scully, C. and M. Carrozzo (2008). "Oral mucosal disease: Lichen planus." Br J Oral Maxillofac Surg **46**(1): 15-21.
32. Wilson, J. M. and Y. G. Jungner (1968). "Principles and practice of mass screening for disease." Bol Oficina Sanit Panam **65**(4): 281-393.
33. Sankaranarayanan, R., K. Ramadas, G. Thomas, R. Muwonge, S. Thara, B. Mathew, B. Rajan and G. Trivandrum Oral Cancer Screening Study (2005). "Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial." Lancet **365**(9475): 1927-1933.
34. Mignogna, M. D. and S. Fedele (2005). "Oral cancer screening: 5 minutes to save a life." Lancet **365**(9475): 1905-1906.
35. Burzynski, N. J., F. J. Firriolo, J. M. Butters and C. L. Sorrell (1997). "Evaluation of oral cancer screening." J Cancer Educ **12**(2): 95-99.
36. Bouquot, J.E., (1986) "Common oral lesions found during a mass screening examination." J Am Dent Assoc **112**(1):50–57.
37. Malaovalla, A. M., S. Silverman, N. J. Mani, K. F. Bilimoria and L. W. Smith (1976). "Oral cancer in 57,518 industrial workers of Gujarat, India: a prevalence and followup study." Cancer **37**(4): 1882-1886.
38. Thomson, P. J. (2002). "Field change and oral cancer: new evidence for widespread carcinogenesis?" Int J Oral Maxillofac Surg **31**(3): 262-266.
39. Fist, S. (2003). "The oral brush biopsy: separating fact from fiction." Oral Surg Oral Med Oral Pathol Oral Radiol Endod **96**(6): 654-655.
40. Eisen, D. and S. Frist (2005). "The relevance of the high positive predictive value of the oral brush biopsy." Oral Oncol **41**(7): 753-755; author reply 756.

References

41. Sciubba, J. J. (1999). "Improving detection of precancerous and cancerous oral lesions. Computer-assisted analysis of the oral brush biopsy. U.S. Collaborative OralCDx Study Group." J Am Dent Assoc **130**(10): 1445-1457.
42. Zhang, L., M. Williams, C. F. Poh, D. Laronde, J. B. Epstein, S. Durham, H. Nakamura, K. Berean, A. Hovan, N. D. Le, G. Hislop, R. Priddy, J. Hay, W. L. Lam and M. P. Rosin (2005). "Toluidine blue staining identifies high-risk primary oral premalignant lesions with poor outcome." Cancer Res **65**(17): 8017-8021.
43. Gandolfo, S., M. Pentenero, R. Broccoletti, M. Pagano, M. Carrozzo and C. Scully (2006). "Toluidine blue uptake in potentially malignant oral lesions in vivo: clinical and histological assessment." Oral Oncol **42**(1): 89-95.
44. Warnakulasuriya, K. A. and N. W. Johnson (1996). "Sensitivity and specificity of OraScan (R) toluidine blue mouthrinse in the detection of oral cancer and precancer." J Oral Pathol Med **25**(3): 97-103.
45. Onofre, M. A., M. R. Sposto and C. M. Navarro (2001). "Reliability of toluidine blue application in the detection of oral epithelial dysplasia and in situ and invasive squamous cell carcinomas." Oral Surg Oral Med Oral Pathol Oral Radiol Endod **91**(5): 535-540.
46. Zhang, L., M. Williams, C. F. Poh, D. Laronde, J. B. Epstein, S. Durham, H. Nakamura, K. Berean, A. Hovan, N. D. Le, G. Hislop, R. Priddy, J. Hay, W. L. Lam and M. P. Rosin (2005). "Toluidine blue staining identifies high-risk primary oral premalignant lesions with poor outcome." Cancer Res **65**(17): 8017-8021.
47. Gandolfo, S., M. Pentenero, R. Broccoletti, M. Pagano, M. Carrozzo and C. Scully (2006). "Toluidine blue uptake in potentially malignant oral lesions in vivo: clinical and histological assessment." Oral Oncol **42**(1): 89-95.
48. Epstein, J. B., M. Gorsky, S. Lonky, S. Silverman, Jr., J. D. Epstein and M. Bride (2006). "The efficacy of oral lumenoscopy (ViziLite) in visualizing oral mucosal lesions." Spec Care Dentist **26**(4): 171-174.
49. Kerr, A. R., D. A. Sirois and J. B. Epstein (2006). "Clinical evaluation of chemiluminescent lighting: an adjunct for oral mucosal examinations." J Clin Dent **17**(3): 59-63.
50. Farah, C. S. and M. J. McCullough (2007). "A pilot case control study on the efficacy of acetic acid wash and chemiluminescent illumination (ViziLite) in the visualisation of oral mucosal white lesions." Oral Oncol **43**(8): 820-824.

51. Epstein, J. B., M. Gorsky, S. Lonky, S. Silverman, Jr., J. D. Epstein and M. Bride (2006). "The efficacy of oral lumenoscopy (ViziLite) in visualizing oral mucosal lesions." Spec Care Dentist **26**(4): 171-174.
52. Huber, M.A., S.A. Bsoul and G.T.Terezhalmy (2004). "Acetic acid wash and chemiluminescent illumination as an adjunct to conventional oral soft tissue examination for the detection of dysplasia: a pilot study" Quintessence Int **35**(5):378-384.
53. Oh, E. S. and Laskin, D. M. (2007). "Efficacy of the ViziLite system in the identification of oral lesions." J Oral Maxillofac Surg **65**(3): 424-426.
54. Ram, S. and C. H. Siar (2005). "Chemiluminescence as a diagnostic aid in the detection of oral cancer and potentially malignant epithelial lesions." Int J Oral Maxillofac Surg **34**(5): 521-527.
55. J.R. Lakowicz, Principles of Fluorescence Spectroscopy, Plenum Press, New York, 1983
56. Betz, C. S., H. Stepp, P. Janda, S. Arbogast, G. Grevers, R. Baumgartner and A. Leunig (2002). "A comparative study of normal inspection, autofluorescence and 5-ALA-induced PPIX fluorescence for oral cancer diagnosis." Int J Cancer **97**(2): 245-252.
57. Leunig, A., C. S. Betz, M. Mehlmann, H. Stepp, S. Arbogast, G. Grevers and R. Baumgartner (2000). "Detection of squamous cell carcinoma of the oral cavity by imaging 5-aminolevulinic acid-induced protoporphyrin IX fluorescence." Laryngoscope **110**(1): 78-83.
58. Leunig, A., M. Mehlmann, C. Betz, H. Stepp, S. Arbogast, G. Grevers and R. Baumgartner (2001). "Fluorescence staining of oral cancer using a topical application of 5-aminolevulinic acid: fluorescence microscopic studies." J Photochem Photobiol B **60**(1): 44-49.
59. Schantz, S. P., V. Kolli, H. E. Savage, G. Yu, J. P. Shah, D. E. Harris, A. Katz, R. R. Alfano and A. G. Huvos (1998). "In vivo native cellular fluorescence and histological characteristics of head and neck cancer." Clin Cancer Res **4**(5): 1177-1182.
60. Harris, D. M. and J. Werkhaven (1987). "Endogenous porphyrin fluorescence in tumors." Lasers Surg Med **7**(6): 467-472.
61. Onizawa, K., N. Okamura, H. Saginoya, H. Yusa, T. Yanagawa and H. Yoshida (2002). "Analysis of fluorescence in oral squamous cell carcinoma." Oral Oncol **38**(4): 343-348.
62. Onizawa, K., N. Okamura, H. Saginoya and H. Yoshida (2003). "Characterization of autofluorescence in oral squamous cell carcinoma." Oral Oncol **39**(2): 150-156.
63. Onizawa, K., H. Saginoya, Y. Furuya, H. Yoshida and H. Fukuda (1999). "Usefulness of fluorescence photography for diagnosis of oral cancer." Int J Oral Maxillofac Surg **28**(3): 206-210.
64. Kolli, V. R., H. E. Savage, T. J. Yao and S. P. Schantz (1995). "Native cellular fluorescence of neoplastic upper aerodigestive mucosa." Arch Otolaryngol Head Neck Surg **121**(11): 1287-1292.

65. Kolli, V. R., H. E. Savage, T. J. Yao and S. P. Schantz (1995). "Native cellular fluorescence of neoplastic upper aerodigestive mucosa." Arch Otolaryngol Head Neck Surg **121**(11): 1287-1292.

66. Chen, C. T., C. Y. Wang, Y. S. Kuo, H. H. Chiang, S. N. Chow, I. Y. Hsiao and C. P. Chiang (1996). "Light-induced fluorescence spectroscopy: a potential diagnostic tool for oral neoplasia." Proc Natl Sci Counc Repub China B **20**(4): 123-130.

67. Chen, C. T., H. K. Chiang, S. N. Chow, C. Y. Wang, Y. S. Lee, J. C. Tsai and C. P. Chiang (1998). "Autofluorescence in normal and malignant human oral tissues and in DMBA-induced hamster buccal pouch carcinogenesis." J Oral Pathol Med **27**(10): 470-474.

68. Gillenwater, A., R. Jacob, R. Ganeshappa, B. Kemp, A. K. El-Naggar, J. L. Palmer, G. Clayman, M. F. Mitchell and R. Richards-Kortum (1998). "Noninvasive diagnosis of oral neoplasia based on fluorescence spectroscopy and native tissue autofluorescence." Arch Otolaryngol Head Neck Surg **124**(11): 1251-1258.

69. Nazeer Shaiju, S., S. Ariya, R. Asish, P. Salim Haris, B. Anita, G. Arun Kumar and R. S. Jayasree (2011). "Habits with killer instincts: in vivo analysis on the severity of oral mucosal alterations using autofluorescence spectroscopy." J Biomed Opt **16**(8): 087006.

70. Rigas, B., S. Morgello, I. S. Goldman and P. T. Wong (1990). "Human colorectal cancers display abnormal Fourier-transform infrared spectra." Proc Natl Acad Sci U S A **87**(20): 8140-8144.

71. Jackson, M., L. P. Choo, P. H. Watson, W. C. Halliday and H. H. Mantsch (1995). "Beware of connective tissue proteins: assignment and implications of collagen absorptions in infrared spectra of human tissues." Biochim Biophys Acta **1270**(1): 1-6.

72. Bellisola, G. and C. Sorio (2012). "Infrared spectroscopy and microscopy in cancer research and diagnosis." Am J Cancer Res **2**(1): 1-21.

73. Wu, J. G., Y. Z. Xu, C. W. Sun, R. D. Soloway, D. F. Xu, Q. G. Wu, K. H. Sun, S. F. Weng and G. X. Xu (2001). "Distinguishing malignant from normal oral tissues using FTIR fiber-optic techniques." Biopolymers **62**(4): 185-192.

74. Fukuyama, Y., S. Yoshida, S. Yanagisawa and M. Shimizu (1999). "A study on the differences between oral squamous cell carcinomas and normal oral mucosas measured by Fourier transform infrared spectroscopy." Biospectroscopy **5**(2): 117-126.

75. Schultz, C. P., K. Z. Liu, P. D. Kerr and H. H. Mantsch (1998). "In situ infrared histopathology of keratinization in human oral/oropharyngeal squamous cell carcinoma." Oncol Res **10**(5): 277-286.

76. Sukuta, S. and R. Bruch (1999). "Factor analysis of cancer Fourier transform infrared evanescent wave fiberoptical (FTIR-FEW) spectra." Lasers Surg Med **24**(5): 382-388.

77. Raman C.V. and Krishnan K.S. (1928). “A change of wavelength in light scattering. Part I.” Indian J Phys **2,**399.

78. Raman C.V. and Krishnan K.S. (1928). “A change of wavelength in light scattering.” Nature **121,**501.

79. Hanlon, E. B., R. Manoharan, T. W. Koo, K. E. Shafer, J. T. Motz, M. Fitzmaurice, J. R. Kramer, I. Itzkan, R. R. Dasari and M. S. Feld (2000). "Prospects for in vivo Raman spectroscopy." Phys Med Biol **45**(2): R1-59.

80. Nie, S. and S. R. Emory (1997). "Probing Single Molecules and Single Nanoparticles by Surface-Enhanced Raman Scattering." Science **275**(5303): 1102-1106.

81. Kneipp, J., H. Kneipp, W. L. Rice and K. Kneipp (2005). "Optical probes for biological applications based on surface-enhanced Raman scattering from indocyanine green on gold nanoparticles." Anal Chem **77**(8): 2381-2385.

82. Kim, J. H., J. S. Kim, H. Choi, S. M. Lee, B. H. Jun, K. N. Yu, E. Kuk, Y. K. Kim, D. H. Jeong, M. H. Cho and Y. S. Lee (2006). "Nanoparticle probes with surface enhanced Raman spectroscopic tags for cellular cancer targeting." Anal Chem **78**(19): 6967-6973.

83. Yazdi, Y., N. Ramanujam, R. Lotan, M. F. Mitchell, W. Hittelman and R. Richards-Kortum (1999). "Resonance Raman Spectroscopy at 257 nm Excitation of Normal and Malignant Cultured Breast and Cervical Cells." Appl. Spectrosc. **53**(1): 82-85.

84. Taleb, A., J. Diamond, J. J. McGarvey, J. R. Beattie, C. Toland and P. W. Hamilton (2006). "Raman microscopy for the chemometric analysis of tumor cells." J Phys Chem B **110**(39): 19625-19631.

85. Chan, J. W., D. S. Taylor, T. Zwerdling, S. M. Lane, K. Ihara and T. Huser (2006). "Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells." Biophys J **90**(2): 648-656.

86. Koljenovic, S., T. C. Bakker Schut, J. P. van Meerbeeck, A. P. Maat, S. A. Burgers, P. E. Zondervan, J. M. Kros and G. J. Puppels (2004). "Raman microspectroscopic mapping studies of human bronchial tissue." J Biomed Opt **9**(6): 1187-1197.

87. Nan, X., J. X. Cheng and X. S. Xie (2003). "Vibrational imaging of lipid droplets in live fibroblast cells with coherent anti-Stokes Raman scattering microscopy." J Lipid Res **44**(11): 2202-2208.

88. Cheng, J. X., S. Pautot, D. A. Weitz and X. S. Xie (2003). "Ordering of water molecules between phospholipid bilayers visualized by coherent anti-Stokes Raman scattering microscopy." Proc Natl Acad Sci USA **100**(17): 9826-9830.

89. Ramsay, A. D. (1999). "Errors in histopathology reporting: detection and avoidance." Histopathology **34**(6): 481-490.

References

90. Abbey, L. M., G. E. Kaugars, J. C. Gunsolley, J. C. Burns, D. G. Page, J. A. Svirsky, E. Eisenberg, D. J. Krutchkoff and M. Cushing (1995). "Intraexaminer and interexaminer reliability in the diagnosis of oral epithelial dysplasia." Oral Surg Oral Med Oral Pathol Oral Radiol Endod **80**(2): 188-191.
91. Tietz NW,ed. (1976) Fundamentals of clinical biochemistry. Philadelphia:saunders,103-119
92. C. Chatfield and A. J. Collins, Introduction to Multivariate Analysis, Chapman & Hall, London, 1980.
93. B. F. J. Manly, Multivariate Statistical Methods: A Primer, Chapman & Hall, London, 1994.
94. Ellis, D. I. and R. Goodacre (2006). "Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy." Analyst **131**(8): 875-885.
95. Polat, K. and S. Gunes (2008). "Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer." Expert Syst. Appl. **34:**214–221.
96. Wang, L. and B. Mizaikoff, (2008). "Application of multivariate data-analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy." Anal. Bioanal. Chem.**391**:1641–1654.
97. Korenius, T., J. Laurikkala, and M. Juhola, (2007). "On principal component analysis, cosine and Euclidean measures in information retrieval."Inform.Sci.**177**: 4893-4905.
98. PLSplus/IQ User's Guide, Galactic Industries Corporation, Salem, New Hampshire (1999).
99. German, M. J., A. Hammiche, N. Ragavan, M. J. Tobin, L. J. Cooper, S. S. Matanhelia, A. C. Hindley, C. M. Nicholson, N. J. Fullwood, H. M. Pollock and F. L. Martin (2006). "Infrared spectroscopy with multivariate analysis potentially facilitates the segregation of different types of prostate cell." Biophys J **90**(10):3783-3795.
100. Kemsley, E .K. (1996). "Discriminant analysis of high-dimensional data: a comparison of principal components analysis and partial least squares data reduction methods." Chemom. Intell. Lab.Syst.**33:**47–61.
101. Hastie, T.; Tibshirani, R.; Friedman, J.The Elements of Statistical Learning, 2nd ed.; Springer: New York, 2010 R. A. Eisenbeis and R. B. Avery, Discriminant Analysis and Classification Procedures: Theory and Applications, Heath, Lexington, MA (1972).
102. Barnes, R. J., M. S. Dhanoa and S. J. Lister (1993). "Correction to the description of standard normal variate (SNV) and detrend (DT) transformations in practical spectroscopy with applications in food and beverage analysis,"J. Near Infrared Spectrosc.**1**:185–186.

References

103. Harvey, T. J., E. Gazi, A. Henderson, R. D. Snook, N. W. Clarke, M. Brown and P. Gardner (2009). "Factors influencing the discrimination and classification of prostate cancer cell lines by FTIR microspectroscopy." Analyst **134**(6): 1083-1091.

104. Fearn, T.Handbook of Vibrational Spectroscopy; Chalmers, J.M.,Griffiths, P. R, Eds.; Wiley Publications: Chichester, 2002

105. Devijver, Pierre A.; and Kittler, Josef; Pattern Recognition: A Statistical Approach, Prentice-Hall, London, GB, 1982

106. Stone, Mervyn (1977). "Asymptotics for and against cross-validation". Biometrika **64** (1): 29–35.

107. Efron, Bradley; Tibshirani, Robert (1997). "Improvements on cross-validation: The .632 + Bootstrap Method". Journal of the American Statistical Association **92** (438): 548–560

108. Kenneth Levenberg (1944). "A Method for the Solution of Certain Non-Linear Problems in Least Squares". Quarterly of Applied Mathematics **2**: 164–168.

109. Donald Marquardt (1963). "An Algorithm for Least-Squares Estimation of Nonlinear Parameters". *SIAM* Journal on Applied Mathematics **11** (2): 431–441.

110. Philip E. Gill and Walter Murray (1978). "Algorithms for the solution of the nonlinear least-squares problem". SIAM Journal on Numerical Analysis **15**(5): 977–992

111. Collette, T. W. and T. L. Williams (2002). "The role of Raman spectroscopy in the analytical chemistry of potable water." J Environ Monit **4**(1): 27-34.

112. Lepot, L., K. De Wael, F. Gason and B. Gilbert (2008). "Application of Raman spectroscopy to forensic fibre cases." Sci Justice **48**(3): 109-117.

113. Izake, E. L. (2010). "Forensic and homeland security applications of modern portable Raman spectroscopy." Forensic Sci Int **202**(1-3): 1-8.

114. Bersani, D. and J. M. Madariaga (2012). "Applications of Raman spectroscopy in art and archaeology." Journal of Raman Spectroscopy **43**(11): 1523-1528.

115. Lyon, L. A., C. D. Keating, A. P. Fox, B. E. Baker, L. He, S. R. Nicewarner, S. P. Mulvaney and M. J. Natan (1998). "Raman spectroscopy." Anal Chem **70**(12): 341R-361R.

116. Verrier, S., A. Zoladek and I. Notingher (2011). "Raman micro-spectroscopy as a non-invasive cell viability test." Methods Mol Biol **740**: 179-189.

117. Motz, J. T., M. Fitzmaurice, A. Miller, S. J. Gandhi, A. S. Haka, L. H. Galindo, R. R. Dasari, J. R. Kramer and M. S. Feld (2006). "In vivo Raman spectral pathology of human atherosclerosis and vulnerable plaque." J Biomed Opt **11**(2): 021003.

118. Nogueira, G. V., L. Silveira, A. A. Martin, R. A. Zangaro, M. T. Pacheco, M. C. Chavantes and C. A. Pasqualucci (2005). "Raman spectroscopy study of atherosclerosis in human carotid artery." J Biomed Opt **10**(3): 031117.

119. Chen, P., A. Shen, W. Zhao, S. J. Baek, H. Yuan and J. Hu (2009). "Raman signature from brain hippocampus could aid Alzheimer's disease diagnosis." Appl Opt **48**(24): 4743-4748.

120. Schipper, H. M., C. S. Kwok, S. M. Rosendahl, D. Bandilla, O. Maes, C. Melmed, D. Rabinovitch and D. H. Burns (2008). "Spectroscopy of human plasma for diagnosis of idiopathic Parkinson's disease." Biomark Med **2**(3): 229-238.

121. Maher, J. R., M. Takahata, H. A. Awad and A. J. Berger (2011). "Raman spectroscopy detects deterioration in biomechanical properties of bone in a glucocorticoid-treated mouse model of rheumatoid arthritis." J Biomed Opt **16**(8): 087012.

122. Dingari, N. C., I. Barman, G. P. Singh, J. W. Kang, R. R. Dasari and M. S. Feld (2011). "Investigation of the specificity of Raman spectroscopy in non-invasive blood glucose measurements." Anal Bioanal Chem **400**(9): 2871-2880.

123. Gniadecka, M., O. Faurskov Nielsen, D. H. Christensen and H. C. Wulf (1998). "Structure of water, proteins, and lipids in intact human skin, hair, and nail." J Invest Dermatol **110**(4): 393-398.

124. Alfano, R.,R. C.H. Liu, W.L. Sha, H.R. Zhu, D.L. Atkins, J. Cleary, R. Prudente and E. Clemer (1991).“Human Breast Tissues Studied by IR Fourier Transform Raman Spectroscopy” Lasers Life Sci.,**4**(23) 23-28.

125. Redd, D. C. B., Z. C. Feng, K. T. Yue and T. S. Gansler (1993). "Raman Spectroscopic Characterization of Human Breast Tissues: Implications for Breast Cancer Diagnosis." Appl. Spectrosc. **47**(6): 787-791.

126. Frank, C. J., D. C. Redd, T. S. Gansler and R. L. McCreery (1994). "Characterization of human breast biopsy specimens with near-IR Raman spectroscopy." Anal Chem **66**(3): 319-326.

127. Frank, C. J., R. L. McCreery and D. C. Redd (1995). "Raman spectroscopy of normal and diseased human breast tissues." Anal Chem **67**(5): 777-783.

128. Hanlon, E. B., R. Manoharan, T. W. Koo, K. E. Shafer, J. T. Motz, M. Fitzmaurice, J. R. Kramer, I. Itzkan, R. R. Dasari and M. S. Feld (2000). "Prospects for in vivo Raman spectroscopy." Phys Med Biol **45**(2): R1-59.

129. Haka, A. S., K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari and M. S. Feld (2005). "Diagnosing breast cancer by using Raman spectroscopy." Proc Natl Acad Sci USA **102**(35): 12371-12376.

130. Chowdary, M. V. P., K. K. Kumar, J. Kurien, S. Mathew and C. M. Krishna (2006). "Discrimination of normal, benign, and malignant breast tissues by Raman spectroscopy." Biopolymers **83**(5): 556-569.

131. Kalyan Kumar, K., M. V. P. Chowdary, S. Mathew, L. Rao, C. Murali Krishna and J. Kurien (2008). "Raman spectroscopic diagnosis of breast cancers: evaluation of models." Journal of Raman Spectroscopy **39**(9): 1276-1282.

132. Gniadecka, M., P. A. Philipsen, S. Sigurdsson, S. Wessel, O. F. Nielsen, D. H. Christensen, J. Hercogova, K. Rossen, H. K. Thomsen, R. Gniadecki, L. K. Hansen and H. C. Wulf (2004). "Melanoma diagnosis by Raman spectroscopy and neural networks: structure alterations in proteins and lipids in intact cancer tissue." J Invest Dermatol **122**(2): 443-449.

133. Nijssen, A., K. Maquelin, L. F. Santos, P. J. Caspers, T. C. Bakker Schut, J. C. den Hollander, M. H. Neumann and G. J. Puppels (2007). "Discriminating basal cell carcinoma from perilesional skin using high wave-number Raman spectroscopy." J Biomed Opt **12**(3): 034004.

134. Mahadevan-Jansen, A., M. F. Mitchell, N. Ramanujam, A. Malpica, S. Thomsen, U. Utzinger and R. Richards-Kortum (1998). "Near-infrared Raman spectroscopy for in vitro detection of cervical precancers." Photochem Photobiol **68**(1): 123-132.

135. Krishna, C. M., N. B. Prathima, R. Malini, B. M. Vadhiraja, R. A. Bhatt, D. J. Fernandes, P. Kushtagi, M. S. Vidyasagar and V. B. Kartha (2006). "Raman spectroscopy studies for diagnosis of cancers in human uterine cervix." Vibrational Spectroscopy **41**(1): 136-141.

136. Krishna, C. M., G. D. Sockalingum, B. M. Vadhiraja, K. Maheedhar, A. C. Rao, L. Rao, L. Venteo, M. Pluot, D. J. Fernandes, M. S. Vidyasagar, V. B. Kartha and M. Manfait (2007). "Vibrational spectroscopy studies of formalin-fixed cervix tissues." Biopolymers **85**(3): 214-221.

137. Vidyasagar, M. S., K. Maheedhar, B. M. Vadhiraja, D. J. Fernendes, V. B. Kartha and C. M. Krishna (2008). "Prediction of radiotherapy response in cervix cancer by Raman spectroscopy: a pilot study." Biopolymers **89**(6): 530-537.

138. Yamazaki, H., S. Kaminaka, E. Kohda, M. Mukai and H. O. Hamaguchi (2003). "The diagnosis of lung cancer using 1064-nm excited near-infrared multichannel Raman spectroscopy." Radiat Med **21**(1): 1-6.

139. Stone, N., P. Stavroulaki, C. Kendall, M. Birchall and H. Barr (2000). "Raman spectroscopy for early detection of laryngeal malignancy: preliminary results." Laryngoscope **110**(10 Pt 1): 1756-1763.

140. Stone, N., C. Kendall, J. Smith, P. Crow and H. Barr (2004). "Raman spectroscopy for identification of epithelial cancers." Faraday Discuss **126**: 141-157; discussion 169-183.

References

141. Crow, P., A. Molckovsky, N. Stone, J. Uff, B. Wilson and L. M. WongKeeSong (2005). "Assessment of fiberoptic near-infrared raman spectroscopy for diagnosis of bladder and prostate cancer." Urology **65**(6): 1126-1130.

142. Kendall, C., N. Stone, N. Shepherd, K. Geboes, B. Warren, R. Bennett and H. Barr (2003). "Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus." J Pathol **200**(5): 602-609.

143. Koljenovic, S., T. C. Bakker Schut, R. Wolthuis, B. de Jong, L. Santos, P. J. Caspers, J. M. Kros and G. J. Puppels (2005). "Tissue characterization using high wave number Raman spectroscopy." J Biomed Opt **10**(3): 031116.

144. Krafft, C., S. B. Sobottka, G. Schackert and R. Salzer (2006). "Raman and infrared spectroscopic mapping of human primary intracranial tumors: a comparative study." Journal of Raman Spectroscopy **37**(1-3): 367-375.

145. Morris, M. D., W. F. Finney, R. M. Rajachar and D. H. Kohn (2004). "Bone tissue ultrastructural response to elastic deformation probed by Raman spectroscopy." Faraday Discuss **126**: 159-168; discussion 169-183.

146. Awonusi, A., M. D. Morris and M. M. Tecklenburg (2007). "Carbonate assignment and calibration in the Raman spectrum of apatite." Calcif Tissue Int **81**(1): 46-52.

147. Esmonde-White, K. A., G. S. Mandair, F. Raaii, J. A. Jacobson, B. S. Miller, A. G. Urquhart, B. J. Roessler and M. D. Morris (2009). "Raman spectroscopy of synovial fluid as a tool for diagnosing osteoarthritis." J Biomed Opt **14**(3): 034013.

148. Morris, M. D. and G. S. Mandair (2011). "Raman assessment of bone quality." Clin Orthop Relat Res **469**(8): 2160-2169.

149. Esmonde-White, K. A., F. W. Esmonde-White, M. D. Morris and B. J. Roessler (2013). "Effects of pathology dyes on Raman bone spectra." J Biomed Opt **18**(5): 57002.

150. Crow, P., A. Ritchie, M. Wright, C. Kendall and N. Stone (2002). "Evaluation of Raman spectroscopy to provide a real time, optical method for discrimination between normal and abnormal tissue in the prostate." European Urology Supplements **1**(1): 92.

151. Crow, P., N. Stone, C. A. Kendall, J. S. Uff, J. A. Farmer, H. Barr and M. P. Wright (2003). "The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma in vitro." Br J Cancer **89**(1): 106-108.

152. Crow, P., N. Stone, C. A. Kendall, R. A. Persad and M. P. Wright (2003). "Optical diagnostics in urology: current applications and future prospects." BJU Int **92**(4): 400-407.

153. Chowdary, M. V., K. K. Kumar, K. Thakur, A. Anand, J. Kurien, C. M. Krishna and S. Mathew (2007). "Discrimination of normal and malignant mucosal tissues of the colon by Raman spectroscopy." Photomed Laser Surg **25**(4): 269-274.

154. Krishna, C. M., G. D. Sockalingum, R. A. Bhat, L. Venteo, P. Kushtagi, M. Pluot and M. Manfait (2007). "FTIR and Raman microspectroscopy of normal, benign, and malignant formalin-fixed ovarian tissues." Anal Bioanal Chem **387**(5): 1649-1656.

155. Kalyan Kumar, K., A. Anand, M. V. P. Chowdary, Keerthi, J. Kurien, C. Murali Krishna and S. Mathew (2007). "Discrimination of normal and malignant stomach mucosal tissues by Raman spectroscopy: A pilot study." Vibrational Spectroscopy **44**(2): 382-387.

156. Haka, A. S., Z. Volynskaya, J. A. Gardecki, J. Nazemi, J. Lyons, D. Hicks, M. Fitzmaurice, R. R. Dasari, J. P. Crowe and M. S. Feld (2006). "In vivo margin assessment during partial mastectomy breast surgery using raman spectroscopy." Cancer Res **66**(6): 3317-3322.

157. Bitar, R., M. A. Martins, D. Ribeiro, C. Carvalho, E. A. P. Santos, L. N. Z. Ramalho, F. Ramalho, H. Martinho and A. A. Martin (2008). "In vivo Raman spectroscopy for breast cancer: diagnosis in animal model." Proc. SPIE **6853**.

158. Lieber, C. A., S. K. Majumder, D. L. Ellis, D. D. Billheimer and A. Mahadevan-Jansen (2008). "In vivo nonmelanoma skin cancer diagnosis using Raman microspectroscopy." Lasers Surg Med **40**(7): 461-467.

159. Kirsch, M., G. Schackert, R. Salzer and C. Krafft (2010). "Raman spectroscopic imaging for *in vivo* detection of cerebral brain metastases." Anal Bioanal Chem **398**(4): 1707-1713.

160. Beljebbar, A., S. Dukic, N. Amharref and M. Manfait (2010). "Ex vivo and in vivo diagnosis of C6 glioblastoma development by Raman spectroscopy coupled to a microprobe." Anal Bioanal Chem **398**(1): 477-487.

161. Bergholt, M. S., W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. So, A. Shabbir and Z. Huang (2013). "Fiber-optic Raman spectroscopy probes gastric carcinogenesis in vivo at endoscopy." J Biophotonics **6**(1): 49-59.

162. Bergholt, M. S., W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. So and Z. Huang (2010). "Raman endoscopy for in vivo differentiation between benign and malignant ulcers in the stomach." Analyst **135**(12): 3162-3168.

163. Teh, S. K., W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh and Z. Huang (2010). "Near-infrared Raman spectroscopy for optical diagnosis in the stomach: identification of Helicobacter-pylori infection and intestinal metaplasia." Int J Cancer **126**(8): 1920-1927.

164. Bergholt, M. S., W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Yan So and Z. Huang (2011). "In vivo diagnosis of gastric cancer using Raman endoscopy and ant colony optimization techniques." Int J Cancer **128**(11): 2673-2680.

165. Huang, Z., M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh and K. G. Yeoh (2010). "In vivo early diagnosis of gastric dysplasia using narrow-band image-guided Raman endoscopy." J Biomed Opt **15**(3): 037017.

166. Huang, Z., S. K. Teh, W. Zheng, K. Lin, K. Y. Ho, M. Teh and K. G. Yeoh (2010). "In vivo detection of epithelial neoplasia in the stomach using image-guided Raman endoscopy." Biosens Bioelectron **26**(2): 383-389.

167. Bergholt, M. S., W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. So and Z. Huang (2011). "Characterizing variability in *in vivo* Raman spectra of different anatomical locations in the upper gastrointestinal tract toward cancer detection." J Biomed Opt **16**(3): 037003.

168. Duraipandian, S., M. Sylvest Bergholt, W. Zheng, K. Yu Ho, M. Teh, K. Guan Yeoh, J. Bok Yan So, A. Shabbir and Z. Huang (2012). "Real-time Raman spectroscopy for in vivo, online gastric cancer diagnosis during clinical endoscopic examination." J Biomed Opt **17**(8): 081418.

169. Day, J. C., R. Bennett, B. Smith, C. Kendall, J. Hutchings, G. M. Meaden, C. Born, S. Yu and N. Stone (2009). "A miniature confocal Raman probe for endoscopic use." Phys Med Biol **54**(23): 7077-7087.

170. Short, M. A., S. Lam, A. McWilliams, J. Zhao, H. Lui and H. Zeng (2008). "Development and preliminary results of an endoscopic Raman probe for potential in vivo diagnosis of lung cancers." Opt Lett **33**(7): 711-713.

171. Mahadevan-Jansen, A., M. F. Mitchell, N. Ramanujam, U. Utzinger and R. Richards-Kortum (1998). "Development of a fiberoptic probe to measure NIR Raman spectra of cervical tissue in vivo." Photochem Photobiol **68**(3): 427-431.

172. Robichaux-Viehoever, A., E. Kanter, H. Shappell, D. Billheimer, H. Jones, 3rd and A. Mahadevan-Jansen (2007). "Characterization of Raman spectra measured in vivo for the detection of cervical dysplasia." Appl Spectrosc **61**(9): 986-993.

173. Kanter, E. M., S. Majumder, E. Vargis, A. Robichaux-Viehoever, G. J. Kanter, H. Shappell, H. W. Jones, 3rd and A. Mahadevan-Jansen (2009). "Multiclass discrimination of cervical precancers using Raman spectroscopy." J Raman Spectrosc **40**(2): 205-211.

174. Kanter, E. M., E. Vargis, S. Majumder, M. D. Keller, E. Woeste, G. G. Rao and A. Mahadevan-Jansen (2009). "Application of Raman spectroscopy for cervical dysplasia diagnosis." J Biophotonics **2**(1-2): 81-90.

175. Kanter, E. M., S. Majumder, G. J. Kanter, E. M. Woeste and A. Mahadevan-Jansen (2009). "Effect of hormonal variation on Raman spectra for cervical disease detection." Am J Obstet Gynecol **200**(5): 512 e511-515.

176. Vargis, E., N. Brown, K. Williams, A. Al-Hendy, B. C. Paria, J. Reese and A. Mahadevan-Jansen (2012). "Detecting biochemical changes in the rodent cervix during pregnancy using Raman spectroscopy." Ann Biomed Eng **40**(8): 1814-1824.

177. Duraipandian, S., W. Zheng, J. Ng, J. J. Low, A. Ilancheran and Z. Huang (2013). "Non-invasive analysis of hormonal variations and effect of postmenopausal Vagifem treatment on women using in vivo high wave number confocal Raman spectroscopy." Analyst **138**(14): 4120-4128.

178. Puppels, G. J., F. F. de Mul, C. Otto, J. Greve, M. Robert-Nicoud, D. J. Arndt-Jovin and T. M. Jovin (1990). "Studying single living cells and chromosomes by confocal Raman microspectroscopy." Nature **347**(6290): 301-303.

179. Puppels, G. J., H. S. Garritsen, G. M. Segers-Nolten, F. F. de Mul and J. Greve (1991). "Raman microspectroscopic approach to the study of human granulocytes." Biophys J **60**(5): 1046-1056.

180. Puppels, G. J., J. H. Olminkhof, G. M. Segers-Nolten, C. Otto, F. F. de Mul and J. Greve (1991). "Laser irradiation and Raman spectroscopy of single living cells and chromosomes: sample degradation occurs with 514.5 nm but not with 660 nm laser light." Exp Cell Res **195**(2): 361-367.

181. Puppels, G. J., H. S. Garritsen, J. A. Kummer and J. Greve (1993). "Carotenoids located in human lymphocyte subpopulations and natural killer cells by Raman microspectroscopy." Cytometry **14**(3): 251-256.

182. Hawi, S. R., W. B. Campbell, A. Kajdacsy-Balla, R. Murphy, F. Adar and K. Nithipatikom (1996). "Characterization of normal and malignant human hepatocytes by Raman microspectroscopy." Cancer Lett **110**(1-2): 35-40.

183. Uzunbajakava, N., A. Lenferink, Y. Kraan, E. Volokhina, G. Vrensen, J. Greve and C. Otto (2003). "Nonresonant confocal Raman imaging of DNA and protein distribution in apoptotic cells." Biophys J **84**(6): 3968-3981.

184. Krafft, C., T. Knetschke, A. Siegner, R.H.W. Funk, R. Salzer (2003). “Mapping of single cells by near infrared Raman microspectroscopy.” Vibrational Spectroscopy **32(**1):75-83.

185. Krafft, C., T. Knetschke, A. Siegner, R.H.W. Funk, R. Salzer (2005). “Identification of organelles and vesicles in single cells by Raman microspectroscopic mapping.” Vibrational Spectroscopy **38**(1-2): 85-93.

186. Crow, P., B. Barrass, C. Kendall, M. Hart-Prieto, M. Wright, R. Persad and N. Stone (2005). "The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines." Br J Cancer **92**(12): 2166-2170.

187. Matthaus, C., S. Boydston-White, M. Miljkovic, M. Romeo and M. Diem (2006). "Raman and infrared microspectral imaging of mitotic cells." Appl Spectrosc **60**(1): 1-8.

188. Matthaus, C., T. Chernenko, J. A. Newmark, C. M. Warner and M. Diem (2007). "Label-free detection of mitochondrial distribution in cells by nonresonant Raman microspectroscopy." Biophys J **93**(2): 668-673.

189. Yu, C., E. Gestl, K. Eckert, D. Allara and J. Irudayaraj (2006). "Characterization of human breast epithelial cells by confocal Raman microspectroscopy." Cancer Detect Prev **30**(6): 515-522.

190. Claudia Scalfi-Happ, Andrea Jauss, Wolfram Ibach, Olaf Hollricher, Simone Fulda, Carmen Hauser, Rudolf Steiner, Angelika Rück (2007). "Confocal Raman microscopy as a diagnostic tool for investigation of living neuroblastoma tumour cells." Medical Laser Application, **22**(3),157-164

191. Schulze, H. G., S. O. Konorov, N. J. Caron, J. M. Piret, M. W. Blades and R. F. Turner (2010). "Assessing differentiation status of human embryonic stem cells noninvasively using Raman microspectroscopy." Anal Chem **82**(12): 5020-5027.

192. Zoladek, A., F. C. Pascut, P. Patel and I. Notingher (2011). "Non-invasive time-course imaging of apoptotic cells by confocal Raman micro-spectroscopy." Journal of Raman Spectroscopy **42**(3): 251-258.

193. Nijssen, A., T. C. Bakker Schut, F. Heule, P. J. Caspers, D. P. Hayes, M. H. Neumann and G. J. Puppels (2002). "Discriminating basal cell carcinoma from its surrounding tissue by Raman spectroscopy." J Invest Dermatol **119**(1): 64-69.

194. Koljenovic, S., L. P. Choo-Smith, T. C. Bakker Schut, J. M. Kros, H. J. van den Berge and G. J. Puppels (2002). "Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy." Lab Invest **82**(10): 1265-1277.

195. Koljenovic, S., T. B. Schut, A. Vincent, J. M. Kros and G. J. Puppels (2005). "Detection of meningioma in dura mater by Raman spectroscopy." Anal Chem **77**(24): 7958-7965.

196. Koljenovic, S., T. C. Bakker Schut, J. P. van Meerbeeck, A. P. Maat, S. A. Burgers, P. E. Zondervan, J. M. Kros and G. J. Puppels (2004). "Raman microspectroscopic mapping studies of human bronchial tissue." J Biomed Opt **9**(6): 1187-1197.

197. Krafft, C., S. B. Sobottka, K. D. Geiger, G. Schackert and R. Salzer (2007). "Classification of malignant gliomas by infrared spectroscopic imaging and linear discriminant analysis." Anal Bioanal Chem **387**(5): 1669-1677.

References

198. Krafft, C., D. Codrich, G. Pelizzo and V. Sergo (2008). "Raman mapping and FTIR imaging of lung tissue: congenital cystic adenomatoid malformation." Analyst **133**(3): 361-371.

199. Beljebbar, A., O. Bouche, M. D. Diebold, P. J. Guillou, J. P. Palot, D. Eudes and M. Manfait (2009). "Identification of Raman spectroscopic markers for the characterization of normal and adenocarcinomatous colonic tissues." Crit Rev Oncol Hematol **72**(3): 255-264.

200. Tan, K. M., C. S. Herrington and C. T. A. Brown (2011). "Discrimination of normal from pre-malignant cervical tissue by Raman mapping of de-paraffinized histological tissue sections." Journal of Biophotonics **4**(1-2): 40-48.

201. Shetty, G., C. Kendall, N. Shepherd, N. Stone and H. Barr (2006). "Raman spectroscopy: elucidation of biochemical changes in carcinogenesis of oesophagus." Br J Cancer **94**(10): 1460-1464.

202. Kamemoto, L. E., A. K. Misra, S. K. Sharma, M. T. Goodman, H. Luk, A. C. Dykes and T. Acosta (2010). "Near-infrared micro-Raman spectroscopy for in vitro detection of cervical cancer." Appl Spectrosc **64**(3): 255-261.

203. Christoph Krafft, Mehrnaz Alipour Diderhoshan, Peter Recknagel, Milos Miljkovic, Michael Bauer and Jürgen Popp (2011). “Crisp and soft multivariate methods visualize individual cell nuclei in Raman images of liver tissue sections.”Vibrational Spectroscopy, 55(1):90-100.

204. Halina Abramczyk, Beata Brozek-Pluska, Jakub Surmacki, Joanna Jablonska-Gajewicz, Radzisław Kordek (2012). “Raman ‘optical biopsy’ of human breast cancer.” Progress in Biophysics and Molecular Biology, 108(1-2)74-81

205. Cals, F. L. J., T. C. Bakker Schut, S. Koljenović, G. J. Puppels and R. J. B. de Jong (2013). "Method development: Raman spectroscopy-based histopathology of oral mucosa." Journal of Raman Spectroscopy **44**(7): 963-972.

206. Bakker Schut, T. C., M. J. Witjes, H. J. Sterenborg, O. C. Speelman, J. L. Roodenburg, E. T. Marple, H. A. Bruining and G. J. Puppels (2000). "In vivo detection of dysplastic tissue by Raman spectroscopy." Anal Chem **72**(24): 6010-6018.

207. Venkatakrishna, K., Jacob Kurien, Keerthilata M. Pai, Manna Valiathan, Nirmala Nagesh Kumar, C. Murali Krishna, G. Ullas and V.B. Kartha (2001). “Optical pathology of oral tissue: A Raman Spectroscopy diagnostic method.” Current Science **80**:101-105.

208. Krishna, C. M., G. D. Sockalingum, J. Kurien, L. Rao, L. Venteo, M. Pluot, M. Manfait and V. B. Kartha (2004). "Micro-Raman spectroscopy for optical pathology of oral squamous cell carcinoma." Appl Spectrosc **58**(9): 1128-1135.

209. Malini, R., K. Venkatakrishna, J. Kurien, K. M. Pai, L. Rao, V. B. Kartha and C. M. Krishna (2006). "Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: a Raman spectroscopy study." Biopolymers **81**(3): 179-193.

210. Oliveira, A. P., R. A. Bitar, L. Silveira, R. A. Zangaro and A. A. Martin (2006). "Near-infrared Raman spectroscopy for oral carcinoma diagnosis." Photomed Laser Surg **24**(3): 348-353.

211. Yaogai Hu, Tao Jiang and Zhengyu Zhao (2008)."Discrimination of Squamous Cell Carcinoma of the Oral Cavity Using Raman Spectroscopy and Chemometric Analysis." Intelligent Networks and Intelligent Systems, First International Conference on Intelligent Networks and Intelligent Systems 633-636.

212. Sunder, N. S., Rao Nirmala N, Kartha V B, Ullas G and Kurien Jacob (2011). "Laser Rramanspectroscopy: A novel diagnostic tool for oral cancer" J Orofac Sci. **3(**2)15-19.

213. Guze, K., M. Short, S. Sonis, N. Karimbux, J. Chan and H. Zeng (2009). "Parameters defining the potential applicability of Raman spectroscopy as a diagnostic tool for oral disease." J Biomed Opt **14**(1): 014016.

214. Bergholt, M. S., W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. So and Z. Huang (2011). "Characterizing variability in *in vivo* Raman spectra of different anatomical locations in the upper gastrointestinal tract toward cancer detection." J Biomed Opt **16**(3): 037003.

215. Ghanate, A. D., S. Kothiwale, S. P. Singh, D. Bertrand and C. M. Krishna (2011). "Comparative evaluation of spectroscopic models using different multivariate statistical tools in a multicancer scenario." J Biomed Opt **16**(2): 025003.

216. Matousek, P. and N. Stone (2009). "Emerging concepts in deep Raman spectroscopy of biological tissue." Analyst **134**(6): 1058-1066.

217. A. Nancy (2006). "Tencate's Oral Histology: Development, Structure and Function." Chap. 12, Mosby, Inc., an affiliate of Elsevier Inc., St. Louis, MO.

218. Krishna, C. M., J. Kurien, S. Mathew, L. Rao, K. Maheedhar, K. K. Kumar and M. V. Chowdary (2008). "Raman spectroscopy of breast tissues." Expert Rev Mol Diagn **8**(2): 149-166.

219. Mo, J., W. Zheng and Z. Huang (2010). "Fiber-optic Raman probe couples ball lens for depth-selected Raman measurements of epithelial tissue." Biomed Opt Express **1**(1): 17-30.

220. Yamamoto, Y. S., Y. Oshima, H. Shinzawa, T. Katagiri, Y. Matsuura, Y. Ozaki and H. Sato (2008). "Subsurface sensing of biomedical tissues using a miniaturized Raman probe: study of thin-layered model samples." Anal Chim Acta **619**(1): 8-13.

221. F. S. Parker (1983). "Applications of Infrared, Raman and Resonance Raman Spectroscopy in Biochemistry", Plenum Press, New York and London.

222. Movasaghi, Z., S. Rehman and I. U. Rehman (2007). "Raman Spectroscopy of Biological Tissues." Applied Spectroscopy Reviews **42**(5): 493-541.

223. Shin, D. M., N. Voravud, J. Y. Ro, J. S. Lee, W. K. Hong and W. N. Hittelman (1993). "Sequential increases in proliferating cell nuclear antigen expression in head and neck tumorigenesis: a potential biomarker." J Natl Cancer Inst **85**(12): 971-978.

224. Pindborg, J. J. and P. N. Wahi (1997). Histological typing of cancer and precancer of the oral mucosa, Springer.

225. Silverman, S., Jr. and M. Gorsky (1990). "Epidemiologic and demographic update in oral cancer: California and national data-1973 to 1985." J Am Dent Assoc **120**(5): 495-499.

226. Zarbo, R. J. and J. D. Crissman (1988). "The surgical pathology of head and neck cancer." Semin Oncol **15**(1): 10-19.

227. Bousquet, O. and P. A. Coulombe (1996). "Cytoskeleton: missing links found?" Curr Biol **6**(12): 1563-1566.

228. Coulombe, P. A. and M. B. Omary (2002). "'Hard' and 'soft' principles defining the structure, function and regulation of keratin intermediate filaments." Curr Opin Cell Biol **14**(1): 110-122.

229. Moll, R., W. W. Franke, D. L. Schiller, B. Geiger and R. Krepler (1982). "The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells." Cell **31**(1): 11-24.

230. Schweizer, J., P. E. Bowden, P. A. Coulombe, L. Langbein, E. B. Lane, T. M. Magin, L. Maltais, M. B. Omary, D. A. Parry, M. A. Rogers and M. W. Wright (2006). "New consensus nomenclature for mammalian keratins." J Cell Biol **174**(2): 169-174.

231. Hesse, M., A. Zimek, K. Weber and T. M. Magin (2004). "Comprehensive analysis of keratin gene clusters in humans and rodents." Eur J Cell Biol **83**(1): 19-26.

232. Herrmann, H. and U. Aebi (2004). "Intermediate filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular Scaffolds." Annu Rev Biochem **73**: 749-789.

233. Nelson, W. G. and T. T. Sun (1983). "The 50- and 58-kdalton keratin classes as molecular markers for stratified squamous epithelia: cell culture studies." J Cell Biol **97**(1): 244-251.

234. Omary, M. B., N. O. Ku, P. Strnad and S. Hanada (2009). "Toward unraveling the complexity of simple epithelial keratins in human disease." J Clin Invest **119**(7): 1794-1805.

235. Fuchs, E. and K. Weber (1994). "Intermediate filaments: structure, dynamics, function, and disease." Annu Rev Biochem **63**: 345-382.

References

236. Magin, T. M., P. Vijayaraj and R. E. Leube (2007). "Structural and regulatory functions of keratins." Experimental Cell Research **313**(10): 2021-2032.

237. Paramio, J. M. and J. L. Jorcano (2002). "Beyond structure: do intermediate filaments modulate cell signalling?" BioEssays **24**(9): 836-844.

238. Pallari, H. M. and J. E. Eriksson (2006). "Intermediate filaments as signaling platforms." Sci STKE **2006**(366): pe53.

239. Kim, S., P. Wong and P. A. Coulombe (2006). "A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth." Nature **441**(7091): 362-365.

240. Oshima, R. G. (2002). "Apoptosis and keratin intermediate filaments." Cell Death Differ **9**(5): 486-492.

241. Fillies, T., R. Werkmeister, J. Packeisen, B. Brandt, P. Morin, D. Weingart, U. Joos and H. Buerger (2006). "Cytokeratin 8/18 expression indicates a poor prognosis in squamous cell carcinomas of the oral cavity." BMC Cancer **6**: 10.

242. Vaidya, M. M., A. M. Borges, S. A. Pradhan and A. N. Bhisey (1996). "Cytokeratin expression in squamous cell carcinomas of the tongue and alveolar mucosa." Eur J Cancer B Oral Oncol **32B**(5): 333-336.

243. Hendrix M.J., E.A. Seftor, Y.W. Chu, K.T. Trevor and R.E. Seftor (1996). "Role of intermediate filaments in migration, invasion and metastasis." Cancer metastasis reviews.**15**(4):507-525.

244. Alam H, T. K. Samrat, S. N. Dalal and M. M. Vaidya (2011). "Loss of keratins 8 and 18 leads to alterations in 6 4-integrin-mediated signalling and decreased neoplastic progression in an oral-tumour-derived cell line." Journal of Cell Science, **124**, 2096-2111.

245. Alam H, P. Gangadaran, A.V. Bhate, D.A. Chaukar, S.S. Sawant, R. Tiwari, J. Bobade, S.Kannan, A. K. D'cruz, S. Kane and M. M. Vaidya (2011). "Loss of Keratin 8 Phosphorylation Leads to Increased Tumor Progression and Correlates with Clinico-Pathological_Parameters of OSCC Patients" PloS one,**6**(11),e27767.

246. Tatake R. J., N. Rajaram, R. N. Damle, B. Balsara, A. N. Bhisey and S. G. Gangal (1990) "Establishment and characterization of four new squamous cell carcinoma cell lines derived from oral tumors", Journal of Cancer Research and Clinical Oncology, **116**(2),179-186.

247. Lowry, O. H., N. J. Rosebrough, A. L. Farr and R. J. Randall (1951). "Protein measurement with the folin phenol reagent." Journal of Biological Chemistry **193**(1): 265-275.

248. Folch, J., M. Lees and G. H. S. Stanley (1957). "A Simple method for the isolation and purification of total lipids from animal tissues." Journal of Biological Chemistry **226**(1): 497-509.

249. Rouser, G., S. Fleischer and A. Yamamoto (1970). "Two dimensional thin layer chromatographic separation of polar lipids and determination of phospholipids by phosphorus analysis of spots." Lipids **5**(5): 494-496.

250. Samek, O., A. Jonáš, Z. Pilát, P. Zemánek, L. Nedbal, J. Tříska, P. Kotas and M. Trtílek (2010). "Raman Microspectroscopy of Individual Algal Cells: Sensing Unsaturation of Storage Lipids in vivo." Sensors **10**(9): 8635-8651.

251. Notingher, I. (2007). "Raman Spectroscopy Cell-based Biosensors." Sensors **7**(8): 1343-1358.

252. Campbell, C. (2013). "Cell Sensing with Raman Spectroscopy." Encyclopedia of Biophysics. G. K. Roberts, Springer Berlin Heidelberg**:** 245-247.

Appendix I

Publications and reprints

Appendix I

Publications and reprints

1 Atul Deshmukh*, **S.P.Singh***, Pankaj Chaturvedi, C. Murali Krishna, "Raman spectroscopy of normal oral buccal mucosa tissues: A study on intact and incised biopsies" J Biomed Opt16 (12), 2011 (*Equal contribution). (DOI: 10.1117/1.3659680.) [Atul performed maxillofacial surgical procedure on biopsies, S.P.Singh recorded Raman spectra and analyzed the data]

2 **S. P. Singh**, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna,"Raman spectroscopy in head and neck cancers: towards oncological applications" J Cancer Res Ther. S2, 8, 2012. (DOI: 10.4103/0973-1482.92227)

3 **S. P. Singh**, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna, "In vivo Raman spectroscopy methods for oral cancers diagnosis" Proc. of SPIEVol. 8219, 82190K, 2012. (DOI: 10.1117/12.905453)

4 **S. P. Singh**, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna, "In vivo Raman spectroscopic identification of premalignant lesions in oral buccal mucosa" J Biomed Opt, 17(10), 105002,2012. (DOI: 10.1117/1.JBO.17.10.105002)

5 **S. P. Singh**, Hunain Alam, Milind M. Vaidya, C. Murali Krishna,"Raman spectroscopic study of keratin 8 knockdown oral squamous cell carcinoma cells" Proc. of SPIEVol. 8225, 82251C, 2012. (DOI:10.1117/12.905450)

6 **S. P. Singh**, Aditi Sahu, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna, "In vivo Raman spectroscopy of oral buccal mucosa: A study on malignancy associated changes (MAC)/cancer field effects (CFE)" Analyst,138, 4175-4182, 2013 (DOI: 10.1039/C3AN36761D)