Raman Micro-Spectroscopy Studies of Oral Cancerous and Premalignant Conditions

By

Ms. Aditi Kamlesh Sahu

[LIFE09201004007]

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

October, 2016

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 Aditi Sahu entitled, '**Raman micro-spectroscopy studies of oral cancerous and premalignant conditions**' and recommend that it may be accepted as fulfilling the dissertation requirements for the Degree of Doctor of Philosophy.

S. Chiplunkar ----------Date---17/10/2016---

Chairperson: Dr. Shubhada V. Chiplunkar

----------Date---17/10/'16---

Guide/Convener: Dr. C. Murali Krishna

V. Prasanna ----------Date---17/10/16.---

Member 1: Dr. Prasanna Venkatraman

Dibyendu Bhattacharyya ----------Date---17/10/16---

Member 2: Dr. Dibyendu Bhattacharyya

----------Date---18/10/16---

Invitee 1: Dr. Sudhir Nair

Asawari ----------Date---18/10/16---

Invitee 2: Dr. Asawari Patil

A. Chattopadhyay ----------Date---17/10/16---

External Examiner: Dr. Amitabha Chattopadhyay

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis 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 thesis requirement.

Date:
Place: Navi Mumbai

Dr. C. Murali Krishna
(Guide)

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment 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 the 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.

ACTREC, Kharghar, Navi Mumbai.

October 2016.

Aditi Sahu.

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.

ACTREC, Kharghar, Navi Mumbai,

October 2016.

Aditi Sahu.

List of Publications arising from the Thesis

Journal:

1. In vivo sub-site classification and diagnosis of oral cancers using Raman spectroscopy, ***Aditi Sahu***, Arti R. Hole, Atul Deshmukh, Sudhir Nair, Pankaj Chaturvedi, C. Murali Krishna, JIOHS, 9 (5), 2016

2. Oral cancer screening: A serum Raman spectroscopic approach, ***Aditi Sahu*** , Suyash Dhoot, Amandeep Singh, Sharada S. Sawant, Nikhila Nandakumar, Sneha Talathi-Desai, Mandavi Garud, Sandeep Pagare, Sanjeeva Shrivastava, Sudhir Nair, Pankaj Chaturvedi, C. Murali Krishna, JBO, 20 (11), 2015

3. Raman spectroscopy and cytopathology of oral exfoliated cells for oral cancer diagnosis, ***Aditi Sahu,*** Sneha Tawde, Venkatesh Pai, Poonam Gera, Pankaj Chaturvedi, Sudhir Nair, C. Murali Krishna, Ana Met, 7, 2015.

4. "Raman spectroscopy of serum: a study on oral cancers", ***Aditi Sahu***, Sneha Talathi, Sharada Sawant and C. Murali Krishna, Biomed Spectro Imaging, 4(2), 2015.

5. "Recurrence prediction in oral cancers: a serum Raman spectroscopy study", ***Aditi Sahu***, Nikhila Nandakumar, Sharada Sawant and C. Murali Krishna, Analyst 140(7):2294-301, 2015.

6. "Raman spectroscopy and oral exfoliative cytology", ***Aditi Sahu***, Nupur Shah, Manoj Mahimkar, Mandavi Garud, Sandeep Pagare, Sudhir Nair, C. Murali Krishna, Proc SPIE 89262N-89262N-4, 2014.

7. "Classification of oral cancers using Raman spectroscopy of serum", ***Aditi Sahu***, Sneha Talathi, Sharada Sawant, C. Murali Krishna, Proc SPIE 89390E-89390E-7, 2014.

8. "Raman spectroscopy of serum: an exploratory study for detection of oral cancers", ***Aditi Sahu,*** Sharada Sawant, Hitesh Mamgain, C. Murali Krishna., Analyst, 2013.

9. "In vivo Raman spectroscopy of oral buccal mucosa: a study on malignancy associated changes (MAC)/cancer field effects (CFE)", S.P.Singh, ***Aditi Sahu***, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna., Analyst, 2013.

10. "Raman spectroscopy of oral buccal mucosa: a study on age-related physiological changes and tobacco-related pathological changes", ***Aditi Sahu***, Atul Deshmukh, A.D. Ghanate, S.P. Singh, Pankaj Chaturvedi, C. Murali Krishna, Technol Cancer Res Treat. 11, 529-541 (2012).

(Prints and re-prints of the revised, in press and published studies have been enclosed with the thesis).

Conferences attended:

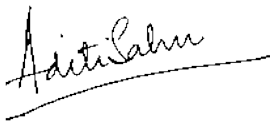
1. Presented a poster entitled "A Raman spectroscopy approach for classification of subsites in oral cancers: A feasibility study", ***Aditi Sahu,*** S.P. Singh, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna (also selected as a proferred talk) at AACR 2011, New Delhi
2. Presented a poster entitled" "Serum based diagnosis of asthma: a Raman spectroscopic approach", ***Aditi Sahu,*** Krishna Dalal, Sarla Naglot, and C. Murali Krishna, at ICORS 2012, Bangalore
3. Presented a poster entitled "To explore utility of Raman spectroscopy in therapeutic drug monitoring of Imatinib", Vikram Gota, ***Aditi Sahu,*** Anand Patil, and C. Murali Krishna at ICORS 2012, Bangalore
4. Poster presentation entitled, "Raman spectroscopy and oral exfoliative cytology: Investigating misclassifications between contralateral normal and tumor sites", ***Aditi Sahu,*** S. Tawde, Poonam Gera, S. Nair, C. Murali Krishna in TMC Platinum Jubilee Conference, Mumbai, 2016
5. Poster presentation entitled, "Serum Raman spectroscopic classification of buccal mucosa and tongue cancers", ***Aditi Sahu,*** Sneha Talathi-Desai, Sharada Sawant, C. Murali Krishna at OWLS, TIFR, Mumbai, 2016.

Other publications:

1. Unique spectral markers discern recurrent Glioblastoma cells from heterogeneous parent population, Ekjot Kaur, ***Aditi Sahu,*** Arti Adhav, Jacinth Rajendra, Rohan Chaubal, Nilesh Gardi, Amit Dutt, Aliasgar Moiyadi, Murali Krishna Chilakapati, Shilpee Dutt, Scientific Reports, 6, 2016

2. "Raman spectroscopy for detection of Imatinib in Plasma: A Proof of Concept," Sanhita Rath*, ***Aditi Sahu*,*** Vikram Gota, P G Martínez-Torres, J L Pichardo-Molina, C. Murali Krishna, JIOHS, 08, 2015.

3. "Serum based diagnosis of asthma using Raman spectroscopy: an early phase pilot study", ***Aditi Sahu,*** Krishna Dalal, Sarla Naglot, Parveen Aggarwal, Murali Krishna C. PLoS One 8(11): e78921, 2013.



Aditi Sahu

Dedicated to

My parents (Mrs. Shobha Sahu and Mr. Kamlesh Kumar Sahu)

Elder sister (Ms. Suchita Sahu)

And Candy…

ACKNOWLEDGEMENTS

Firstly and foremost, I would like to take this opportunity to thank my guide and mentor Dr. C. Murali Krishna. For providing the required technical facilities, challenging and encouraging me to work to the best of my abilities, I express my gratitude to him. He not only provided me diverse opportunities in the form of collaborative projects apart from my main project, but also gave the freedom to execute those. He would not just personally enquire about our well-being during times of sickness, he has also made sure that we receive proper medical treatment by arranging visits to the doctor. In spite of my health problems, he allowed me to continue and also work at my convenience. Thank you, Sir, for everything.

I am thankful to Dr. Shubhada Chiplunkar (Director, ACTREC), Dr. Sudeep Gupta (Deputy Director, ACTREC) for providing me this opportunity to work in this institution and for the use of the excellent infrastructure the institute boasts of. I am thankful to DAE and ACTREC for my fellowship. I am fortunate to have Dr. Shubhada Chiplunkar (Chairperson), Dr. Prasanna Venkatraman and Dr. Dibyendu Bhattacharyya, Dr. Sudhir Nair and Dr. Asawari Patil as doctoral committee members - their expert comments, critical analysis of the results and helpful suggestions have contributed significantly in my understanding of key biological concepts and techniques.

A sincere thanks to all my clinical collaborators - Dr. Sudhir Nair, Dr. Pankaj Chaturvedi, Dr. Devendra Chaukar, Dr. Vikram Gota, Dr. Akshat Malik, Dr. Sandeep Pagare, Dr. Mandavi Garud for providing samples and relevant clinical suggestions as and when needed. I also thank Dr. Asawari Patil and Dr. Poonam Gera for their painstaking help in evaluation of Pap smears. Dr. Poonam also for her immense patience, professional help and friendly discussions that would drive home something more than plain knowledge.

A special thanks to all Chilakapati lab members- Arti Ma'am, Surya, Rubina, Isha, Dr. Atul, Tanmoy and Piyush for their inputs and suggestions during the course of my Ph.D. Thanks to my trainees - Sneha, Nikhila, and Sharika, Amandeep, Abhishek, Venkatesh and Suyash for help during my PhD years. Thanks to Arti Ma'am, for being a constant pillar of support. A special mention about Surya, my senior, for teaching the technicalities and basic concepts related to

Raman spectroscopy and clinical experiments. I owe most of my knowledge on the subject to his friendly lessons; Tanmoy and Piyush for the innumerous healthy critical discussions over tea; and Piyush for being the best-lab mate and friend possible - from help in proof readings at the eleventh hour to valuable help during times of sickness, he did them all. Thanks to Nisha, Nitish and Sandeep, and the newly-joined members Bhagyashree and Siddharth -for providing much needed help in printing, proof-reading. The two new kids helped immensely with the viva arrangements, for which I am thankful. A humble thanks to all the study participants- healthy volunteers and patients for their valuable samples. Thanks to the nursing staff and the blood bank facility for lending help in sample collection.

I would also take this opportunity to thank the staff in ACTREC for taking good care of the facilities. Special thanks to flow cytometry, common instrumentation and imaging facility.

I would also like to mention the contribution of my lovely batchmates in keeping the frustrating 6 years of PhD as joyous as possible. Thanks to Poonam, Richa, Crismita - my ex-roommates, Nikhat, Rajan, Indrajit, Raja, Abira, Piyush and Rushikesh for their personal and professional help. A couple of seniors and juniors who need special mention are Moquit, Shalini, Prathamesh, Ekjot, Subhoshree, and Rupa.

Lastly, very heartfelt thanks to my friends- Manjiri, Anagha, Sagar, Vivian and Swapnil, Kundan Vinay and Akshat; my extended family of Aunty, Uncle, Chandan, Kavita, Vinay bhai and my family for innumerable things. To my parents, I would like to offer my deepest gratitude for their unflinching love and support. It is because of their efforts I could reach this milestone in my life. Thanks also for teaching me to believe in God, this kept me going even in times of failures. And to my Didi, who believed in me even when I did not believe in myself. Her words often gave me the much needed solace and drove me to work harder.

It is due to the professional and emotional support from all of these people that I could complete my stated objectives and compile this thesis.

Aditi Sahu.

LIST OF CONTENTS

PREFACE

Oral cancers are the sixteenth most common cancer in the world and the most common cancer in Indian males. Oral cancers are associated with low disease free-survival rates which can be attributed to delays in diagnosis and recurrences. Visual inspection followed by biopsy and histopathology of suspicious are the gold standard for screening and diagnosis of oral cancers. While visual inspection may be useful for diagnosis in high risk-populations biopsy is an invasive procedure. As early diagnosis may lead to better patient survival and improve quality of life, several diagnostic adjuncts for early diagnosis are being investigated. Raman spectroscopy, a vibrational spectroscopic approach based on inelastic scattering of light, has shown promise in cancer diagnosis. In the present study, the potential of ex vivo and in vivo Raman spectroscopy in oral cancer diagnosis and screening were explored.

Chapter 1 provides a general introduction to the work presented in the thesis. In this chapter, an introduction to cancer, oral cancers, their epidemiology, causative factors, pre-cancerous conditions and current diagnostic techniques have been described. In the next section, literature review on specific applications of optical spectroscopy in oral cancer diagnosis has been provided. An introduction into Raman spectroscopy, Raman Effect, theory of Raman scattering, instrumentation, literature review on biomedical applications, especially in oral cancers has been included in the subsequent section. In the final section, the aim and objectives of the study have been stated.

In the next chapter, use of minimally invasive samples like serum and exfoliated cells for oral cancer diagnosis has been discussed. These less invasive samples are associated with several advantages like accessibility, low cost, multiple sampling, and transportation from primary screening centers to centralized facility for analysis. For the sake of clarity, they have been discussed as two different chapters (Chapter 2 on serum and 3 on exfoliated cells).In the second chapter, feasibility of oral cancer diagnosis using serum has been explored. Results from two pilot studies on resonance and conventional Raman have been presented. After promising findings in the pilot studies, the scope of the study was broadened by including premalignant and disease control samples. The applicability of serum Raman spectroscopy in screening of oral

cancers was subsequently investigated. In the final section, recurrence prediction in oral cancer patients using serum was explored in before and after surgery serum samples.

The study on Raman oral exfoliative cytology has been described in Chapter 3. Raman spectroscopy of exfoliated cells from suspicious lesions may help in preliminary evaluation of these lesions and prevent unnecessary biopsies. In this chapter, the feasibility of differentiating cells from healthy and tumor sites was first explored in a proof of concept study. In the succeeding section, the potential of classifying exfoliated cells from healthy and pre-cancer sites was explored.

In Chapter 4, the feasibility of classifying serum from patients with cancers at two different subsites- buccal mucosa and tongue was explored using both the resonance and conventional Raman spectroscopic approaches.

In Chapter 5, in vivo studies for sub-site classification and oral cancer diagnosis at different subsites have been discussed. Previous Raman spectroscopic studies have demonstrated potential of Raman spectroscopy in classifying normal, premalignant and tumor conditions and also in identifying cancer field effects at buccal mucosa. Other studies have indicated subsite classification between various anatomical sites in oral cavity. In this study, the subsite differences between buccal mucosa, lip and tongue were investigated in healthy and pathological (contralateral, premalignant and tumor) conditions. Although the study aimed to understand spectroscopic differences between seven different subsites- buccal mucosa, lip, tongue, floor of mouth, retromolar trigone, hard palate, lip and gingival, premalignant and tumor accruals could not be made for all mentioned subsites. Therefore, the study was limited to only buccal mucosa, lip, tongue- the 3 most commonly affected subsites in the Indian subcontinent. After investigating subsite differences, effect of these differences on healthy versus pathological classification was explored at pooled and individual subsites. In the concluding section, detection of physiological aging at these 3 subsites was explored and the effect of physiological aging on healthy versus pathological classification was also assessed.

In the final chapter, the conclusions drawn from this thesis and future directions have been presented.

Synopsis

Homi Bhabha National Institute

SYNOPSIS OF Ph. D. THESIS

1. Name of the Student: Aditi Kamlesh Sahu

2. Name of the Constituent Institution: ACTREC-TMC

3. Enrolment No.: LIFE09201004007

4. Title of the Thesis: "Raman micro-spectroscopy studies of oral cancerous and premalignant conditions"

5. Board of Studies: Life Sciences

Introduction

Oral cancer is a major health problem worldwide, with an annual incidence estimate of approximately 275,000 cases[1]. Oral cancers form a significant health burden in developing countries like India where they account for over 30% of all cancers with 80, 000 new cases reported each year[2]. Tobacco (both smoking and smokeless) and alcohol are major etiological factors[3]. Despite advancement in surgical and treatment modalities, disease-free survival rates have remained low for several decades, attributed to late diagnosis. Studies have shown that early detection leads to better prognosis and increased 5-year survival rate to 90%[4,5]. Unfortunately, there is a paucity of tools suitable for early diagnosis and mass screening. Visual

examination, followed by biopsy and histopathology is the standard screening procedure. While visual examination has been associated with low sensitivity, biopsy is limited by ethical and practical considerations, lower patient compliance, sampling errors and inconsistency in interpretation especially in early/premalignant stages[6,7]. Several adjunct techniques like tissue staining (vital iodine and toluidine blue), brush biopsy followed by cytomorphometry, light-based screening aids have been explored[8,9]. Biomarker identification using salivary and serum biomarkers like DNA, RNA and protein have also been explored[10-13]. None of these approaches have been clinically established till date. Recent studies suggest utility of holistic approaches in disease diagnosis and prognosis prediction[14]. The holistic information provided by optical spectroscopic techniques has generated interest in them as potential cancer screening tools.[15]. Raman spectroscopy (RS), an optical spectroscopic technique based on inelastic scattering, yields molecular finger-print of samples. RS has been extensively explored for diagnosis of several cancers including oral cancers[16-18]. Ex vivo RS studies have shown potential in discriminating healthy, premalignant, inflammatory and malignant tissues[19]. Subsequent in vivo studies have indicated possibility of acquiring in vivo spectra in clinically implementable time, distinguished normal, pre-malignant and oral cancer at buccal mucosa subsite and also demonstrated anatomic differences in sub-sites of oral cavity[20-23]. In vivo Raman spectroscopic detection of normal, pre-cancer and cancer at all oral sub-sites can help develop a non-invasive approach for mass-screening and diagnosis of oral cancer. Less invasive samples like serum and exfoliated cells may be more practical for screening and diagnostic applications. Previous studies have shown feasibility of detecting breast, head and neck cancer, cervical, colorectal, nasopharyngeal, ovarian cancers using serum[24-29]. Exfoliated cells have also been used as

samples for identifying different conditions in oral cavity using IR micro-spectroscopy[30]. However, little literature is available on Raman spectroscopy of oral cancer and pre-cancers using less invasive samples. The study therefore focuses on the detection of oral cancer and pre-cancer conditions using less invasive samples like serum and exfoliated cells in addition to the in vivo approach.

Aim and objectives: Development of Raman micro-spectroscopy as a method for early diagnosis and sub-site classification of oral cancers using minimally invasive samples like serum and exfoliated cells and non-invasive (*in vivo*) approach.

Following are the specific objectives of the thesis:

Objectives:

1) To explore feasibility of classifying normal, oral cancer and precancerous conditions using less invasive samples like serum and exfoliated cells.

2) To correlate Raman spectra of serum with histopathology of sub-sites in oral cancers.

3) To explore potential of Raman spectroscopy in diagnosis and sub-site classification of oral cancers *in vivo.*

Objective 1: To explore feasibility of classifying normal, oral cancer and precancerous conditions using less invasive samples like serum and exfoliated cells.

Serum studies

Ex vivo samples like serum are ideal samples for diagnostic applications owing to accessibility, possibility of multiple sampling, storage, transport and analysis at a centralized facility. Although serum biomarkers have been used for diagnosis of several disorders, including cancers,

no established biomarker is available for oral cancer diagnosis. Thus, oral cancer diagnosis using RS of serum was explored.

Exploring resonance Raman for serum-based diagnosis of oral cancer

With the onset of malignancy, concentrations of anti-oxidants like β-carotene are known to decrease. The qualitative and quantitative differences in β-carotene along with other biomolecules like DNA, protein and amino acids were investigated in normal (n=16) and oral cancer (n=54) sera using Raman microscope (WITec alpha300RS, GmbH, Ulm, Germany). Spectral acquisition parameters were: λ_{ex}= 532 nm, objective- 50X (Nikon), laser power-20 mW, acquisition time: 5 s and averaged over 4 accumulations. The acquired Raman spectra were corrected for CCD response, spectral contaminations from the substrate, first-derivatized, normalized and subjected to multivariate analysis PCA –principal component analysis and PC-LDA-principal component linear discriminant analysis using both spectra- and patient-wise approaches. More robust and practical, patient wise-approach was selected for future data analysis. Mean spectral comparisons indicate differential contributions of proteins, DNA, and amino acids and β-carotene in the analyzed groups. Apart from other biomolecules, intense β-carotene features played an important role in distinguishing healthy and oral cancer. PC-LDA yielded ~78% efficiency of classification. ***(Analyst, 2013).***

Exploring serum-based diagnosis of oral cancer, pre-cancer and disease controls using red laser

After confirming the feasibility of classifying normal and oral cancer sera using 532 nm excitation, classification between oral cancer and normal sera using 785 nm excitation- the most

commonly employed excitation for biological was explored on 246 subjects (n=126 healthy controls, n= 120 oral cancers). Spectra were acquired using Raman microprobe (Horiba Jobin-Yvon, France) which consisted of a laser (785 nm, Process Instruments) and HE-785 spectrograph (Horiba-Jobin-Yvon, France) coupled with CCD (Synapse, Horiba-Jobin-Yvon). Spectral acquisition parameters were- objective: 40X (Nikon), laser power: 30 mW, integration time: 15 s and average: 4. Preprocessed spectra were subjected to PCA and PC-LDA. Mean spectral features indicate differences in amino acids, β-carotene, amide III and DNA between the groups. PC-LDA demonstrated ~87% efficiency of classification between the normal and cancer groups. Model building and test predictions yielded prediction efficiency of ~86% for both groups. Thus, potential of serum RS in oral cancer diagnosis was observed. Following the encouraging findings in cancer diagnosis, feasibility of oral pre-cancer detection was explored by analyzing normal and premalignant cohorts (n=47). PC-LDA results indicate normal and pre-cancer could be distinguished with ~77% efficiency. The lowered efficiency may be attributed to the less severe biochemical changes in the circulation of pre-cancer subjects. Cancer and pre-cancer may have been detected as any general abnormal condition using serum. To confirm cancer-specific detection, disease control group (n=46, Malaria and Glioma) was also included. Healthy and disease controls were correctly predicted with an average efficiency of about ~86%. Thus serum RS could differentiate between oral cancer, pre-cancer and disease controls in the binary classification systems.

To enable screening-related applications of serum RS, a 2-step classification system was developed. In the first step, normal were tested against all abnormals in the ‘normal versus

abnormal (disease control, premalignant and cancer)' classification system. If a sample is classified as abnormal in the first step, the type of abnormality- disease control, premalignant and cancer would be characterized in the second step. In the first normal versus abnormal classification, 83% normal and 69% abnormal were correctly predicted. The sensitivity to detect each abnormal condition was 85%, 83% and 79% for disease, premalignant and cancer, respectively. Although high specificity (83%) and lower sensitivity (69%) rates were observed, these rates are comparable to the existing screening procedures like mammography, Pap smear and faecal-occult-blood test (FOBT)[31-33]. Thus, potential utility of serum RS in detection of oral pre-cancer and cancers and prospectively for screening applications was confirmed in the present study. ***(SPIE Proc, 2014; Biomed Spectra Imaging, 2015; Analyst, communicated)***

Serum-based RS for recurrence detection in oral cancers

As potential of serum RS in diagnosis of oral cancers was demonstrated, the feasibility of detecting recurrence - a major cause of low disease free survival rates in oral cancers was also explored. Feasibility of differentiating recurrence (n=10) and non-recurrence (n=12) patient serum samples collected before surgery and one week after surgery was explored in this retrospective study. Mean spectra indicate DNA and protein may be potential spectral markers for recurrence. Although no classification could be observed in before surgery samples, after surgery samples could be differentiated with ~78% efficiency. Prospectively, RS of post-surgery serum samples may have the potential to predict recurrence in clinics. ***(Analyst, 2015)***

<u>Exfoliated cell studies</u>

Raman exfoliative cytology has shown potential in cervical cancer applications[34]. Raman oral exfoliative cytology followed by cytological analysis (Pap staining) for diagnosis of oral pre-cancer and cancer was therefore, explored. Standardization of several parameters like sample collection, device for sample collection, cell count, spectra acquisition from cell pellet, fixation of sample after spectra recording and protocol for Pap staining was first undertaken.

Proof of concept: Raman oral exfoliative cytology for oral cancer diagnosis

After necessary standardizations, the feasibility of Raman oral exfoliative cytology along with cytopathology for oral cancer diagnosis was evaluated on 70 specimens- exfoliated cells from 15 healthy volunteers (HV), 15 healthy tobacco users (HT), 20 contralateral or disease control (DC) and 20 tumor (T) sites of oral-cancer patients. After sample processing and RBC lysis protocol (to remove blood contamination observed due to high vascularity of tumors), Raman spectra were acquired from the obtained cell pellet subsequent to which pellet was subjected to Pap staining. Spectral findings demonstrate that with increase in severity of pathology from HV to T, higher DNA and changes in secondary structure of proteins were noticed. PCA and PC-LDA indicate that healthy (HV and HT) are mostly distinct from cancer groups DC and T. Some misclassifications in HT and DC were observed. These findings also correlate very well with cytopathological findings ***(Anal Methods, under review).***

Raman oral exfoliative cytology for oral pre-cancer diagnosis

After successful discrimination of healthy and cancer groups, the feasibility of oral pre-cancer diagnosis was explored. In this study, exfoliated cells from HV (n=35), HT (n=45) and premalignant subjects (n=51) were investigated. Spectral features indicate intensity-related differences and shifts in protein and DNA between the groups. Preliminary PC-LDA findings

indicate tendency of classification of HV cells, and major overlap between HT and PML cells. Presence of cytological similarity- orangeophilic, parakeratotic cells and anucleate squames may be responsible for observed misclassification. Use of better sample collection procedures to permit adequate representation even from basal and parabasal layers- the primary site for early malignancy changes is warranted for future studies. ***(SPIE Proc., 2014, J Biomed Opt, communicated)***

Objective 2: To correlate Raman spectra of serum with histopathology of sub-sites in oral cancers.

Clinico-epidemiological studies suggest that the biological characteristics of buccal mucosa and tongue cancers differ: although similar histopathologically (squamous cell carcinoma), buccal mucosa and tongue cancers demonstrate varying prognosis, aggressiveness, metastasis to lymph nodes and overall survival. Different biological markers such as the size of tumor, oncogene mutation expression, and apoptotic signals have been detected for buccal and tongue tumors. Raman spectral differences were first investigated in the resonance Raman study. A tendency of classification between serum from buccal mucosa (n=14) and tongue cancer (n=40) was observed. Prominent spectral differences in protein, DNA and β-carotene levels were observed between the two groups. PCA showed two slightly overlapping clusters while PC-LDA yielded ~69% classification efficiency between the groups ***(Analyst, 2013)***. The preliminary findings of the resonance Raman study were confirmed on a larger cohort of n= 120 cancer patients employing 785 nm excitation. Spectra were acquired from 62 buccal mucosa and 58 tongue cancer subjects using Raman microprobe. Spectral acquisition details were- λ_{ex}: 785 nm

objective: 40X, laser power: 30 mW, integration time: 15 s and average: 4. Minor spectral variations in the peaks corresponding to proteins: Phe, amide III, amide I, CH2 bending, and DNA- DNA backbone and bases were observed in this study. PC-LDA findings demonstrate an average classification efficiency of ~71% between the 2 groups. Thus, feasibility of distinguishing buccal mucosa and tongue cancers was confirmed in this study ***(Proc. SPIE 2014, Biomed Spectro Imaging, 2015, SPIE 2016, communicated).***

Objective 3: To explore potential of Raman spectroscopy in diagnosis and sub-site classification of oral cancers *in vivo.*

In vivo RS for early cancer diagnosis is associated with numerous merits: non-invasiveness, rapidity and objectivity. The utility of in vivo RS in identifying inherent anatomical differences at subsites and the influence of these anatomical differences on healthy and pathological classification were explored in this study. Spectra were acquired from 72 healthy subjects (HV) and 85 oral cancer patients (CNT, PML, TUM) using fiber-optic Raman probe (Fiber probe-Inphotonics Inc., USA) from healthy, contralateral normal, premalignant and tumor on 7 different subsites, buccal mucosa, lip, hard palate, floor of mouth (FOM), retro-molar trigone (RMT), gingiva and tongue in the oral cavity. The pre-processed spectra were subjected to multivariate analysis PCA and PC-LDA.

Exploring anatomical differences at subsites in healthy subjects

Spectral differences between the various probed subsites were identified on PCA and PC-LDA. As cancers and pre-cancers of buccal mucosa, lip and tongue were most commonly encountered in this study; anatomical differences and influence of anatomical variability were evaluated only on these subsites. In healthy subjects, buccal mucosa and tongue classified as almost distinct

entities, but lip misclassified with both these sites. This could be attributed to the fact that lip shares some anatomical features with both, buccal mucosa and tongue ***(AACR, 2011)***.

Raman spectroscopy based diagnostics: exploring classification between healthy, contralateral, premalignant and tumor conditions

As RS could distinctly identify sub-sites buccal mucosa, lip and tongue because of their anatomical characteristics, the observed inter-anatomical variability may influence RS based-oral cancer diagnostics. The potential of in vivo RS in discriminating healthy and pathological (contralateral normal, premalignant and malignant) was thus undertaken for both, pooled subsites and individual subsites. In the first step, efficiency of RS in classifying healthy vs. pathology at 3 pooled subsites-buccal mucosa, lip and tongue was undertaken. Overall, classification efficiencies of 98%, 60%, 30%, and 76% were observed for healthy, contralateral normal, premalignant and malignant, respectively. In the next step, PC-LDA for individual subsites was also carried out. For buccal mucosa, classification efficiency of 99%, 70%, 56% and 72% for healthy, contralateral normal, premalignant and malignant was observed. Similar trends were observed for subsites tongue and lip. Thus, PC-LDA of individual sub-sites yielded enhanced overall outcomes. Nevertheless, as healthy could be correctly predicted with ~98% efficiency even for pooled subsites, use of this one standard model may be a more desirable and practical approach for preliminary screening ***(Biochem Biophys Res Commun, communicated)***

Age-related effects at subsites and influence on healthy and pathological classification

Age is an important factor that may confound diagnosis, since ageing is associated with micro-structural changes in oral subsites. Aging-related physiological changes were therefore investigated at the 3 subsites- buccal mucosa, lip and tongue. Age-related changes could be

identified clearly at the subsite buccal mucosa; no distinct patterns for lip and tongue were identified. These age-related changes however did not have any bearing on healthy and pathological classification at any of the subsites. ***(Technol Cancer Res Treat, 2012)***

Summary

The present study aimed to identify Raman spectral signatures to classify oral pre-cancers, cancers from controls using less-invasive (serum and exfoliated cells) and non-invasive (in vivo) approaches. Briefly, findings of the study are: **A. Serum studies:** 1) Serum RS can classify premalignant, disease and oral cancer distinctly from healthy controls; prospectively may be employed for oral cancer screening 2) Buccal and tongue cancers can be distinguished with ~71% efficiency; **B. Exfoliated cells:** 3) Exfoliated cells from healthy and oral cancer sites can be distinguished; 4) Exfoliated cells from premalignant lesions can be distinguished from healthy mucosa, improvement in cell collection mechanisms crucial; C. **In vivo:** 5) Anatomical differences at buccal mucosa, lip and tongue are apparent in healthy conditions, 6) Efficacy of in vivo RS in differentiating healthy, contralateral, premalignant and tumor conditions was demonstrated at individual and pooled subsites- buccal mucosa, lip and tongue. The 'pooled subsites' standard model may serve as a more useful preliminary screening tool. Thus, the present study suggests that with use of samples like serum and exfoliated cells and in vivo measurements, RS may have potential in overall management of oral cancer after further validation in defined study designs.

References:

1 Warnakulasuriya, S. Global epidemiology of oral and oropharyngeal cancer. *Oral oncology* **45**, 309-316 (2009).
2 Coelho, K. R. Challenges of the oral cancer burden in India. *Journal of cancer epidemiology* **2012** (2012).
3 Khandekar, S., Bagdey, P. & Tiwari, R. Oral cancer and some epidemiological factors: A hospital based study. *Indian J Community Med* **31**, 157-159 (2006).

4 Prasad, L. Burden of oral cancer: An Indian scenario. *Journal of Orofacial Sciences* **6**, 77 (2014).
5 Khan, Z. U. An Overview of Oral Cancer in Indian Subcontinent and Recommendations to Decrease its Incidence. (2012).
6 Sankaranarayanan, R. *et al.* Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial. *The Lancet* **365**, 1927-1933 (2005).
7 Lane, P. M. *et al.* Simple device for the direct visualization of oral-cavity tissue fluorescence. *Journal of biomedical optics* **11**, 024006-024006-024007 (2006).
8 Messadi, D. V. Diagnostic aids for detection of oral precancerous conditions. *International journal of oral science* **5**, 59-65 (2013).
9 Mehrotra, R. & Gupta, D. K. Exciting new advances in oral cancer diagnosis: avenues to early detection. *Head & neck oncology* **3**, 1-9 (2011).
10 Liu, C. J. *et al.* Increase of microRNA miR-31 level in plasma could be a potential marker of oral cancer. *Oral diseases* **16**, 360-364 (2010).
11 Sidransky, D. Emerging molecular markers of cancer. *Nature Reviews Cancer* **2**, 210-219 (2002).
12 Sawant, S. S., Zingde, S. M. & Vaidya, M. M. Cytokeratin fragments in the serum: their utility for the management of oral cancer. *Oral oncology* **44**, 722-732 (2008).
13 Malati, T. Tumour markers: an overview. *Indian Journal of Clinical Biochemistry* **22**, 17-31 (2007).
14 Bro, R. *et al.* Forecasting individual breast cancer risk using plasma metabolomics and biocontours. *Metabolomics*, 1-5 (2015).
15 Swinson, B., Jerjes, W., El-Maaytah, M., Norris, P. & Hopper, C. Optical techniques in diagnosis of head and neck malignancy. *Oral Oncology* **42**, 221-228 (2006).
16 Ellis, D. I., Cowcher, D. P., Ashton, L., O'Hagan, S. & Goodacre, R. Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool. *The Analyst* **138**, 3871-3884 (2013).
17 Fenn, M. B. *et al.* Raman spectroscopy for clinical oncology. *Advances in Optical Technologies* **2011** (2011).
18 Evers, D., Hendriks, B., Lucassen, G. & Ruers, T. Optical spectroscopy: current advances and future applications in cancer diagnostics and therapy. *Future oncology* **8**, 307-320 (2012).
19 Malini, R. *et al.* Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: a Raman spectroscopy study. *Biopolymers* **81**, 179-193 (2006).
20 Guze, K. *et al.* Parameters defining the potential applicability of Raman spectroscopy as a diagnostic tool for oral disease. *Journal of biomedical optics* **14**, 014016-014016-014019 (2009).
21 Singh, S., Deshmukh, A., Chaturvedi, P. & Krishna, C. M. In vivo Raman spectroscopic identification of premalignant lesions in oral buccal mucosa. *Journal of biomedical optics* **17**, 1050021-1050029 (2012).
22 Krishna, H., Majumder, S. K., Chaturvedi, P. & Gupta, P. K. Anatomical variability of in vivo Raman spectra of normal oral cavity and its effect on oral tissue classification. *Biomedical Spectroscopy and Imaging* **2**, 199-217 (2013).
23 Singh, S., Deshmukh, A., Chaturvedi, P. & Krishna, C. M. in *SPIE BiOS*. 82190K-82190K-82196 (International Society for Optics and Photonics).
24 Pichardo-Molina, J. *et al.* Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients. *Lasers in medical science* **22**, 229-236 (2007).
25 Harris, A. T. *et al.* Potential for Raman spectroscopy to provide cancer screening using a peripheral blood sample. *Head & neck oncology* **1**, 1-8 (2009).
26 González-Solís, J. L. *et al.* Cervical cancer detection based on serum sample Raman spectroscopy. *Lasers in medical science* **29**, 979-985 (2014).
27 Feng, S. *et al.* Nasopharyngeal cancer detection based on blood plasma surface-enhanced Raman spectroscopy and multivariate analysis. *Biosensors and Bioelectronics* **25**, 2414-2419 (2010).
28 Owens, G. L. *et al.* Vibrational biospectroscopy coupled with multivariate analysis extracts potentially diagnostic features in blood plasma/serum of ovarian cancer patients. *Journal of biophotonics* **7**, 200-209 (2014).
29 Lin, D. *et al.* Colorectal cancer detection by gold nanoparticle based surface-enhanced Raman spectroscopy of blood serum and statistical analysis. *Optics express* **19**, 13565-13577 (2011).
30 Papamarkakis, K. *et al.* Cytopathology by optical methods: spectral cytopathology of the oral mucosa. *Laboratory Investigation* **90**, 589-598 (2010).
31 Nanda, K. *et al.* Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Annals of internal medicine* **132**, 810-819 (2000).
32 Niv, Y. & Sperber, A. D. Sensitivity, specificity, and predictive value of fecal occult blood testing (Hemoccult II) for colorectal neoplasia in symptomatic patients: a prospective study with total colonoscopy. *The American journal of gastroenterology* **90**, 1974-1977 (1995).
33 Ferrini, R., Mannino, E., Ramsdell, E. & Hill, L. Screening mammography for breast cancer: American College of Preventive Medicine practice policy statement. *American journal of preventive medicine* **12**, 340-341 (1996).

34 Rubina, S., Amita, M., Bharat, R. & Krishna, C. M. Raman spectroscopic study on classification of cervical cell specimens. *Vibrational Spectroscopy* **68**, 115-121 (2013).

Publication in Referred Journal:

a. Published

- "Raman spectroscopy of oral buccal mucosa: a study on age-related physiological changes and tobacco-related pathological changes", ***Aditi Sahu***, Atul Deshmukh, A.D. Ghanate, S.P. Singh, Pankaj Chaturvedi, C. Murali Krishna., Technol Cancer Res Treat. 11, 529-541 (2012).
- "In vivo Raman spectroscopy of oral buccal mucosa: a study on malignancy associated changes (MAC)/cancer field effects (CFE)", S.P.Singh, ***Aditi Sahu***, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna., Analyst, 2013.
- "Raman spectroscopy of serum: an exploratory study for detection of oral cancers", ***Aditi Sahu,*** Sharada Sawant, Hitesh Mamgain, C. Murali Krishna., Analyst, 2013.
- "Serum based diagnosis of asthma using Raman spectroscopy: an early phase pilot study", ***Aditi Sahu***, Krishna Dalal, Sarla Naglot, Parveen Aggarwal, Murali Krishna C. PLoS One 8(11): e78921, 2013.
- "Recurrence prediction in oral cancers: a serum Raman spectroscopy study", ***Aditi Sahu***, Nikhila Nandakumar, Sharada Sawant and C. Murali Krishna, Analyst 140(7):2294-301, 2015.
- "Raman spectroscopy of serum: a study on oral cancers", ***Aditi Sahu***, Sneha Talathi, Sharada Sawant and C. Murali Krishna, Biomedical Spectroscopy and Imaging, 4(2), 20-15.

b. Accepted: NA

c. Communicated:

- Raman spectroscopy and cytopathology of oral exfoliated cells for oral cancer diagnosis, ***Aditi Sahu,*** Sneha Tawde, Venkatesh Pai, Poonam Gera, Pankaj Chaturvedi, Sudhir Nair, C. Murali Krishna, Analytical Methods
- Oral cancer screening: A serum Raman spectroscopic approach, ***Aditi K. Sahu*** , Suyash Dhoot, Amandeep Singh, Sharada S. Sawant, Nikhila Nandakumar, Sneha Talathi-Desai, Mandavi Garud, Sandeep Pagare, Sanjeeva Shrivastava, Sudhir Nair, Pankaj Chaturvedi, C. Murali Krishna, Analyst
- Diagnosis of cancers of oral cavity using in vivo Raman spectroscopy, ***Aditi K. Sahu***, Arti R. Hole, Atul Deshmukh, Sudhir Nair, Pankaj Chaturvedi, C. Murali Krishna, Biochemical and Biophysical Research Communications

d. Other Publications and Conference proceedings:

- "Raman spectroscopy and oral exfoliative cytology", ***Aditi Sahu***, Nupur Shah, Manoj Mahimkar, Mandavi Garud, Sandeep Pagare, Sudhir Nair, C. Murali Krishna, Proc SPIE 89262N-89262N-4, 2014.
- "Classification of oral cancers using Raman spectroscopy of serum", ***Aditi Sahu***, Sneha Talathi, Sharada Sawant, C. Murali Krishna, Proc SPIE 89390E-89390E-7, 2014.

Conference attended / Oral/ poster presentations:

- "A Raman spectroscopy approach for classification of subsites in oral cancers: A feasibility study", Aditi Sahu, S.P. Singh, Atul Deshmukh, Pankaj Chaturvedi, C. Murali Krishna, *AACR 2011*, New Delhi.
- "Serum based diagnosis of asthma: a Raman spectroscopic approach", Aditi Sahu, Krishna Dalal, Sarla Naglot, and C. Murali Krishna, *ICORS 2012*, Bangalore
- "To explore utility of Raman spectroscopy in therapeutic drug monitoring of Imatinib", Vikram Gota, Aditi Sahu, Anand Patil, and C. Murali Krishna *ICORS 2012*, Bangalore
- "Raman spectroscopy and oral exfoliative cytology", Aditi Sahu, Nupur Shah, Manoj Mahimkar, Mandavi Garud, Sandeep Pagare, Sudhir Nair, C. Murali Krishna, Proc SPIE 89262N-89262N-4, 2014.
- "Classification of oral cancers using Raman spectroscopy of serum", Aditi Sahu, Sneha Talathi, Sharada Sawant, C. Murali Krishna, Proc SPIE 89390E-89390E-7, 2014.

Signature of student: Aditi Sahu Date: 25/06/2015

Doctoral committee:

S. No.	Name	Designation	Signature	Date
1.	Dr. Shubhada V. Chiplunkar	Chairperson	S.V. Chiplunkar	26/6/15
2.	Dr. C. Murali Krishna	Convener		27/6/15
3.	Dr. Prasanna Venkatraman	Member		30.6.15
4.	Dr. Dibyendu Bhattacharyya	Member		2/7/15
5.	Dr. Sudhir Nair	Invitee/Member		6/7/15
6.	Dr. Asawari Patil	Invitee/Member		7/7/15

Forwarded through:

S.V. Chiplunkar

Dr. S.V. Chiplunkar
Director ACTREC
Chairperson,
Academic & Training Programme

Dr. S. V. Chiplunkar
Director
Advanced Centre for Treatment, Research & Education in Cancer (ACTREC)
Tata Memorial Centre
Kharghar, Navi Mumbai 410 210.

Dr. K. Sharma
Director, Academics
T.M.C.

Prof. K.S. Sharma
DIRECTOR - ACADEMICS, TMC
Mumbai - 400 012

ABBREVIATIONS

SCC- squamous cell carcinoma
OSCC- oral squamous cell carcinoma
AJCC- American joint committee on cancer
SEER- Surveillance, Epidemiology and End Results
HPV- human papilloma virus
HSV- herpes simplex virus
EBV- Epstein-Barr virus
CIS- carcinoma-in-situ
OSMF- Oral sub-mucous fibrosis
LOH- loss of heterozygosity
TB- Toluidine blue
DNA- deoxy ribonucleic acid
RNA- ribonucleic acid
CRP- C-reactive protein
CEA- Carcinoembryonic antigen
TPA- Tissue polypeptide antigen
TPS- Tissue polypeptide specific
CA- cancer antigen
MMP- Matrix metalloproteinase
NADH- nicotinamide adenine dinucleotide (reduced form)
NAD^+- nicotinamide adenine (oxidized)
PCA- Principal Component Analysis
PPV- positive predictive value
SFM- spectral filtering modulation
PC-LDA- principal component linear discriminant analysis
FTIR- Fourier-transform infrared spectroscopy
SVM- support vector machine
ESS- Elastic scattering spectroscopy
DRS- Diffuse reflectance spectroscopy
IR- infrared
CCD- charged coupled device
CW- Continuous-wave
Nd:YAG- neodymium-doped yttrium aluminium garnet
mW- milli Watt
NIST- National Institute of Standards and Technology
HCA- hierarchal cluster analysis
ANN- artificial neural network
PLS- partial least square algorithm
LDA- linear discriminant analysis
QDA- quadratic discriminant analysis
FDA- factorial discriminant analysis
PLS-DA- partial-least-square discriminant analysis
LOOCV- Leave one out cross validation
SERS- Surface-enhanced Raman spectroscopy
DCDR- Drop coating deposition Raman
RRS- resonance Raman spectroscopy
CARS- coherent anti-Stokes Raman spectroscopy
SRS- stimulated Raman spectroscopy
SORS- spatially offset Raman spectroscopy
SESORS- surfaced-enhanced spatially offset Raman spectroscopy
SIRS- systemic immune response syndrome

ICU- Intensive Care Unit
DMBA- 7,12-dimethybenzanthracene
RBC- red blood cell
BCC-basal cell carcinoma
SRM- standard reference materials
LP- long-pass
SP- short-pass
SCA- Sickle cell anemia
ATR- attenuated total reflectance
NIR- near infrared
CFE- cancer-field-effects
MAC- malignancy-associated-changes
PSA- prostate specific antigen
HCG- human chorionic gonadotropin
TMH- Tata memorial hospital
IRB- Institutional Review Board
OPD- outpatient department
µl- micro-liter
SRM- standard reference material
cfNA- Cell free nucleic acids
CaF_2- calcium fluoride
BM- buccal mucosa
HNSCC- head and neck squamous cell carcinoma
EGFR- epidermal growth factor receptor
ACP- acid phosphatase
MRC- minimal residual cancer
FC- field cancerization
Pap- Papanicolaou
AgNOR- Argyrophilic nucleolar organiser region
CIN- cervical intraepithelial neoplasia
N:C- nuclear:cytoplasmic
HV- healthy volunteers
HT- healthy tobacco or habit control
DC- Disease control
T- Tumor
OG- Orange G
EA- Eosin-Azure
EC- Exfoliative cytology
OPL- oral premalignant
HSP- heat shock protein
MRP- mitochondrial ribosomal protein
OCT- optical coherence tomography
HRME- high-resolution micro endoscopy
RMT- retro-molar trigone
FOM- floor of mouth
ACTREC- Advanced Center for Treatment, Research, Education in Cancer
ANSI- American National Standards Institute
FOBT- fecal-occult-blood test
MRI- magnetic resonance imaging

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

"Tumors destroy man in a unique and appalling way, as flesh of his own Flesh which has somehow been rendered proliferative, rampant, predatory And ungovernable. They are the most concrete and formidable of Human maladies, yet despite more than 70 years of experimental study, They remain the least understood!"

Francis Peyton Rous, Tumor virologist, Nobel lecture- 1966

1.1. Cancer

Life expectancy has shown a global increase, as death toll from major diseases has shown a substantial decrease in the last decade. In most countries, the reduced age-sex specific mortality was associated with a shift towards higher deaths caused by non-communicable disease and injuries, including cancer[35]. In developing countries like India, the non-communicable diseases such as cardiovascular disease, cancer, and diabetes killed 38 million people in the year 2013, significantly more than in the high-income countries[36 37].

Cancer is a complex group of diseases characterized by cells which are no longer completely responsive to the signals that regulate cellular differentiation, survival, proliferation and death. Consequently, these cells accumulate within the tissue and cause local damage and inflammation. There are over 200 different types of cancers. Globally, cancers have emerged as the second-leading cause of death after cardiovascular diseases. The global proportion of cancer-related deaths increased from 12% in 1990 to 15% in 2013. During the same tenure, the number of cancer incidence almost doubled in India[35].

To tackle this significant health problem in the coming decades, while the goals of health policy include prevention of premature deaths and reduce health care costs, the major goal of clinical medicine is to improve detection systems and therapeutic options "to prevent complications and catastrophic events." [38,39]

1.2. Oral cancer:

Oral cancer or oral cavity cancer is a subtype of head and neck cancer and is defined as any 'abnormal, cancerous tissue growth' located in the oral cavity[40]. Oral cancer commonly involves the subsites buccal mucosa, tongue, floor of the mouth, gingiva, lips, palate, maxilla or mandible. These cancers at distinct subsites are heterogeneous and also have variable predisposing factors, prevalence, and treatment outcomes. Squamous cell carcinoma (SCC) is the most common type of malignant neoplasm in the oral cavity[41].

It is well established that oral squamous cell carcinoma (OSCC) develops as a result of accumulation of several molecular and biochemical cellular alterations[42] and changes in the underlying fibrovascular stroma including neo-vascularization[43]. As initial cellular abnormalities accumulate, they are reflected in alterations of the clinical appearance of the affected epithelial tissues[44], known as pre-cancer. Pre-cancer refers to a clinically apparent precancerous lesion, which is a benign, morphologically altered tissue that has a greater than normal risk of malignant transformation. The clinical significance of oral precancerous lesions lies in their intimate association with malignant transformation into OSCC[45,46]. Tobacco use, including smokeless tobacco, and excessive alcohol intake are the major etiological factors and are estimated to account for about 90% of oral cancers[47]. Despite the accessibility and amenability of oral cavity to clinical inspection during visual examination, known etiological factors, the presence of preceding premalignant lesions with mostly well-defined clinical diagnostic features, most oral cancers are detected in advanced stages[48].

1.2.1. Oral cavity:

The oral cavity represents the first structure of the digestive tract. The oral cavity as a whole can be subdivided into 3 major areas- oral cavity proper, oropharynx and the vestibule. The vestibule is the space between the cheeks and lip laterally and dentition medially. The oral cavity proper is defined as the region from the skin-vermillion junction of the lips to the hard and soft palate superiorly and to the line of circumvallate papillae inferiorly and the palatoglossal arch posteriorly. The oropharynx lies posterior to the palatoglossal arch and includes the posterior one-third of the tongue, palatine tonsils, soft palate and posterior wall. The oral cavity is lined by the oral mucosa which is a stratified squamous epithelium. The major functions of the oral cavity include, a) entry of food, b) mechanical processing of food through the action of teeth, tongue and palatal surfaces, c) lubrication by mixing with mucus and salivary gland secretions, d) limited digestion of carbohydrates and lipids, e) maintenance of protective barrier and provide immune defense, f) facilitate speech and swallowing. The oral cavity is composed of distinct anatomic subsites. Lips, buccal mucosa, the upper and lower alveolar ridges with their attached gingiva, the retromolar trigone, the hard palate, the floor of the mouth, and the anterior two-thirds of the tongue majorly constitute the oral cavity, as shown in Figure 1.1.

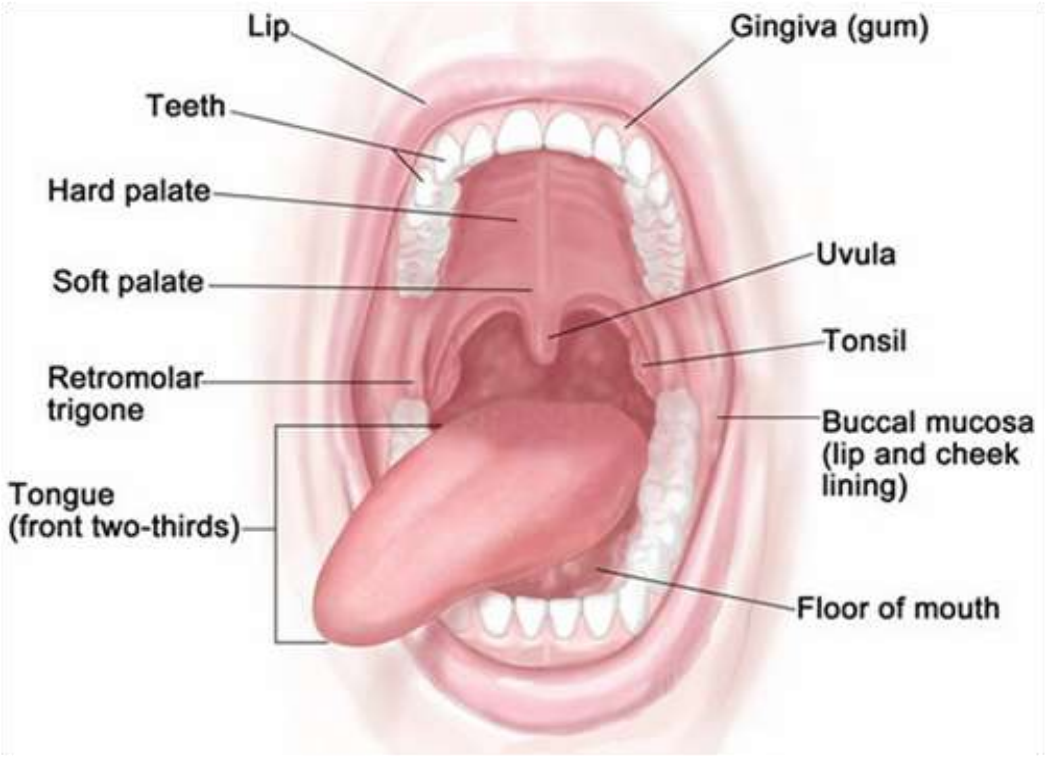


Figure 1.1 Oral cavity diagrammatic representation (Adapted from http://seer.cancer.gov/i/factsheets/oralcav-lg.jpg)

The upper and lower lips begin at the junction of the skin with the vermillion border and form the anterior boundary of the oral cavity. The buccal mucosa includes the lining of inner cheeks and lips extending from the line of contact of the opposite lips to the line of attachment of upper and lower alvelolar ridges and pterygomandibular raphe and forms the lateral borders of the oral cavity. The upper attached gingiva along with the attached gingival mucosa constitute dental surfaces of the maxillary bone while the lower alveolar ridges refer to mucosa-covered alveolar processes of the mandible. The mucosa of upper and lower ridges is tightly attached to the underlying bone. The retromolar trigone is the triangular stretch of mucosa over a flat, bony surface posterior to the last mandibular molar tooth and it represent the posterior area of the vestibule. The roof of the oral cavity is formed by the hard and soft palate. The hard palate is

formed by the palatine processes of the maxillary bone and the horizontal plates of the palatine bones. The mucosa at this site is tightly adherent to the bone. The tongue along with the floor of mouth comprises the inferior boundary of the oral cavity. The tongue can be divided into two parts: anterior two-thirds oral tongue and posterior one-third tongue base. The dorsal surface of the tongue is covered with specialized mucosa which contains surface projections called papillae. The floor of mouth is the semilunar space over the mylohyoid and geniohyoid muscle that is lined by a very thin layer of non-keratinized mucosa.

1.2.2. Epidemiology:

Oral cancer is the sixteenth most common cancer reported globally, with an annual incidence of over 300,000 cases, of which ~60% are reported from developing countries. There exists variability in the incidence of oral cancer in different parts of the world, directly proportional to prevalence of risk factors in that demographic area. Because of ethnic, social and life-style related factors, highest incidence rates are found in three developing countries (India, Pakistan and Brazil) and one developed country (France)[49]. In fact, it is estimated that over 90% of the global smokeless tobacco burden is in South Asia- around 100 million people use smokeless tobacco in India and Pakistan alone [50,51]. Due to a known dose-response relationship between tobacco consumption and development of oral cancer, chronic tobacco abusers are at high risk for development of oral cancer [52]. Oral cancer is the most common cancer in Indian males, while the third most common cancer in Indian females. Oral cancers account for over 30% of all cancers in India, while only 3% of malignancies in the U.S. population[2,53,54].

1.2.3. Symptoms

Oral cancers may have an early symptomatic phase in the form of persistent red or white patches (erythroplakia/leukoplakia) or non-healing ulcers. Other symptoms may include lump or swelling in the oral soft tissues, hoarseness, difficulty in chewing or swallowing (dysphagia), difficulty in opening mouth (trismus), chronic ear-pain (otalgia),numbness, oral or nose bleeding (epistaxis), change in denture fittings and cervical lymphadenopathy[55].

1.2.4. Diagnosis and treatment

Biopsy followed by histopathology of suspicious lesions found during clinical examination is the gold standard for diagnosis of oral cancers. Removal of lesions with mild or severe dysplasia is advocated while mild dysplasia is followed up for reversal or progression. Treatment for oral cancer includes surgery, radiotherapy, and chemotherapy; surgery combined with chemotherapy and radiotherapy improves overall survival. Approximately one-third of patients treated with surgery and adjuvant therapy experience recurrence (loco-regional, relapse, second primary and second field tumors) and/or distant metastasis. In spite of the advancement in surgical and treatment modalities, low disease free survival rates have been observed for several decades. The main reasons for the dismal survival rates include- diagnosis mainly in advanced stages, recurrence, inadequate access to health services and lack of primary knowledge about causative factors. Early detection of oral cancer may ensure better patient survival rates and improved quality of life [56,57].

1.2.5. Staging and Histopathological grading

Staging of oral cancers is an important pre-requisite for treatment decisions and prediction of prognosis. Staging is done as per the criteria laid down by the American Joint Committee on

Cancer (AJCC) manual[58]. The TNM system of cancer staging is prescribed by the AJCC. In TNM, T stands for extent of primary tumor, N stands for absence/presence and extent of regional lymph node metastasis and M stands for absence or presence of distant metastasis. The use of numerical digits indicate progressive extent of the malignant disease, for example, T1 represents early cancers while T4 represents last or terminal stage cancers. There exist four classification systems for each cancer site: Clinical, Pathologic, Retreatment and Autopsy classification. For oral cancers, stage of disease at diagnosis is one of the key determinants of patient survival. In a Surveillance, Epidemiology and End Results (SEER) program report, the five-year survival for patients with localized disease (T1 or T2 stage) was 81.9 percent which dropped to 46.4 percent for patients with regional spread (T3) and to 21.1 percent for those with distant metastases (T4)[59].

Histological grading of oral SCC is an important parameter for clinical prediction of prognosis. Microscopic classification of oral SCC is based on subjective assessment of degree of keratinization, cellular and nuclear polymorphism and mitotic activity. The three grades of SCC are- well-differentiated, moderately-differentiated and poorly-differentiated. While well- and moderately-differentiated are classified as low grade with better prognosis, poorly-differentiated and undifferentiated are classified as high grade. These grades may not reflect the actual prognosis because of the associated limitations such as histological heterogeneity, subjective methods of assessment and non-reliance on functional features[41,60,61].

1.2.6. Risk factors/Etiological agents

Tobacco abuse is the major risk factor for oral cancer development. Alcohol consumption, along with viral infections such as human papilloma virus (HPV), poor dietary habits are also known to be significant risk factors[62,63]. In fact, tobacco use, alcohol consumption and poor diet account for over 90 percent of all cancers[64]. In case of tobacco, all forms of use-including smoke and smokeless contribute to the development of cancer in the oral cavity.

Studies indicate a strong, definitive association between use of tobacco products and development of oral cancer. Conventionally, around 75% patients above the age of fifty and diagnosed with oral cancer have used tobacco in one form or the other. Several forms of tobacco abuse persist; these include both smoking (cigarettes, pipes, cigars, and bidis) and smokeless tobacco (chewing tobacco, snuff, and betel quid). The carcinogenicity of tobacco lies in the presence of numerous carcinogens present in the tobacco smoke or the water-soluble components that leach into the saliva[65]. Smoking tobacco is inherently linked to several cancers, including lung, esophagus, bladder, pancreas and oral cavity[66]. Smoking is practiced mainly by means of cigarettes or bidis (home-made cigarettes used extensively by rural population in the Indian sub-continent)[67]. Epidemiological studies indicate that the risk of developing oral cancer is five to nine times greater for smokers than for nonsmokers[68,69]. Cigarette smoke contains more than 60 carcinogenic compounds. Both bidi and cigarette smokers pose a substantial risk for development of oral cancer[70]. The sites for oral cancer development induced by tobacco smoking are not fixed; however lip is the most commonly affected site in pipe users[64]. Smokers palate or stomatitis nicotiana lesions, predominantly found in the hard palate of chronic smokers, are associated with low malignant potential[64,71,72]. Reverse smoking (where the lit end of cigarette is

kept inside the oral cavity) had led to incidences of palatal cancer in India and Latin America[73,74].

Smokeless tobacco (without combustion) is another form of tobacco consumption. They account for significant proportion of oral cancers in South Asian countries; these countries account for over 90% of global smokeless tobacco burden[75,76]. This form of tobacco abuse involves placing the tobacco in close contact with the mucous membranes through which nicotine is absorbed into the circulation to give pharmacological effects. Smokeless tobacco contains about 28 known carcinogens which majorly include the nonvolatile alkaloid-derived tobacco-specific N-nitrosamine and N-nitrosamino acids[77]. Chewing tobacco and snuff are two main types of smokeless tobacco. Snuff is made from fine-cut or ground tobacco and is usually dry or moist while chewing tobacco comes in the form of loose leaf, plug or twist. They can be used alone or in combination with other ingredients[78]. Some common forms of smokeless tobacco include betel quid (Areca nut, betel leaf/inflorescence, slaked lime, catechu, condiments, with or without tobacco), Khaini (tobacco and lime), Mishri (burned tobacco), Zarda (boiled tobacco), Gadakhu (tobacco and molasses), Mawa (tobacco, lime and areca), Nass (tobacco, ash, cotton oil), Shammah (tobacco, ash, lime) and Toombak (tobacco and sodium bicarbonate) used predominantly in South Asian countries and Middle-East[77,79]. The pre-cancerous lesions due to regular use of smokeless tobacco include leukoplakia, erythroplakia, snuff dippers' lesion, tobacco pouch keratosis which develop at the site of product application. Additionally, use of areca nut in mixtures or in betel quid leads to a progressive pre-cancerous condition called oral sub-mucous fibrosis (OSMF)[80].

Alcohol, either alone or in combination with tobacco, has shown to have carcinogenic potential. Moderate-to-heavy drinkers have shown to have a three to nine times greater risk of developing oral cancer than non-drinkers. The synergistic effect of alcohol and smoking has also been demonstrated; it has been shown that cohort of patients with both smoking and drinking habits have over one hundred times greater risk for developing a cancer[68,69,81-85].

Infectious agents like viruses are also known to be associated with oral and oropharyngeal cancer. It is postulated that about 15 percent of head and neck cancers may have a viral etiology. HPV have shown a strong association with oral cancers, particularly in non-tobacco using female population below 40 years of age. High-risk HPVs like HPV-16 and HPV-18 have been detected in about 22 percent and 14 percent of subjects, respectively[86]. Other viruses like herpes simplex virus (HSV) and Epstein-Barr virus (EBV) have also been speculated to have a role in oral carcinogenesis[87-90].

Other factors such as low intake of fruits and vegetables, familial and genetic predisposition[91], occupational risks[92], poor oral hygiene and malformed/ irregular dentition may also lead to oral cancer[63,93,94]. Adequate intake of fruits and vegetables has shown to have an inverse relationship with risk of oral cancer[95,96].

1.2.7. Oral pre-cancer lesions

The most common precancerous lesions present clinically as white, red or a mix of white and red mucosal changes. These clinical conditions are known as leukoplakia or erythroplakia. There are other pathological conditions that are considered precancerous including oral lichen planus and

OSMF. The presence of dysplasia in these conditions indicates an increased risk of malignant transformation.

1.2.7.1. Leukoplakia

According to the World Health Organization (WHO), leukoplakia is defined as "a white patch or plaque that cannot be characterized clinically or pathologically as any other disease."[72] The term is based only on clinical diagnosis of exclusion; it does not have a histopathological basis[97]. Leukoplakia is known to have a strong association with tobacco abuse[98]. Leukoplakia is seen most frequently in middle-aged and older adults, especially men, with an increasing prevalence seen with age[99]. The most common sites for leukoplakia development are buccal mucosa, alveolar mucosa, and lower lip; however, lesions in the floor of mouth, lateral tongue, and lower lip are most likely to show dysplastic or malignant changes[100]. Photographic representation of leukoplakia in the oral cavity is presented in Figure 1.2.

Leukoplakia may be homogenous or non-homogenous. Homogenous leukoplakia is characterized by uniformly flat, white and thin lesions with surface fissures (because of keratin) and non-homogenous lesions are typically white and red lesions with irregular (speckled leukoplakia), papillary (verrucous leukoplakia) or nodular features[98,101]. Histologic examination may reveal hyperkeratosis, dysplasia, carcinoma-in-situ (CIS) or invasive SCC in these lesions[102]. The rate of progression to malignancy has been reported to be between 3.6% and 17.5%[103,104]. Leukoplakias with an intermixed red component (speckled leukoplakia or erythro-leukoplakia) are at a greater risk for showing dysplasia or carcinoma.

Proliferative verrucous leukoplakia is a variant with thick and papillary surface without known risk factors like tobacco and is seen most commonly in women. It is characterized by multi-focal development and has a strong potential for malignant transformation[105-107].

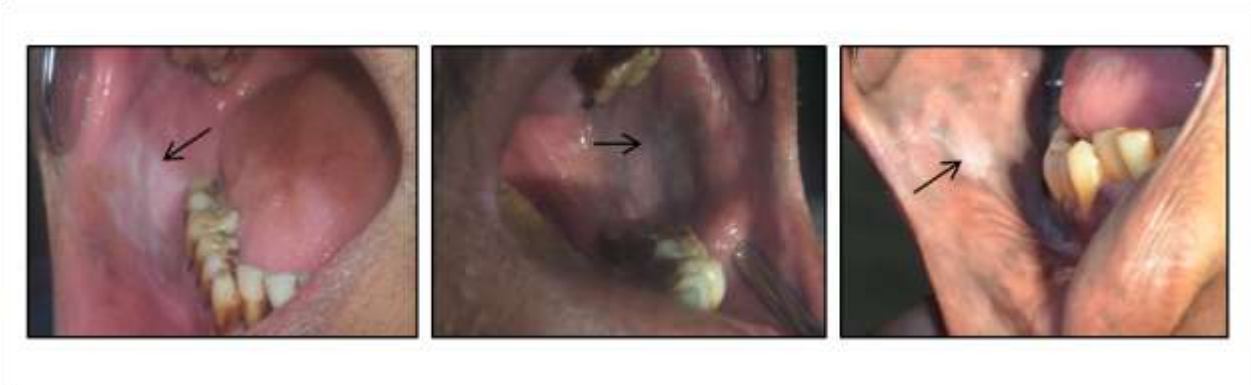


Figure 1.2. Photographic representations of leukoplakia lesions on buccal mucosa (Arrows indicate location of the leukoplakia lesions)

1.2.7.2. Erythroplakia

Erythroplakia is a clinical term that refers to a red patch that cannot be defined clinically or pathologically as any other condition[72]. Erythroplakia is most commonly seen in older males as erythematous lesions with a soft, velvety texture[44,108]. Tobacco consumption and alcohol are etiological factors and floor of mouth, lateral tongue, soft palate and retromolar trigone are most common afflicted sites[62,109]. Although less commonly encountered, erythroplakia demonstrate a greater propensity of malignant transformation. Upon histological analysis, 51% lesions demonstrated invasive squamous cell carcinoma (SCC), with 40% demonstrating carcinoma in situ, and 9% exhibiting mild-moderate dysplasia[110]. Photographic representation of erythroplakia in the oral cavity is presented in Figure 1.3.

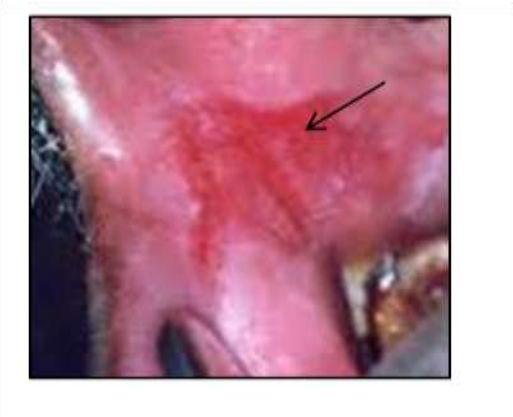


Figure 1.3. Photographic representation of an erythroplakia lesion on buccal mucosa (Arrow indicates location of the erythroplakia lesion)

1.2.7.3. Oral sub-mucous fibrosis

OSMF is a chronic, debilitating progressive mucosal condition characterized by inflammation and subsequent juxta-epithelial fibrosis of the oral cavity. It involves a fibro-elastic transformation induced due to chronic inflammatory reactions. It leads to epithelial atrophy, fibrosis in lamina propria and deeper connective tissues, and eventually to stiffness and trismus[111]. The pathogenesis is multi-factorial while the strongest risk factor for OSMF is use of Areca nut[112]. The most common site for OSMF development is buccal mucosa, however other sites may also be involved. OSMF is associated with symptoms such as burning sensation and/or intolerance to spicy food in the early stages. The symptoms worsen with time; fibrosis develops in the later stages and affects mouth opening[102,113]. The disease often manifests with diffuse involvement of the oral cavity that clinically appears as whitish mucosa lacking elasticity. Dysplasia has been reported in 7-26% of OSMF tissues, and malignant transformation rate is approximately 7%[114]. A cytology study demonstrated presence of class I cytology ie,

characteristics indicative of benign atypical cytologic changes in oral smears from OSMF patients[115]. Photographic representation of OSMF in the oral cavity is presented in Figure 1.4.

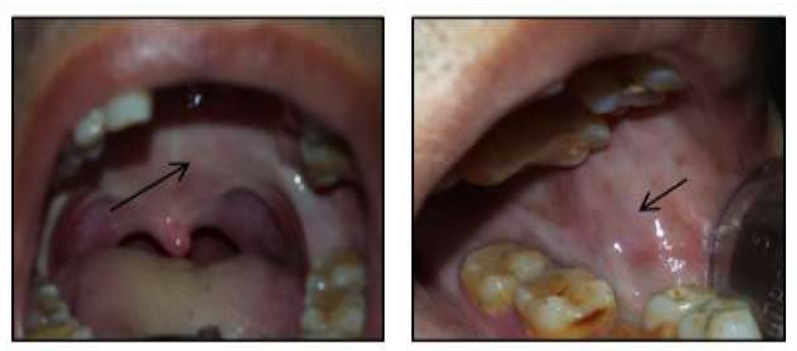


Figure 1.4. Photographic representations of OSMF at hard palate and buccal mucosa (Arrows indicate location of the areas of OSMF)

1.2.7.4. Tobacco pouch keratosis

Tobacco pouch keratosis or TPK is a tobacco-related mucosal alteration associated with use of smokeless tobacco in the form of snuff or chewing of tobacco. Most common sites include buccal and labial vestibule. The degree of mucosal alteration depends on the tobacco habit. Early lesions appear mostly as thin, white wrinkled areas or as hyperkeratotic, granular patches. Advanced lesions exhibit greatly thickened zones of grayish white mucosa with well-developed folds and fissures. Histopathologically, hyperkeratosis of oral epithelium along with incidences of minor dysplasia is observed[62,116,117]. Photographic representation of TPK in the oral cavity is presented in Figure 1.5.

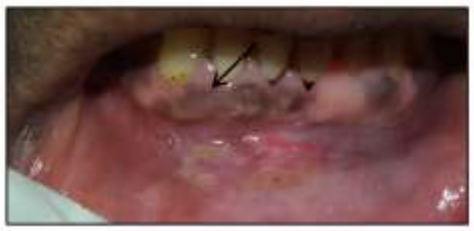


Figure 1.5. Photographic representation of a TPK lesion in the labial vestibule (Arrow indicates location of the TPK lesion)

1.2.7.5. Lichen planus

Lichen planus is a chronic, inflammatory, T-cell mediated auto-immune disorder characterized by six distinct clinical forms present mostly on tongue dorsum, buccal mucosa or gingiva. It usually presents as white striations, papules or plaques in oral cavity along with occasional erythematous lesions, erosions or blisters. In few instances, lichen planus may undergo malignant transformation[118,119]. Photographic representation of a reticular type of lichen planus lesion on buccal mucosa is presented in Figure 1.6.

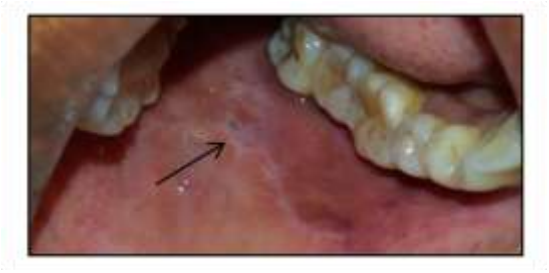


Figure 1.6. Photographic representation of a reticular-type lichen planus lesion on buccal mucosa (Arrow indicates location of the lichen planus lesion)

1.2.7.6. Common Benign lesions

Several benign conditions that manifest in the oral cavity may resemble pre-cancerous lesions. Some important examples of benign changes include- candidiasis, aphthous ulcers, cold sores, hairy tongue, frictional hyperkeratosis, mucocele. Adequate clinical experience can differentiate these conditions from cancerous changes.

1.2.8. Screening and diagnostic methods

Visual inspection followed by biopsy and histopathology is the gold standard for screening and diagnosis of oral cancers. A study has demonstrated positive effect of visual screening on oral cancer mortality in a cluster-randomized controlled trial in India in high-risk populations[6]. Adjunct techniques like Toluidine blue staining, oral cytology, tissue fluorescence (VelScope) and chemiluminescence (Vizilite) based methods, are being explored as complementary techniques to visual inspection[8]. Optical spectroscopic techniques like IR (infra red), fluorescence and Raman have also been explored for oral cancer diagnosis.

1.2.8.1. Brush biopsy and molecular analysis

Interest in oral exfoliated cells has been re-kindled in the last two decades owing to improvements in sample collection and staining procedures. Earlier studies could not achieve the expected sensitivity and specificity[120] rates as the sampling devices could not sample the deeper layers of the lesions. The few abnormal cells that may have been exfoliated were manually investigated on a slide where they were usually outnumbered by normal cells. Brush biopsy refers to a method of collection of exfoliated cells from the oral cavity using a brush, instead of the original wooden spatula. Oral CDx- the oral brush biopsy (transepithelial sample) coupled with cytomorphometry[121] has led to renewed interest in the field. Much higher sensitivity and

specificity values for detection of dysplasia in suspicious lesions were observed using this approach[122]. Further, brush biopsy can be coupled to molecular tests like DNA-image cytometry (to measure ploidy status) to yield additional information regarding the DNA content of the cells[123]. Genetic cytology for allelic imbalance and loss of heterozygosity (LOH) at 3p, 9p, 11q and 17p along with immunohistochemical staining for p53 and mutational analysis for TP53 showed promise in prediction of malignant transformation[124,125].

1.2.8.2. Vital-tissue staining

Staining of suspicious tissues with agents such as Toluidine blue (TB) may help in identification of dysplasia. Toluidine blue is a metachromatic dye with special affinity towards DNA because of its acidophilic nature. Malignant and dysplastic cells have higher DNA content and thus stain more intensely with this dye. This is the basis for differentiating normal and abnormal mucosal areas using TB. However, studies have shown high sensitivity but lower specificity for malignant lesions while low sensitivity and specificity rates for pre-malignant lesions[126,127]. TB has also helped in margins assessment during surgical resection[128]. Another study has suggested concerted use of TB and brush biopsy for improved detection of oral premalignant lesions[129].

1.2.8.3. Serum-based diagnostic markers

Serum tumor marker based approach may help in the early detection and improved patient monitoring for oral cancers. Therefore, DNA markers like TP53, microsatellite alterations, presence of HPV and EBV DNA, RNA markers like cytokeratins, microRNAs (miR21, miR181b, miR345) and protein markers CD44, telomerase, Matrix metalloproteinase (MMP)-2, MMP-9, C-reactive protein (CRP), squamous cell carcinoma antigen (SCC-ag),

Carcinoembryonic antigen (CEA), cytokeratins like Tissue polypeptide antigen (TPA, Cytokeratin 8), Tissue polypeptide specific antigen (TPS, Cytokeratin 18), Cyfra 21-1 (Cytokeratin 19) have been explored in serum from oral cancer patients[10-13,130]. Before clinical use, the individual sensitivity and specificity of these molecular markers have to be validated over a large sample size.

1.2.8.4. Saliva-based methods

Salivary biomarkers have been investigated for non-invasive diagnosis of oral cancer. High-throughput methods like microarray, transcriptomics and proteomics are being employed to detect limiting concentration of salivary biomarkers. Salivary concentrations of Cyfra 21-1, TPS and CA (cancer antigen) -125 were shown to be higher in oral cancer patients. Additionally, seven RNAs and two microRNA markers in saliva of oral cancer patients were found to have discriminatory potential. Another study demonstrated IL8, IL1B, DUSP1, HA3, OAZ1, S100P, and SAT as potential salivary RNA biomarkers that yielded high sensitivity (91%) and specificity (91%) in distinguishing OSCC from the controls[131-134].

1.2.8.5. Light-based methods

Vizilite plus and MicroLux DL are based on the principle of chemiluminescence wherein rinse using 1% acetic acid is followed by direct visual examination using a blue-white light source. While Vizilite uses a disposable chemiluminescent light, MicroLux DL employs a reusable, battery-powered source. Oral mucosal abnormalities appear as aceto-white against a lightly bluish normal mucosa. VelScope is a portable device for direct visualization of oral cavity using intense blue excitation light that employs changes in inherent tissue autofluorescence. Using a

selective (narrow-band) filter, dark regions representing abnormal areas with loss of fluorescence can be observed interspersed between green fluorescence emitting normal mucosa. Identafi employs multi-spectral fluorescence and reflectance technology to enhance visualization of mucosal abnormalities. White light is used for conventional examination, violet light to detect changes in tissue autofluorescence while amber light is used to detect morphological abnormalities.

1.2.8.6. Spectroscopic-based methods

Spectroscopic techniques like autofluorescence, diffuse reflectance and Raman spectroscopy are increasingly being explored for early oral cancer diagnosis.

1.3. Optical spectroscopy:

Optical spectroscopy involves study of light-tissue interaction. The optical spectrum derived from any tissue contains information about the histological and biochemical make up of that tissue. There has been increasing interest in the use of optical spectroscopy systems to be able to provide tissue diagnosis in real-time, non-invasively and in situ[18]. Tissue-light interaction has provided a valuable adjunct to biomedical diagnostics. In view of the inherent accessibility of oral cavity, development of effective, non-invasive diagnostic modalities based on optical spectroscopy can help in early diagnosis of oral pre-cancers and cancers. Techniques such as fluorescence spectroscopy, reflectance spectroscopy, elastic scattering spectroscopy, infrared spectroscopy and Raman spectroscopy are increasingly being investigated for oral cancer applications[15].

1.3.1. Fluorescence spectroscopy:

Steady-state fluorescence measurements from small and large tissue regions have been investigated. For these measurements, sources of both intrinsic (endogenous fluorophore or auto-fluorescence) and extrinsic fluorescence (exogenous fluorophores) have been considered. Both approaches are associated with certain advantages and disadvantages[135]. Auto-fluorescence spectroscopy involves tissue excitation through a fiber, and fluorescence is recorded by a spectrograph, while the reflected light is filtered out. Auto fluorescence based oral cancer diagnosis began in the late 1970s. Naturally occurring fluorophores such as collagen, elastin, keratin and NADH (nicotinamide adenine dinucleotide) after excitation with a suitable wavelength produce auto-fluorescence. Changes in blood concentration, collagen content, nuclear size distribution, or epithelial thickness can alter the concentration, scattering and absorption efficiencies of the natural fluorophores depending on the disease status. For example, the epithelial layer shields the strongly fluorescing collagen layer, leading to weak fluorescence during hyperplasia. An increase in cell metabolism in response to malignant changes can affect the balance between the fluorescent NADH (increase) and non-fluorescent NAD+ (decrease)[136]. In case of oral cavity, Harris and Werkhaven et al., reported difference in auto-fluorescence based on porphyrin band between healthy and tumor mucosa[137]. These findings were established later and fluorescence was attributed to microorganisms living on ulcerating or necrotic surfaces [138,139]. In another study on 130 patients, 91.1% sensitivity and 84.3% specificity was obtained in distinguishing malignant and benign lesions[140].

Chen et al. using laser-induced fluorescence reported positive predictive values (PPV) of 94% and 93%, for abnormal and normal tissues, respectively at 300 nm[141]. In the consequent study on differentiating normal, hyperkeratosis, benign and SCC, classification efficiency of 86%, 87% and 89% were obtained for hyperkeratosis, normal and dysplasia, respectively[142]. Gillenwater et al. have reported in vivo auto-fluorescence spectra from oral mucosa of 8 healthy volunteers and 15 patients with premalignant/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. Sensitivity of 82% and specificity of 100% were reported[143]. A study by Shaiju et al. involved habitual tobacco users and premalignant lesions to show that autofluorescence spectroscopy in combination with linear discriminant analysis can identify tobacco habit induced oral cavity disorders. They carried out quantification of hemoglobin concentration and porphyrin levels. Feasibility of using a single system to analyze the changes in fluorophores/chromophores within the tissue was carried out with the spectral filtering modulation (SFM) effect. Principal component-based linear discriminant analysis (PC-LDA) was used for distinguishing spectra. Their findings suggest that lower collagen level and increased redox value can be a prognostic marker for oral cancer risk assessment[144]. In a recent study, time-resolved laser-induced fluorescence spectroscopy was employed for diagnosis of premalignant and malignant lesions in oral cavity. Findings demonstrate that addition of time-resolved fluorescence-derived parameters significantly improved the capability of fluorescence spectroscopy–based diagnostics in the hamster buccal pouch model of oral cancer[145]. Fluorescence spectroscopy in the discrimination of normal oral mucosa, oral cancer, and potentially malignant disorders was also investigated in 115 individuals in a recent study. Spectra

were classified using decision tree algorithm and compared to histopathology. The specificity and sensitivity obtained were 93.8% and 88.5%, respectively at 406 nm excitation.[146] The same group evaluated the efficacy of fluorescence spectroscopy in surgical demarcation. Findings indicate potential in discrimination of normal and altered mucosa[147]. Fluorescence spectroscopic characterization of salivary metabolites of oral cancer patients was also recently investigated. The fluorescence excitation spectrum discriminate normal and oral cancer saliva with 84.1% sensitivity and 93.2% specificity[148].

1.3.2. Fourier-transform infrared spectroscopy:

IR spectroscopy is a vibrational spectroscopic method based on absorption of infra-red light. Sample is exposed to IR radiation and some of this radiation is absorbed and remaining is transmitted. The detector measures the amount of energy at each frequency which has passed through the sample. This results in a spectrum which is a plot of intensity vs. frequency. To overcome the limitations of long output times, 'interferometers' were introduced; this approach is termed as Fourier-transform infrared spectroscopy or FTIR. Fourier-transform infrared spectroscopy (FTIR) has been employed to study the structure and functions of cellular components in tissues[149]. One of the first studies was carried out by Schultz et al to analyze whether changes in tissue biochemistry induced by well-differentiated SCC can be detected by infrared spectroscopy. DNA and keratin content in tissues predominantly differentiated between normal and SCC biopsies. This study demonstrated that infrared micro-spectroscopy in combination with bivariate statistics could detect cellular changes in epithelial cancers[150].

Fukuyama et al in 1998 investigated the differences of Fourier transform infrared (FTIR) spectra between OSCC and normal gingival epithelium or normal subgingival tissue. Normal and tumor spectra show differences in the bands at 1431, 1482, 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[151]. Wu et al. have shown normal and tumor oral tissues discrimination on the basis of lipid and protein contents. 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 weak C=O band. Authors also suggested that the most common difference between normal and tumor sections was C=O band at 1745 cm^{-1}. Spectral findings were correlated with histological slides[152]. A recent study has explored FTIR-based spectral biomarker selection to discriminate leukoplakia and oral cancer on paraffin-embedded tissue section from 47 subjects. Spectral markers were selected using mean spectral differences, Mann-Whitney's U test and forward feature selection techniques. Differences were evaluated using support vector machine (SVM). Results indicate altered glycogen and keratin content in histological sections and these features could successfully discriminate leukoplakia and OSCC[153].

1.3.3. Elastic scattering spectroscopy

Elastic scattering spectroscopy or ESS is a wavelength dependent spectroscopy that yields information about the structural and morphological changes in tissues. ESS is sensitive to the size and shape of dense sub-cellular organelles like nucleus, nucleolus, even to chromatin content and nuclear:cytoplasmic ratio and therefore can detect changes in these structures. As malignancy is accompanied by a concomitant alteration in these structures at the cellular level,

ESS is a promising tool for detection of neoplasia. The tissue is interrogated by short pulses of white light, delivered by a thin optical fiber placed in contact with the tissue surface while the light scattered by the tissue is collected by a second fiber placed adjacent to the first fiber. ESS has been explored for oral cancer diagnosis by several groups. In one of the first studies, a trimodal spectroscopic approach including intrinsic fluorescence, diffuse reflectance and elastic scattering were employed to probe 91 tissue sites from 15 patients with varying degrees of malignancy (normal, dysplastic, and cancerous sites) and 8 healthy volunteers in vivo (Muller et al)[154]. A sensitivity and specificity of 96% and 96%, respectively, in distinguishing cancerous/dysplastic (mild, moderate, and severe) from normal tissue was achieved. In addition, the authors were able to distinguish dysplastic from cancerous tissue with a sensitivity of 64% and a specificity of 90%. A study by Jerjes et al explored feasibility of identifying metastasis in cervical nodes of oral cancer patients[155]. One hundred and thirty lymph nodes were examined from 13 patients who underwent neck dissection. The nodes were formalin fixed, bivalved and subjected to ESS and then routinely processed. On comparison with histopathology, ESS yielded a sensitivity of 98% and a specificity of 68%. In the next study by Jerjes et al[156], bony resection margins from formalin fixed archived materials were assessed by ESS and the results correlated with the histopathological diagnosis. Spectra were acquired from the mandibular specimens of 21 patients, of which 231 spectra were taken from histologically positive sites and 110 spectra from normal tissue. Using LDA, a sensitivity of 87% and a specificity of 80% were obtained. These results imply that ESS may identify tumour involvement of resection margins. In the subsequent study by the same group (Sharwani et al), diagnosis of premalignant and malignant oral lesions was investigated along with a corresponding histopathological analysis[157]. Twenty-

five oral sites from 25 patients with oral leukoplakia were examined by ESS and surgical biopsies were acquired. LDA yielded a sensitivity of 72% and a specificity of 75%. These results suggest that ESS may be able to identify dysplasia in oral tissues.

1.3.4. Diffuse reflectance spectroscopy

Diffuse reflectance (DR) results due to single and multiple backscattering events of the excitation light. Diffuse reflectance spectroscopy or DRS in the visible wavelength range is sensitive to the absorption and scattering properties of epithelial tissue. These properties reflect their underlying physiological and morphological properties like nuclear size, distribution, epithelial thickness and collagen content, and the amount of oxy and deoxy-haemoglobin. On excitation using visible light, oxygenated (HbO_2) and deoxygenated hemoglobin (Hb), arising from blood vessels in the stroma are dominant absorbers while light scattering is caused mainly by cell nuclei and other organelles in epithelium and stroma, and collagen and cross-links in stroma. During neoplastic transformation, stromal layer absorption increases due to angiogenesis while scattering in stroma decreases due to degradation of extracellular matrix. On the contrary, epithelial scattering increases due to hyperplasia, increase in nuclear size and DNA content. DRS can be assembled so as to acquire information from thickness of epithelial layer or penetrate deeper to the stromal layer and acquire tissue information from both epithelial and stromal layers. DRS is especially useful because of attributes such as cost-effectiveness, rapidity and high sensitivity. Several studies have demonstrated utility of DRS in identifying malignant changes in oral and cervical epithelium. Autofluorescence and diffuse reflectance spectra were used to differentiate normal, benign, premalignant and malignant lesions by De Veld et al[158].

Potential of DRS in oral cancer detection using diffuse reflectance spectral ratio R540/R575 of oxygenated hemoglobin bands was shown by N. Subhash et al in 2006[159]. Further studies highlighted the possibility of differentiating normal oral mucosal areas from hyperplastic and dyspastic areas[160]. A multi-spectral imaging camera system that records diffuse reflectance (DR) images of the oral lesion at 545 and 575 nm with white light illumination to scan entire oral lesion in single sitting was developed by the same group[161]. The feasibility of differentiating early oral cancer from normal oral tissue was also shown in hamsters using a diffusing probe in the 500-800 nm spectral range[162]. A portable, easy-to-use and low cost, yet accurate and reliable DRS device that can aid in the screening and diagnosis of oral and cervical cancer has been developed by Yu et al[163].

1.4. Raman Spectroscopy

Raman spectroscopy is a vibrational spectroscopy based on inelastic scattering of light. Inelastic scattering of light, also known as Raman Effect, was discovered by Sir C. V. Raman after seminal experiments on scattering by employing sunlight as the source, telescope as the collector and eye as the detector[164]. This effect was discovered in the year 1928 for which Raman received the Nobel Prize in 1930. When a sample is irradiated with intense monochromatic light, phenomena such as absorption, scattering and reflection occur. Most of the scattered photons have the same frequency of the incident light (Rayleigh scattering), while a small proportion (one in ten million) are inelastically scattered i.e. with a frequency different from the incident photons; this phenomenon is termed as Raman Effect.

When the frequency of the scattered light is lower than the frequency of incident photon, the process is called Stokes shift. If the frequency of scattered photon is higher than incident photon, the process is called Anti-Stokes shift. The scattered photon frequency will be: Stokes ($\nu_0 - \nu_r$) or anti-Stokes ($\nu_0 + \nu_r$), where ν_0 is the incident light frequency and ν_r is the characteristic molecular vibrational frequency. The different types of scattering are illustrated in the simplified energy diagram shown in Figure 1.7. The energy difference between the incident and scattered photon (Raman shift) is represented as wavenumber (cm^{-1}).

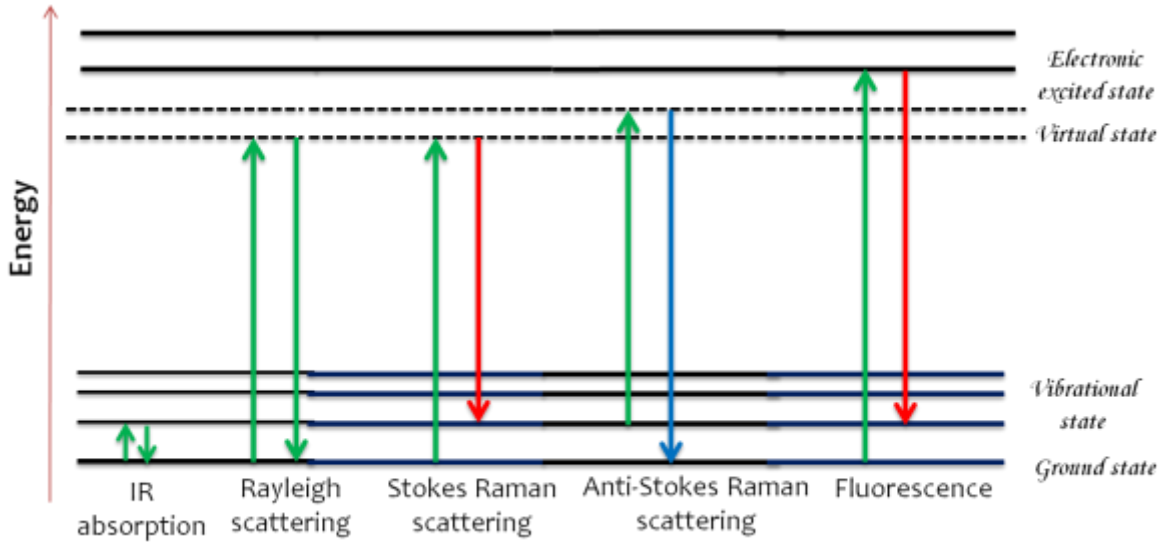


Figure 1.7. Jablonski energy diagram

As mentioned previously, IR spectroscopy is the vibrational spectroscopy approach based on absorption of light. IR and Raman spectroscopy are sister techniques: vibrations strong in IR spectrum (associated with strong dipole moments) are typically weak in RS (dependent on molecular polarizibility). In spite of being information-rich, IR spectroscopic methodologies are less suitable for in vivo and in situ studies because water, the major component of biological

tissues, is highly absorptive in the mid-IR range; sample thickness is another hindrance which can affect applicability of IR-spectroscopy.

In contrast, Raman spectroscopy with features like non-destructiveness, minimal sample preparation needs, rapidity, objectivity and negligible interference with water is more suited for biological applications. The Raman spectra can provide key information about the structure of molecules as the position and intensity of features reflect the molecular structure. Thus, the chemical identity of the sample can be determined. The contribution of molecules in the Raman spectrum is proportional to their relative abundance in the sample, which is the basis for determining the quantity of a sample. Therefore, as no two molecules give exactly the same Raman spectrum, and the intensity of the scattered light is proportional to the amount of material present, RS provides both qualitative and quantitative information regarding the sample allowing spectral interpretation, creation of spectral libraries, sample molecular quantitation and application of chemometrics[165-167].

1.4.1. Theory of Raman Spectroscopy:

Raman scattering can be explained on the basis of classical theory or quantum mechanical theory. The classical wave theory assumes light to be an electromagnetic radiation, which contains an oscillating electric field. This electric field interacts with a molecule through its polarizability. Polarizability (α) may be defined as the ability of the electron cloud of the molecule to undergo deformation. A vibration can give rise to a Raman spectrum only if the polarizability of the molecule changes during vibration. When exposed to electromagnetic

radiation, a dipole moment is induced in the molecule by the external field. A dipole moment can be defined as the magnitude of charge multiplied by the distance between the two charges.

According to the classical theory, Raman scattering can be explained as follows:

The electric field strength *(E)* of the electromagnetic wave (laser beam) fluctuates with time *(t)* as shown by Eq. (1):

$$E = E_0 cos 2\pi \vartheta_0 t,$$

Where E_0 is the vibrational amplitude and v_0 is the frequency of the laser. If a diatomic molecule is irradiated by this light, an electric dipole moment P is induced:

$$P = \alpha E = \alpha E_0 cos 2\pi \vartheta_0 t.$$

Here, α is as proportionality constant and is called *polarizability*. If the molecule is vibrating with a frequency v_m, the nuclear displacement q is written

$$q = q_0 cos 2\pi \vartheta_m t,$$

where q_0 is the vibrational amplitude. For a small amplitude of vibration, α is a linear function of q. Thus, α can be written as

$$\alpha = \alpha_0 + \left(\frac{\partial \alpha}{\partial q}\right)_0 q_0 + \ldots\ldots$$

Here, α_0 is the polarizability at the equilibrium position, and $(d\alpha/dq)_0$ is the rate of change of α with respect to the change in q, evaluated at the equilibrium position.

Combining equations (2), (3) and (4),

$$P = \alpha E_0 cos2\pi\vartheta_0 t$$

$$= \alpha_0 E_0 cos2\pi\vartheta_0 t + \left(\frac{\partial\alpha}{\partial q}\right)_0 qE_0 \cos 2\pi\vartheta_0 t$$

$$= \alpha_0 E_0 cos2\pi\vartheta_0 t + \left(\frac{\partial\alpha}{\partial q}\right)_0 q_0 E_0 \cos 2\pi\vartheta_0 t \, cos2\pi\vartheta_m t$$

$$= \alpha_0 E_0 cos2\pi\vartheta_0 + \frac{1}{2}\left(\frac{\partial\alpha}{\partial q}\right)_0 q_0 E_0 [cos\{2\pi(\vartheta_0 + \vartheta_m)t\} + \cos\{2\pi(\vartheta_0 - \vartheta_m)t\}]$$

According to classical theory, the first term represents an oscillating dipole that radiates light of frequency v_0 (Rayleigh scattering), while the second term corresponds to the Raman scattering of frequency $v_0 + v_m$ (anti-Stokes) and $v_0 — v_m$ (Stokes). If $(d\alpha/dq)_0$ is zero, the vibration is not Raman-active. Therefore, to be Raman active, the rate of change of polarizability (α) with the vibration must not be zero[168].

According to the quantum theory, energy levels of atoms and molecules cannot be continuous but are discrete, that is, vibrational energy cannot be continuously variable, but can have only discrete values. These values are fixed by type of vibration.

When a molecule is irradiated with frequency much higher than the vibrational modes of the molecule, the incident photon and the molecule momentarily form a virtual state. The virtual state is unstable; and photon separate after their momentary interaction, both evolving into new states, which may differ from their initial states.

On transition from virtual state to initial ground state, a scattered photon of the same energy is emitted. This is the Rayleigh scattered photon. The transition from virtual state to an excited state (energy higher than initial state) of molecular vibration results in a scattered photon of energy lower than the incident photon. This is the Stokes Raman scattered photon.

The transition from a virtual state to a state at lower energy than the initial state results in scattered photon having more energy than the incident photon, giving anti-Stokes Raman lines. In the quantum theory, the intensities are proportional to the number of molecules exited to the virtual state. In other words, the intensities are proportional to the population of initial levels, that is, proportional to population distribution according to Maxwell-Boltzman distribution.

$$\frac{anti-Stokes\ intensity}{Stokes\ intensity} = \frac{(\nu+\nu_\nu)^4}{(\nu-\nu_\nu)^4} e^{-(h\nu_\nu/kT)}$$

The exponential term, which results from Maxwell-Boltzman equation, is not predicted by the classical theory, but follows naturally in quantum mechanical treatment of the phenomenon and is in accordance with experimental findings.

1.4.2. Instrumentation:

Raman spectroscopy is an inherently weak process: only 1 in 10 million photons are Raman scattered. Thus, sophisticated instrumentation i.e. powerful excitation source, high throughput spectrograph and sensitive detection systems are a pre-requisite. The introduction of low-noise charged coupled device (CCD) detector technology; highly efficient imaging spectrographs and compact semiconductor laser excitation sources enabled extensive Raman spectroscopic applications in diverse areas. Raman instrumentation can be of two main types: Interferometric

and Dispersive Raman spectrometers. To reduce fluorescence, near-infra red lasers are utilized. As scattering efficiency decreases with increase in laser wavelength, excitation of 1064 nm requires interferometer coupled with detector. An interferometer converts the Raman signal into an interferogram, permitting the detector to collect the entire Raman spectrum simultaneously. Dispersive Raman spectroscopy with CCDs coupled to grating-spectrographs, are more commonly employed.

Typically, a dispersive Raman spectrometer is made up of (i) excitation source, (ii) optical system, (iii) spectrograph and (iv) detection system. A schematic of major components of Raman spectrometer is shown in Figure 1.8.

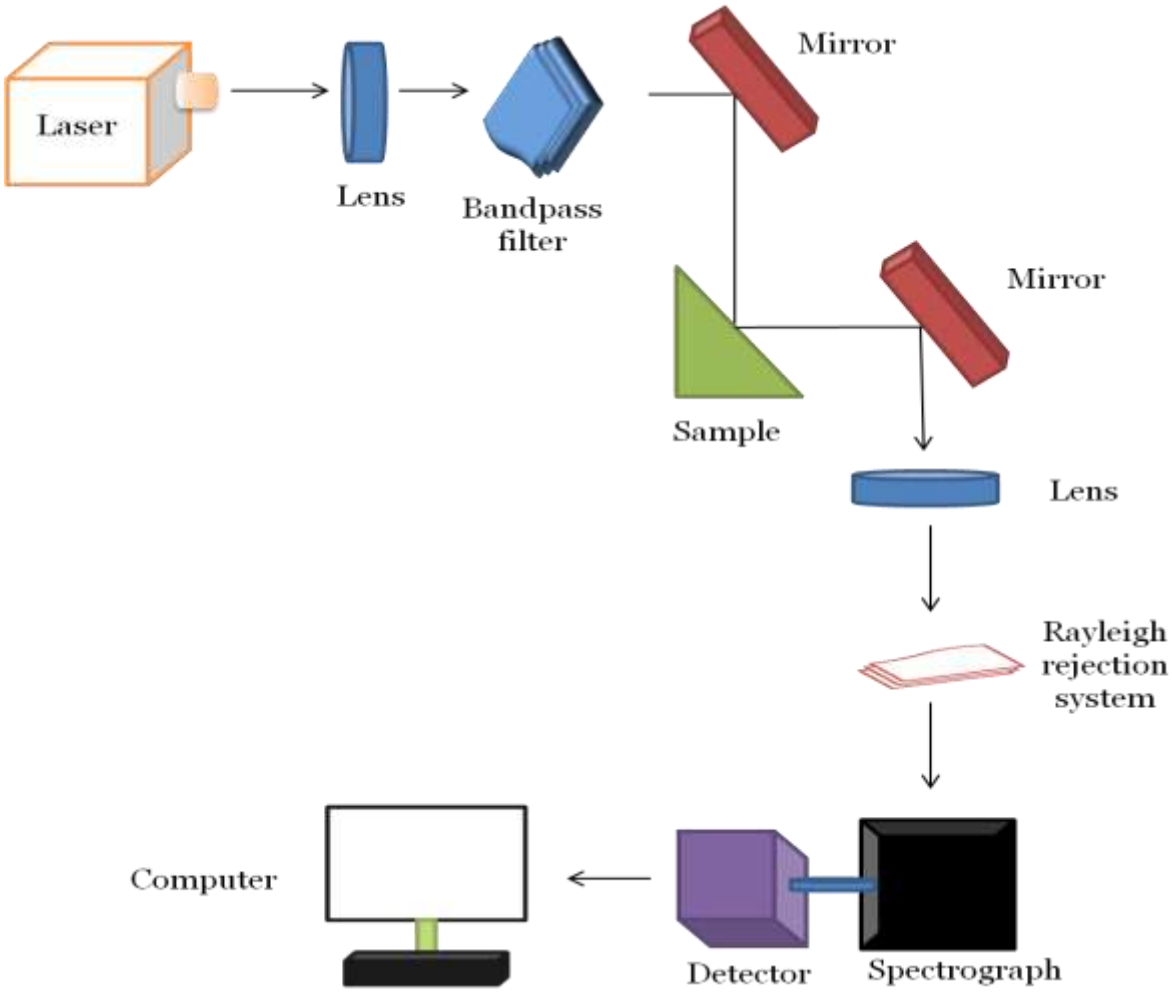


Figure 1.8. Schematic representation of Raman instrument system

a) Excitation source:

As Raman spectroscopy measures a shift in wavelength (or frequency), it is imperative that a monochromatic excitation source be employed. Prior to the advent of lasers, the major excitation source for Raman spectroscopy was typically a mercury lamp, fitted with a filter to transmit a single/limited wavelength range of light. After the introduction of lasers, a continuous-wave laser is usually the most commonly employed excitation source for Raman instruments. Early gas

lasers emitted in the visible region. Continuous-wave (CW) lasers such as Ar+ (351.1-514.5 nm), Kr+ (337.4-676.4 nm), and He-Ne (632.8 nm) are commonly used for Raman spectroscopy. More recently, pulsed lasers such as Nd: YAG (neodymium-doped yttrium aluminium garnet) and diode lasers are being commonly employed.

b) Optical systems

Being an inherently weak phenomenon, Raman spectroscopy requires proper focusing of laser beam onto the sample as well as proper collection of the scattered photons for higher efficiency. Excitation and collection can be achieved by combinations of lenses through optical configurations like 90 and 180 degree scattering geometry. A collecting lens and a focusing lens are used as part of collection optics, to enable efficient collection of the inherently weak Raman scattering. As lenses play a very important role in both excitation and collection to obtain maximum light throughput, F number should be in accordance with the wavelength selector.

Mechanisms to ensure excitation of sample using single wavelength of light while collection of only Raman scattered radiation and filtering out elastically scattered Rayleigh signal are also employed. Major examples include monochromators and interference filters. The advent of optical filters have eliminated the need for cumbersome, low throughput, multistage scanning spectrometers, and enabled rapid spectral data acquisition. Different categories of filters include-

Long-pass filters: A long-pass (LP) filter is an optical interference or coloured glass filter that attenuates shorter wavelengths and transmits longer wavelengths over the active range of the target spectrum. 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.

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

Band-pass filters: Band-pass filters only transmit a range of wavelength band, and block all other wavelengths. Such a filter can be made by combining a LP- and a SP filter.

Notch filters: Holographic notch filters have a sharp, discrete absorption due to which high laser attenuation in a very narrow bandwidth is obtained. These filters enable measurements to be made for both the Stokes and anti-Stokes Raman scattering.

c) Spectrograph

The main function of the spectrograph is to disperse the light 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 (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. The function of grating 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. Diffraction gratings are majorly used to generate wavelength dependent interference patterns in spectrographs. They provide better resolution and dispersion as compared to prism based instruments. Gratings can be transmission or reflection based. The traditional method of grating manufacture involves mechanically ruled gratings which contain parallel

grooves inscribed on an optically polished surface. Modern instruments contain holographic gratings which are made by recording groove patterns on a photosensitive surface.

d) Detection systems

The inherently weak Raman signals need sensitive detection techniques. Various detection devices such as photon counting, photodiode array and CCDs may be used. In recent times, CCDs are the most commonly employed detectors for Raman spectroscopy. A CCD is a silicon-based semiconductor, arranged as an array of photosensitive elements with each capable of generating photoelectrons and storing them as a small charge. The charge stored on each pixel is a function of number of photons hitting that pixel. 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. One of the most typical dimensions of a CCD is 1024 × 256 pixels. 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.

1.4.3. Variants of Raman spectroscopy:

As already mentioned, Raman spectroscopy is an inherently weak process- 1 in 10 million photons are Raman scattered. To enhance the overall efficiency and increase the range of applications, several modifications to the conventional Raman spectroscopic methodology have been introduced. Raman microscope refers to an assembly of spectrometer coupled to a

microscope. Raman microscopes are extremely useful for micro-analysis of samples because of smaller spot-size; these facilitate bio-fluids and cell-based experiments. Remote measurements such as in vivo biomedical and in situ bioreactor-based applications are enabled by use of fiber-optic probe based Raman spectroscopy. Fiber-optic probes are employed to carry the excitation light to sample and scattered signals to detector.

Some important examples of Raman spectroscopy variants for signal enhancement include surface-enhanced Raman spectroscopy (SERS), drop coating deposition Raman (DCDR) spectroscopy, resonance Raman spectroscopy (RRS), coherent anti-Stokes Raman spectroscopy (CARS), stimulated Raman spectroscopy (SRS), spatially offset Raman spectroscopy (SORS) and surface-enhanced spatially offset Raman spectroscopy (SESORS). For SERS, sample needs to be in proximity or adsorbed onto a roughened metal surface, a colloidal solution or a roughened electrode (usually Ag or Au). Enhancements in the range of 10^3 – 10^9 are usually observed, which can enable single molecule level detection[169,170]. RRS is an approach where marked enhancement of Raman signal is observed when excitation wavelength matches or is close to electronic transition state. In case of DCDR, a drop of sample is dried on the slide and spectra are subsequently acquired from the dried specimen, it is adopted for dilute samples like bio-fluids[171-173]. CARS is a multi-photon form of Raman spectroscopy based on a non-linear conversion of two lasers into a coherent high intensity beam in the anti-Stokes region. The resulting emission is usually many orders of magnitude greater than spontaneous Raman scattering, this approach is highly useful for obtaining spectra of fluorescing samples[174,175]. Like CARS, SRS is another form of coherent Raman spectroscopy bestowed with high signal intensity and inherent immunity from fluorescent background. In SRS, two incident fields coherently

excite the sample, and higher signal intensity is obtained when the difference frequency matches the Raman molecular vibration. SRS has been extensively employed for imaging biomolecules and cells, including in vivo imaging[176-178]. SORS involves retrieval of the diffused scattered light that reaches deeper layers based on their spatial properties[179]. The diffused scattered light can collect biochemical information from in-depth regions and enables non-invasive investigation of buried regions in a sample. SESORS refers to collection of SERS signals from tissue depths[180,181].

1.4.4. Data analysis

The extraction of useful information from Raman spectra of complex systems like biological samples may be difficult as they are characterized by parasitic fluorescence and poor signal to noise ratios. Moreover, the differences between normal and pathological conditions may be subtle and difficult to observe in raw spectra. Therefore, careful processing and analysis of the data is crucial to obtain meaningful information.

1.4.4.1. Spectral pre-processing

The first step in analysis of the spectral data- preprocessing removes noise and normalizes the Raman spectra. Preprocessing eliminates noise (unwanted signals) and enhances requisite signals which may have discriminating features. The major pre-processing steps are correction of instrument response, removal of noise from optical elements and environment, and normalization.

CCD response correction is used to obtain Raman spectra free from instrument-induced spectral artifacts. The relative intensity correction for instrument response may be carried out by using either a white light source or a doped-glass fluorescence source such as standard reference material (SRM) manufactured by the National Institute of Standards and Technology (NIST)[182]. A unit of SRM 2241 consists of an optical glass that emits a broadband luminescence spectrum when excited at this laser wavelength. The shape of the luminescence spectrum of this glass relates the relative spectral intensity to the wavenumber (cm^{-1}). Background signals and noise can be removed using various methods. Cosmic spikes and high-frequency noise which are artifacts arising mainly from the detection electronics can be removed by median filtering and wavelet-based denoising[183,184, 185]. Low-frequency noise arises from the ambient light and fluorescence emission from sample. Baseline subtraction methods such as polynomial fittings[186,187], first or second derivative filters[188,189], wavelet transformations[190,191], the shifted-spectra technique[188], and the application of FT frequency filters[188,192] are commonly employed for removal of background signals. An advantage of polynomial fitting over frequency or derivative filtering is that traditional Raman line shapes are preserved, facilitating spectra interpretations. However, the first and second derivative methods[168,193,194] have the advantage of seizing the successive intensity differences of the Raman spectrum (first derivative), or the successive rates of change of the differences in the spectrum (second derivative). An additional advantage of derivatives is that the baseline is removed without introducing bias or human error[195]. Spectra can be interpolated in specific regions for enhanced recognition of subtle spectral differences. Finger-print region and high wavenumber region are the most commonly employed spectral regions for analysis. Raman spectra acquired from the same samples at

different time points may exhibit a large amount of variance due to several factors. These variations can be removed by normalization of the spectra. Normalization can be carried out using several ways; highest peak normalization and vector normalization are the two most commonly employed methods[196,197]. Smoothing of Raman spectra may also be carried out for removal of high-frequency noise; Savitzky-Golay method is one of the most commonly employed methods for smoothing[194].

1.4.4.2. Data classification methods

One of the most important advantages of using optical spectroscopic techniques for disease diagnosis is the fact that the acquired data are amenable to statistical analytical methods. The applications of these tools facilitate the computation of mathematical parameters that help in unbiased classification of healthy and disease. Spectroscopic data can be analyzed using two approaches- univariate or multivariate. The data analytical methods that deal with only one variable at a time are called univariate methods. Analysis such as first and second order derivatives, curve fitting, difference spectra, band intensity/area under the curve ratios facilitate the visualization of band shifts, broadening and change in intensities were the earliest approaches for data classification. But, during spectroscopic measurements, the spectrum generated consists of multiple variables, i.e. wavenumbers. For extraction of maximum information from the spectrum to achieve better characterization of the sample, several important wavenumbers are analyzed at a time. Multiple variables are simultaneously analyzed by subjecting them to multivariate analysis. These statistical analysis techniques help in identifying relationship between the variables. If each variable constitutes a single dimension, data with ‘n’ variables

may have up to 'n' dimensions. Multivariate tools reduce these dimensions by selecting few but important dimensions. Further, patterns in the data can be modeled and used for independent prediction of new data. Unsupervised and supervised multivariate mathematical tools like PCA, hierarchical cluster analysis- HCA, linear discriminant analysis- LDA, PC-LDA, artificial neural network- ANN, Genetic algorithms, and partial least square algorithm-PLS can be used for data analysis.

Unsupervised methods

Unsupervised methods do not require a prior knowledge about the samples. PCA is the most commonly employed unsupervised method. PCA functions on dimensionality reduction and is used to observe trends in the data. The main function of PCA is to identify trends, pattern and outliers in the data set. PCA uses a linear transformation to represent the original data into a new coordinate space, as shown in Figure 1.9.

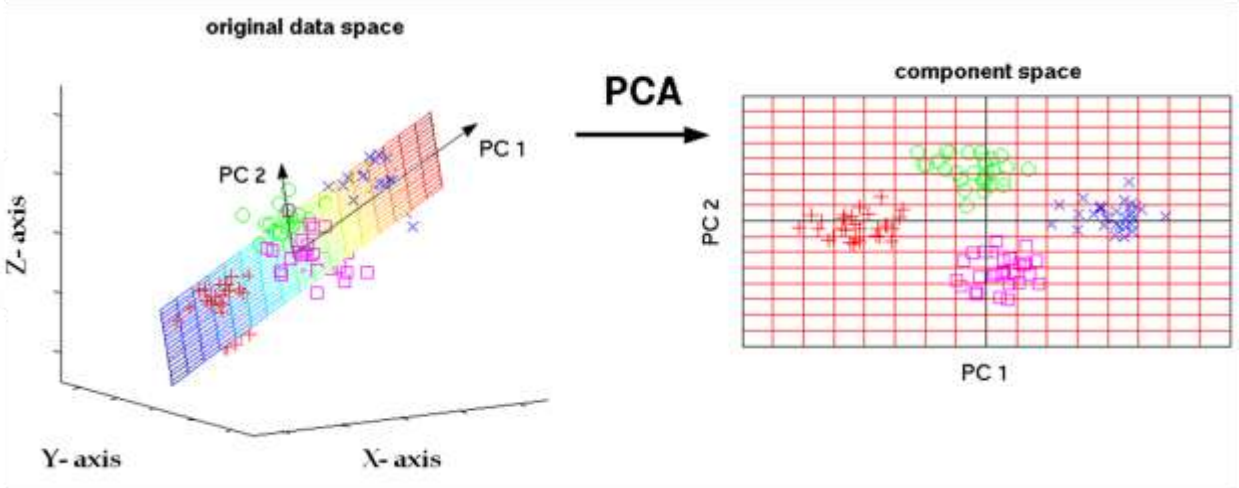


Figure 1.9. PCA transformation (Modified from Ph. D. thesis, Matthias Scholz, Approaches to analyze and interpret biological profile data, University of Potsdam, Germany, 2006).

The dimensions of this new space are in the directions of maximum variance of the features. The first principal component captures the dimension of maximum variance; the second principal component captures the dimension of the second greatest variance, and, similarly, the subsequent principal components capture the dimensions of successively decreasing variances. It is common to choose only a subset of these principal components as features. Usually, the first few principal components are used; the reduced number of features might capture the differences between classes more effectively, thus increasing the classification accuracy. HCA, a clustering method, is another commonly employed unsupervised classification method. The results of HCA are easy to interpret and can be easily visualized as a dendrogram. While performing HCA, the number of clusters to be assumed using analysis is a major matter of concern.

Supervised methods

Supervised methods require a set of labeled data for analysis. These labeled data points are first used to train the algorithm using known information about the data. After subsequent training to the algorithm and building of standard models, data points not included in the data set for modeling can be classified by assigning one of the classes in the algorithm. LDA is a discrimination method based on determination of linear discriminant functions, which maximize the ratio of between-class variance and minimize the ratio of within-class variance. In LDA, classes are supposed to follow a multivariate normal distribution and be linearly separated. LDA, like PCA, can be considered a feature reduction method in the sense that both, LDA and PCA, determine a smaller dimension hyper-plane on which the points will be projected from the higher dimension. However, whereas PCA selects a direction that retains maximal structure among the

data in a lower dimension, LDA selects a direction that achieves maximum separation among the given classes. The number of variables used for analysis should be less than the number of samples. In LDA, discrimination boundaries have to be linear. This particular constraint can be reduced by using quadratic discriminant analysis (QDA), factorial discriminant analysis (FDA) or partial-least-square discriminant analysis (PLS-DA). These discrimination techniques are specialized for data points which are well separated, and on features which follow a normal distribution. If there is a correlation between the different variables, the function may not be able to bring out the expected separation. To avoid this, transformation of the features using PCA or Partial Least squares (PLS) may be carried out first. Other supervised methods include support vector machine (SVM) - a binary classifier, ANNs- computational models inspired by the functionality of the central nervous system, and genetic algorithms- a class of evolutionary algorithms which mimic the principles of natural selection for classification and optimization.

For supervised methods like LDA, cross validation of model is important to avoid over fitting or under fitting. This involves dividing the original data set into a training and a validation set, where the training set is used for learning the parameters and the validation set is used for evaluating the performance of the classification technique. Leave one out cross validation (LOOCV) is the commonly employed validation method. LOOCV involves using a single observation from the original sample as the validation data, and the remaining observations as training data. This is repeated such that each observation in the sample is used once as the validation data and averaged over the rounds[198-207].

1.4.5. Raman applications:

Due to attributes like sensitivity, high information content, and non-destructive nature, RS has been extensively applied in fields of chemistry, biology, geology, pharmacology, forensics, pharmaceuticals and material sciences. Some examples include applications in analytical chemistry for process monitoring; detection of pollutants in environmental science; forensics applications that include detection of gunpowder residues, blood, semen; identification of hazardous chemicals in security; industries like pharmaceutical, petrochemical, semiconductor and in research areas -geology, nuclear science, material science, art and archaeology[208-210]. Although Raman spectroscopy had been widely used for chemical and molecular analysis for many years, its application to biomedical problems is relatively recent, facilitated by advancing technology in areas of lasers, detectors and data analysis methods. Thus, diverse potential biological and biomedical applications such as bacterial identification, chemical hazards and illicit substance detection, transdermal quantitation of analytes in blood, as well as food and product authentication have been possible[211-214]. Further, as normal to disease transition of a tissue is accompanied by several biochemical changes, identification of diseased conditions using Raman spectroscopy is a promising approach. Some notable disease related ex vivo applications include study of atherosclerotic plaques, bladder dysfunction, Alzheimer disease, Parkinson's disease, rheumatoid arthritis and osteogenesis imperfecta along with several clinico-oncological applications[215-218]. RS has shown potential in ex vivo detection of potentially pre-cancerous tissues and real-time tissue evaluation with a high degree of spatial resolution[219-221].

Specific applications of Raman spectroscopy include-

1.4.5.1.Serum Raman spectroscopy: Raman spectroscopy has been employed for disease diagnosis of both non-cancer diseases and cancer. Some important studies in disease diagnosis are as follows:

Serum for disease diagnosis (non-cancer)

Serum Raman spectroscopy has been investigated for detection of Pre-eclampsia, which is a pregnancy-related disorder characterized by high blood pressure, proteinuria and associated with organ dysfunction, multiple maternal and adverse fetal effects. In the study carried out by Basar et al, Raman spectra of serum samples from pre-eclamptic and healthy pregnant women were measured, and analyzed using band component analysis and PC-LDA. Band component analysis indicated decreased intensity of 11 Raman bands and increased intensity of 8 Raman bands in the pre-eclamptic samples- especially in lipid-related bands. PC-LDA exhibits a clear separation between the pre-eclamptic and control groups with 78% specificity and 90% sensitivity [222]. It is critical to distinguish between sepsis and systemic immune response syndrome (SIRS) which are life threatening conditions that occur as a major complication in patients hospitalized in an Intensive Care Unit (ICU). In a study by Neugebauer et al, plasma samples obtained from 70 patients from an ICU were analyzed: 39 SIRS samples and 31 sepsis samples. Mean spectra comparisons indicate a highly protein dominated spectrum with bands corresponding to amide I, amide III, phenylalanine, tyrosine, tryptophan, histidine, arginine. It was concluded that the differences could arise from varying protein concentrations (such as cytokines and chemokines) Overall sensitivity and specificity of 0.82 and 1.00 for distinguishing sepsis and SIRS was obtained[223]. Hepatitis C infection is a predominant liver disease caused by the hepatitis C virus

which can manifest as a serious, lifelong illness. In the study carried out by Saade et al[224], the most significant differences among normal and hepatitis C spectra were observed for phenylalanine, lipids and phospholipids. A discriminant line based on mean Mahalanobis distance among both groups was drawn: 15 out of 17 normal and 11 out of 12 hepatitis C sera were correctly classified yielding sensitivity of 88% and the specificity of 92%. Diagnosis of neurological disorders like Alzheimer's and Parkinson's disease has also been investigated using serum Raman spectroscopy. Alzheimer's disease is a neurological disorder characterized by dementia that also causes problems with memory, thinking and behavior and is the most widespread neurodegeneration-induced dementia in the elderly population worldwide. A study by Schipper et al[225] in 2008 explored near-infra red spectroscopy and RS to detect and quantify oxidative substrate modifications in blood plasma samples from Parkinson's disease and normal elderly control subjects. Another study by P. Carmona in 2013[226] was carried out to explore if Raman and infrared spectroscopy of blood plasma can differentiate Alzheimer's disease (AD) from normal aging of healthy controls. AD could be differentiated from normal aging of elderly control persons with a sensitivity of 89% and specificity of 92%. The specificity increases to 100% for the detection of mild AD. In the study published by E. Ryzhikova et al[227], three groups were analyzed: 20 AD (Alzheimer's disease), 18 other dementia (OD) and 10 age-and sex-matched healthy control subjects without any neurological disorder. ANN and genetic algorithms yielded 95% accuracy to differentiate AD patients from OD and healthy elderly volunteers. Globally, dengue fever is the most important arboviral disease. Two RS studies have shown the possibility of dengue detection using blood. In the first study by Rehman et al[228], RS was used to diagnose dengue infected patients from serum and whole blood after excitation at 442 nm and

532 nm. In the consequent study by M. Saleem et al[229], the prediction of dengue virus infection using partial least-squares (PLS) regression was explored and 100% accuracy was obtained. Sickle cell anemia (SCA) is a type of anemia which is characterized by a defect in haemoglobin: the oxygen carrying molecule in red blood cells. In a study carried out by Filho et al, spectral differences between hemoglobin S (HbS) and hemoglobin A (HbA) were evaluated[230]. In a previous study by Wood et al.[231], resonance Raman spectra of oxygenated and deoxygenated functional erythrocytes suggested that this approach can be as an analytical and diagnostic tool for a variety of erythrocyte disorders. In another study[232], Raman spectral differences in the shift positions of 882 and 1,373 cm^{-1} in the HbS and in the positions of 1,547 and 1,622 cm^{-1} in the HbA spectra, which correspond to valine and glutamate, respectively were observed. The discrimination was evaluated based on PCA and Mahalanobis distance and 100 % accuracy was obtained. Serum based diagnosis of asthma- an airways disease characterized by variable airflow obstruction was carried out by Sahu et al[233] using 44 asthma subjects of different grades. Differences like change in protein structure, increase in DNA specific bands and increased glycosaminoglycans-like features were more prominent with increase in asthma severity. Grade-wise classification and treatment response could be determined.

Serum for cancer diagnosis

Diagnosis of several cancers has been investigated using serum Raman spectroscopy. One of the first studies on cancer diagnosis was carried out by Li et al[234]. The next report on serum Raman spectroscopy was reported for breast cancers. Breast cancer is one of the major cancers affecting women worldwide. It is known that early diagnosis leads to improved prognosis and disease free

survival rates. In this study by Pichardo et al[24], ratio of band intensities was analyzed; seven band ratios were found significant which corresponded to proteins, phospholipids, and polysaccharides. LDA could differentiate between the groups. Another study by Bhattacharjee et al[235] in an animal model of breast cancer has shown that serum from tumor-bearing Sprague-Dawley rats is different before and after surgical excision, and both groups are also different from controls. Colorectal cancer has one of the lowest survival rates. Gold nanoparticles based SERS for non-invasive colorectal cancer detection was explored by D. Lin et al.[29]. SERS measurements were performed on patients and healthy volunteers and tentative assignments suggested cancer specific biomolecular changes such as increase in the relative amounts of nucleic acid, a decrease in the percentage of saccharide and protein content in cancer serum. PC-LDA yielded a diagnostic sensitivity of 97.4% and specificity of 100%. Another study by Li et al explored classification between normal and colorectal cancer serum using selected parameters and regression-discriminant analysis. LDA was subsequently applied to the PCs and a diagnostic accuracy of 88% and 83% was obtained[236]. Pancreatic cancer (PC) is also one of the most lethal malignancies with a 5-year survival rate of only 6%. G. Wang et al[237] carried out a study on detection of biomarker MUC4 for pancreatic cancer detection using SERS. Higher SERS response for MUC4 was observed in cancer patients. SERS assay outperformed conventional assays with respect to limits of detection, readout time, and required sample volume. Head and neck cancers are the sixth most common cancer worldwide. In a study by Harris et al [25], potential of a peripheral blood sample to achieve diagnosis of head and neck cancer was evaluated using 20 head and neck cancer patients and 20 patients with respiratory diseases. Conventional LDA approaches showed an accuracy of around 65% while genetic evolutionary algorithm obtained an

accuracy of 83%. S. Feng et al[27] carried out a study for SERS-based specific identification of nasopharyngeal cancers. Silver nanoparticles (Ag NP) as the SERS-active nanostructures were directly mixed with blood plasma to enhance the Raman scattering signals. SERS measurements were performed on patients and healthy volunteers. Tentative assignments in the SERS spectra suggest interesting cancer specific biomolecular differences, including an increase in the relative amounts of nucleic acid, collagen, phospholipids and phenylalanine and a decrease in the percentage of amino acids and saccharide contents in nasopharyngeal cancer patients. PC-LDA yielded high sensitivity (90.7%) and specificity (100%). Cervical cancer is one of the most common cancers in females in the developing countries. Raman spectroscopy of cervical cancer, pre-cancer, and healthy volunteer serum samples was investigated. Main molecular differences among the control and cervical cancer samples were glutathione, tryptophan, β carotene, and amide III as also observed in PCA[26]. Ovarian cancer is the second most common gynecological malignancy in females. In the study carried out by Owens et al[28], ATR (attenuated total reflectance)-FTIR and RS were simultaneously explored. Mean spectra comparisons indicate shifts in DNA, phosphate bond, cytosine, thymine and protein, lipids. PC-LDA could differentiate normal and ovarian samples using 6 factors. SVM yielded a classification efficiency of 74% for Raman while an efficiency of 93% for ATR-FTIR spectroscopy.

1.4.5.2. Exfoliated-cell Raman spectroscopy for disease diagnosis

Exfoliated cells from cervix and oral have been investigated using Raman and Infra-red micro-spectroscopy. While IR micro-spectroscopy has been explored for diagnostic applications, Raman mapping experiments were carried out to understand cellular organization using label-

free methods. This is because IR has lower spatial resolution and much higher speed in contrast to the very high spatial resolution of Raman imaging. Raman image of exfoliated human squamous epithelial buccal cell was obtained by raster scanning the laser beam focused to a spot of about 300 nm diameter over the cell and collecting an entire Raman spectrum from each spot. Unsupervised HCA was used to convert the hyper-spectral dataset into a pseudo-color image. Findings indicate distinct separation of cytoplasm spectra from nucleus spectra. Cytoplasmic spectra demonstrated superposition of protein and phospholipid at certain spots, which could be attributed to lipid droplets, Golgi apparatus or vacuoles. This study indicated the potential of Raman imaging in studying intracellular transport and exchange[238]. Raman micro-spectroscopy for diagnostic applications was eventually explored on cervical exfoliated cells. In a study by Vargis et al, near infra-red micro-spectroscopy of human exfoliated cervical cells could detect HPV infection in patients. As per existing literature, HPV infected cells show increased nucleic acid levels, moderate to low change in protein levels and decreased lipids. A similar spectral profile was observed for HPV infected cells[239]. In another study carried out by Rubina et al, 98 cervical exfoliated cell specimens were obtained from normal and cervical cancer subjects. After initial influence of blood contamination on PC-LDA findings, samples were analyzed after RBC (red blood cells) lysis treatment to remove blood influence. The absence of heme and fibrin bands confirmed the effective removal of blood from the samples. Post RBC lysis, PC-LDA resulted in ~80% classification efficiency[34].

1.4.5.1. In vivo Raman spectroscopy for disease diagnosis

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 vivo Raman studies were undertaken for intraoperative tumor margin assessment in 9 patients undergoing partial mastectomy procedure in 2006[240]. 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 have demonstrated feasibility of transcutaneous spectral acquisition from the DMBA (7,12-dimethybenzanthracene) -induced tumors and their margins in Sprague-Dawley rats[241]. Due to the obvious accessibility to skin, in vivo Raman spectroscopic analysis is a very practical, rapid and 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 SCC with efficiency of 95% was achieved[242]. Application of Raman spectroscopy in brain cancer diagnosis is widely reported. In vivo studies to this extent have been conducted in animal models by Kirsch et al[243] and Beljebbar et al[244]. Several studies for gastric and esophageal cancer detection that coupled Raman spectroscopy with endoscopy and white light imaging have been carried out[245-249,250] . Studies for developing a miniature confocal fiber-optic Raman probe useful during endoscopy-guided biopsies for better lesion identification have also been undertaken[251]. To enable in vivo applications in lung cancer diagnosis, several groups are focusing on developing an optimized endoscopic probe for bronchoscopic diagnosis[252]. 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[253-256]. Classification between normal, cancerous and different pathological grades along with subtle variations due to hormonal effects and Vagifem treatment have also been studied[257], [258] In another study, Duraipandian explored the combined Raman /Autofluorescence spectroscopy modality for improving early diagnosis of cervical precancer in vivo at colposcopy. Better diagnostic accuracy was obtained using confocal Raman spectroscopy[259]. Buckley et al demonstarted in 2014 that SORS can noninvasively detect a known compositional, genetic bone abnormality, osteogenesis imperfecta[260]. A study conducted by Rubina et al[261] in 2014 confirmed utility of in vivo Raman spectroscopy in detection of cervical cancers and proposed use of vagina as a better internal control. Recent studies conducted by Bergholt et al have characterized the in vivo Raman spectroscopic properties of normal colorectal tissues and assess distinctive biochemical variations at anatomical sites following which real-time in vivo diagnosis of adenomatous polyps was explored[262,263]. Significant

progress in detection of brain cancer in vivo using Raman spectroscopy has been made. Studies by Jerymn et al[264] first demonstrated intra-operative acquisition of Raman spectra from normal and cancerous areas of the brain. In the next step, the same group attempted to design a Raman spectroscopy system to maximize the volume of resected cancer tissue during glioma surgery. Findings indicate successful distinguishing of necrosis from tumor and normal brain tissue with an accuracy of 87%[265].

1.4.6. Raman spectroscopy in oral cancers:

The first Raman spectroscopic applications in oral cancers were investigated by Schut et al in 2000. This group explored the in vivo classification of normal and dysplastic tissue in a rat model. Raman spectra of normal and dysplastic tissues were obtained with 100 s integration time and sensitivity and specificity of 100% for detecting high-grade dysplasia/ CIS were observed[266]. This animal study was followed by a study on human oral cancers biopsies by Venkatakrishna et al in 2001. One hundred forty spectra from 49 biopsies were acquired using 785 nm excitation and SpeX Triax 320 spectrometer. Using PCA based multivariate analysis, sensitivity and specificity of 85% and 90% was observed[267]. A subsequent study by the same group in 2004 (Krishna et al) demonstrated the suitability of formalin-fixed tissues for Raman spectroscopy[268]. Malini et al. in the year 2006 carried out an extended study in discriminating normal, cancerous, precancerous and inflammatory conditions using 785 nm excitation and 30 s integration time. Lipid rich features in normal conditions while predominant protein features were observed in all pathological conditions, including tumors. Classification between different groups was explored using PCA coupled with multi-parametric 'limit test' and high sensitivity and specificity was achieved[19]. In the same year, Oliveira et al explored the feasibility of differentiating normal, dysplastic and SCC tissues in a hamster model of oral carcinogenesis. FT-Raman spectroscopy and 1064-nm Nd:YAG laser line demonstrated that DMBA induced oral cancer in the buccal pouch of hamsters can be identified distinctly. Major differences between normal and malignant spectra were attributed to composition, conformational, and structural changes of proteins, and possible increase in their content in the malignant epithelia and 91% and 69% sensitivity and specificity were observed[269]. Hu et al acquired spectra of 66 human oral mucosa tissues (43

normal and 23 malignant) using confocal Raman micro-spectroscopy in 2008. PCA along with calculation of areas under 1004, 1156, 1360 1587 and 1660 cm^{-1} band were used as a classification method after preprocessing spectra using wavelet-based analysis[270]. Another study by Sunder et al in 2011 was carried out to evaluate the applicability of near infrared (NIR) Raman spectroscopy in differentiating normal epithelium and different grades of oral cancer. Findings 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, 1440 and 1450 cm^{-1} [271].

These successful ex vivo studies paved the way for human in vivo oral cancer studies. The first in vivo Raman spectroscopic study was for identifying site wise variations in the human oral cavity and reported by Guze et al in 2009. In this study, reproducibility of Raman spectroscopic signature of normal oral mucosa among anatomic oral sites and among subjects of different races and gender was evaluated. 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 were acquired with 1 s acquisition time. A fiber-optic probe of 1.8 mm diameter and 0.75 m length consisting of 200 μm excitation fiber surrounded by 27 ultralow 100 μm collection fibers 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 and gender does not have any influence on spectra; also that different oral cavity sites can be discriminated based on degree of keratinization[20]. A similar study was conducted by Bergholt et al. in the year 2011 where they 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}). PLS-DA were employed to assess the inter-anatomical variability. Findings suggest that anatomical characteristics of different sites have influence on the in vivo Raman spectra and these sites can be grouped together based on anatomical and spectral similarity to develop diagnostic algorithms[272]. Singh et al reported in 2012 the first in vivo spectral acquisition from oral cancer patients in clinically implementable time[273,23]. Subsequently, studies to detect pre-malignant conditions were carried out. Results suggest that premalignant conditions can be objectively discriminated against normal and tumor conditions[21]. In an another study, the origin of Raman signals in tissues was also investigated by the same group[274]. Singh et al conducted studies to detect malignancy-associated-changes (MAC) or cancer-field-effects (CFE) in oral cancer patients[275]. Studies by Krishna et al investigated spectral differences in different oral cavity subsites. Their findings suggested clustering of sub-sites into 4 major anatomical clusters- a) outer lip, and lip vermillion, b) buccal mucosa c) hard palate d) dorsal, lateral and ventral tongue and soft palate. Further, the authors also suggest the use of anatomy-matched algorithms to increase discrimination between healthy and abnormal conditions[22,276]. Singh et al also conducted a study to correlate spectral and biochemical markers[277].

Raman imaging to understand biochemical variations in normal and oral malignant buccal mucosa was explored by Behl et al. Epithelium and stromal regions of normal cells, and cellular components of normal and tumor sections could be classified by PCA[278]. Potential of Raman spectroscopy in surgical demarcation was investigated by two recent studies. In the first study by Barroso et al, differential water content in malignant and surrounding normal tissue was used as a basis for identifying surgical margins using Raman bands of OH- and CH- stretching vibrations

in high wavenumber region. The water content in squamous cell carcinoma were significantly higher than surrounding healthy tissue. Thus, tumor tissue could be detected with a sensitivity of 99% and a specificity of 92% after using a cutoff water content value of 69%. In another study by Cals et al[279], Raman imaging of normal and tissue sections from 10 oral cancer patients was carried out and 127 pseudo-color Raman images were generated. These images were linked to the histopathological evaluation of same sections and spectra were annotated based on histopathological findings. LDA was used to build models for tumor and surrounding healthy tissue. Raman spectroscopy successfully differentiated between tumor and surrounding healthy tissues. Another study by Carvalho et al explored rapid screening of OSCC using cell lines. Raman spectra were acquired from nuclei, nucleoli and cytoplasm of epithelial cancer, dysplastic and normal epithelial cell lines. PCA could differentiate between normal and cancer cell line. Spectra from cytoplasmic regions could differentiate dysplastic and cancer cell lines. Spectral features demonstrate increased DNA features from nuclear and nucleolar regions while higher lipids from cytoplasmic regions[280]. A similar study was performed in high wavenumber region by the same group. This approach yielded more efficient discrimination than the fingerprint region in cell cultures[281]. A recent study by Kumar et al have shown sequential cancer progression in hamster buccal pouch (HBP) model over fourteen week period of carcinogenesis both in vivo and ex vivo, complemented by histopathological analyses[282]. Spectral changes showed lipid dominance in early stages while later stages and tumors showed increased protein to lipid ratio, nucleic acids, PCA showed progressive increase in classification with increase in DMBA treatment; further, changes in control HBP which suggests that RS may be able to identify micro-heterogeneity which may often go unnoticed in conventional biochemical assays..

Apart from several ex vivo and in vivo findings, studies employing minimal invasive or non-invasive approaches like blood, saliva and urine have also been explored for oral cancer diagnosis. SERS studies on 10 saliva samples were carried out by Kho et al in 2005. After centrifugal treatment to saliva for removal of debris and unwanted particles, saliva was deposited on gold particle films and spectra were acquired with He-Ne laser excitation. Additional peaks at 1097 cm^{-1} and 1627 cm^{-1}, as compared to conventional Raman, were observed in cancer saliva samples[283]. Blood plasma was explored for oral cancer diagnosis by Rekha et al in 2013 using 28 plasma samples. Spectra were acquired from Labram HR 800 in the spectral range 800-1800 cm^{-1}. Marked spectral differences were observed between cancer and normal plasma corresponding to protein, amino acid, glucose and lipids. Successful classification between the groups was observed using PCA and LDA[284]. A recent study for oral cancer diagnosis using urine was carried out by Elumalai et al. where characterization of the metabolites of human urine of normal subjects and oral cancer patients in the finger print region (500–1800 cm^{-1}) was explored. PC-LDA findings yield sensitivity and specificity of 98.6% and 87.1%, respectively to discriminate healthy and cancer patients[285].

1.5. Aim and Objectives:

Oral cancers are the sixteenth most common cancers in the world. Low disease-free survival rates associated with oral cancers rates are attributed mainly to delays in diagnosis and recurrence. Early diagnosis of oral cancers can improve overall survival rates. Due to limitations of existing methodologies and absence of an established clinical marker, several adjuncts for early diagnosis are being explored. Optical spectroscopy is also being investigated for oral cancer diagnosis. Several ex vivo and in vivo studies have demonstrated potential of Raman spectroscopy for detection of oral cancers. After confirming feasibility of acquiring in vivo spectra in clinically implementable time, normal, pre-malignant and oral cancer classification and detection of even early changes like cancer-field effects at buccal mucosa subsite were demonstrated. Additionally, anatomic differences have also been detected in sub-sites of oral cavity. In vivo Raman spectroscopic detection of normal, pre-cancer and cancer at all oral sub-sites can help develop a non-invasive approach for mass-screening and diagnosis of oral cancer. Therefore, in vivo subsite classification and oral cancer diagnosis at different subsites was explored in the present study.

The applications of in vivo RS are limited by need for on-site instrumentation and strict experimental conditions, especially in low resource settings and in highly populous countries like India. Ex vivo samples like serum and exfoliated cells may be more practical for oral cancer screening and diagnostic applications. Apart from multiple advantages like accessibility, less invasive procedure, low cost, and multiple sampling, these samples can also be transported from primary screening centers to a centralized facility for analysis. Previous studies have shown

feasibility of detecting several diseases, including cancers using serum Raman spectroscopy. Exfoliated cells have also been employed for detection of cervical cancers and studying intracellular structure and function. Diagnosis of oral cancer and pre-cancers using Raman spectroscopy of less invasive samples was also investigated in this study.

The studies reported in this thesis therefore aimed to develop Raman spectroscopy as a method for early diagnosis and achieve sub-site classification of oral cancers using minimally invasive samples like serum and exfoliated cells and also using in vivo approach.

To achieve the stated aim, following specific objectives were undertaken-

Objectives:

1) To explore feasibility of classifying normal, oral cancer and precancerous conditions using less invasive samples like serum and exfoliated cells.

2) To correlate Raman spectra of serum with histopathology of sub-sites in oral cancers.

3) To explore potential of Raman spectroscopy in diagnosis and sub-site classification of oral cancers in vivo.

Chapter 2

Raman spectroscopy of serum for diagnosis of oral cancers

Introduction:

Serum tumor markers can play an important role in the management of cancer patients by contributing to diagnosis, staging and risk assessment, evaluation of response to therapy and early detection of relapse. Such markers arise from the presence of a tumor: can be secreted or shed by the tumor in excess of the normal tissue, can be uniquely specific to the tumor phenotype, often as embryonic, fetal (*i.e*. alpha fetoprotein-AFP), can be unique extracellular matrix or cell adhesion molecules, receptors, growth factors, cytokines, or products of abnormal metabolism or mediators of body's own defense in response to tumor (antibodies)[286,287]. Some examples of currently employed tumor markers in clinical practice include prostate specific antigen (PSA), cancer antigen -125 (CA-125), AFP, calcitonin, β2-microglobulin, β-human chorionic gonadotropin (HCG), carcinoembryonic antigen (CEA) and Cyfra 21-1 for prostate, ovarian, liver, medullary thyroid, myeloma/lymphomas, testicular and lung cancer, respectively[287-292]. Recent studies have shown the utility of blood sample-based liquid biopsy in cancer diagnosis and monitoring treatment response by detecting DNA fragments, or by detecting blood telomere length for predicting onset of cancer[293,294].

A serum tumor marker based approach may help in the early detection and improved patient monitoring for oral cancers. Therefore, as previously mentioned DNA, RNA and protein markers have been explored in serum and saliva from patients[10,11,13,130]. Before clinical use, the individual sensitivity and specificity of these markers have to be validated over a large sample size. Oral carcinogenesis is also accompanied with several biochemical changes. Some reported alterations

include: increase in total sialic acids, lipid bound sialic acids, total protein, cell free nucleic acids (DNA, RNA), and decrease in total lipids, cholesterol, high density lipoprotein, antioxidants like beta carotene, vitamins A, C,E, GSH, Zn, glutathione, serum albumin and an altered plasma free amino acids profile[295-301]. A panel constituting multiple biomarkers, rather than a particular marker, may better reflect the disease status. Individual detection of various markers in the panel requires tedious sample preparation, numerous reagents, long output times, extensive analysis and thus increases overall cost and time for diagnosis. As an alternative, measurement of global changes in the metabolic profile of a sample may also aid in cancer diagnosis. A recent study that investigated entire metabolome of blood samples has shown potential in prediction of breast cancer with 80% efficiency[14].

Optical spectroscopic techniques like IR, fluorescence and Raman spectroscopy also yield global molecular changes in the sample. Fluorescence spectroscopy has been employed for oral and breast cancer detection using serum[302-304]. IR spectroscopy of serum has been explored for detection of various diseases like Alzheimer's, SCA and leukemia, breast, ovarian and endometrial cancers[305-310]. Raman spectroscopy is associated with advantages like non-destructive and non-intrusive analysis, high sensitivity to even subtle biochemical perturbations, no sample preparation, rapid and objective and has therefore been used for detection of several cancers like breast, head and neck cancer, cervical, colorectal, pancreatic, nasopharyngeal and ovarian cancers[24-29,237]. Therefore, feasibility of identifying oral cancer using serum Raman spectroscopy was explored in this study.

In this chapter, detection of oral cancers using serum Raman spectroscopy has been explored. In the first two sections, pilot studies to explore feasibility of oral cancer detection using both, resonance Raman and conventional Raman were explored. After demonstrating potential in classifying normal and oral cancer sera, a large cohort study after inclusion of premalignant and disease control samples for exploring screening applications was carried out in the next section. In the final section, feasibility of recurrence detection in oral cancer was also investigated. Spectral differences were first investigated across the groups and subsequently multivariate tool PC-LDA was employed for data analysis in all sections.

2. 1 Exploring resonance Raman for serum-based diagnosis of oral cancer

Literature suggests that anti-oxidants like carotenoids and beta-carotene play a role in reducing incidence of cancers[311]. Additionally, an association between high circulatory levels of beta carotene and lower risk of malignancy has been observed[312,313]. There is increasing evidence that oxidative stress contributes during promotion stages of carcinogenesis, and at this stage, anti-oxidants may play an important role in prevention of carcinogenesis progression [314,315]. Oral cancers are known to progress through multiple steps of carcinogenesis in which multiple genetic events lead to transformation from normal to pre-cancer to cancer. Increase in oxidative stress has been implicated in development of oral cancer in tobacco and alcohol users. Several studies have indicated decreased beta carotene levels in circulation of oral pre-cancer and cancer subjects, and further observed their decreasing levels with disease progression[316,317]. Beta carotene is known to have strong absorption around 400-500 nm[318]. Resonance Raman studies using similar excitation wavelengths have shown three intense fundamental Raman peaks for beta carotene[319-321]. In this section, differences between normal and oral cancer sera with respect

to beta-carotene content along with other biomolecules were explored using resonance Raman spectroscopy. Data analysis was carried out in several 2-model systems such as- normal and cancer, normal and buccal mucosa cancer, and normal and tongue cancer. As multiple spectra were acquired from each sample, two approaches for data analysis were possible: spectra-wise and patient-wise. Both these approaches were explored in this section.

2.1.1. Materials and Methods

2.1.1.1. Subject details

A sample size of 54 patients harboring primary oral squamous cell carcinoma of tongue (n=40) and buccal mucosa (n=14) with confirmed clinical and histopathological diagnoses of oral squamous cell carcinoma were included. Samples were collected randomly from patients presenting with oral cancer in different subsites, who visited the outpatient department (OPD) of Tata Memorial Hospital (TMH), Mumbai, India. Blood samples were collected after overnight fasting, prior to any surgery-related interventions. All recruited patients were cases without prior anticancer treatment, history of malignancy, and second primary cancers. Patient's history like age, sex, symptoms, tobacco chewing/smoking, and alcohol consumption habits was obtained from the hospital records and by using a questionnaire. Blood samples were also collected from normal individuals (n=16) as the control group after overnight fasting. This group comprised of genetically unrelated healthy individuals and was chosen from patients' own relations, friends and colleagues. The subjects were declared 'normal' without clinical history as reported by themselves. Subject details are summarized in Table 2.1. This study was approved by the human

ethics committee, and written informed consent was obtained from each subject before enrollment.

Sr. no.	Subject category	Subjects recruited	Age range/Mean age	Gender	Spectra recorded
1.	Normal	16	27-59 years/38 yr	6F,10M	160
2.	Buccal mucosa cancer	14	36-60 years/46 yr	4F, 10 M	128
3.	Tongue cancer	40	31-63 years/ 44 yr	7F, 33M	403

Table 2.1 Subject accrual summary for resonance Raman study on serum

2.1.1.2. Sample collection

Five ml venous blood samples were collected from subjects with the help of a sterile injection. Both serum and plasma were explored as samples for oral cancer diagnosis. Serum was selected for all future studies to circumvent additional confounding factors like clotting factors present in plasma. Samples were placed standing for 30 minutes to allow clot formation and then centrifuged at 3500 rpm for 10 minutes to obtain serum. After removing the fat body with the help of a microtip, samples were centrifuged again at 3500 rpm for 10 minutes. The obtained serum was aliquoted in different tubes and stored at -80ºC till use. One of the aliquots was utilized for Raman spectroscopic analysis.

2.1.1.3. Raman spectroscopy

Frozen serum samples, after passive thawing, were subjected to Raman spectroscopy by placing 30 µl sample on glass slide and spectra was recorded using a commercial Raman microscope WITec alpha300RS (WITec GmbH, Ulm, Germany), as shown in Figure 2.1. Instrument was calibrated for spectrum using an Argon mercury lamp. The serum samples were excited with the 532 nm laser focused through a 50X Nikon objective (Numerical aperture 0.55). The laser power at the sample was 20 mW and Raman signals were collected by spectrograph with a grating of 600 grooves/mm. As specified by the manufacturer, spectral resolution is ~ 3.8 cm^{-1}. Each spectrum in the 0-3600 cm^{-1} range was acquired for 5 s and averaged over 4 accumulations. On an average, 11 spectra were acquired from each sample, to give rise to a total of 691 spectra from 70 samples. The serum samples of cancer and normal groups were run randomly.

Figure 2.1. Photographic representation of Raman microscope employed for resonance Raman studies

2.1.1.4. Spectral pre-processing and Data analysis

Spectral pre-processing of the data was carried out using Labspec 5.0 software (Horiba Jobin-Yvon, France) as per already existing protocol[322,323]. The acquired Raman spectra were first corrected for instrument response by employing a NIST standard- SRM 2241(NIST, Gaithersburg, MD, USA). The resulting intensity-instrument-response corrected spectra are free from instrument-induced spectral artifacts and comparable across instrumentation of different manufacturers[182]. The spectral contributions from the optical elements, i.e. the background signals were removed by subtracting the background spectra acquired using only the substrate (glass slides) to measured sample spectra. To remove interference of the slow moving background, first derivatives of spectra were computed using the Savitzky-Golay method (window size 3). Interpolation was carried out in the spectral range of 800-1800 cm^{-1}, since this region is an important constituent of the finger-print region. Correction for spectral differences due to relative intensity changes or variation in sample thickness was carried out using Vector normalization. These first derivative, interpolated and vector normalized spectra were then subjected to PC-LDA. As described previously, LDA can be used in conjunction with PCA (PC-LDA) to increase the efficiency of 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, significant principal components ($p<0.05$) were selected as input for LDA. 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[324,325]. During PC-LDA, use of less than 10 PC factors for LDA is the best compromise on the information being included and data noise being excluded[326]. PC-LDA models were validated by Leave-one-out cross-validation

(LOOCV). Algorithms for these analyses were implemented in MATLAB (Mathworks Inc.) based in-house software[327].

The results of PC-LDA are depicted in the form of a confusion matrix, where all diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions. The confusion matrix is generated to understand separation between the groups obtained by accounting for contribution of all factors selected for analysis. Scree plots depict the variance (or % correct classifications) accounted for by the total number of factors selected for analysis. PC-LDA results are also depicted in the form of scatter plots, generated by plotting various combinations of scores of factors. Plotting different combinations of factor scores gives a visual understanding of classification pattern in the data.

Since multiple spectra were acquired from each sample, two approaches for data analysis are possible: spectra-wise, where each spectrum is treated as an independent sample and all spectra are individually analyzed by PC-LDA; or patient-wise, where all spectra acquired from a single sample are averaged such that each sample is represented by a single spectrum. These mean spectra can then be subjected to 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, 3) and vector normalized spectra were used for spectral comparisons.

2.1.2. Results and Discussion

2.1.2.1. Spectral features

Vector-normalized average spectra of normal, buccal mucosa and tongue cancer are presented in Figure 2.2. Contributions of proteins, DNA, β-carotene and several amino acids were observed in the mean spectra of all groups. Differences in the form of intensity related variations and shifts were observed across these mean spectra. Highly intense Raman bands in the ~1155 and 1523 cm^{-1} regions that can be assigned to β-carotene, could be due to resonance Raman effect. These β-carotene specific bands were observed in all sera with highest relative intensity in normal, followed by tongue and buccal cancer mean spectra.

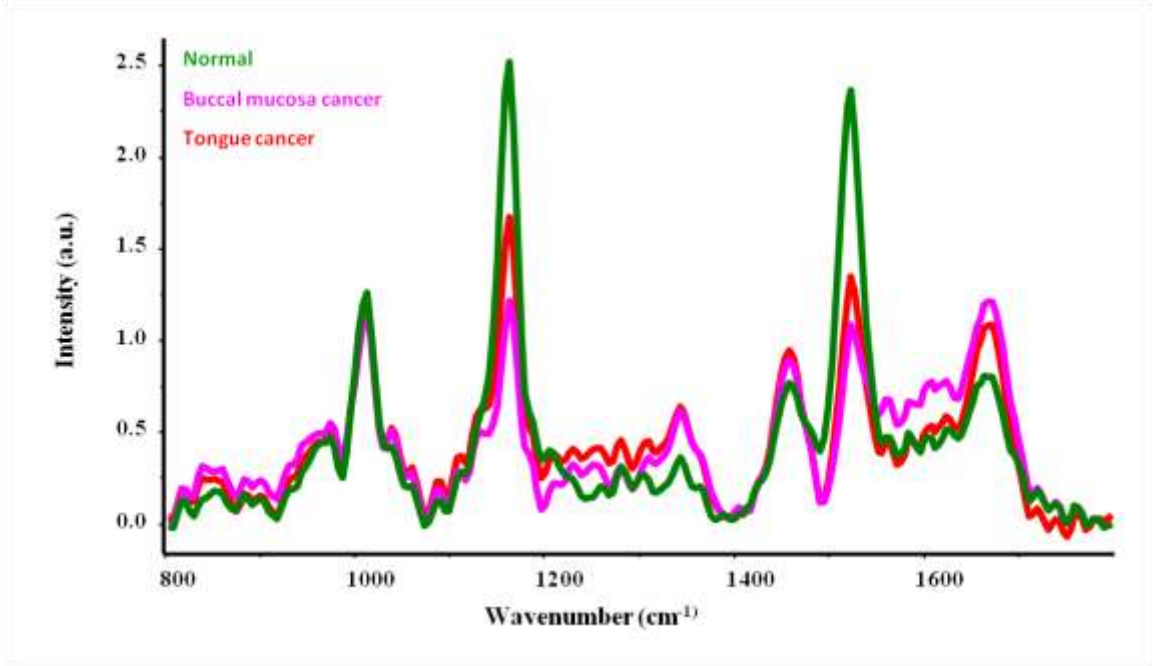


Figure 2.2. Mean Raman spectra of serum from normal, buccal mucosa cancer and tongue cancer groups for spectral comparisons in resonance Raman study

The differences between DNA and protein levels across the groups analyzed support the available data on increased cell free DNA and increased expression of some proteins during cancer conditions. Cell free nucleic acids (cfNA) like DNA, RNA and mRNA have been detected in circulation of several cancer patients. The reason for increased cell free nucleic acid levels in cancer patients' blood remains largely unknown; however two main mechanisms have been proposed: apoptosis and necrosis, or release of intact cells in the bloodstream and their subsequent lysis[297,298]. It has been previously described that the amount of β carotene present in blood decreases during cancer conditions. Reports suggest significantly reduced plasma levels of β-carotene in women with histopathologically diagnosed cervical dysplasia and an inverse association between β-carotene plasma levels and increasingly severe graded cervical histopathology. Also, increased breast cancer risk was observed in subjects with lower levels of β-carotene. Epidemiological studies have suggested that high endogenous levels of pro-oxidants and deficiencies in antioxidants level are likely to be important risk in the progression of pre-cancer to cancer. Lower levels of β-carotene in sera of pathological conditions like oral cancer, oral sub-mucous fibrosis and leukoplakia, as compared to healthy controls have also been reported. Thus, the higher antioxidant levels of β carotene in normal group observed in this study are in concordance with the existing reports[299,300,328]. Redistribution or translocation of plasma free amino acids (PFAAs) to support visceral or tumor protein synthesis is an essential feature observed in cancer patients. Clinical data from 13 studies have shown that levels of phenylalanine, tyrosine and tryptophan may be different for different cancers. Oral cancer specific PFAA is not reported, as per the available literature. Further, PFAA profile can differ between the early and late stages of cancer[301]. As per our findings, the overall spectral profiles

indicate an increase in Phe concentrations in normal groups while increased Trp and Tyr levels in cancer conditions. Thus, mean cancer serum spectra indicate that higher DNA, proteins and decreased beta carotene are characteristics of cancer.

2.1.2.2. Multivariate analysis

PC-LDA was employed for multivariate analysis. Since multiple spectra were acquired from each sample, two approaches for data analysis were possible: spectra-wise, where each spectrum is treated as an independent sample and all spectra are individually analyzed by PC-LDA; or patient-wise, where all spectra acquired from a single sample are averaged such that each sample is represented by a single spectrum. These mean spectra can then be subjected to PC-LDA followed by LOOCV. The utility of both these approaches was explored in this section. Spectra-wise approach, where each spectrum is treated as an independent entity, was first adopted to explore if sample heterogeneity can influence classification between normal and cancer. Patient-wise approach, where all spectra from a sample are averaged such that each patient is represented by a single spectrum, was adopted to eliminate sample heterogeneity, if any. Thus, PC-LDA using both approaches, spectra-wise and patient-wise was carried out and validated using leave-one-spectrum-out and leave-one-patient-out, respectively. Classification between normal and cancer groups was explored in various 2-model systems: normal and cancer serum, buccal mucosa cancer and normal serum, tongue cancer and normal serum.

a) Normal and cancer serum

The vector normalized spectra from the normal and cancer groups were imported into MATLAB and PC-LDA was carried out on this 2-model system. Five factors accounting for 85% correct

classifications were utilized for PC-LDA (Figure 2.3a). Scores of factors 1 and 2 were used to visualize classification between the groups, as seen in the scatter plot (Figure 2.3b). Two almost distinct clusters for cancer and normal groups were observed.

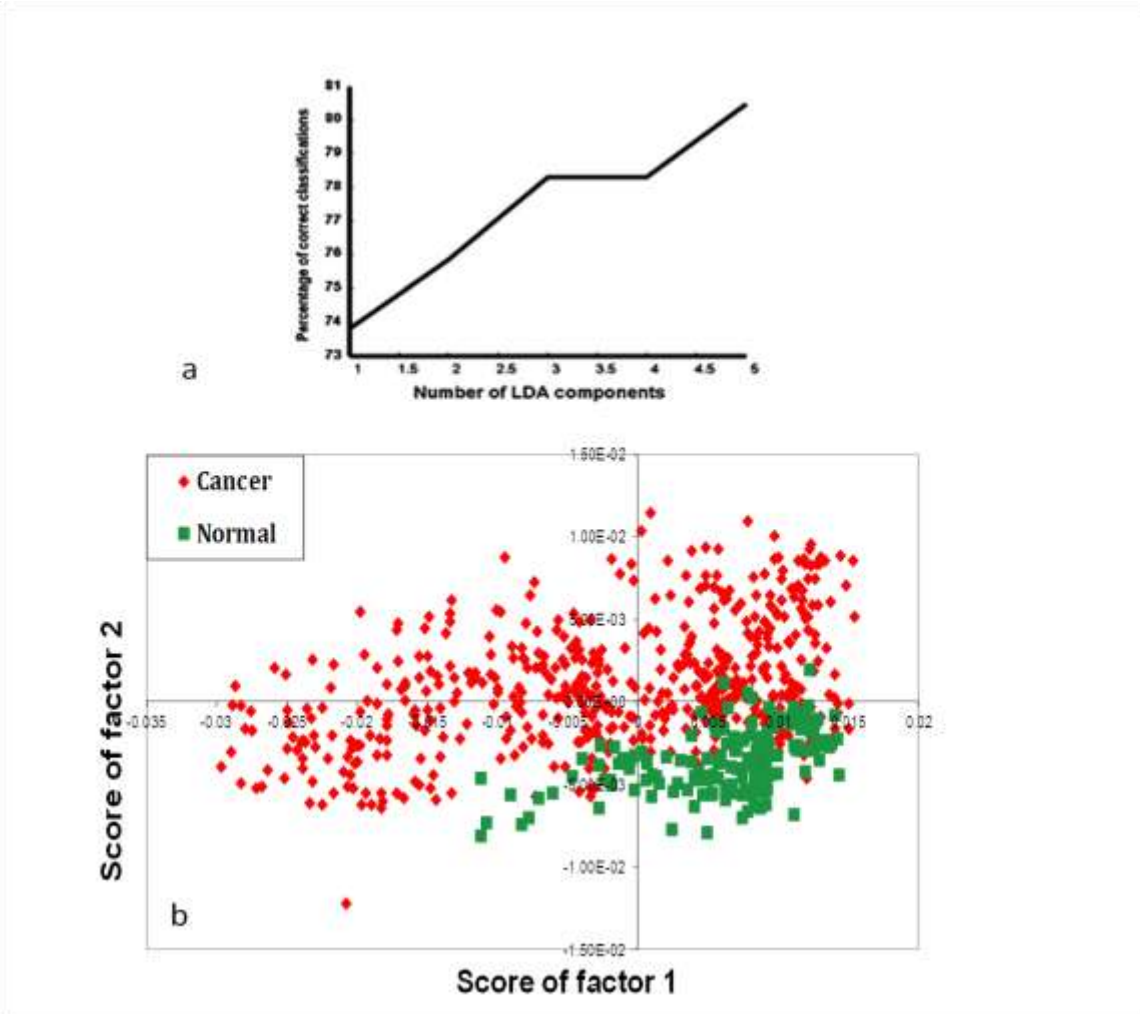


Figure 2.3. PC-LDA for exploring differences between normal and cancer serum in resonance Raman study - Spectra-wise analysis, a) Scree plot, b) Scatter plot

As shown in the confusion matrix in Table 2.2, 404/531 cancer spectra and 151/160 normal spectra were correctly classified. Thus, a leave-one- spectrum-out cross validated classification

efficiency of 76% for cancer and 94% for normal was observed, where classification efficiency is the percentage of spectra from each group that are correctly classified.

	Cancer	Normal	Total	% efficiency
Cancer	404	127	531	76
Normal	9	151	160	94

Table 2.2. PC-LDA for exploring differences between normal and cancer serum using Spectra-wise approach- confusion matrix of LOOCV results in resonance Raman study

In the PC-LDA for the patient-wise analysis, 3 factors accounting for ~82% correct classifications were employed (Figure 2.4a). The PC-LDA scatter plot for scores of factors 2 and 3 presented in Figure 2.4b shows a distinctive separation between the normal and cancer groups.

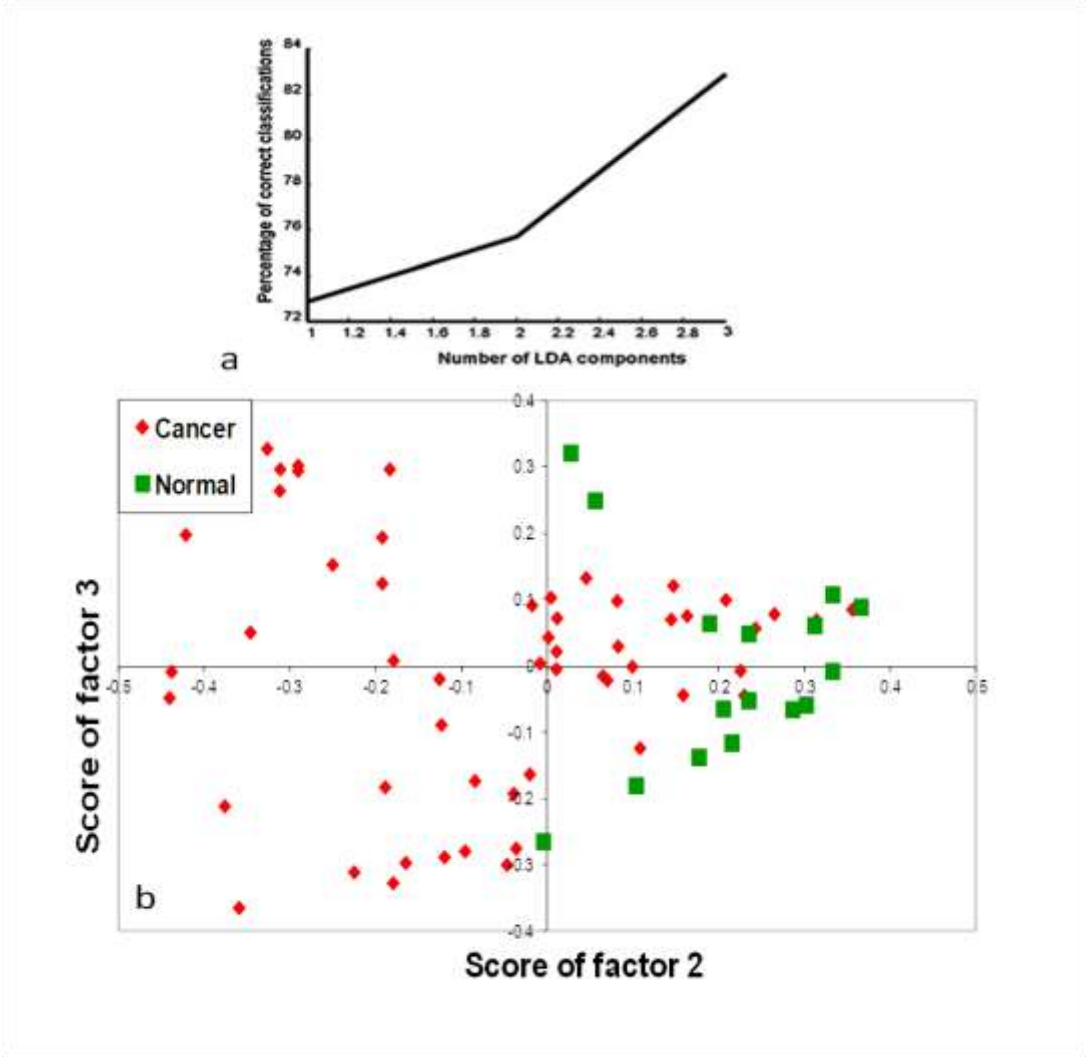


Figure 2.4. PC-LDA for exploring differences between normal and cancer serum in resonance Raman study - Patient-wise analysis, a) Scree plot, b) Scatter plot

As presented in Table 2.3, leave-one-patient-out indicates correct predictions for 44/54 cases for cancer while 12/16 correct predictions for the normal group, yielding a classification efficiency of 81% and 75%, respectively.

	Cancer	Normal	Total	% efficiency
Cancer	44	10	54	81
Normal	4	12	16	75

Table 2.3. PC-LDA for exploring differences between normal and cancer serum using Patient-wise approach - confusion matrix of LOOCV results in resonance Raman study

b) Normal and buccal mucosa cancer serum

To explore the classification of normal with the individual cancers (buccal mucosa and tongue), PC-LDA was first carried out for normal and buccal mucosa serum spectra using 4 factors accounting for 95% correct classifications for the spectra-wise analysis (Figure 2.5a). Scatter plot shown in Figure 2.5b indicates two clusters with minor overlap.

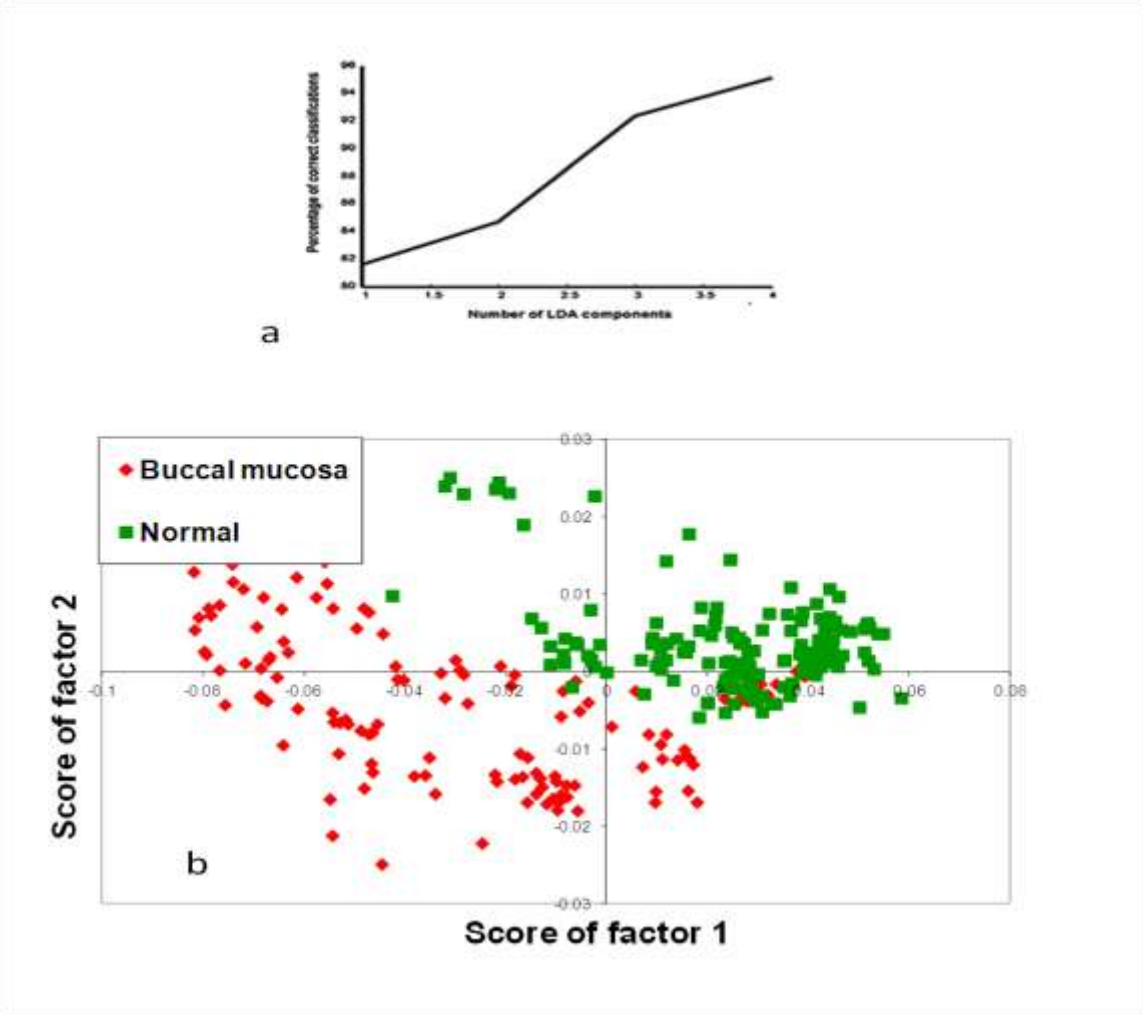


Figure 2.5. PC-LDA for exploring differences between normal and buccal mucosa cancer serum in resonance Raman study - Spectra-wise analysis, a) Scree plot, b) Scatter plot

The classification performance of leave-one-spectrum-out cross validation, shown in Table 2.4, shows a classification efficiency of 91% for buccal mucosa cancer and 98% for the normal group.

	Buccal mucosa cancer	Normal	Total	% efficiency
Buccal mucosa cancer	117	11	128	91
Normal	4	156	160	98

Table 2.4. PC-LDA for exploring differences between normal and buccal mucosa cancer serum using Spectra-wise approach- confusion matrix of LOOCV results in resonance Raman study

For the patient-wise analysis, 4 factors that contributed ~93% correct classifications were employed for PC-LDA (Figure 2.6a). As shown in Figure 2.6b, two almost distinct clusters for buccal mucosa cancer and normal groups were observed in the scatter plot plotted using scores of factors 2 and 3.

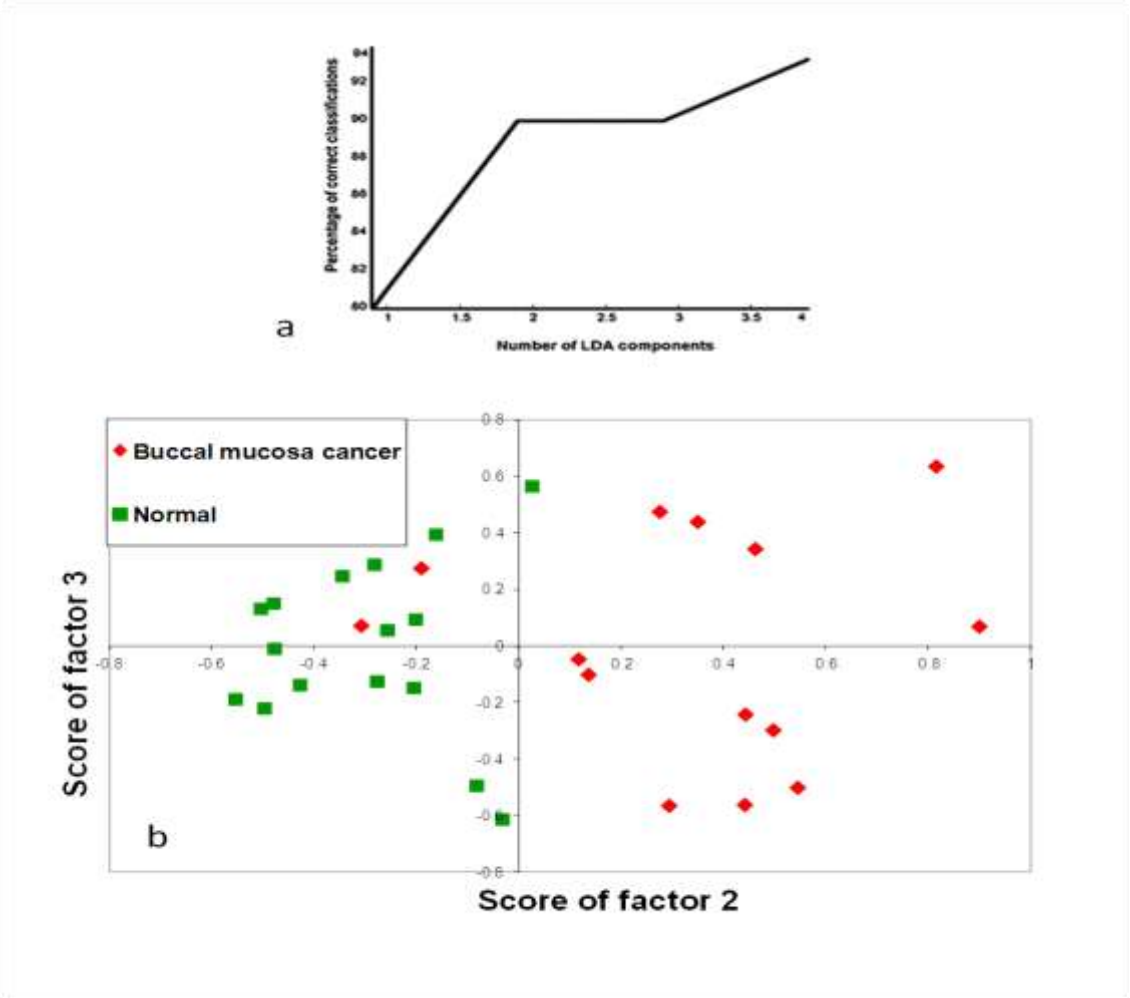


Figure 2.6. PC-LDA for exploring differences between normal and buccal mucosa cancer serum in resonance Raman study - Patient-wise analysis, a) Scree plot, b) Scatter plot

Leave-one-patient-out indicates 12/14 and 13/16 correct classifications for buccal mucosa and cancer groups, yielding a classification efficiency of 86% and 81%, respectively, as summarized in Table 2.5.

	Buccal mucosa cancer	Normal	Total	% efficiency
Buccal mucosa cancer	12	2	14	86
Normal	3	13	16	81

Table 2.5. PC-LDA for exploring differences between normal and buccal mucosa cancer serum using Patient-wise approach- confusion matrix of LOOCV results in resonance Raman study

c) Normal and tongue cancer serum

The feasibility of differentiating normal and tongue cancer was explored by using 6 factors accounting for 79% correct classifications in the PC-LDA for spectra-wise analysis (Figure 2.7a). The PC-LDA scatter plot for scores of factors 2 and 3 presented in Figure 2.7b shows slightly overlapping clusters for normal and cancer groups.

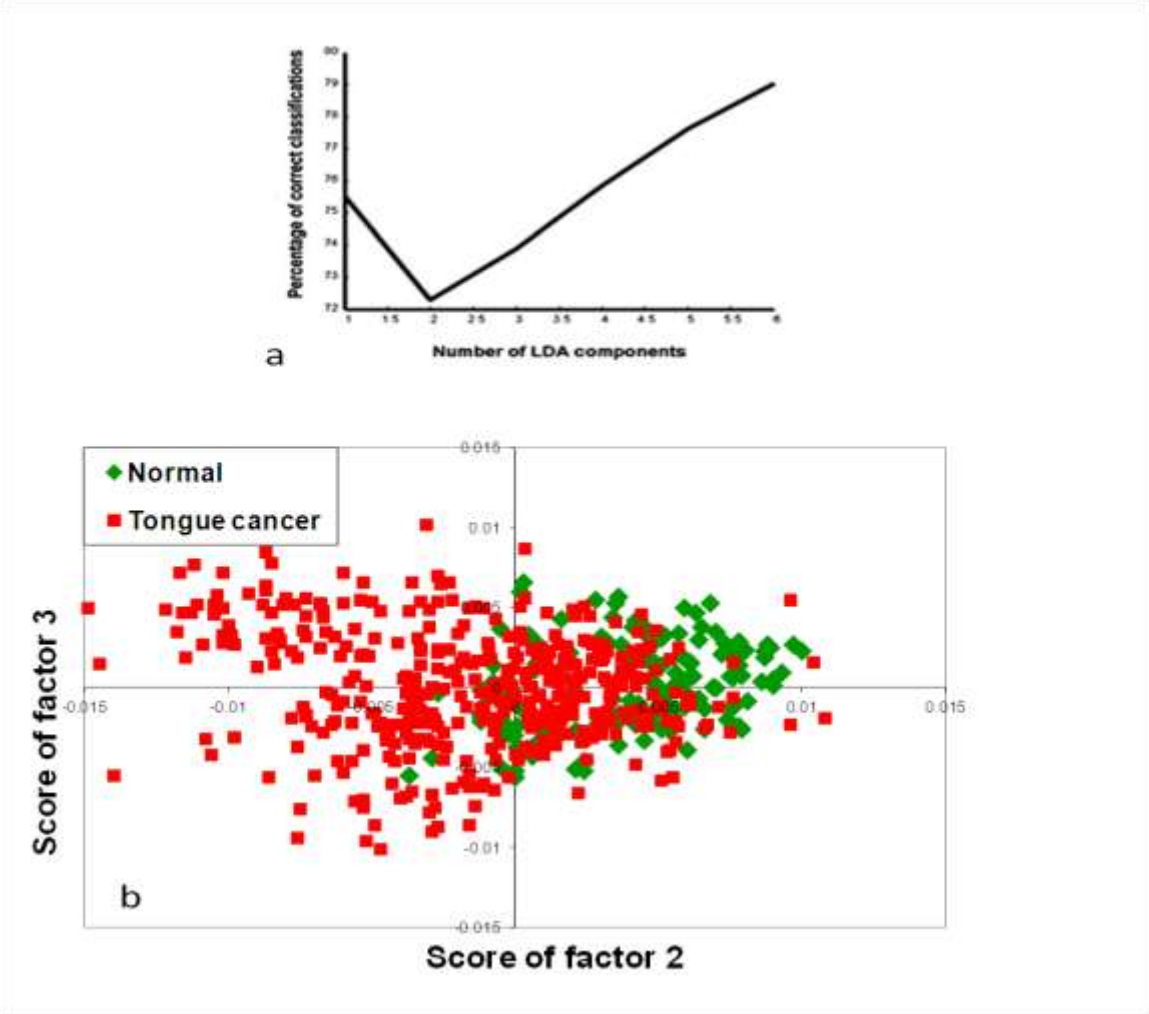


Figure 2.7. PC-LDA for exploring differences between normal and tongue cancer serum in resonance Raman study - Spectra-wise analysis, a) Scree plot, b) Scatter plot

Table 2.6 details the results of the LOOCV of spectra-wise analysis, which yields a classification efficiency of 76% for the tongue cancer group and 85% for normal group.

	Normal	Tongue cancer	Total	% efficiency
Normal	136	24	160	85
Tongue cancer	98	305	403	76

Table 2.6. PC-LDA for exploring differences between normal and tongue cancer serum using Spectra-wise approach- confusion matrix of LOOCV results in resonance Raman study

In the patient-wise analysis, 4 factors accounting for ~82% correct classifications were employed (Figure 2.8a). As seen in Figure 2.8b, scores of factors 1 and 2 were used to visualize classification between the groups that yield two clusters with minor overlap for tongue cancer and normal groups.

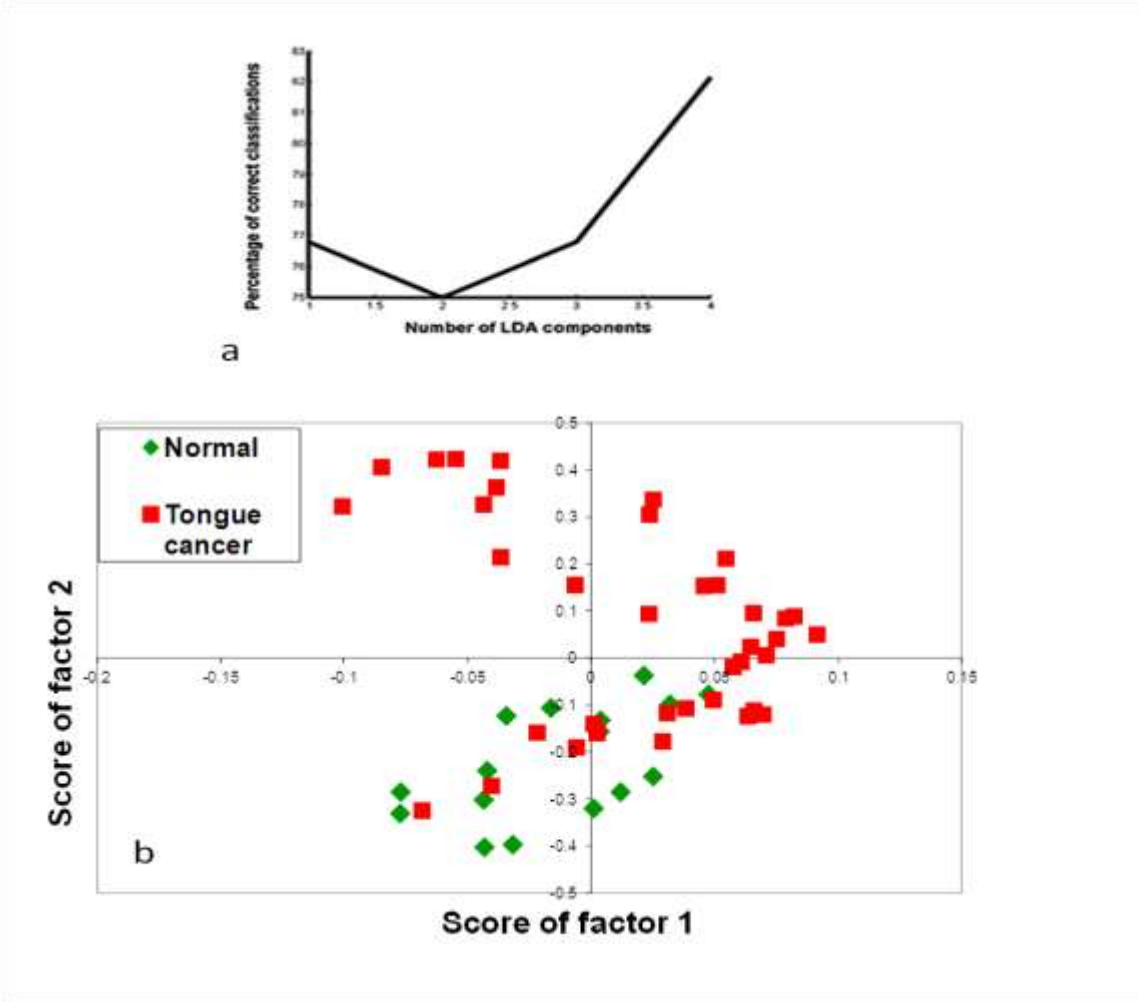


Figure 2.8. PC-LDA for exploring differences between normal and tongue cancer serum in resonance Raman study - Patient-wise analysis, a) Scree plot, b) Scatter plot

LOOCV yields a classification efficiency of 78% and 82% for tongue cancer and normal, respectively (Table 2.7).

	Normal	Tongue cancer	Total	% efficiency
Normal	13	3	16	82
Tongue cancer	9	31	40	78

Table 2.7. PC-LDA for exploring differences between normal and tongue cancer serum using Patient-wise approach- confusion matrix of LOOCV results in resonance Raman study

It was observed that along with higher average classification efficiency in the spectra-wise approach, heterogeneity in the samples was also evident. Spectra from a given sample (eg. Normal) misclassified across different groups i.e. both, cancer and normal in this approach. Thus, assigning the sample into a particular group becomes difficult. To eliminate such issues, patient-wise approach with a slightly reduced average classification efficiency was adopted to achieve unambiguous prediction of samples.

The patient-wise approach gives a better representation of the sample, and was therefore adopted for prospective studies.

2.2. Exploring serum-based diagnosis of oral cancer using conventional Raman spectroscopy

Feasibility of classifying normal and oral cancer sera was confirmed by resonance Raman using 532 nm excitation. In this study, the spectral differences between normal and cancer groups could majorly be attributed to beta-carotene. In the next step, this feasibility of classification was explored using 785 nm excitation- the most widely used excitation for biomedical applications. Biological samples are plagued by strong fluorescence background that potentially swamps the weak Raman signals. Further, it is known that in addition to beta carotene, several other biomolecules like DNA, amino acids, lipids and proteins also differ in normal and cancer conditions. In view of attributes like low parasitic fluorescence, photo-degenerative effects and to explore other molecular differences between normal and cancer serum, 785 nm excitation was explored for oral cancer and normal classification using a similar cohort of samples in this pilot study. In this section, data analysis was carried out using 2 group systems, normal and cancer, normal and buccal mucosa cancer, and normal and tongue cancer and adopting the patient-wise approach.

2.2.1. Materials and Methods

2.2.1.1. Subject details

A sample size of 69 oral cancer patients harboring either buccal mucosa cancer (n=36) or tongue cancer (n=33) from OPD of TMH, Mumbai, India were included in the study after seeking an informed written consent. As stated in section 2.1.1.1, blood samples were collected after over-night fasting and collection of demographic data from each subject. Blood samples were also

obtained from healthy volunteers (n=17) as the 'normal' group. Like in the previous study, control group comprised of healthy individuals chosen from patients' own relations, friends and colleagues without any clinical history. Subject accrual summary is presented in Table 2.8.

Sr. No.	Subjects	Samples analyzed	Spectra recorded
1.	Normal	17	51
2.	Buccal mucosa cancer	36	108
3.	Tongue cancer	33	99
Total		86	258

Table 2.8. Subject accrual summary for conventional Raman study on serum

2.2.1.2. Sample collection

Serum samples were separated as stated in section 2.1.1.2.

2.2.1.3. Raman spectral acquisition

After passive thawing, samples were subjected to Raman spectroscopy by placing 30 µl volume on calcium fluoride (CaF_2) window and spectra were recorded using Fiber Optic Raman microprobe (Horiba-Jobin-Yvon, France), as shown in Figure 2.9. Briefly, this system consists of a laser (785 nm, Process Instruments) as an excitation source and HE 785 spectrograph (Horiba-Jobin-Yvon, France) coupled with CCD (Synapse, Horiba-Jobin-Yvon) as dispersion and detection elements, respectively. Optical filtering of unwanted noise, including Rayleigh signals,

is accomplished through 'Superhead', the other component of the system. Optical fibers were employed to carry the incident light from the excitation source to the sample and also to collect the Raman scattered light from the sample to the detection system. Raman microprobe was assembled by coupling a 40X microscopic objective (Nikon, Japan) to the superhead. Spectral acquisition details were: excitation wavelength (λ_{ex}) = 785 nm, laser power = 30 mW, spectra were integrated for 10 seconds and averaged over 6 accumulations. Three spectra were recorded from each sample to give a total of 258 spectra: 108 buccal mucosa, 99 tongue cancer and 51 healthy spectra, respectively.

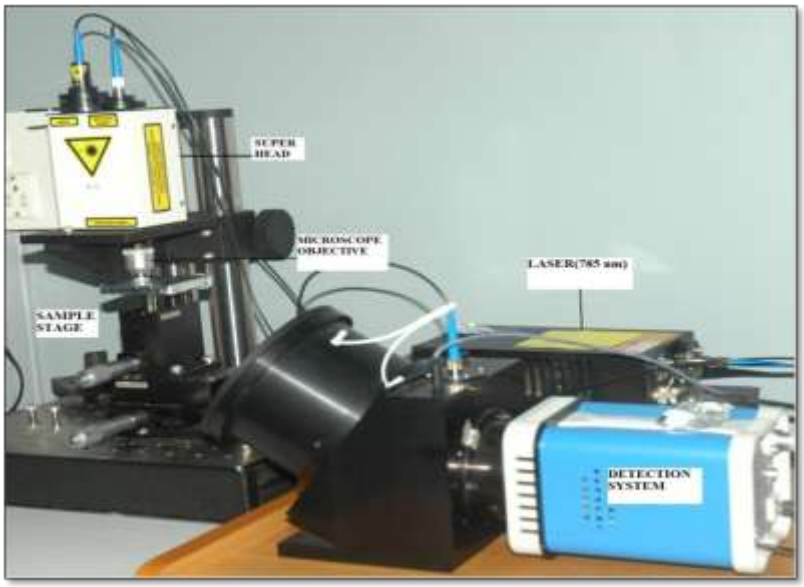


Figure 2.9. Photographic representation of microprobe employed in the conventional Raman studies

2.2.1.4. Spectral pre-processing and data analysis

As described in section 2.1.1.4, spectra were processed for CCD response correction, background subtraction, first derivatized, interpolated in the finger-print region (800-1800 cm^{-1})

and normalized before subjecting them to multivariate analysis. Vector-normalized spectra from the normal, buccal mucosa and tongue cancer groups were analyzed using supervised PC-LDA using the patient-wise approach. PC-LDA findings were validated using LOOCV.

2.2.2. Results and Discussion

2.2.2.1. Spectral features

Vector-normalized spectra of healthy, buccal mucosa and tongue cancer serum are presented in Figure 2.10, which indicate differences in amide I (around 1660 cm^{-1}), amide III (1260 cm^{-1}), δCH_2 deformation (around 1450 cm^{-1}), amino acids like Phe (1008 cm^{-1}), Trp (1560 cm^{-1}, 1622 cm^{-1}), DNA (1080 cm^{-1}, 1340 cm^{-1}) and carotenoids (1162 cm^{-1}, 1530 cm^{-1}) between the cancer and normal groups.

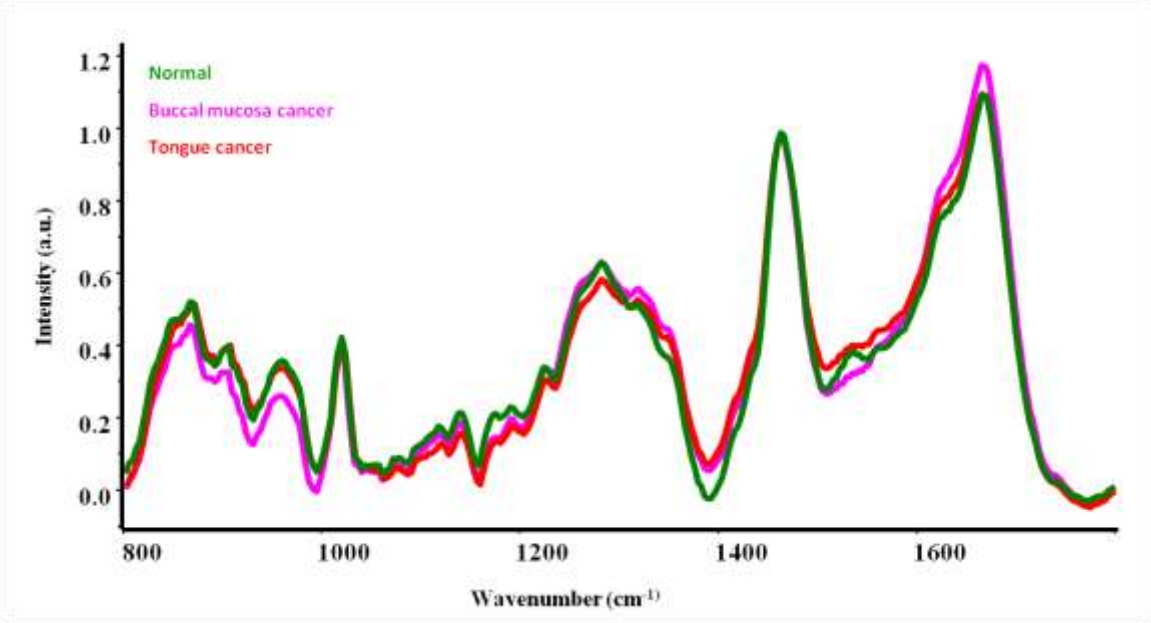


Figure 2.10. Mean Raman spectra of serum from normal, buccal mucosa cancer and tongue cancer groups for spectral comparisons in conventional Raman study

The spectral differences indicate predominance of protein, DNA, minor changes in lipid content and an altered amino acid profile in the cancer sera. Increased levels of circulating cell free nucleic acids (DNA, RNA and microRNA) as well as dyslipidemia have been reported for several cancers. A decreasing beta carotene level in the cancer sera with respect to the healthy controls has also been observed, which again coincides with our previous study and literature reports[19-24].

2.2.2.2. Multivariate analysis

PC-LDA was employed for classification using patient-wise approach, which was further validated by LOOCV. Patient-wise approach was selected for data in the previous section because of better representation of sample. The data was analyzed in the following 2-group systems: cancer vs. normal, buccal mucosa cancer vs. normal, tongue cancer vs. normal.

a) Normal and cancer serum

To evaluate the feasibility of classification, PC-LDA was carried out using 5 factors accounting for 85% correct classifications (Fig 2.11a). Scores of factor 1 and 2 were used to visualize the classification between the groups, as observed in the scatter plot in Figure 2.11b. Two separated clusters were obtained for each of the groups.

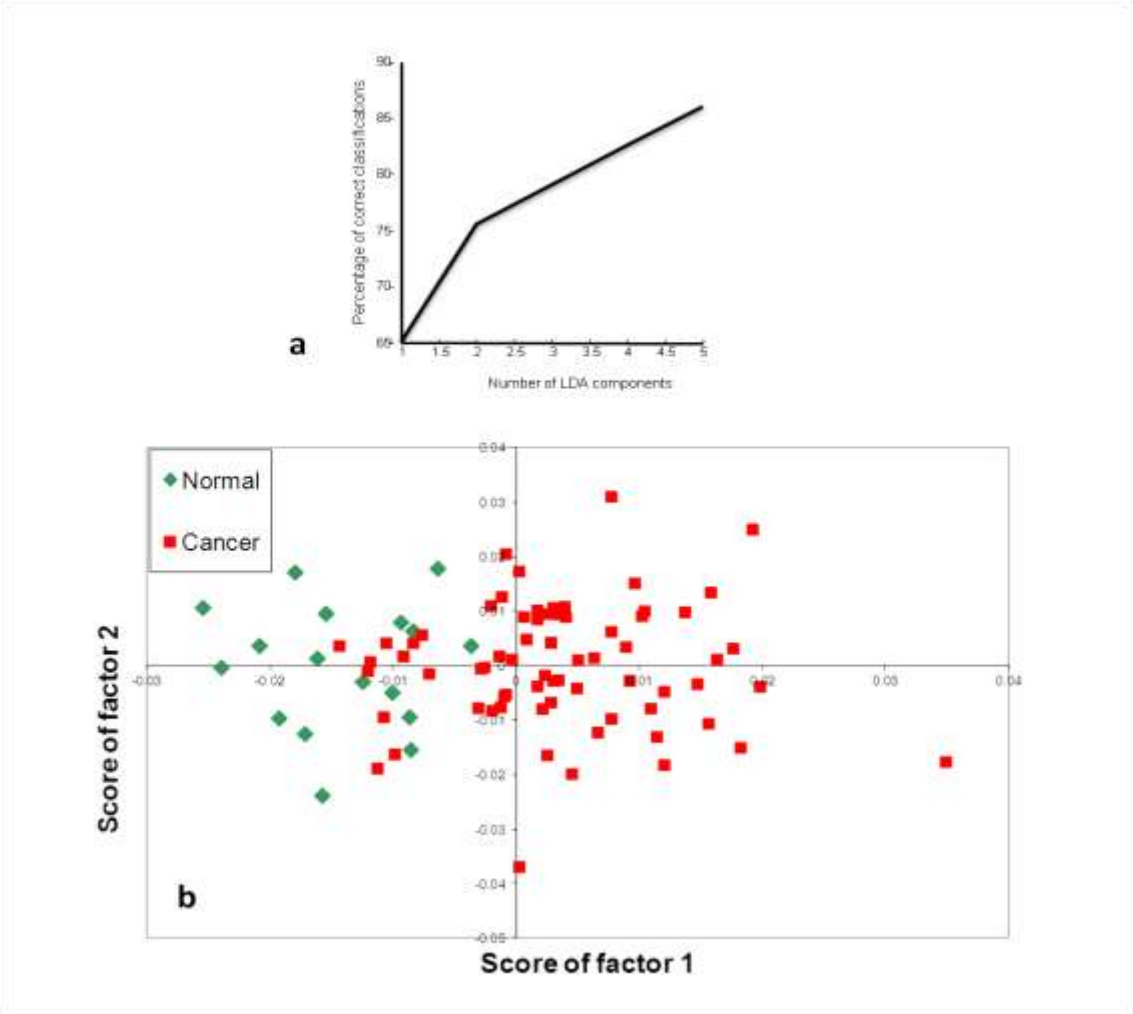


Figure 2.11. PC-LDA for exploring differences between serum from normal and cancer subjects in conventional Raman study, a) Scree plot, b) Scatter plot

LOOCV confusion matrix indicates 13/17 normal spectra and 56/69 spectra were correctly classified (Table 2.9), to yield a classification efficiency of 76% and 81%, respectively.

	Normal	Cancer	Total	% efficiency
Normal	13	4	17	76
Cancer	13	56	69	81

Table 2.9. PC-LDA for exploring differences between normal and cancer serum - confusion matrix of LOOCV results in conventional Raman study

b) Normal and buccal mucosa cancer serum

Spectra from sera of 36 buccal mucosa cancer and 17 healthy subjects were subjected to PC-LDA using 5 factors accounting for 91% correct classifications (Fig 2.12a). Scatter plot using scores of factor 1 and 2 shown in Fig 2.12b indicates two distinct clusters for the groups.

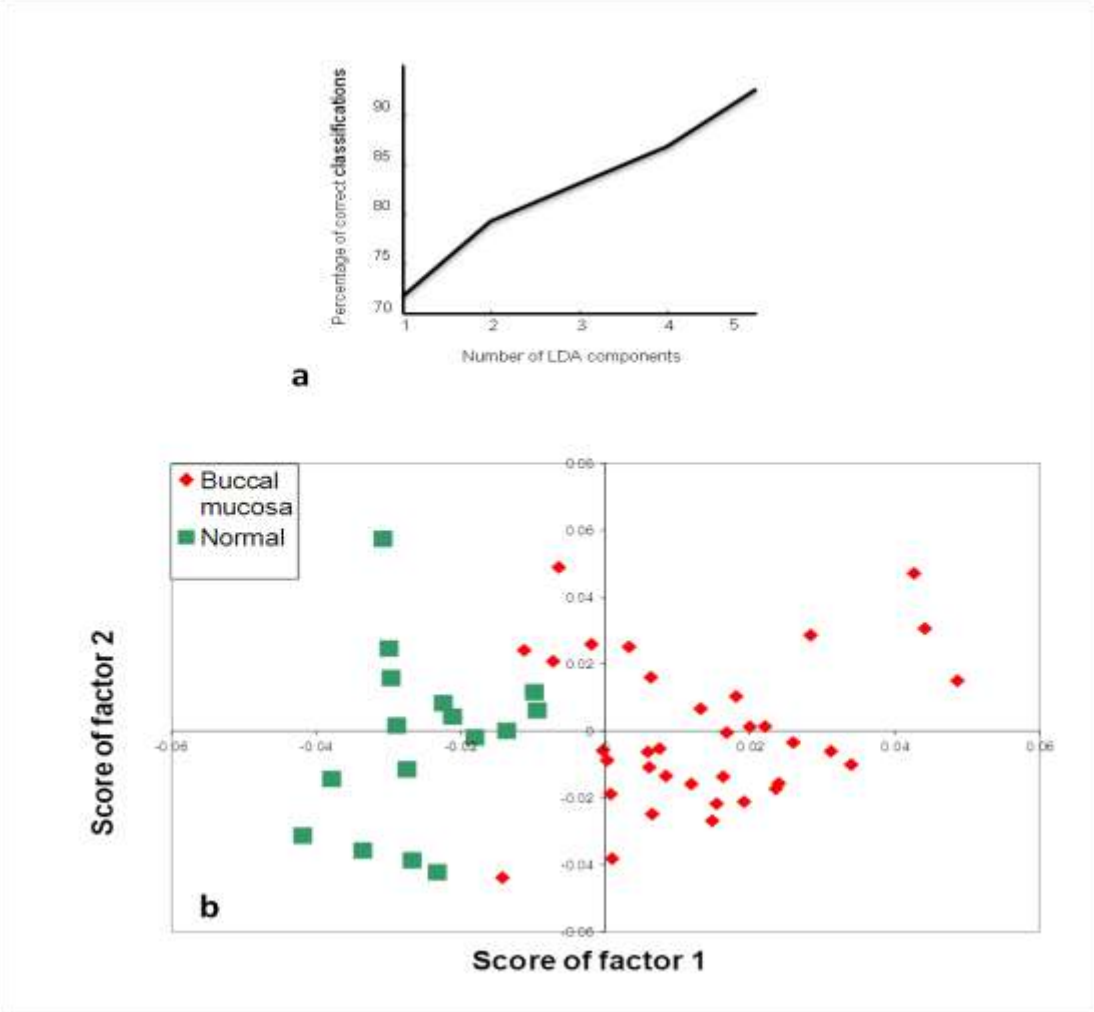


Figure 2.12. PC-LDA for exploring differences between serum from normal and buccal mucosa cancer subjects in conventional Raman study, a) Scree plot, b) Scatter plot

LOOCV yields 31/36 (86%) and 13/17 (76%) correct classifications for buccal mucosa cancer and normal groups, respectively (Table 2.10).

	Buccal mucosa cancer	Normal	Total	% efficiency
Buccal mucosa cancer	31	5	36	86
Normal	4	13	17	76

Table 2.10. PC-LDA for exploring differences between normal and buccal mucosa cancer serum - confusion matrix of LOOCV results in conventional Raman study

c) Normal and tongue cancer serum

PC-LDA using 4 factors corresponding to 87% correct classifications was carried out (Fig 2.13a). Scores of factor 1 and 2 were explored for classification; the scatter plot indicates slightly overlapping clusters for tongue cancer and normal groups seen in Figure 2.13b.

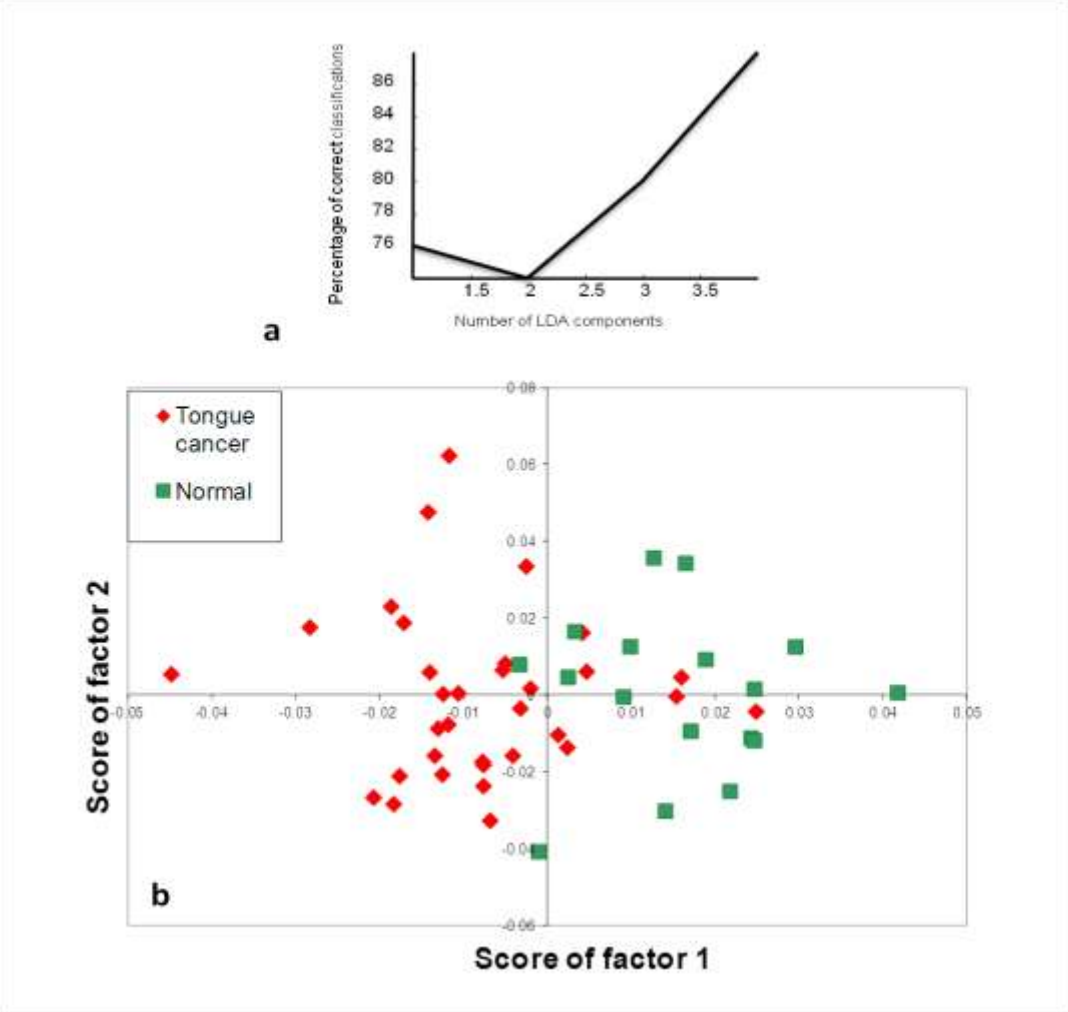


Figure 2.13. PC-LDA for exploring differences between serum from normal and tongue cancer subjects in conventional Raman study, a) Scree plot, b) Scatter plot

The LOOCV confusion matrix yields 26/33 (79%) and 11/17 (65%) true predictions for tongue cancer and normal groups, as shown in Table 2.11.

	Tongue cancer	Normal	Total	% efficiency
Tongue cancer	26	7	33	79
Normal	6	11	17	65

Table 2.11. PC-LDA for exploring differences between normal and tongue cancer serum - confusion matrix of LOOCV results in conventional Raman study

Thus, like the resonance Raman studies, potential of serum Raman spectroscopy in differentiating normal and oral cancers was observed in this study. Almost similar classification rates (~78%) were obtained for distinguishing normal and oral cancer serum in both studies. The findings of the present study re-affirm the applicability of Raman spectroscopy in oral cancer management, irrespective of the excitation wavelength employed.

2.3. Serum Raman spectroscopy for oral cancer screening using conventional Raman

Screening, an important tool in the overall management of cancers, has lead to early detection and decreased mortality rates for several cancers. Blood-test based screening is more practical and desirable, particularly for mass screening in populous nations like India and other developing countries where oral cancer is endemic.

Diagnosis of oral cancers was demonstrated in both pilot studies with a classification efficiency of ~78% between normal and oral cancer groups. Before envisaging clinical or screening-related applications, the efficacy of serum RS in early pre-cancer (pre-malignant) and specific cancer detection and evaluation of results on large sample size was evaluated. Serum RS based diagnosis of premalignant conditions for oral cancer is hitherto unreported. Although serum RS has shown to detect several cancers, the specific detection of any cancer has also not been demonstrated. In this section, detection of oral pre-malignant conditions (leukoplakia, OSMF and TPK) and specific detection of oral cancer using a disease control group were first explored. The classification between normal and oral cancer groups was also evaluated on a large cohort. Finally, the overall clinical utility of serum RS was investigated using specific classification models for screening-applications.

The following approach was used for data analysis- In the first step, the feasibility of classifying normal, premalignant, oral cancer and disease control groups using binary classification systems was explored. After exploring classification between these groups, the potential of serum RS in screening-related applications was investigated using 2 models: a single all-encompassing 4-model; and a normal versus abnormal coupled with type of abnormality deduction model. PC-

LDA was employed to build the standard models followed by cross validation using LOOCV and adopting patient-wise approach.

2.3.1. Materials and Methods

2.3.1.1. Subject details

Subject categories of normal, disease control, oral pre-malignant and oral cancer were included in the present study. Subjects who were clinically normal (without any apparent current disease), with or without tobacco/alcohol habits were categorized under the 'Normal' group. The 'Disease control' group was included as a non-oral cancer malignancy control. Glioma patients belonging to grade II-IV (n=35) were included in this group. Patients with confirmed diagnosis of malaria were also included in the study as an additional disease control group, however only few cases (n=11) could be accrued. 'Oral pre-malignant' (OPL) category comprised of subjects with clinically evident mucosal changes, either homogenous or heterogeneous leukoplakia patch, TPK lesion or palpable OSMF. 'Oral cancer' category consisted of subjects with primary histopathologically confirmed oral cancer at sub-sites buccal mucosa (BM) or tongue. Patients with a clinical history of anti-cancer treatment chemo or radiotherapy, second malignancy, recurrence were excluded from the present study. Both oral cancers and oral pre-malignant subjects were found to have long term tobacco abuse history. After preliminary screening, serum samples were collected from 339 subjects during 2010-2014. 126 subjects satisfied the criterion for Normal, 120 for oral cancer, 47 for oral-premalignant and 35 for disease control. Malaria samples were not included in the data analysis due to the small sample size. Therefore, 328 subjects were used for data analysis. The subjects for these categories were screened and selected

from OPD services of Tata Memorial Center, Mumbai, and D. Y. Patil Dental College and Hospital, Navi Mumbai. The Normal (healthy or control) subjects were selected from the patients' relatives, friends and staff members of these institutes. Samples were collected after seeking appropriate consent from the participating subjects. The subject accrual summary is presented in Table 2.12

Sr. No.	Category	Subjects recruited	Spectra analyzed
1.	Normal	126	126
2.	Disease control	35	35
3.	Premalignant	47	47
4.	Oral cancer	120	120

Table 2.12. Subject accrual summary for oral cancer screening studies on serum

2.3.1.2. Sample details

Samples were collected and processed as described in section 2.1.1.2.

2.3.1.3. Raman spectral acquisition

Spectra were acquired using HE-785 commercial Raman instrument (Horiba-Jobin-Yvon, France) as elaborated in section 2.2.1.3.

2.3.1.4. Spectral pre-processing and data analysis

Spectra were pre-processed and used as input for multivariate analysis as described in section 2.1.1.4. Average spectra were also computed, as described in the same section.

2.3.2. Results

2.3.2.1. Spectral features

The smoothed (Savitzky-Golay method, 3), normalized average spectra for all groups are shown in Figure 2.14a-d. Major spectral features were observed at 830 and 850 (Tyr doublet), 1007 (Phe), 1085-1100 (DNA backbone), 1162 and 1530 (β-carotene), 1267 (amide III), 1318 and 1342 (DNA bases), 1451 (CH_2 bending), 1556 and 1622 (Trp) and 1656 cm^{-1} (amide I). Tentative assignments have been made as per the available literature[329,330]. Thus, spectra are characterized by presence of contributions mainly from amino acids, beta carotene, DNA and proteins. Intensity-related variations in amino acids, amide III and DNA peaks were observed across the different groups.

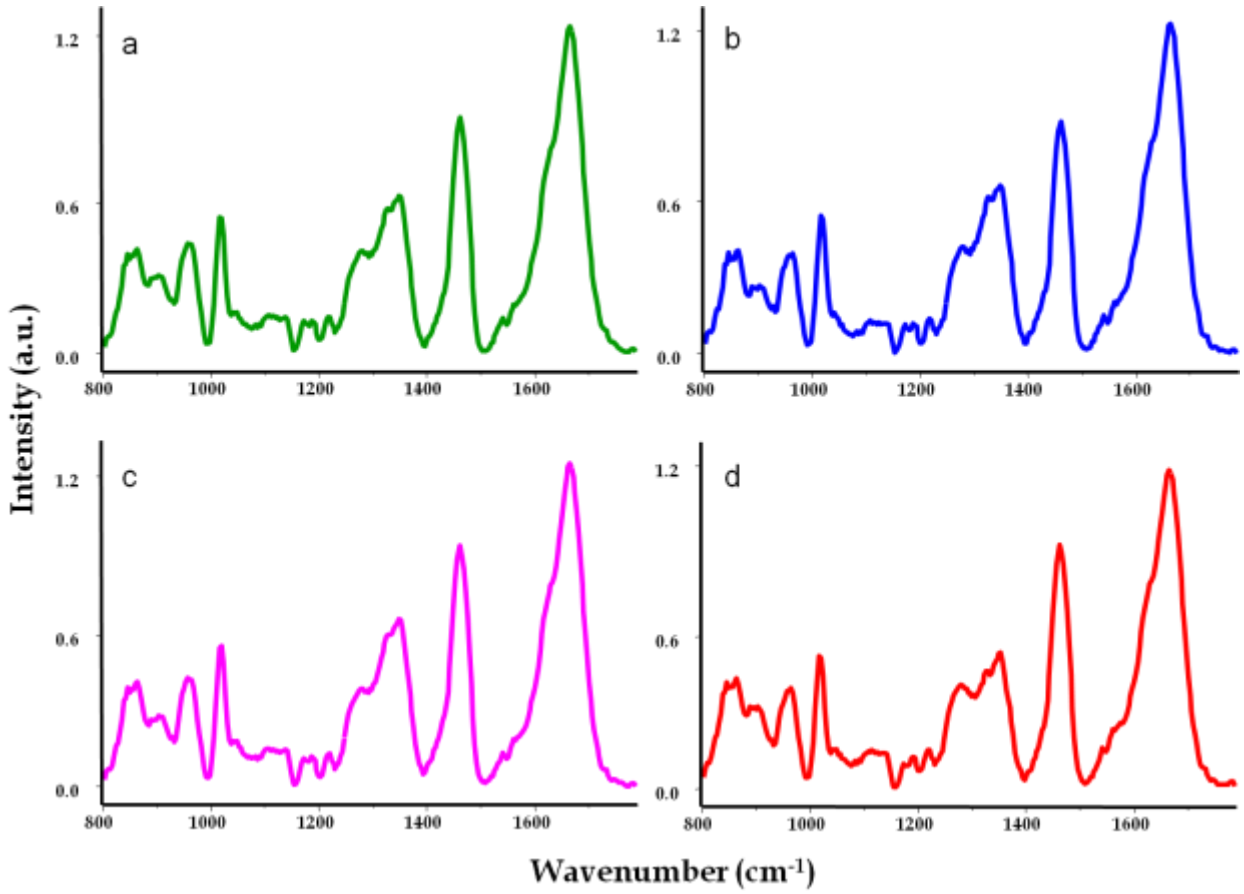


Figure 2.14. Mean Raman spectra of serum from a) normal, b) disease control, c) premalignant, d) oral cancer, for spectral comparisons

Oral cancer spectra showed higher Tyr, Trp (doublet at 830 and 850, 1552), amide III (1270) and CH_2 deformation (1450) but slightly lower amide I (1660) and sharper DNA bases (1342). Most of these spectral variations corroborate our previous findings[331,332]. With respect to normal group, premalignant spectra showed higher Phe (1008), lower amide III (1270), higher DNA bases (1320, 1342), higher CH_2 deformation (1450) and slightly lower amide I (1660). The disease control spectrum showed higher Phe (1008), amide III (1270), DNA bases features (1320, 1342), lower CH_2 deformation (1450) and higher amide I (1660). It has been reported that appearance of premalignant changes in the oral cavity are associated with increase in serum

concentrations of enzyme lactate dehydrogenase (LDH), β_2-microglobulin, changes in lipid profile, serum glyco-conjugates like sialic acid, lipid-bound sialic acid, serum antioxidants like Vitamin A, E, β-carotene, trace elements like iron, zinc, copper[333-335]. In case of disease control group that comprised of glioma cases, changes in proteomic profiles, serum gangliosides, trace elements like copper, zinc, lead, manganese, cobalt, and increase in YKL-40 levels have been linked with glioma development[336-339].

2.3.2.2. Multivariate analysis

The feasibility of classifying normal, premalignant, oral cancer and disease control (non-oral cancer) groups using binary classification systems was first explored. PC-LDA was employed to build the standard models followed by cross validation using LOOCV. In the subsequent step, the potential of serum RS in screening-related applications was investigated using 2 models. The first model was a single all-encompassing 4-model system. In the second screening-model, a 2-step approach was developed. Normal versus abnormal classification in the first step followed by type of abnormality deduction in the second step was employed as the second model. Patient-wise approach was employed for data analysis.

a) Exploring classification between normal and oral premalignant groups

PC-LDA for normal versus premalignant subjects was carried out using 5 factors (~7% classification efficiency) as shown in Figure 2.15a. The scatter plot indicates slightly overlapping clusters for normal and premalignant groups (Figure 2.15b).

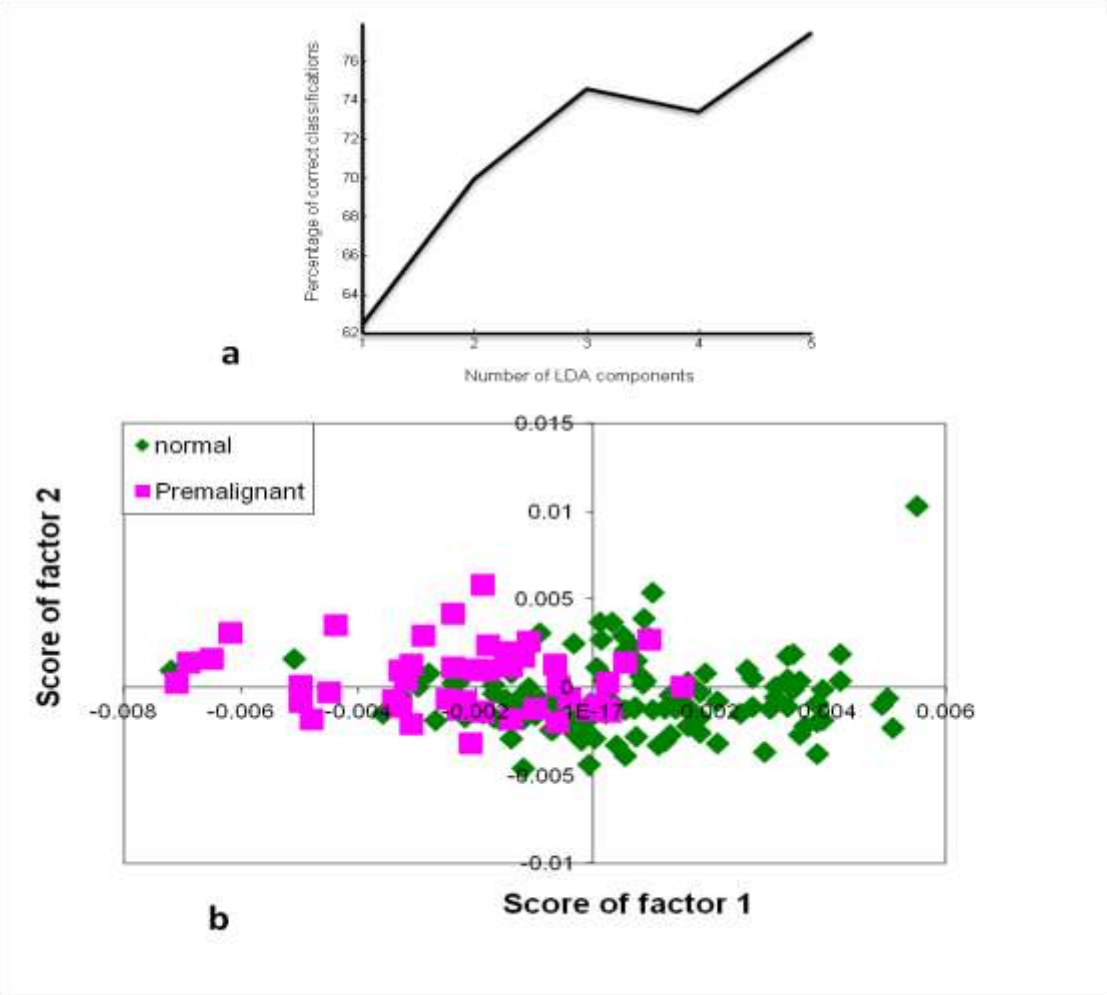


Figure 2.15. PC-LDA for exploring differences between serum from normal and premalignant subjects, a) Scree plot, b) Scatter plot

LOOCV confusion matrix indicates low efficiency in identifying these groups, only 95/126 normal (~75%) and 37/47 premalignant conditions (~79%) were correctly predicted (Table 2.13). This indicates similarity in normal and premalignant serum samples, in light of lack of appreciable cancer-related changes.

	Normal	Premalignant	Total	% efficiency
Normal	95	31	126	75.3
Premalignant	10	37	47	78.7

Table 2.13. PC-LDA for exploring differences between normal and premalignant serum-confusion matrix of LOOCV results.

b) Exploring classification between normal and disease control groups

PC-LDA for normal versus disease control was carried out using 5 factors (Figure 2.16a). The PC-LDA scatter plot shown in Figure 2.16b indicates two almost distinct clusters for normal and glioma.

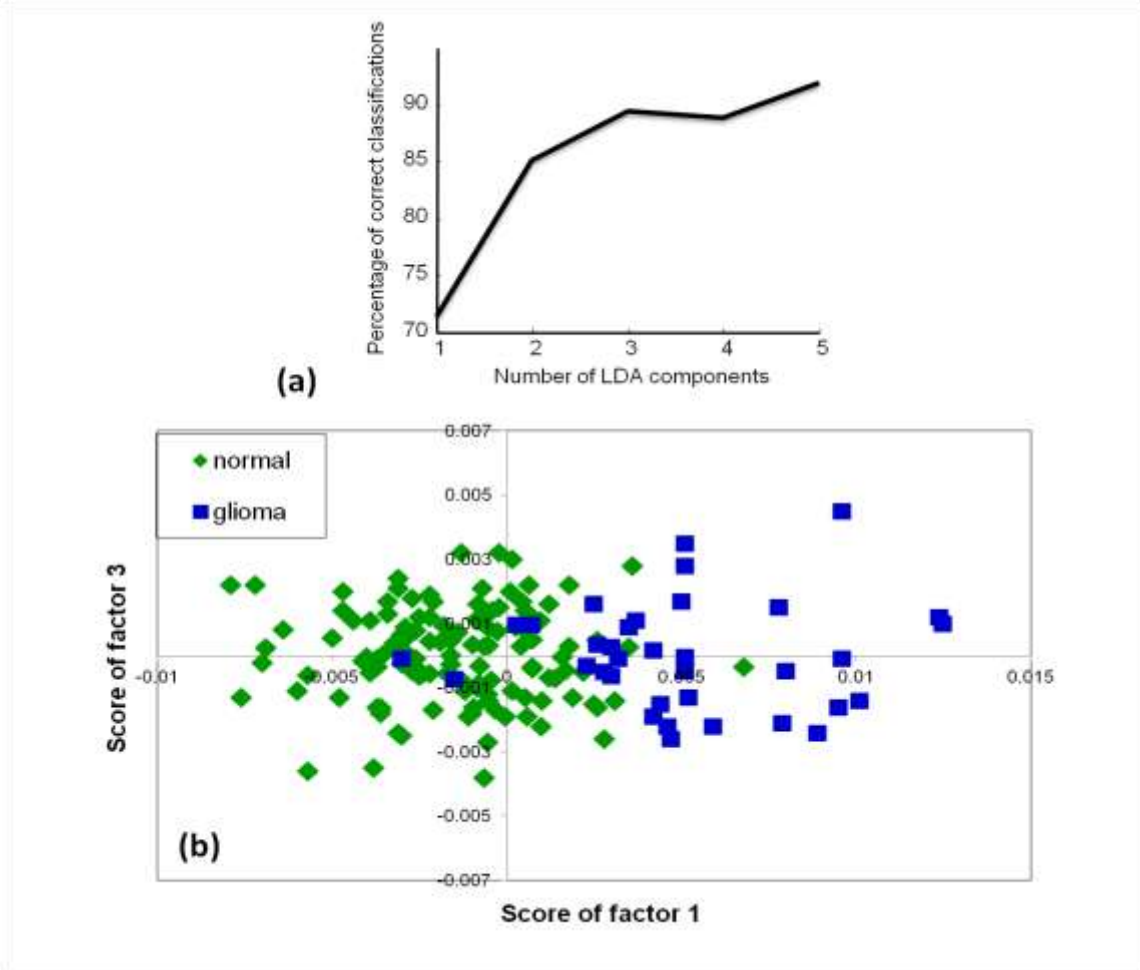


Figure 2.16. PC-LDA for exploring differences between serum from normal and disease control subjects, a) Scree plot, b) Scatter plot

The LOOCV confusion matrix in Table 2.14 indicates ~89% overall classification efficiency for identifying both disease and normal samples. Thus, healthy and disease samples can be identified with high sensitivity and specificity.

	Normal	Disease	Total	% efficiency
Normal	116	10	126	92.06
Disease	5	30	35	85.7

Table 2.14. PC-LDA for exploring differences between normal and disease control serum-confusion matrix of LOOCV results.

c) Exploring classification between normal and oral cancer groups

PC-LDA of normal versus oral cancer was carried out using 6 factors (Figure 2.17a). Scatter plot shown in Figure 2.17b demonstrates two distinct clusters each for normal and oral cancer.

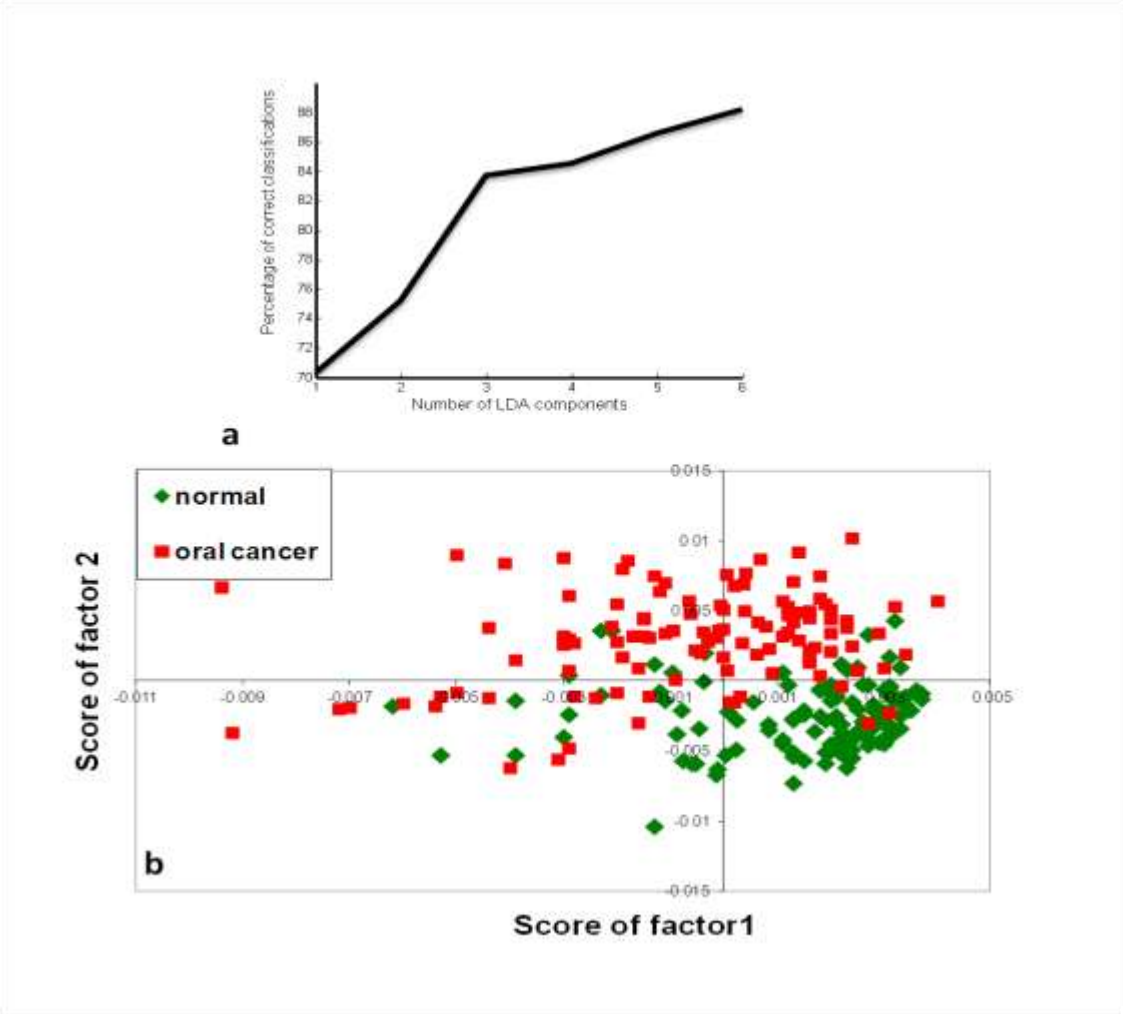


Figure 2.17. PC-LDA for exploring differences between serum from normal and oral cancer subjects, a) Scree plot, b) Scatter plot

Table 2.15a indicates LOOCV efficiency of 90% to identify normal while 84% for detection of cancer samples. To substantiate these findings, standard model of normal versus oral cancer was built using n=60 samples. The remaining 126 samples of normal and oral cancer were used as independent test dataset on this model. The LOOCV confusion matrix for the standard model and the test predictions are shown in Table 2.15b and 2.15c, respectively. In the standard model, normal and cancer could be identified with 88% and 83% efficiency, respectively. As shown in

Table 2.15c, 57/66 normal samples (86%) and 52/60 cancer samples (87%) were predicted correctly. Thus, test predictions indicate that normal and oral cancer samples can be identified with high sensitivity and specificity.

	Normal	Cancer	Total	% efficiency
Normal	113	13	126	89.6
Cancer	19	101	120	84.1

a

	Normal	Cancer	Total	% efficiency
Normal	53	7	60	88.3
Cancer	10	50	60	83.3

b

	Model N	Model C	Total	% efficiency
Test N	57	9	66	86.3
Test C	8	52	60	86.6

c

Table 2.15. PC-LDA for exploring classification between normal and oral cancer serum samples a) LOOCV confusion matrix of normal versus oral cancer, b) LOOCV confusion matrix of standard model for normal versus oral cancer, c) Test prediction using independent test data

d) Exploring cancer-specific diagnosis for oral cancers

Our studies have shown successful classification between normal and oral cancer samples. To determine if RS can differentiate between two different types of cancers, a non-oral cancer group i.e. glioma was also incorporated in this study. Additionally, specific diagnosis of cancers is important during cancer screening. Thus, the feasibility of differentiating these two different types of cancers was explored in 2-model and 3-model systems. Oral cancers included in this study are oral squamous cell carcinomas while gliomas are brain cancers that originate in the glial cells. Different serum biochemistry changes are reported in these cancers, as previously mentioned in the spectral analysis sections.

2-model system

PC-LDA of oral cancer vs. glioma samples was carried out using 10 factors to explore cancer-specific differences in serum (Figure 2.18a). The scatter plot shown in Figure 2.18b shows two distinct clusters for oral cancer and glioma.

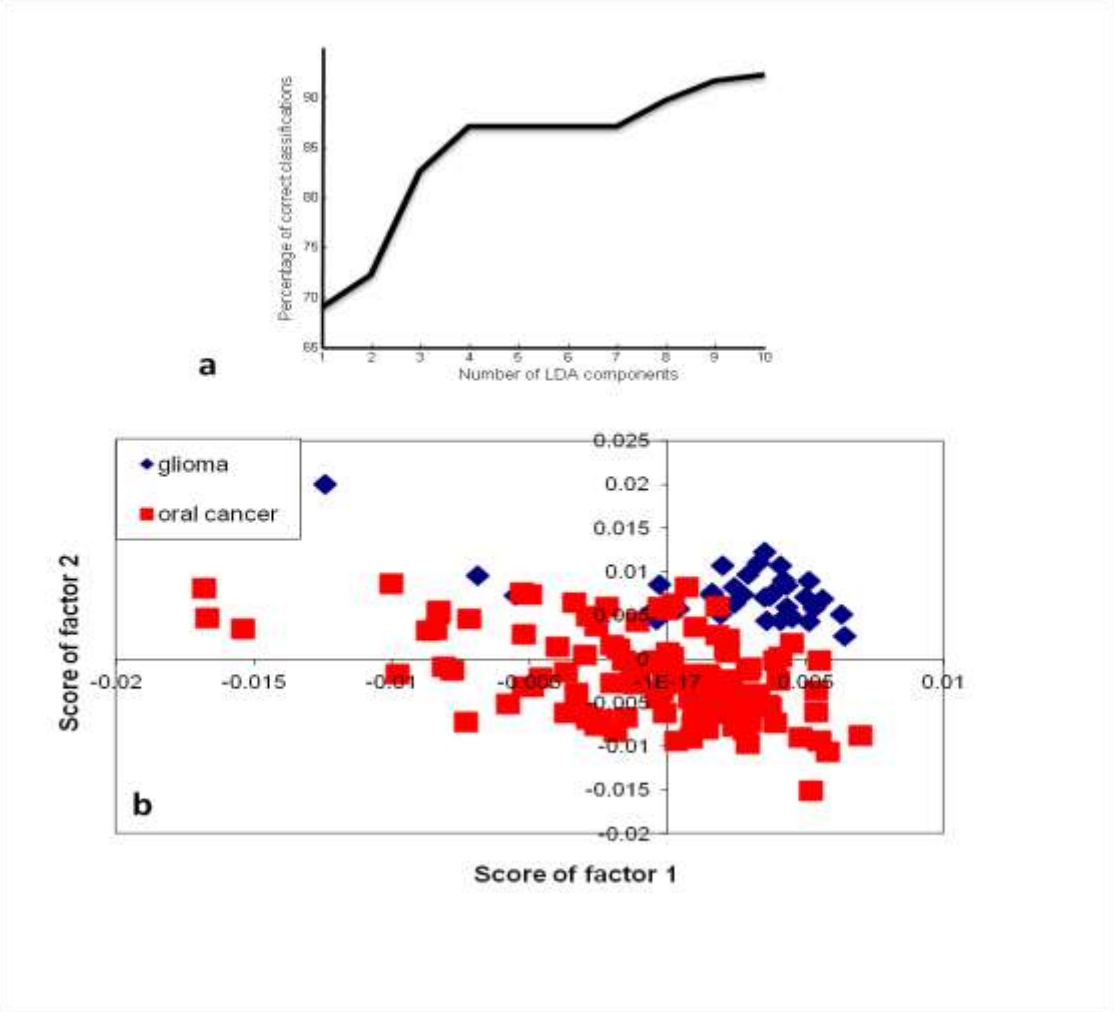


Figure 2.18. PC-LDA for exploring cancer specific differences between serum from glioma and oral cancer subjects, a) Scree plot, b) Scatter plot

The LOOCV table (Table 2.16) demonstrates correct predictions of 31/35 glioma samples and 107/120 oral cancer samples. Thus, these cancers were discriminated with ~89% efficiency.

	Glioma	Oral cancer	Total	% efficiency
Glioma	31	4	35	88.5
Oral cancer	13	107	120	89.1

Table 2.16. PC-LDA for exploring cancer specific differences between glioma and oral cancer serum- confusion matrix of LOOCV results

These results indicate that different cancers have different Raman spectroscopic signatures and may be identified distinctly based on serum spectral characteristics.

3-model system

Distinction between normal and cancer, and simultaneous detection of type of cancer (oral cancer or glioma) was also evaluated using a 3-model system. PC-LDA was carried out using 10 factors accounting for ~79% efficiency (Figure 2.19a). PC-LDA scatter plot shows 3 slightly overlapping clusters corresponding to normal, oral cancer and disease control (Figure 2.19b).

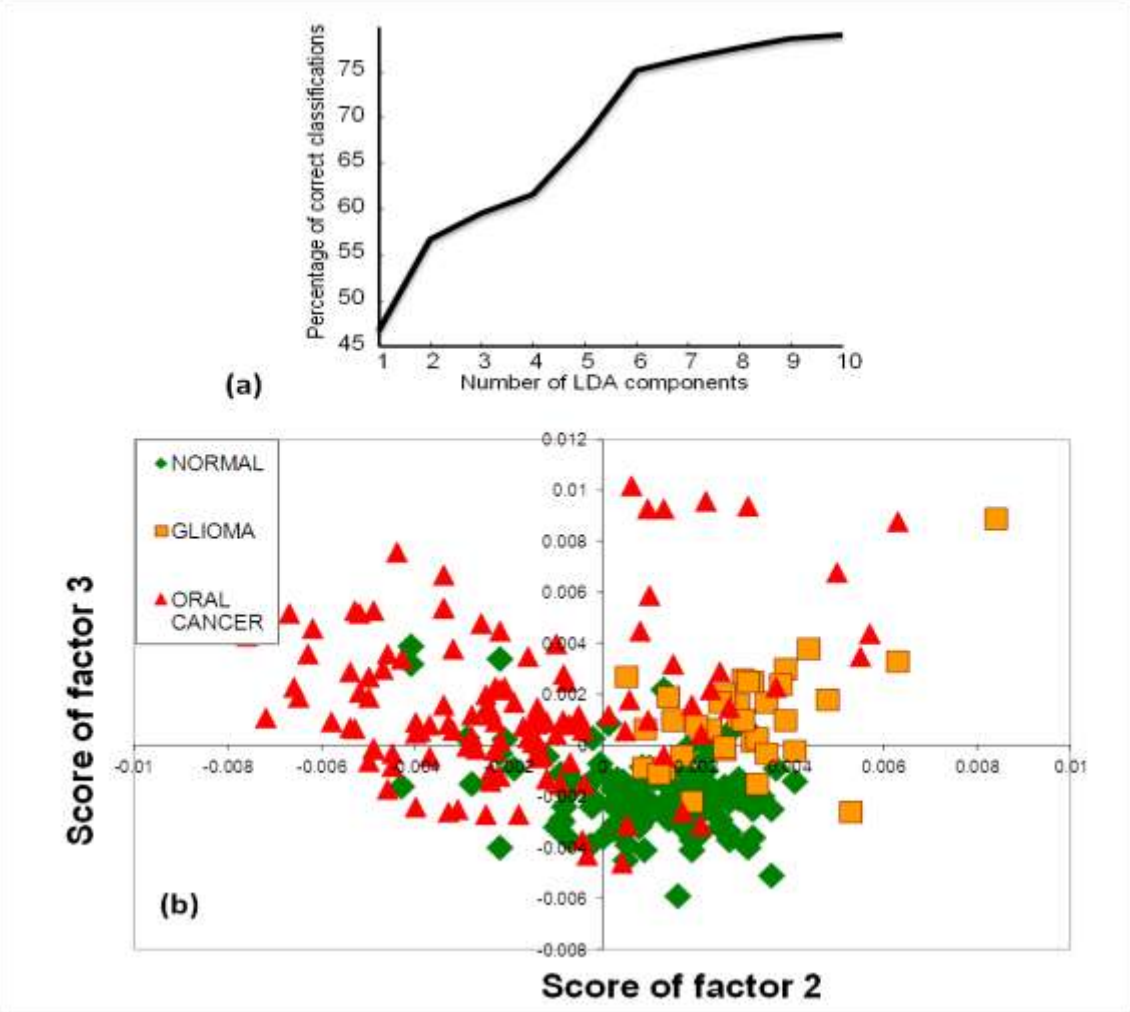


Figure 2.19 PC-LDA for exploring differences between serum from normal, disease control and oral cancer subjects, a) Scree plot, b) Scatter plot

The LOOCV confusion matrix shown in Table 2.17 demonstrates the feasibility of identifying normal with ~80% efficiency. Glioma and oral cancer could be identified with 77% and 75% efficiency, respectively.

	Normal	Glioma	Oral cancer	Total	% efficiency
Normal	100	15	11	126	79.3
Glioma	8	27	0	35	77.1
Oral cancer	17	13	90	120	75

Table 2.17 PC-LDA for exploring cancer specific differences between normal, oral cancer and glioma serum- confusion matrix of LOOCV results

These results indicate that different cancers have different Raman spectroscopic signatures and may be identified distinctly based on serum spectral characteristics, even in a robust 3-model system.

e) Investigating serum RS for screening applications

Two classification models were evaluated to confirm the potential of serum RS in oral cancer screening; first a single 4-group model and second a normal versus abnormal model followed by determining the type of abnormality.

Exploring classification between normal, disease control, oral premalignant and cancer sera using a single 4-model system

In the first screening model, a 4-group model approach was used to explore classification between normal, disease control, premalignant and oral cancer. This is an ideal model for screening applications, as all possible groups constitute this model. Thus, a single step analysis

will be required to ascertain the identity of the sample during screening. PC-LDA was carried out using 7 factors which accounted for ~64% correct classifications. The scree plot and PC-LDA scatter plot are shown in Figure 2.20.

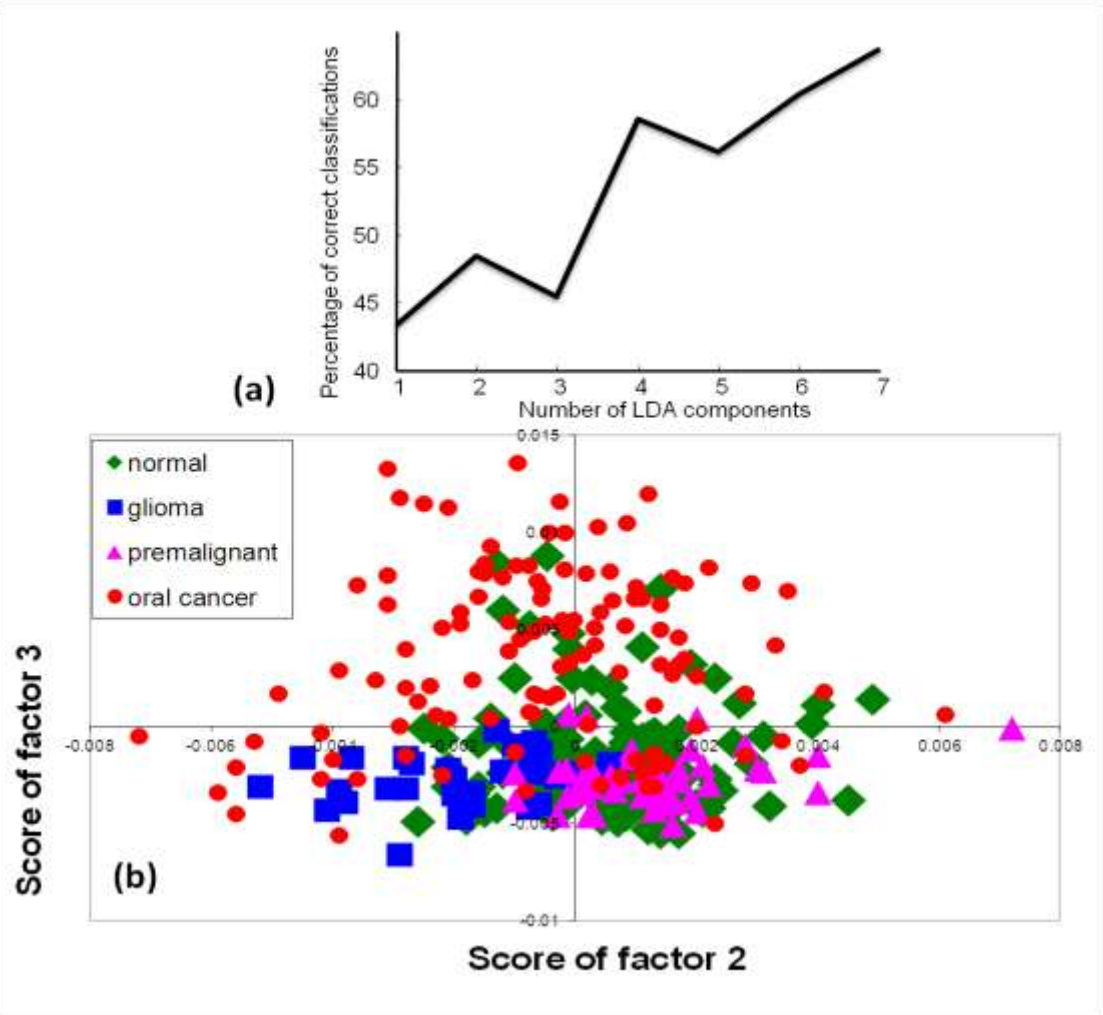


Figure 2.20. PC-LDA for exploring oral cancer screening applications- Model I, Single 4-group model of normal, pre-malignant, disease control and oral cancer a) Scree plot, b) Scatter plot

As shown in Table 2.18, several misclassifications between normal, disease and premalignant groups were observed. About 15% and 25% normal misclassified with disease and premalignant,

respectively while only 10% misclassified with cancer group. High misclassifications of normal with premalignant group could be due to less severe changes in premalignant group, as opposed to major biochemical changes in cancer group. In case of cancer, 15% misclassified with normal, 10% with disease and 8% with premalignant groups. Thus, overall a specificity of ~90% and sensitivity of ~85% was observed for normal versus cancer classification. Premalignant and disease controls could also be identified with 71% and 66% efficiency.

	Normal	Disease	Premalignant	Oral cancer	Total	% efficiency
Normal	63	17	33	13	126	50
Disease	7	25	3	0	35	71.4
Premalignant	11	5	31	0	47	65.9
Oral cancer	17	11	9	83	120	69.1

Table 2.18. PC-LDA for exploring oral cancer screening applications- Model I, Single 4-group model for classification between normal, disease control, premalignant and oral cancer- confusion matrix of LOOCV results

Exploring classification between normal and abnormal

In the second classification model, feasibility to identify a healthy/normal serum sample against all other pathological/abnormal conditions was explored. Thus, in this first step, classification between normal group and a combined abnormal group (disease + premalignant + oral cancer) was explored. In this classification system, if a sample would be classified as abnormal, the

second step would be carried out to detect the type of abnormality: disease, premalignant or malignant. This could also be an ideal approach for screening-related applications; where healthy samples are excluded and abnormal samples are subjected to further round of confirmatory analyses. The first normal versus abnormal PC-LDA classification was carried out using 6 factors accounting for 71% classification efficiency (Figure 2.21a). The scatter plot shown in Figure 2.21b indicates two slightly overlapping clusters corresponding to normal and abnormal.

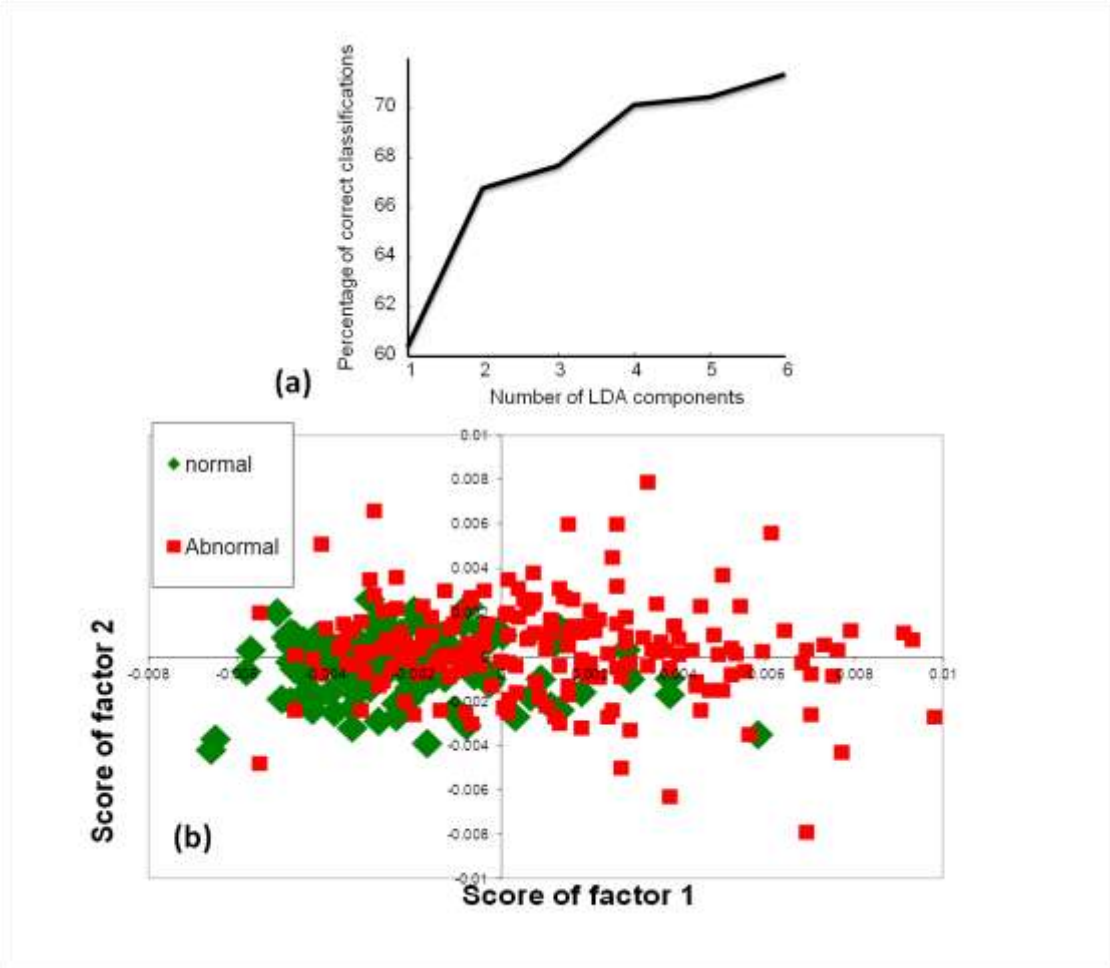


Figure 2.21. PC-LDA for exploring oral cancer screening applications- Model IIa, Normal versus all abnormal subjects, a) Scree plot, b) Scatter plot

As shown in Table 2.19, normal could be identified with ~80% efficiency while abnormal or pathological could be identified with 64% efficiency. Thus, the sensitivity and specificity for this approach was found to be ~64% and ~80%, respectively.

	Normal	Abnormal	Total	% efficiency
Normal	101	25	126	80.1
Abnormal	73	129	202	63.8

Table 2.19. PC-LDA for exploring oral cancer screening applications- Model IIa, Normal versus all abnormal- confusion matrix of LOOCV results

The feasibility to differentiate the different abnormal conditions was explored in the next step. PC-LDA was carried out using 7 factors accounting for ~78% efficiency (Figure 2.22a). The scatter plot shown in Figure 2.22b indicates two overlapping clusters for disease control and premalignant while a slightly distinct cluster for oral cancer.

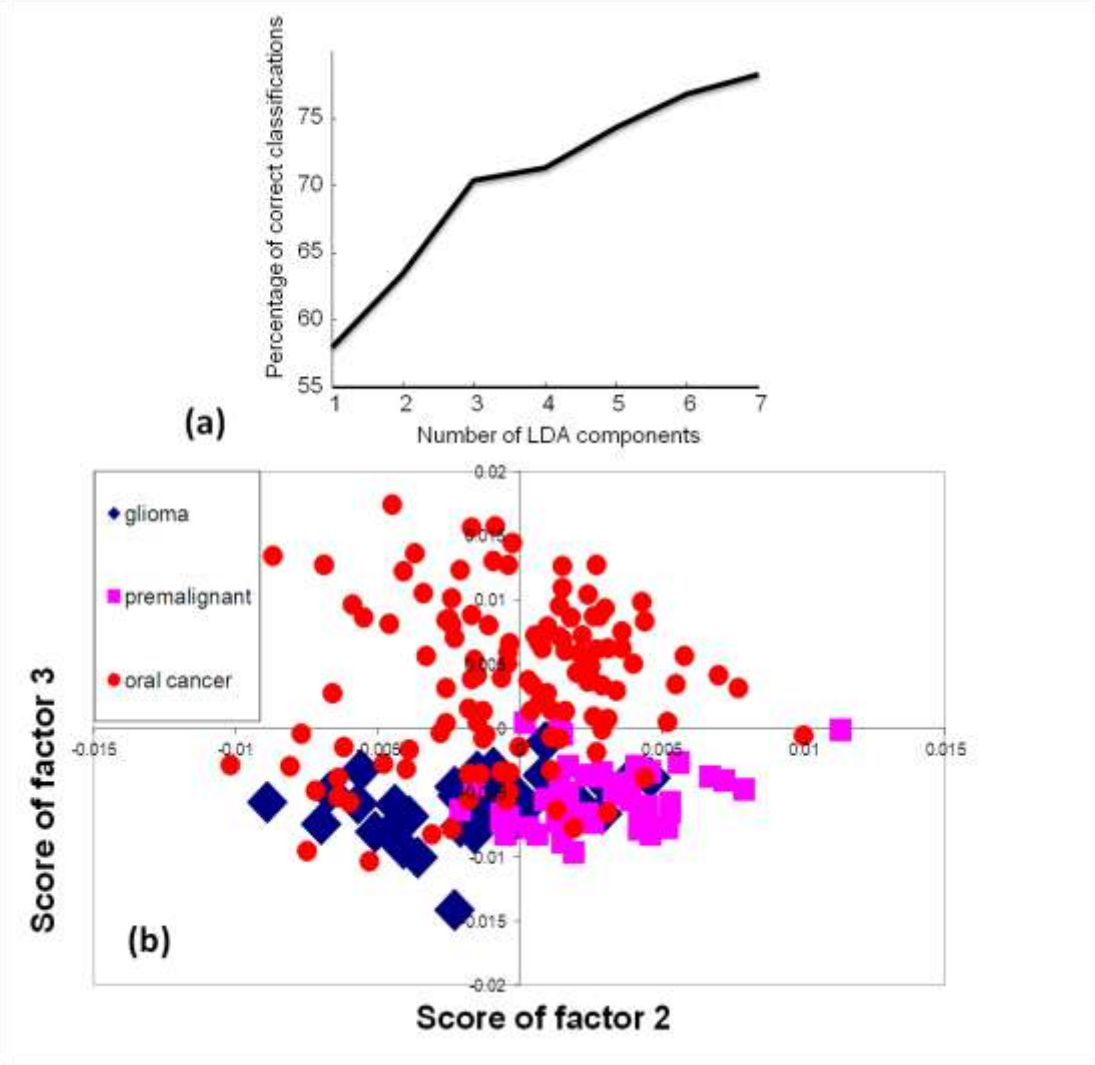


Figure 2.22. PC-LDA for exploring oral cancer screening applications- Model IIb, Determining type of abnormality- pre-malignant, disease control and oral cancer a) Scree plot, b) Scatter plot

LOOCV indicates that 26/35 disease subjects could be correctly classified, 36/47 premalignant and 91/120 cancers were correctly classified (Table 2.20). The sensitivity to detect each abnormal condition was 74%, 77% and 75% for disease, premalignant and cancer, respectively.

	Glioma	Premalignant	Oral cancer	Total
Glioma	26	8	1	35
Premalignant	9	36	2	47
Oral cancer	14	15	91	120

Table 2.20. PC-LDA for exploring oral cancer screening applications- Model IIb, Determining type of abnormality- pre-malignant, disease control and oral cancer - confusion matrix of LOOCV results

Most disease misclassifications were observed with pre-malignant samples and vice versa. About 10% cancer spectra were misclassified as disease control group. This could be attributed to the fact that both groups have subjects suffering from cancer, either oral cancer or glioma. Few premalignant (2/47) and malignant (15/120) misclassifications were observed between these two classes. The premalignant subjects (n=2) classifying as malignant may be indicative of imminent malignant transformation, though no follow up of these subjects was conducted in the present study. The lower rates of identifying premalignant as cancer could be attributed to the fact that only about 4% Leukoplakia and 5-10% OSMF transform into malignancy [45,56]. During screening applications, such samples can nevertheless be termed as suspicious and followed up for malignancy after primary screening. Overall in this analysis, disease could be identified with 74%, premalignant with 77% and oral cancer with ~75% efficiency. But, the identification of abnormal condition against normal in the first step had a lower sensitivity rate of ~64% and higher specificity rate of 80%. For an ideal cancer screening test, higher sensitivity is desirable- to eliminate possibility of losing any malignant cases to false negative diagnosis. Although the

lower specificity rates may also be a concern, the samples falsely classified ‘positives’ will only be treated as ‘suspicious’; and all these ‘suspiciously-abnormal’ samples will be deemed as ‘Cancer’ only after several rounds of confirmatory tests.

2.4. Exploring serum-based RS for recurrence detection in oral cancers

For head and neck cancers, which include oral cancers, approximately one-third of patients treated with surgery and adjuvant therapy experience recurrence ((loco-regional, relapse, second primary and second field tumors) and/or distant metastasis. The rates of oral cancer recurrence in patients administered standard treatment vary from 18 to 76%[340], while the overall 5-year survival of 50% rate has not improved in decades. Early detection of recurrence is clinically important[341], as it can lead to improved prognosis, optimized treatment and enhanced overall outcomes. The existing methods such as presence of cervical lymph nodes metastasis, extra capsular spread and positive histopathological margin are inadequate.

Recent genetic and molecular studies on 3p14 and 9p21 chromosomal loss; p53 mutations in surgical margins have shown identification of recurrence-prone patients. A study by Reis et al has elucidated a 4-gene signature (MMP1, COL4A1, P4HA2 and THBS2) in histologically normal margins that may be predictive of oral cancer recurrence. Another study has identified 4 sub-groups of HNSCC, the subgroup with epidermal growth factor receptor (EGFR)-associated profile and activation of NF-κβ signaling genes had poor prognosis[342-345]. Serum tumor markers may also hold promise in identifying recurrence. The tumor/recurrence-related markers with potential in determining prognosis include CEA for colorectal, CA 15-3, CEA, cMethDNA, serum testosterone levels for breast, AFP for liver, CA-125 for ovarian, PSA and acid phosphatase (ACP) for prostate cancer[287,346-351]. Several studies have also demonstrated the presence of cell-free DNA: host, tumor or virus associated as diagnostic and prognostic markers of cancers like colorectal, cervical, nasopharyngeal cancers[352-355]. Another recent study has

shown utility of HPV-DNA in blood and saliva to predict recurrence in HPV-associated oral cancer patients[356]. No definite marker for recurrence prediction in non-HPV oral cancers has been established till date. Further, literature suggests that a single marker may not be efficient in detection of recurrence. A multiplex panel of several proteins and nucleic acids is being investigated for recurrence detection of several cancers[357-359]. Proteomic profiling of serum for detection of tumor/recurrence markers has been carried out for several cancers. In this context, an approach encompassing proteomics, genomics and metabolomics may be ideal for recurrence detection, and one such approach is Optical spectroscopy. As potential of serum RS in diagnosis of oral cancers was demonstrated, the feasibility of detecting recurrence was also explored.

2.4.1. Materials and Methods

2.4.1.1. Subject details

Patients harboring primary oral squamous cell carcinoma of the oral cavity who visited the TMH, Mumbai OPD were screened for this retrospective study. A criterion of recurrence and non-recurrence was devised as follows: Subjects who reported a recurrence within 2 years of follow up were referred to as 'Recurrence subjects', while the subjects with no reported recurrence for up to 2 years of follow up were called as 'Non-recurrence subjects'. Ten subjects (n=10) fulfilled the criteria for recurrence (mean time for development of recurrence: 6 months), while n=12 fulfilled the criteria for non-recurrence. Thus, a total of 22 subjects were included in this study.

2.4.1.2. Sample collection

Blood was collected from these patients at 2 time points: before and after surgery. Blood samples collected after overnight fasting, prior to any surgery-related interventions was termed "before surgery" while blood collected 1 week post-surgery (before any adjuvant cancer treatment like chemoradiotherapy) was termed "after surgery". All recruited patients were cases without prior anticancer treatment, history of malignancy, and second primary cancers. Serum was separated as described in section 2.1.1.2.

2.4.1.3. Raman spectral acquisition

Spectra were acquired by placing 30µl volume of serum samples on CaF_2 window, recorded using fiber-optic Raman microprobe (Horiba-Jobin-Yvon, France) as described previously in section 2.2.1.3.

2.4.1.4. Spectral pre-processing and data analysis

The acquired Raman spectra were pre-processed as detailed in section 2.1.1.4. Multivariate analysis was carried out using PC-LDA followed by LOOCV. Average spectra were also computed.

2.4.2. Results

2.4.2.1. Spectral features

Major spectral features shown in Figure 2.23a-d include 830 and 850 (Tyr doublet), 1008 (Phe), 1265 (Amide III), 1316, 1320 and 1335 (DNA - related bands), 1450 (CH_2 bending) and 1660 cm^{-1} (Amide I) regions. In before surgery spectra, minor differences between the recurrence and non-recurrence groups were seen at 1260, 1313, 1339, 1450 and 1650 cm^{-1}. These differences

correspond to changes in DNA and protein in these groups. In the after surgery spectra, major differences between the recurrence and non-recurrence groups were observed at 936, 949, 1007, 1126, 1260, 1315, 1335, 1450 and 1657 cm^{-1}. These differences also correspond to changes in DNA and protein across the two groups.

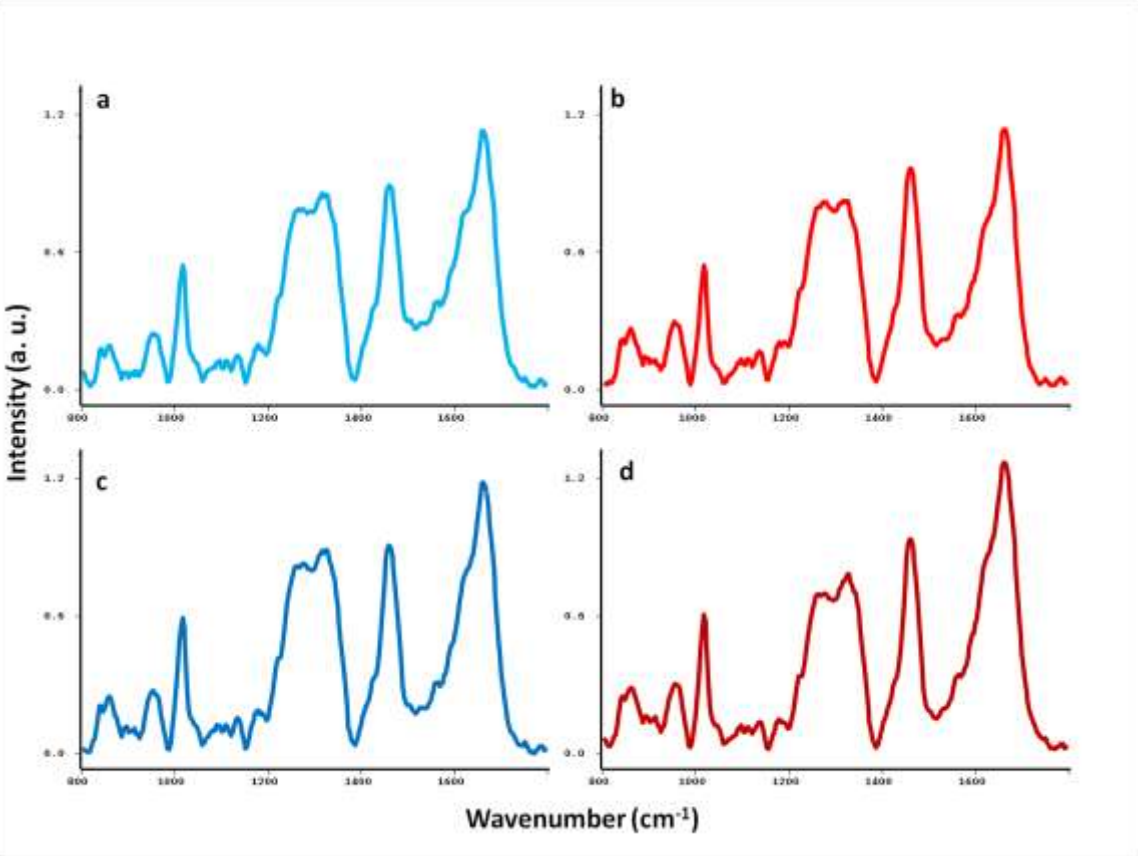


Figure 2.23. Mean Raman spectra of before surgery serum, a) recurrence, b) non-recurrence, and after surgery serum, c) recurrence, d) non-recurrence, for spectral comparisons

The additional DNA and protein signals could originate from either minimal residual cancer or field cancerization. This corroborates with findings that demonstrate high circulating DNA levels in the recurrence group, even higher than the primary cancer group; along with up-regulation of

several proteins in the sera of recurrence patients. Thus, this additional DNA and protein content could be ascribed to recurrence-related factors.

2.4.2.2. Multivariate analysis

PC-LDA followed by LOOCV was employed for data analysis. First, differences between the recurrence and non-recurrence groups were analyzed in *before surgery* sera. Next, the same approach was adopted for *after surgery* sera.

a) Investigating differences in before surgery serum samples

PC-LDA was employed to explore classification between the groups. Three factors were used for the analysis, accounting for ~81% correct classifications (Figure 2.24a). Scores of factor 1 and 2 were employed to obtain scatter plots, as shown in Figure 2.24b. Two overlapping clusters were observed in the scatter plot.

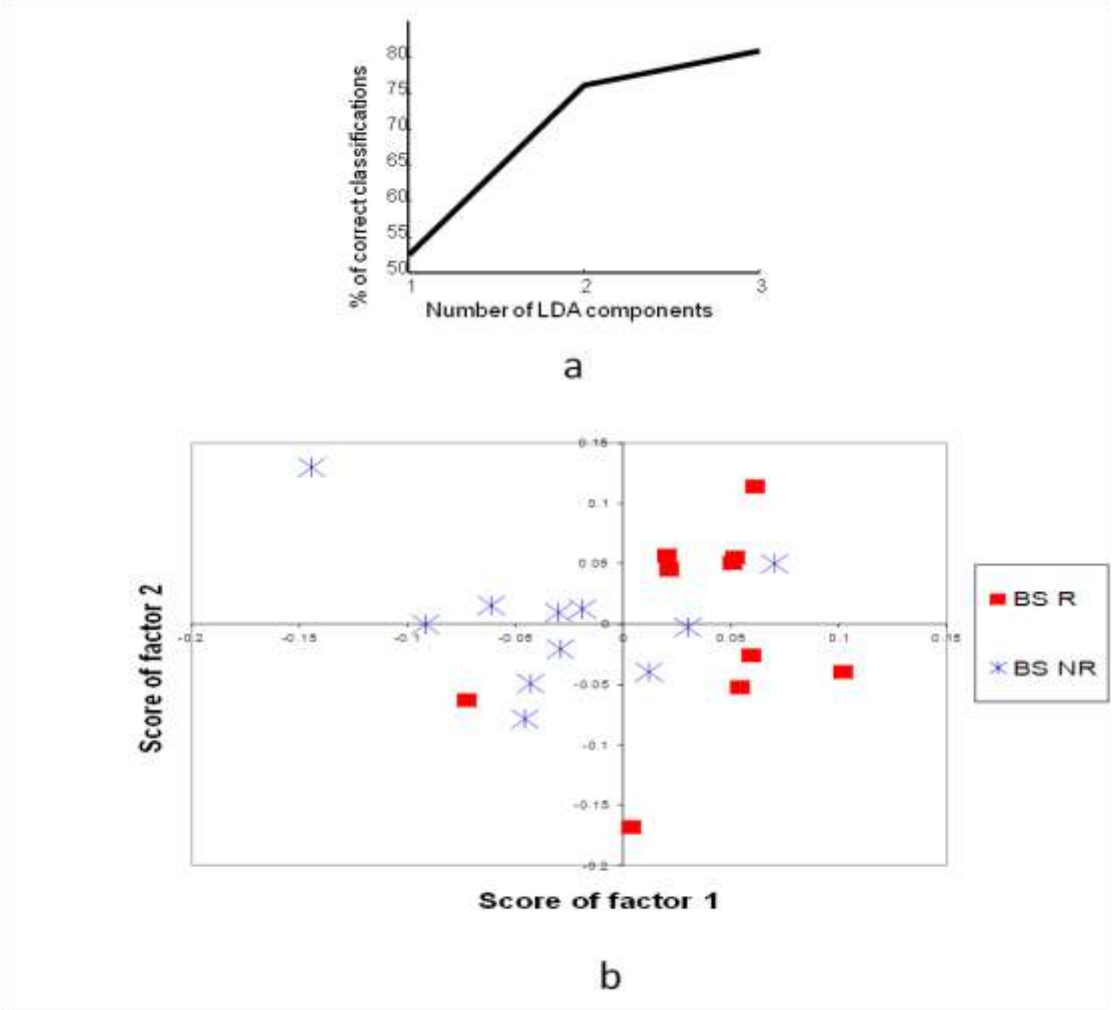


Figure 2.24. PC-LDA for exploring recurrence-specific differences in before surgery serum samples, a) Scree plot b) Scatter plot (BS- before surgery, R-recurrence, NR-non recurrence)

On LOOCV (Table 2.21), 7/10 recurrence spectra and 4/11 non-recurrence spectra were correctly classified, to yield a classification efficiency of 70% and 36%, respectively. A large number of misclassifications of non-recurrence group with the recurrence group were observed. However, a minor tendency of classification was observed for recurrence group.

	BS- R	BS- NR	total	% efficiency
BS- R	7	3	10	70
BS- NR	7	4	11	36.3

Table 2.21. PC-LDA for exploring recurrence-specific differences between before surgery recurrence and non-recurrence serum- confusion matrix of LOOCV results (BS- before surgery, R-recurrence, NR-non recurrence)

b) Investigating differences in after surgery serum samples

Recurrence-related differences could not be detected in before surgery samples. This may be possibly attributed to the presence of additional tumor-associated factors (along with normal serum constituents and recurrence-related factors, if any). These factors may be eliminated from circulation by surgical excision of tumor, and may facilitate detection of recurrence-related factors, if any. Further, recurrence can develop due to two main reasons, minimal residual cancer (MRC) and field cancerization (FC). Thus, recurrence prediction is based on factors arising from MRC, FC or both. As MRC factors can only be detected post-surgical excision of tumor, blood samples collected post-surgery were also analyzed for recurrence detection.

PC-LDA was carried out with 3 factors that accounted for ~ 82 % classifications (Figure 2.25a). The scatter plot for PC-LDA (score of factor 1 vs. score of factor 3) is shown in Figure 2.25b. Two well-separated groups were observed in the scatter plot.

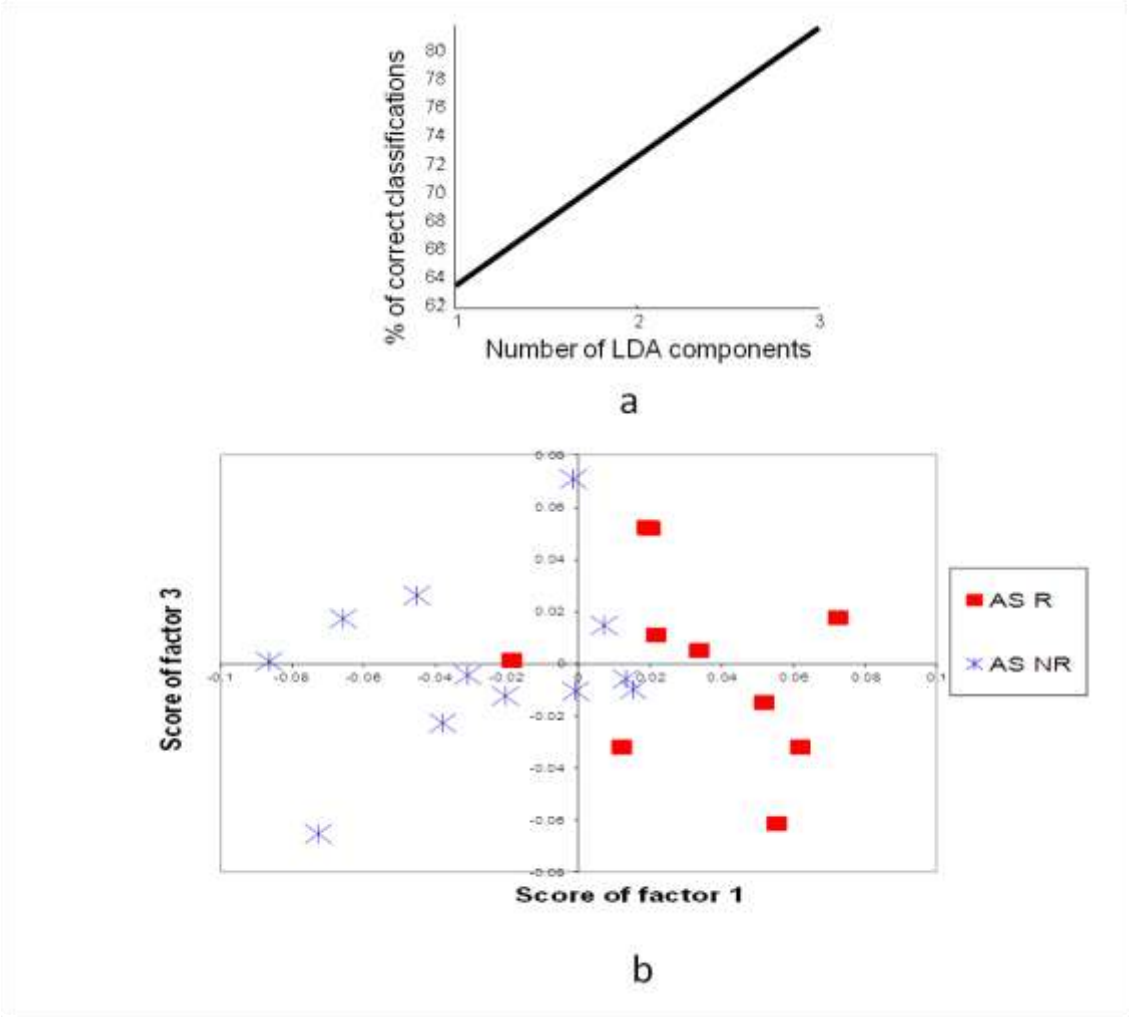


Figure 2.25. PC-LDA for exploring recurrence-specific differences in after surgery serum samples, a) Scree plot b) Scatter plot (AS- after surgery, R-recurrence, NR-non recurrence)

The confusion matrix for LOOCV results, presented in Table 2.22 indicates 8/10 and 8/12 correct predictions for recurrence and non-recurrence, to yield a classification efficiency of 80% and 75%, respectively. Thus, recurrence and non-recurrence groups could be classified with average classification efficiency of ~78%.

	AS- R	AS- NR	total	% efficiency
AS- R	8	2	10	80
AS- NR	4	8	12	75

Table 2.22. PC-LDA for exploring recurrence-specific differences between after surgery recurrence and non-recurrence serum- confusion matrix of LOOCV results (AS- after surgery, R-recurrence, NR-non recurrence)

After removal of tumor, normal serum constituents could be major contributors to Raman spectra. However, some recurrence related factors persisting in sera of recurrence subjects may enable classification between the two groups. Further, as influence of confounding tumor-related factors may have been removed by surgical excision of tumor, these recurrence-related factors may have played a major role in classification of the groups. Thus, blood from oral cancer patients after surgery may have the potential to identify those at high risk of developing recurrence. Recurrence may have been detected by factors arising from MRC, FC or both[360,361]. Additionally, the differences could be attributed to recurrence-related factors like elevated protein and DNA and also corroborated by several molecular studies[353-359,362].

Although the exact underlying reason for differences between the recurrence and non-recurrence groups can only be speculated, the finding that the recurrence-prone oral cancer patients can be identified using serum RS has important clinical implications. Conventionally, oral cancer patients with poor prognosis or lymph node metastasis are administered adjuvant chemo-radiotherapy to ensure a comprehensive treatment and prevent recurrence. If recurrence-prone subjects could be identified by the current methodology one week post surgical excision of

tumor, similar comprehensive treatment regimens can be planned for such patients. Irrespective of histopathological grading and/or nodal metastasis (which may prove to be inadequate), adjuvant treatment post surgery can be administered to such patients. Stringent bi-monthly or monthly follow-ups along with regular imaging modalities to detect even any occult suspicious lesions can be planned. In event of lesion confirmation, treatment options can then be weighed to promote patient life-quality and decrease morbidity.

2.5. Summary

Screening of cancers using a peripheral blood sample may be a more practical and feasible approach. This approach is associated with multiple advantages like accessibility, low invasive procedure, low cost, and multiple sampling[363]. Further, samples can be transported from primary screening centers to a centralized facility for analysis. Studies carried out in this chapter aimed at evaluating serum-based Raman spectroscopy for management of oral cancers.

- In the first pilot study using resonance Raman approach, normal and oral cancer serum samples could be successfully discriminated. Major spectral discriminants were beta-carotene, DNA, protein and amino acids.

- In the consequent pilot study using conventional Raman and 785 nm excitation-the most commonly employed wavelength for biological applications, similar classification efficiency (~78%) was obtained for normal and cancer groups.

- As the two pilot studies demonstrated potential in oral cancer detection, serum Raman spectroscopy of premalignant, disease control and large sample size of normal and oral

cancer samples were evaluated in the subsequent study. While the premalignant samples could be identified with ~75% efficiency, normal and disease controls could be differentiated with a higher efficiency of ~89%. The analysis on a large sample size and independent test data evaluation confirms the discrimination between normal and oral cancer groups with an average efficiency of ~86%.

- Screening applications were also investigated in this study. Using a normal versus abnormal screening model, a sensitivity of 64% and specificity of ~80% was obtained. Although the sensitivity and specificity rates are not very high, they are quite comparable to established screening techniques like Pap test and mammography for cervical and breast cancers, respectively.

- Recurrence detection in oral cancers using serum Raman spectroscopy on a small cohort was explored by using before and after surgery blood samples. Findings demonstrate feasibility of identifying recurrence in after surgery samples.

Chapter 3

Raman oral exfoliative cytology for oral cancer diagnosis

Introduction

Exfoliative cytology is a simple, rapid, and less invasive technique[364]: it is thus well accepted by patients and is suitable for routine application in population screening programmes. This methodology has been used as a tool for screening gynecological malignant diseases, after development of Papanicolaou (Pap) staining by George Papanicolaou[365,366]. Pap staining of oral exfoliated cells was first explored to study keratinization patterns and changes during menstrual cycle[367-369], followed by studies on oral cancer diagnosis[370,371]. Although conventional cytopathology demonstrated potential in early detection of disease in individual exfoliated cells this technique could not achieve the expected sensitivity and specificity at that time[120]. Cytopathological examination involved intimately with this approach is inherently limited to subjective interpretations of nuclear integrity, cellular size and structure, and stain uptake. Further, diagnosis of pre-cancer samples harboring low-grade or high-grade dysplasia only by morphological analysis is difficult as definite cellular alterations may not be present in most cells. After several modifications in sample collection and staining methods, introduction of computerized image analysis, DNA cytometry and addition of molecular tests like detection of p53 mutation and loss of heterozygosity, this area has received renewed interest in the last few years[121,372-374]. Ancillary methods like DNA image cytometry and Argyrophilic nucleolar organiser region (AgNOR) analysis are undertaken to allow more precise cytological diagnosis. Studies combining cytology with DNA cytometric evaluation have shown significantly higher sensitivity and specificity rates. Such multimodal cell analyses are more useful when only few cytologically abnormal (atypical) cells are present. [375-379]

Alternatively, these cells can be subjected to spectroscopic analysis which can yield inherent biochemical signatures. As malignant transformation is associated with numerous cellular biochemical changes, and morphological changes are preceded by biochemical changes, spectroscopic identification of early cancer cells may be possible. Vibrational spectroscopic measurements on exfoliated cells were first carried out in the 1990s. The earliest studies provided evidence for extensive structural changes during carcinogenesis as discerned by IR spectroscopy[380]. Consequent studies evaluated potential of FTIR versus Pap staining and investigated improved methods of sample analysis, including removal of confounding factors like inflammatory and red blood cells and micro-organisms[381-383]. IR micro-spectroscopy studies on oral exfoliated cells indicated sensitivity of spectra to stages of cell maturation and also investigated variance of the spectra of individual human epithelial cells. Spectral cytopathology-IR micro-spectroscopy followed by multivariate data analysis, could detect dysplastic, neoplastic and viral-infected cells. Further, classification of healthy oral squamous cells according to their anatomical origin in the oral cavity was also demonstrated. Raman spectroscopic pellet-based studies on cervical exfoliated cell specimens have shown successful discrimination of normal and abnormal specimens, and also detected HPV infection[34,239]. Recent Raman spectroscopic studies have demonstrated difference between normal and cervical intraepithelial neoplasia (CIN) cytology using ThinPrep approach, while another study using monolayer of cells has discriminated normal, dysplastic and SCC cell lines[280,384]. Raman spectroscopic studies on discrimination of normal, pre-cancer and cancer exfoliated cells from oral cavity are hitherto unreported. Raman oral exfoliative cytology followed by cytological analysis (Pap staining) for diagnosis of oral pre-cancer and cancer was therefore explored. In the first section, a brief

explanation of cytological aspects has been presented. In the second section, proof of principle studies to confirm feasibility of the approach was demonstrated on normal and cancer specimens. In the next step, feasibility of early diagnosis was explored using pre-cancer specimens. Pellet based studies were carried out by acquiring multiple spectra from different areas of the pellet in both studies. Data analysis for both proof-of-concept studies and early diagnosis were carried out using 2 approaches: spectra-wise and patient-wise approaches by PC-LDA followed by LOOCV.

3.1 Cytological diagnosis

After spectra acquisition, cell pellet was smeared onto glass slides, fixed with a fixative agent and subjected to Papanicolaou staining. Thus, cytological diagnosis was carried out for same pellet subjected to Raman spectral acquisition.

Oral epithelium is a stratified squamous epithelium. Specifically, lining mucosa which is a non-keratinizing epithelium covers buccal, labial, and vestibular surfaces, ventral tongue and floor of mouth. The epithelium is composed of a single layer of cuboidal cells, called the stratum basale. The layers adjoining the basal layer are sometimes termed as parabasal cell layer. The next cell layer is called the stratum intermedium where large cells appear oval and somewhat flattened. The third or superficial layer is termed the stratum superficial where cells are polygonal and contain pyknotic nuclei. These three cell layers of the epidermis form the non-keratinized epithelium of the oral mucosa. The epithelia of the oral mucosa are in a constant state of renewal, and the basal and parabasal cells show the highest mitotic activity.

Sample adequacy was judged by the Bethesda system[385,386]. In general, an adequate and representative sample would include superficial, intermediate and occasional basal/ parabasal cells. As there are no standard grading systems like the Bethesda system for evaluation of oral

cytological smears, grading was based on the available literature. The cell populations investigated in the oral smears are listed below and also shown in Figure 3.1.

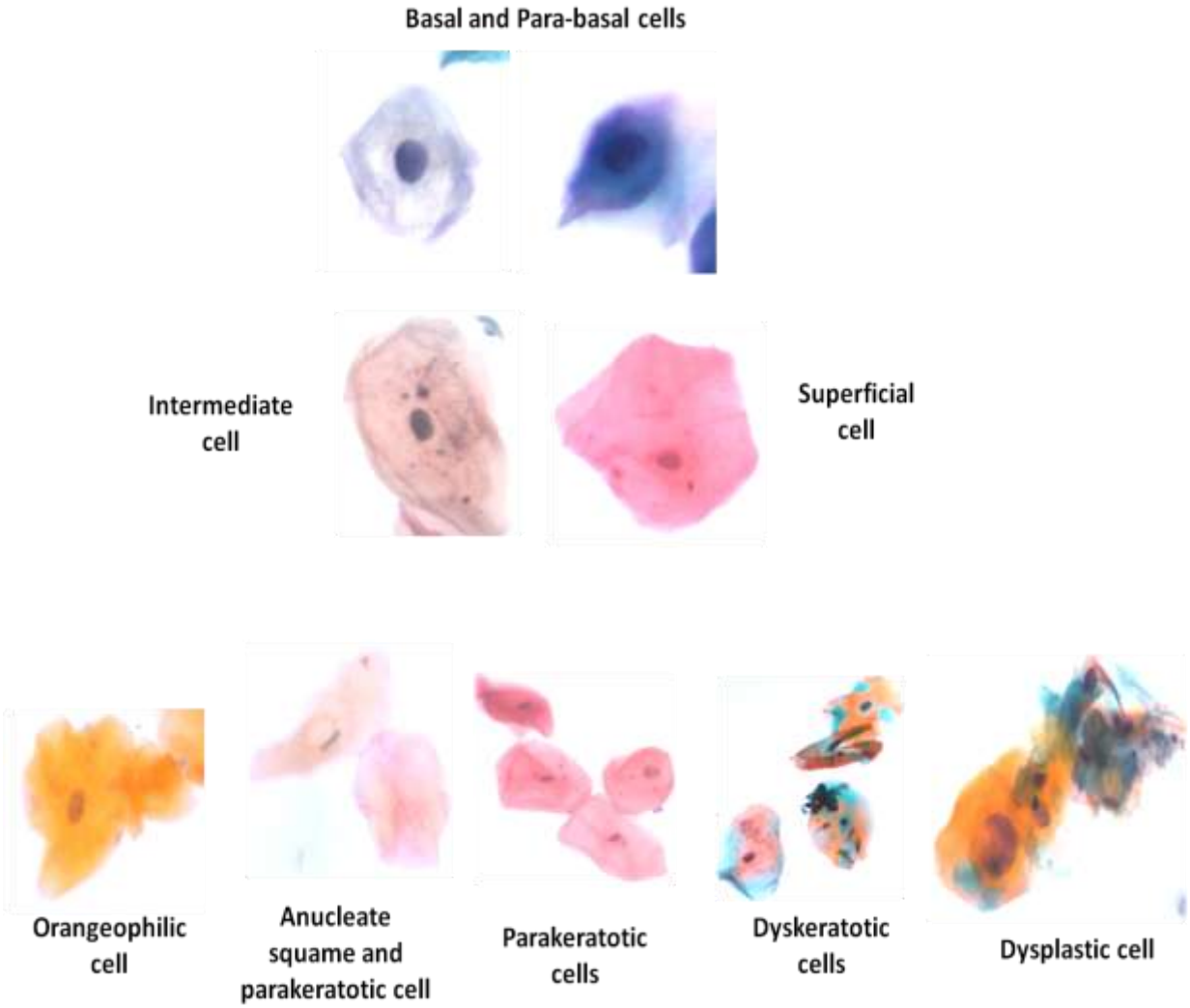


Figure 3.1. Representative images of normal and abnormal cells from clinically normal mucosa and mucosal lesions upon Papanicolaou staining (400X)

- Normal epithelial cells- basal/parabasal, intermediate and superficial.

The basal and parabasal cells are characterized by high nuclear:cytoplasmic (N:C) ratio, however the nuclear features are smooth and chromatin is pale. This distinguishes them from malignant cells which also show high N: C ratio. Superficial and intermediate cells are large polygonal cells due to increase in cytoplasm and with decreased nuclear size.

- Orangeophilic cells- indicates increased keratinization in the epithelium/hyperkeratosis
- Anucleate squames (Orthokeratinized cells) – Large, flat orange cells with degenerated nuclei
- Parakeratotic cells- Mature cells with abundant cytoplasm and small, pyknotic/flattened nuclei.
- Dyskeratotic cells- Abnormal keratinization present in lower layers of the epithelium.
- Dysplasia- Abnormal cells displaying varying degrees of changes in cell size and shape, nuclear size and shape, chromatin and nuclear membrane irregularities.

Cytological features for different oral conditions are described below:

- Benign disorders:

Variable cellular changes are observed in different benign conditions; some examples include presence of ‘corps ronds’ in Darier-White’s disease, keratinized squamous cells with pearls in hereditary benign intraepithelial dyskeratosis.

- Pre-cancer changes

Commonly encountered pre-cancer conditions in the oral cavity include white lesions termed as leukoplakia or red lesions called erythroplakia. On cytological diagnosis, leukoplakia may be benign or pre-malignant. In benign leukoplakia, yellow, yellow-orange stained cells without

nuclei (anucleate squames) are observed. In pre-malignant leukoplakias, nuclear abnormalities in well-differentiated squames are observed. These nuclear changes may be indicative of mild or moderate dysplasia. In the less common red lesions i.e. erythroplakia, small cancer cells with minimal or absent keratinization manifest in the smears. Changes may be indicative of severe dysplasia or carcinoma in situ.

- Malignancy associated changes

Nuclear enlargement and discontinuous nuclear membrane were observed in buccal mucosa of cancer patients. These changes apparently reflect altered mitotic cellular function.

- Malignant changes

- Low-grade changes- Mild enlargement of nucleus with irregular nuclear contours. Slight membrane irregularities may be present, nuclear may be normochromatic. Large, hyperchromatic nuclei in isolated cells may also be apparent.
- High grade changes- Prominent nuclear enlargement with reduced cytoplasm in few cells, hyperchromatic nuclei in synctial groups as moulding.
- Heavily keratinized squamous cancer- Cells are characterized by orange and yellow-staining cytoplasm and large, sometimes pyknotic, dark-staining, irregular nuclei. Ghost cells, cells with heavily keratinized cytoplasm and virtually no residual nuclear material are also frequent.
- Non-ulcerated, invasive, keratinizing carcinoma- Cytological diagnosis may be obscured by abundant ghost cells without perceptible nuclear abnormalities. Relatively minor nuclear abnormalities may be present in few cells that include nuclear enlargement and irregularity of outline with and/or without nuclear hyperchromasia.

- Poorly differentiated SCC- Cytoplasmic keratinization is not prominent, nuclear abnormalities such as large nucleoli, coarse pattern of chromatin distribution, high nucleocytoplasmic ratios are observed.[387]

(Background diathesis which consists of necrotic, inflammatory exudates and stains as pale-orange to pink may also be diagnostic of invasive cancer irrespective of histopathological grade.)

Overall, following major parameters were employed for grading of dysplasia and malignancy:

Oral dysplasia cytological findings[388]

- Hyperchromasia of nuclei
- Increased nuclear to cytoplasmic ratio
- Anisonucleosis and nuclear polymorphism
- Irregularities of nuclear membrane
- Nuclear crowding
- Nuclear moulding, clumping and irregular distribution of chromatin

Oral squamous cell carcinoma cytological findings[388]

- Hyperchromasia of nuclei
- Increased nuclear to cytoplasmic ratio
- Anisonucleosis and nuclear polymorphism
- Irregularities of nuclear membrane
- Nuclear crowding

- Nuclear moulding, clumping and irregular distribution of chromatin
- Dyskeratosis
- Tadpole and strap cells (presence indicative of invasive carcinoma)

3.2 Proof of concept: Raman exfoliative cytology for oral cancer diagnosis

Introduction

Previous IR studies have shown feasibility in differentiating healthy from dysplastic and neoplastic cells. Raman exfoliative cytology for cervical specimens successfully classified normal and abnormal specimens. In the first step, classification between normal and oral cancer specimens was explored to establish proof of concept.

3.2.1 Materials and Methods

3.2.1.1 Exfoliative cytology

Oral exfoliated cells were collected from histopathologically confirmed oral cancer (buccal mucosa) patients with tobacco habits visiting the out-patient department of Head and Neck oncology, Tata Memorial Hospital, Mumbai, India. Cytobrush, cotton swab and wooden stick were employed as possible collection devices for obtaining maximum cell yield. Cell counting was performed using Haemocytometer (Neubauer chamber). Maximum cellular yield was obtained using Cytobrush and was therefore selected as the sampling device. Exfoliated cells were therefore collected from "site of tumor" (referred to as 'T' henceforth) and respective "Contralateral normal or Disease control sites" (referred to as 'DC' henceforth) from the same patient using 'Cytobrush' (Himedia Laboratories, India) in normal saline. Exfoliated specimens

were also collected from both buccal mucosae of healthy subjects unrelated to cancer patients, subjects with tobacco habits were termed as 'healthy tobacco' (HT) while without tobacco habits were termed as 'healthy volunteers' (HV). No clinical abnormalities were found in mucosa of HT and HV subjects. Demographic details of study group are summarized in Table 3.1.

Sr. no.	Category	No. of subjects	Age range/median age	Tobacco habits	No. of spectra
1.	Healthy volunteer (HV)	15	29-55/ 44 years	No	89
2.	Habit control (HT)	15	20-63/ 34 years	Yes	100
3.	Tumor (T)	20	27-70/ 53 years	Yes	116
4.	Contralateral or Disease control (DC)	20	27-70/ 53 years	Yes	86

Table 3.1. Subject accrual summary for proof-of-concept study on exfoliated cells

Samples were transported in ice at 4° C, and immediately processed upon arrival. Exfoliated cells were harvested after vortexing and centrifugation at 6000 rpm for 1 min. Due to high vascularity of tumors, blood contamination was observed in most cancer samples; RBC lysis protocol was therefore adopted as an integral step during cell processing. For sake of uniformity, all samples were treated with RBC lysis protocol after harvesting cell pellet. The protocol is as follows: 1 ml RBC lysis buffer (HiMedia Laboratories, India) was added to each tube, incubated for 15 min at room temperature (mixed in intervals of 5 min), and washed in saline two times. Cell counting was performed using Neubauer's chamber after staining with Trypan blue and the final pellet was employed for spectral acquisition. After spectral acquisition, cells were smeared onto glass slides, fixed and subjected to Pap staining.

3.2.1.2 Raman spectral acquisition

The pellet was placed on a CaF_2 window and spectra recorded using fiber-optic Raman microprobe (Horiba-Jobin-Yvon, France). The details of the instrument have been described in section 2.2.1.3. Spectral acquisition parameters were: laser power: 40 mW, acquisition time: 15 s and average: 3. About 6-7 spectra were recorded from several areas to span the pellet. After spectra recording, cells were smeared onto glass slides and subjected to Papanicolaou staining.

3.2.1.3 Pap staining

After spectral acquisition, cells are fixed in a fixative solution to preserve the cellular integrity and molecular makeup. Various fixatives like methanol, 95% ethanol, Carnoy's fixative, ether:alcohol mixture and PEG:ethanol mixture were evaluated. Ether: alcohol mixture was recommended, but was found to be unsatisfactory. Ninety-five percent ethanol gave satisfactory results and was therefore used as the fixative in this study. Pap stain employs a combination of 3 stains, namely Haematoxylin, Orange G (OG) and Eosin-Azure (EA) [389]. Haematoxylin stains cell nucleus a blue color, while the differentiating stains OG and EA provide a subtle range of orange, bluish-green, and pink hues to the cytoplasm according to the amount of keratin and degree of maturation. Rapid-pap kit (routinely employed for cervical smear specimens) was first explored for Pap staining of oral cell smears after spectra acquisition. Improper nuclear staining was encountered, cytoplasmic differentiation was also not clear. Stains were replaced with new stains and staining was repeated. Exposure to nuclear and cytoplasmic stains were also increased, however the staining was unsatisfactory. Modified Pap staining protocol was adopted, but the staining was unsatisfactory as per the pathologist. The original Pap staining protocol (shown in

Figure 3.2) was eventually employed which yielded satisfactory results. Post Raman-spectral acquisition, cell pellet was smeared onto the glass slide, fixed in 95% ethanol and Pap stained. The slides were read by a certified pathologist subsequent to mounting.

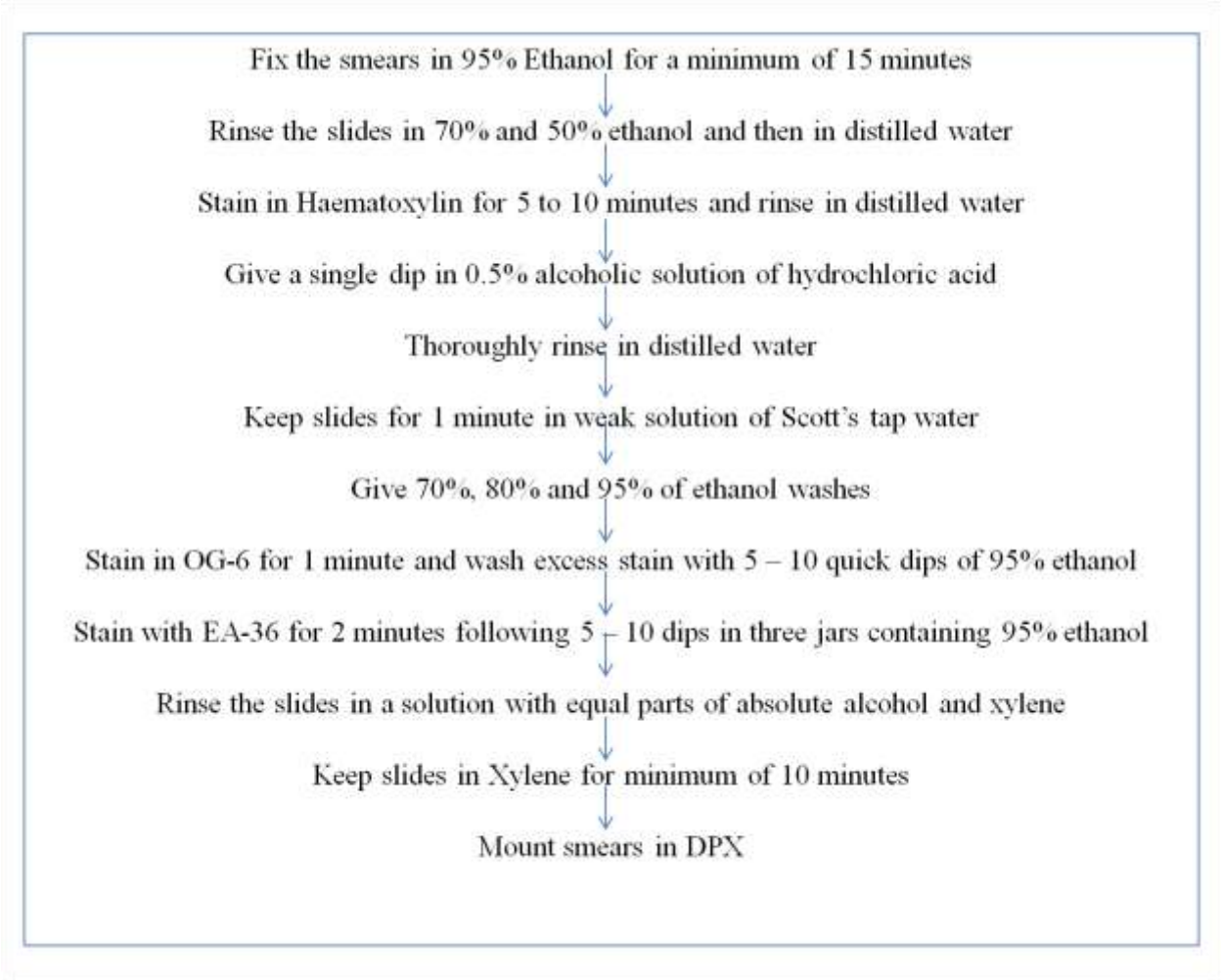


Figure 3.2. Original protocol adopted for Pap staining

3.2.1.4 Spectral pre-processing and data analysis

The acquired Raman spectra were corrected for CCD response and spectral contaminations from substrate and fiber signals. To remove interference of the slow moving background, first derivatives of spectra (Savitzky-Golay method and window size 3) were computed. Spectra were interpolated in the range 800-1800 cm^{-1} as described in section 2.1.1.4. PCA and PC-LDA were employed for data analysis. Average spectra were also computed as described in section 2.1.1.4.

3.2.2 Results and Discussion

3.2.2.1 Cytological findings:

Cytological analysis of the same pellet was carried out after Pap staining by a certified pathologist. Apart from the normal cells of the three epithelial layers, namely stratum basale, intermedium and superficial, slides were scored for presence of orangeophilic cells, parakeratotic cells, anucleate squames which mainly indicate keratinization status of the mucosa; dyskeratotic and dysplastic cells, which may indicate tendency of malignant transformation of the mucosa. Representative slides for the 4 groups, i.e. HV, HT, DC and T are shown in Figure 3.3a-d.

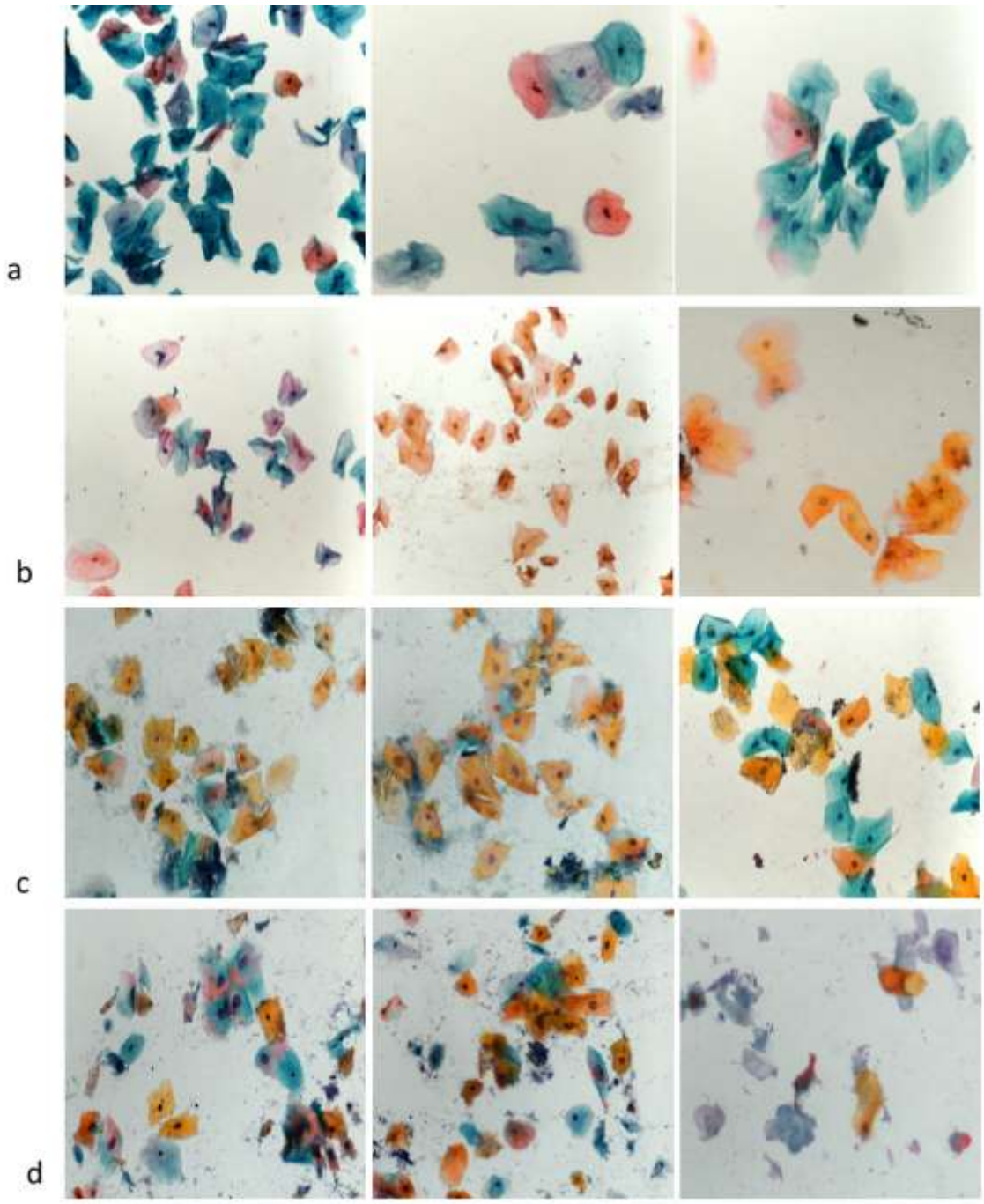


Figure 3.3. Representative images of Pap stained cytological smears (200X) from- a) Healthy volunteers showing cells from superficial and intermediate layers with occasional presence of orangeophilic cells, b) Habit-controls showing increased number of

orangeophilic cells and occasional parakeratotic cells and anucleate squames , c) Disease control or contralateral sites from cancer patients showing increased number of orangeophilic cells with few parakeratotic cells and anucleate squames, and occasional cells with minimally increased nuclear:cytoplasm ratio (indicative of mild dysplasia), d) Tumor showing dysplastic cells with high nuclear :cytoplasm ratio and prominent nucleolus, multinucleation with dirty background (indicative of necrosis) and higher frequency of anucleate squames

In the healthy volunteers (HV) smears, normal proportions of cells from stratum intermedium, stratum superficial, and few cells from stratum basale were observed. A very small percentage of parakeratotic and orangeophilic cells were detected in few cases but no dysplastic or inflammatory cells were found. Representative cytological smears are shown in Figure 3.3a. As in HV, healthy tobacco (HT) smears showed no dysplastic or inflammatory cells but higher numbers of orangeophilic, and occasional parakeratotic cells and anucleate squames were observed (Figure 3.3b). In the oral cancer group- contralateral (DC) and tumor (T) cell smears, higher number of orangeophilic cells, parakeratotic cells and anucleate squames, along with dyskeratotic and dysplastic cells as compared to HT smears were observed. Few instances of dysplasia were observed in DC which were absent in HT. The frequency of dyskeratosis and dysplasia was highest in T (Figure 3.3c-d).

In the HV group, chronic chemical irritation due to tobacco-related products was absent. The small percentage of orangeophilic and parakeratotic cells could possibly be attributed to exposure to some physical or chemical irritants. In case of HT group, chronic tobacco exposure leads to induction of the protective stimuli of the epithelium wherein the epithelial cells produce keratin. Thus, an increase in the number of orangeophilic cells, parakeratotic, anucleate squames

is observed. The oral cancer subjects recruited in the study were also chronic tobacco abusers. Higher frequency and longer duration of tobacco habits in these subjects is observed in terms of increased degree of keratinization (higher numbers of orangeophilic cells, parakeratotic and anucleate squames in both DC and T groups). As expected, from contralateral to tumor condition, frequency of dyskeratosis and dysplasia increased. Variable number of dysplastic cells can also be expected in the contralateral group ascribed to two reasons a) cancer-field effects and b) migration of tumor cells by saliva (micro-metastases)[361]. Thus, overall cytological findings indicate very few parakeratotic, orangeophilic cells in HV, higher numbers in HT and DC, presence of few dysplastic cells in DC and highest numbers of parakeratotic, orangeophilic and dysplastic cells in T[390].

3.2.2.2 Spectral Features:

The average spectra for the four groups are shown in Figure 3.4a-d. Prominent spectral features include bands at 830, 857, 1009, 1093, 1170, bands in the region of 1270-1340, 1450 and 1660 cm^{-1} that can be overall assigned to phenylalanine, DNA-phosphate backbone related features, amide III, CH_2 twisting in proteins and lipids, DNA base features, CH_2 bending in protein and lipids and amide I features from proteins[329,330]. The band at 1170 cm^{-1} can be tentatively assigned to tyrosine and v (C-C) of skeletal structure in proteins[391-393] and also to tobacco-related amine-containing adducts[394-396]. This peak is present in all groups; the higher presence in cancer groups (both contralateral and tumor) may be attributed to increased tobacco exposure groups.

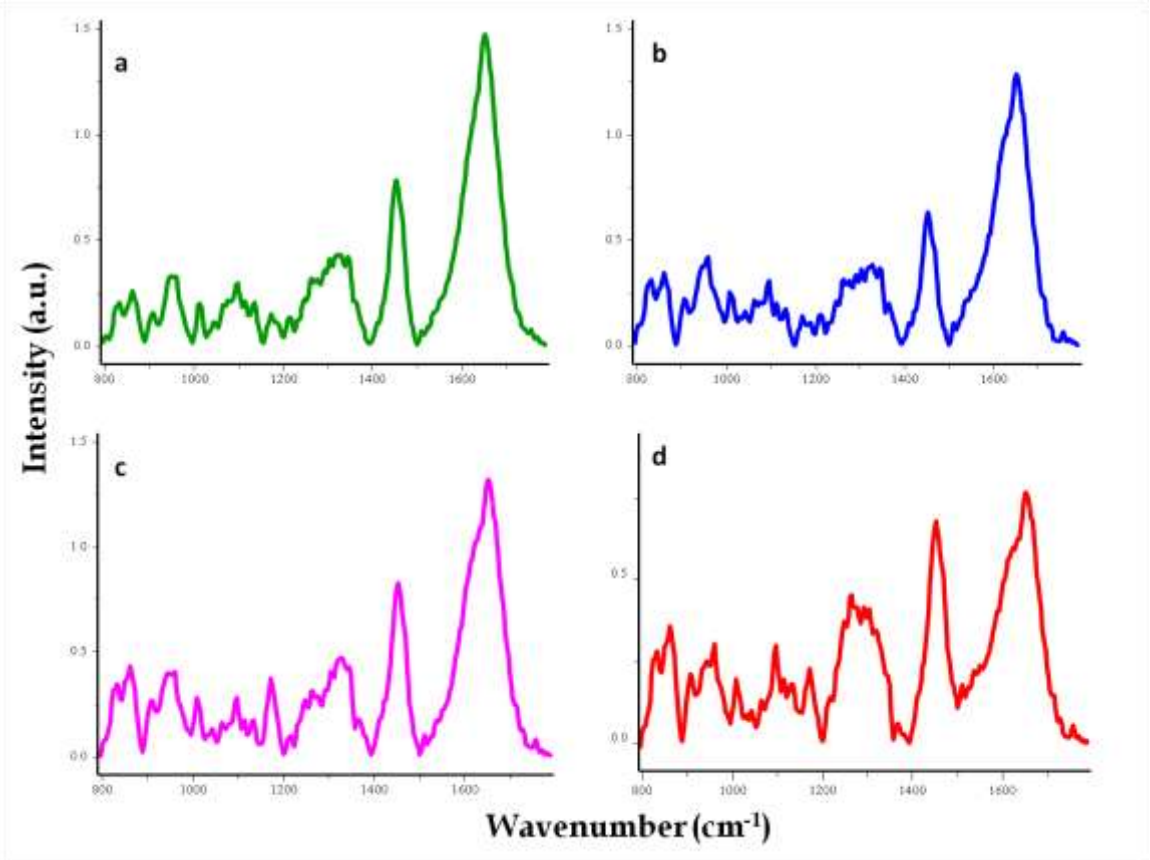


Figure 3.4. Mean Raman spectra of exfoliated cells from, a) healthy volunteers, b) habit-controls, c) disease control, d) tumor, for spectral comparisons

With the increase in severity of pathology in HV to T, higher DNA (1095 and 1325-30 cm^{-1}), higher CH_2 bending (1450 cm^{-1}), higher amide III was observed. Broadening in the amide I region was also encountered with increase in pathological severity. As compared to the other groups, broad amide I, higher CH_2 bending and amide III regions, and nucleic acid bases features around 1320-1340 cm^{-1} were observed in the average tumor spectrum. Thus, increase in DNA content and changes in protein secondary structures could be hallmarks of severe pathological states.

3.2.2.3 Multivariate analysis

3.2.2.3.1 ***Spectra-wise analysis***

A total of 391 spectra (89: Healthy volunteer, 100: habit control, 86: disease control, 116: tumor) from 50 cases (15: Healthy volunteer, 15: habit control, 20: disease control and 20: tumor from same oral cancer patient) were acquired. In the first spectra-wise approach, all spectra from the 50 subjects (total 391 spectra) were used for multivariate analysis and first subjected to PCA to understand trends in data. For PCA, scores of factors 1 and 3 were explored for classification. Factor loadings for factor 1 and 3 are shown in Figure 3.5a-b. The scatter plot in Figure 3.5c indicates slightly overlapping clusters for HV and HT while overlapping clusters are observed between DC and T groups.

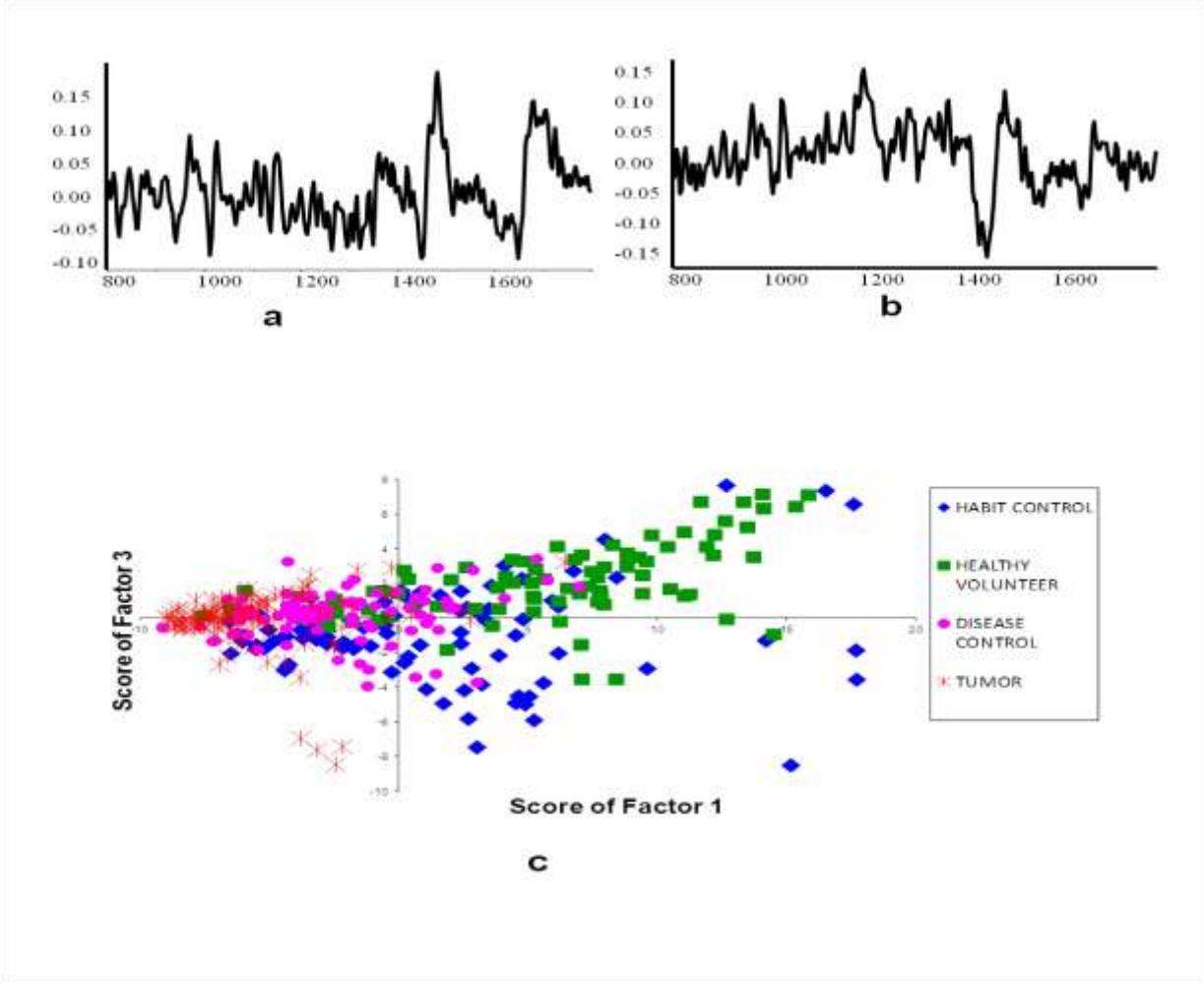


Figure 3.5. PCA for exploring differences between exfoliated cells from healthy volunteers, habit controls, disease control and tumor subjects - Spectra-wise approach, a) Loadings of factor 1, b) Loadings of factor 3, c) Scatter plot.

As PCA is not a classification tool but is used for data compression and visualization to indicate trends in the data, PC-LDA was employed to explore classification between the groups. Six factors (Figure 3.6a) were used for the analysis which accounted for ~68% classifications. Scores of factor 1 and 2 were employed for obtaining scatter plot, as shown in Figure 3.6b. As seen in

the scatter plot, HV and HT form almost different clusters while overlap is observed between DC and T.

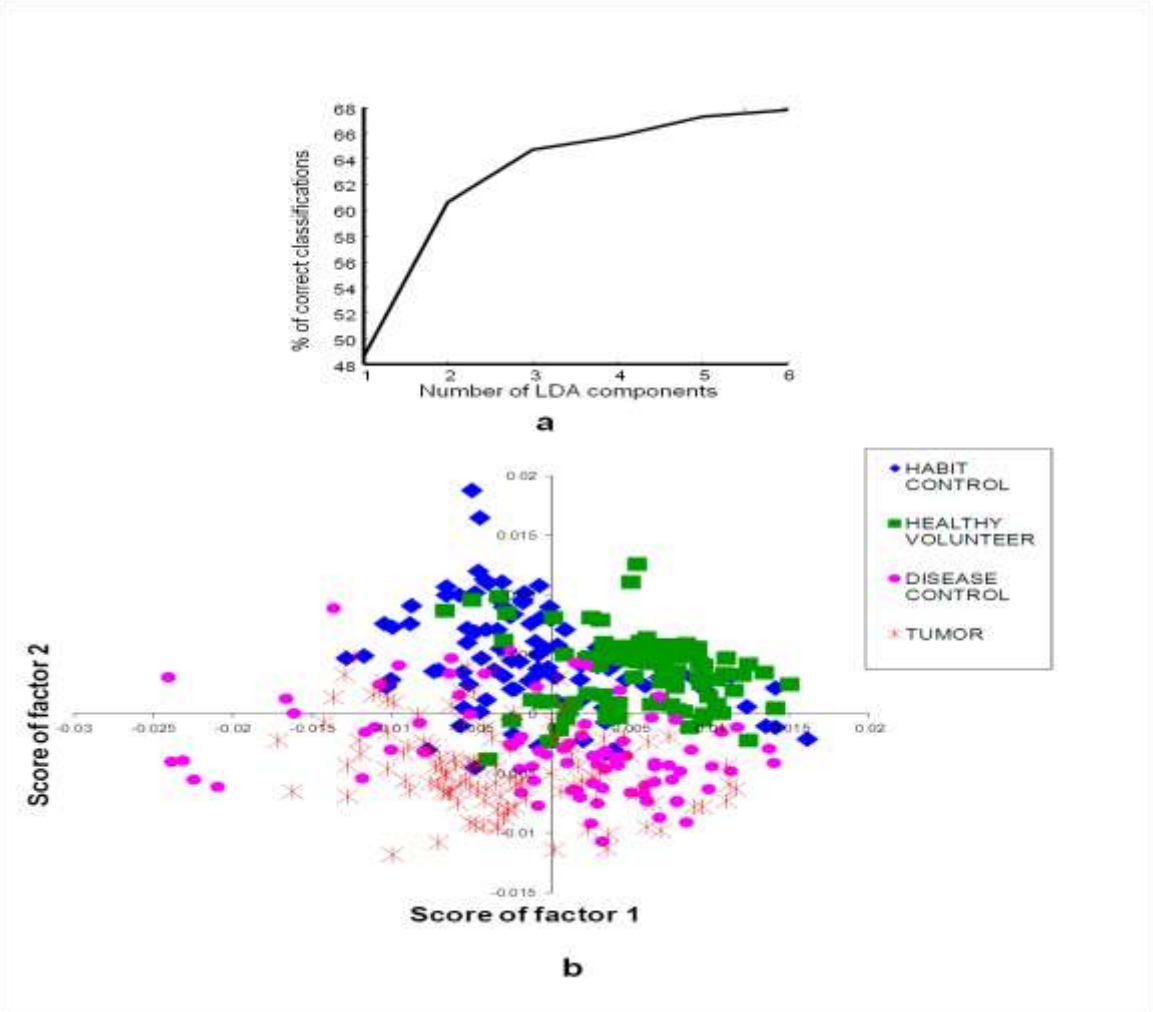


Figure 3.6. PC-LDA for exploring differences between exfoliated cells from healthy volunteers, habit controls, disease control and tumor subjects- Spectra-wise approach, a) Scree plot b) Scatter plot

As seen in the LOOCV confusion matrix (Table 3.2), 69/89 spectra of HV group were correctly classified and 12/89 misclassified with HT. 68/100 HT spectra were correctly classified while 21 spectra misclassified with HV. 58/86 DC spectra were correctly classified while misclassification

of 15 spectra with T was observed. 68/116 spectra of T got classified correctly, while 34 misclassified as DC, 9 as HT and 5 as HV. Thus, misclassifications of HV were seen mainly with HT (14%) while for HT, major misclassifications were with HV (21%), and some with DC (8%). In case of DC and T, major misclassifications were between these groups: 17% DC misclassified with T and 29% T misclassified with DC. Some DC and T also misclassified with HT (~7%) and HV (6% DC and 4% T).

	HEALTHY VOLUNTEER	HABIT CONTROL	DISEASE CONTROL	TUMOR	TOTAL
HEALTHY VOLUNTEER	69	12	4	4	89
HABIT CONTROL	21	68	8	3	100
DISEASE CONTROL	6	7	58	15	86
TUMOR	5	9	34	68	116

Table 3.2. PC-LDA for exploring differences between exfoliated cells from healthy volunteers, habit controls, disease control and tumor subjects using Spectra-wise approach-confusion matrix of LOOCV results

3.2.2.3.2 Patient-wise analysis

A second approach, termed as patient-wise approach, wherein average spectrum of each sample was used for data analysis was also explored. It is known that the laser spot size is 4-5 micron and the penetration is ~40 micron. The laser probing volume accommodates several stacked cells and their intracellular components. Studies suggest that intracellular and cell-to-cell variation can

be detected by Raman spectroscopy[391,397,398]. Thus, average spectrum was calculated to yield a true representative of the sample and also to circumvent the intra-sample heterogeneity. These average spectra were then subjected to multivariate analysis PCA and PC-LDA. PCA was carried out using 10 factors, scores of factor 2 and factor 3 were used to obtain scatter plot. Factor loadings are shown in Figure 3.7a-b. As seen in Figure 3.7c, slightly distinct clusters for healthy, and tumor groups while overlap of contralateral with both healthy tobacco and tumor groups was observed. As compared to the first approach, better classification trends were observed for the patient-wise approach.

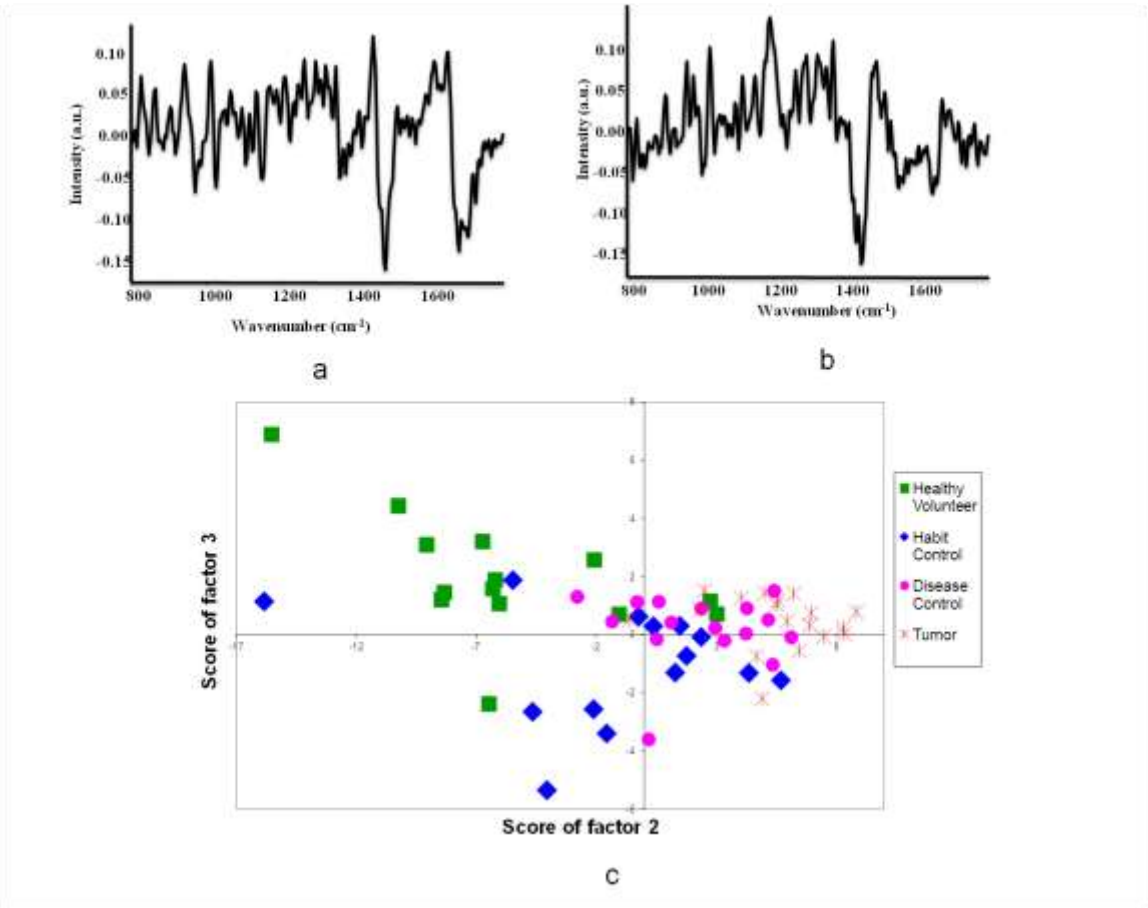


Figure 3.7. PCA for exploring differences between exfoliated cells from healthy volunteers, habit controls, disease control and tumor subjects- Patient-wise approach, a) Loadings of factor 2 b) Loadings of factor 3, c) Scatter plot

PC-LDA was then carried out to build standard models, using 4 factors (Figure 3.8a). The scatter plot shown in Figure 3.8b indicates almost exclusive clusters for healthy and tumor groups.

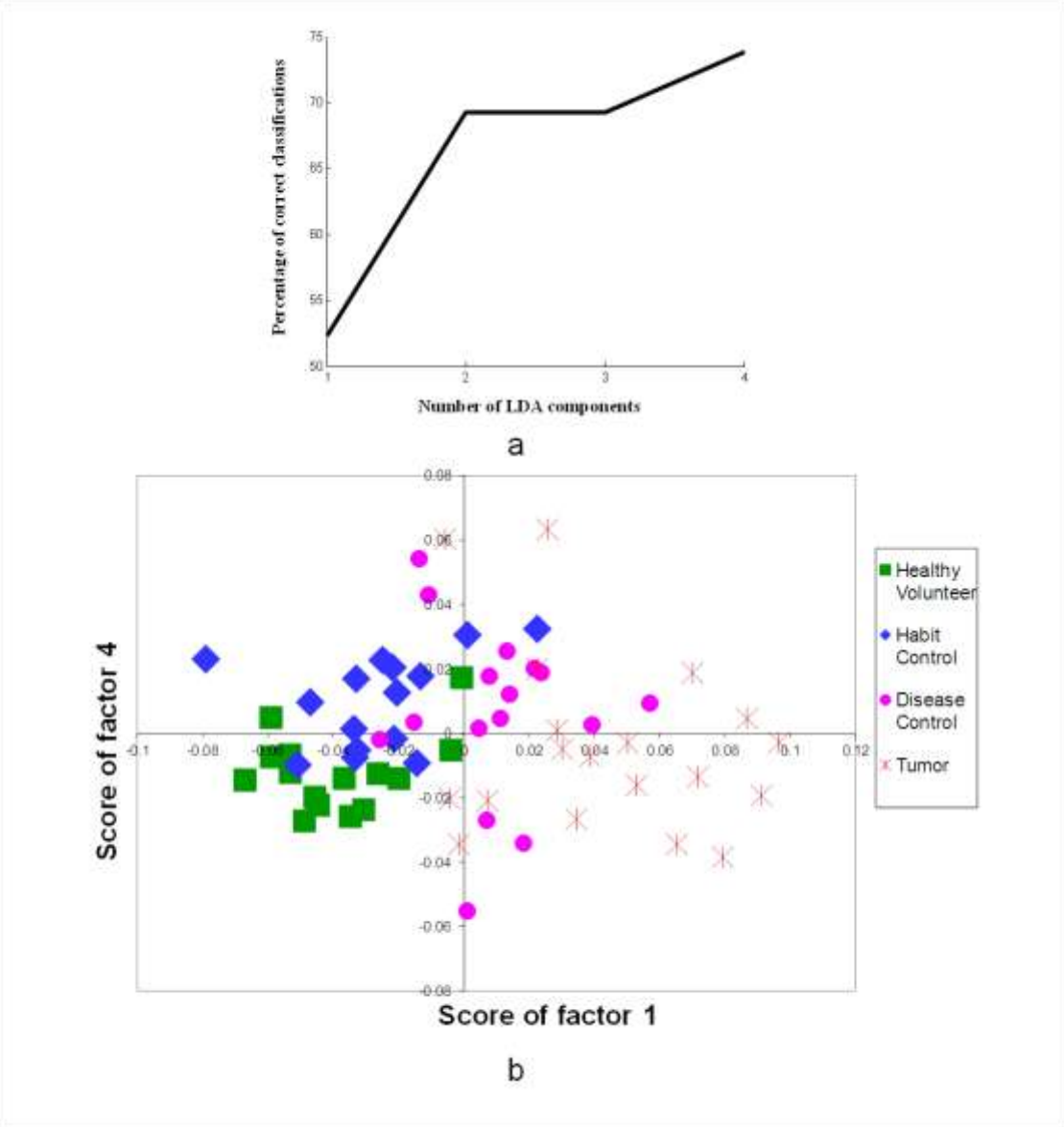


Figure 3.8. PC-LDA for exploring differences between exfoliated cells from healthy volunteers, habit controls, disease control and tumor subjects -Patient-wise approach, a) Scree plot b) Scatter plot

LOOCV findings demonstrate that 13/15 HV, 11/15 HT, 9/16 DC and 12/19 T were correctly classified (Table 3.3). Thus, 86% HV were correctly classified, 6% misclassifications were with

HT and DC, none with T. Similarly for HT, 13% misclassifications were found with HV and DC, none with T. For DC, major misclassifications were with HT (19%), and 12% with HV and T. 21% T misclassified with DC, only 10% and 5% with HT and HV, respectively. Thus, better classification efficiency and lower misclassification rate was obtained for all groups using the patient-wise approach. As averaging leads to true representation of the cell pellet, less-cross talk between samples of different groups was observed.

	HEALTHY VOLUNTEER	HABIT CONTROL	DISEASE CONTROL	TUMOR	TOTAL
HEALTHY VOLUNTEER	13	1	1	0	15
HABIT CONTROL	2	11	2	0	15
DISEASE CONTROL	2	3	9	2	16
TUMOR	1	2	4	12	19

Table 3.3. PC-LDA for exploring differences between exfoliated cells from healthy volunteers, habit controls, disease control and tumor subjects using Patient-wise approach- confusion matrix of LOOCV results

In case of healthy tobacco users, it is known that chronic tobacco exposure leads to biochemical changes in oral mucosa leading to cytological changes in the cells. However, the tobacco exposure in the buccal mucosa area and the response to tobacco exposure may not be uniform for all cells. Thus, presence of normal cells in the HC group smears is expected- also observed during cytological analysis. Minor misclassifications between these two groups can thus be explained. Some misclassifications were also observed between the HT and DC groups. The

contralateral mucosa of the oral cancer patient has also been chronically exposed to tobacco. This persistent and chronic irritation first leads to biochemical changes in the cells (thus similarity with HT) eventually followed by malignant transformation. In case of cancer group, major misclassifications were observed between DC and T. These misclassifications could be attributed to tumor heterogeneity, and collection of other cells (from contralateral site) is also a possibility. Further, as previously stated, contralateral mucosa may have presence of genetically altered cells due to field cancerization or micro-metastases[361,399]. This is evident even in the cytopathological findings, where parakeratotic, orangeophilic cells and minor dysplasia are present in both groups, only higher numbers are evident in the T group[400,401.]

Thus, overall a high classification rate between healthy groups (HV and HT) and cancer groups (DC and T) was observed. Cytological findings indicate overlap between cellular profiles of healthy and cancer groups. The intra-group misclassification could be attributed to this similarity in cell populations.

3.3 Raman oral exfoliative cytology for oral pre-cancer diagnosis

Introduction

After successful discrimination of healthy and cancer groups in the previous section, the feasibility of oral pre-cancer diagnosis was explored. Diagnosis of dysplasia or risk-assessment in oral pre-cancers like leukoplakia, OSMF and TPK can lead to early detection and better management of oral cancers. Exfoliative cytology (EC) is indicated for evaluation of suspicious lesions detected during screening by visual inspection. After a positive result, trans-epithelial sampling and mucosal biopsy may be advised for confirmatory purposes. EC of oral lesions may be advantageous for lesions which are clinically not obviously suspicious for malignancy but nevertheless need surveillance. Tissue biopsy associated with low patient compliance (9%), but brush biopsy has demonstrated nearly cent percent compliance[402]. Studies on dysplasia/neoplasia surveillance in oral lichen planus malignant transformation lead to diagnosis in intraepithelial and micro-invasive phases leading to substantially high 5-year survival rates[403]. Raman spectroscopy of exfoliative cytology from pre-cancer stages may give additional insights. In this section, the utility of Raman spectroscopic analysis of exfoliated cells from healthy non-tobacco habitués, healthy tobacco habitués and oral pre-cancer subjects was evaluated.

3.3.1 Materials and Methods

3.3.1.1 Exfoliative cytology

Oral exfoliated cell specimens were obtained from healthy subjects without tobacco habits i.e. healthy volunteers (n=20, HV), habit controls (n=20, HT) and premalignant subjects with

leukoplakia, TPK or OSMF (n=27, OPL). Subject accrual summary is provided in Table 3.4. Similar tobacco habits were prevalent in both the HT and OPL groups. As described in previous section, cells were collected using Cytobrush (Himedia, India) in normal saline. Exfoliated specimens were also collected from both buccal mucosae. Samples were transported in ice at 4° C, and immediately processed upon arrival. Exfoliated cells were harvested after vortexing and centrifugation at 6000 rpm for 1 min. Cell counting was performed using Neubauer's chamber after staining with Trypan blue and the final pellet was employed for spectral acquisition. After spectral acquisition, cells were smeared onto glass slides, fixed and subjected to Pap staining.

Sr. No.	Category	Subjects accrual	Tobacco habits	No. of spectra
1.	Healthy volunteers (HV)	20	No	218
2.	Habit controls (HT)	20	Yes	213
3.	Oral pre-malignant lesions (OPL) -Leukoplakia -Tobacco pouch keratosis Oral sub-mucous fibrosis	27	Yes	234

Table 3.4. Subject accrual summary for early diagnosis study on exfoliated cells

3.3.1.2 Raman spectral acquisition

The pellet was placed on a CaF_2 window and spectra recorded using fiber-optic Raman microprobe (Horiba-Jobin-Yvon, France) as described in section 2.2.1.3. Spectral acquisition parameters were: laser power: 40 mW, acquisition time: 15 s and average: 3. About 6-7 spectra were recorded from several areas to span the pellet. After spectra recording, cells were smeared onto glass slides and subjected to Papanicolaou staining.

3.3.1.3 Pap staining

Pap staining post-spectral acquisition was carried out as detailed in section 3.2.1.3.

3.3.1.4 Spectral pre-processing and data analysis

As described in section 2.1.1.4, Raman spectra were corrected for CCD response and spectral contaminations from substrate and fiber signals. To remove interference of the slow moving background, first derivatives of spectra (Savitzky-Golay method and window size 3) were computed. Spectra were interpolated in the range 800-1800 cm^{-1} and data analysis was carried out using PCA and PC-LDA for both spectra-wise and patient-wise approaches. Average spectra for spectral comparisons were also computed as described in section 2.1.1.4.

3.3.2 Results and Discussion

3.3.2.1 Cytological findings:

As mentioned in section 3.2.2.1, cytological analysis of the same pellet was carried out after Pap staining and graded for presence of normal cells, orangeophilic cells, parakeratotic cells, anucleate squames, dyskeratotic and dysplastic cells. Representative slides for the 3 groups, i.e. HV, HT, and OPL are showed in Figure 3.9.

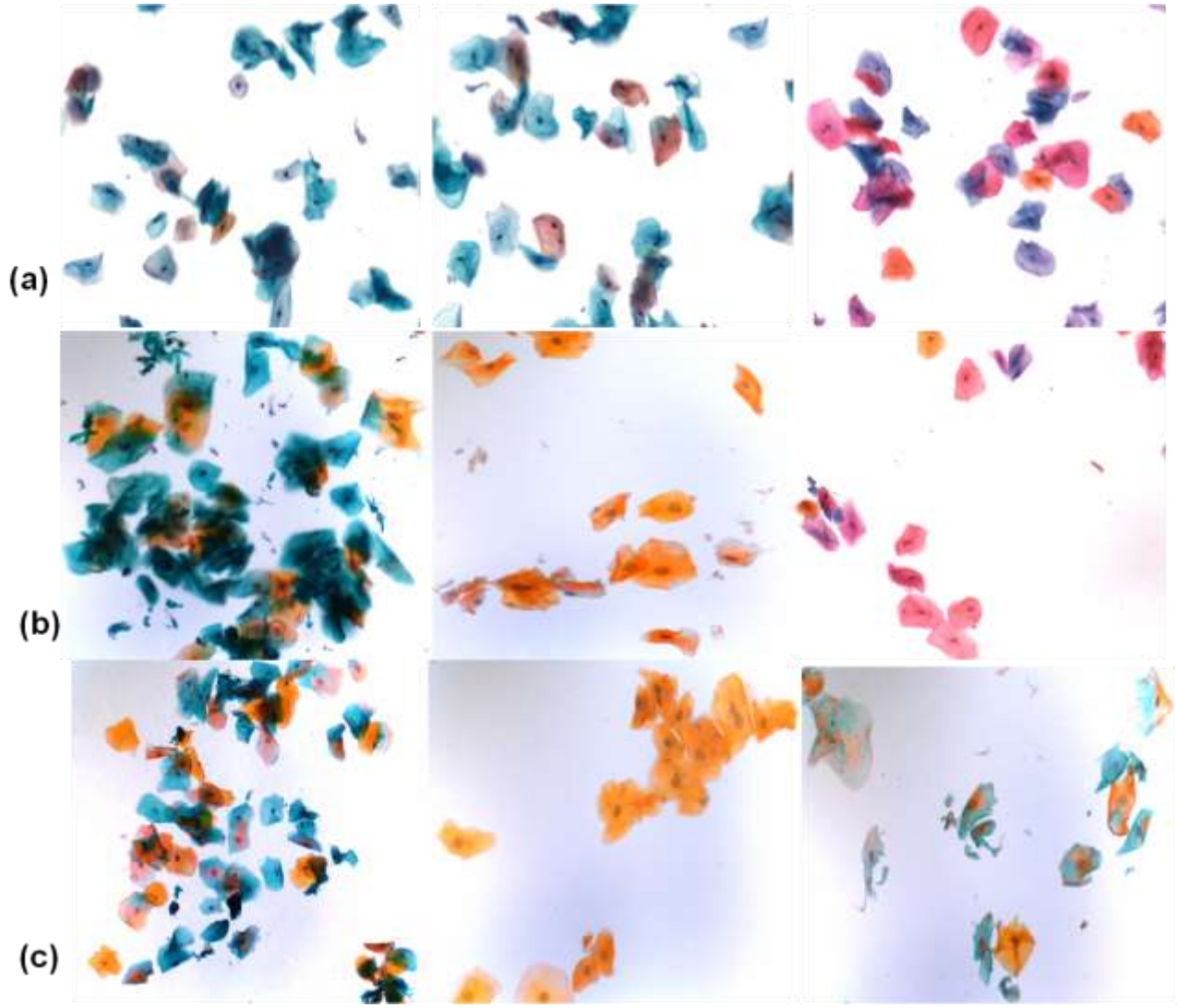


Figure 3.9. Representative images of Pap stained cytological smears (200X) from a)Healthy volunteers showing cells from superficial and intermediate layers with occasional presence of orangeophilic cells, b) Habit-controls showing increased number of orangeophilic cells and occasional parakeratotic cells and anucleate squames , c) Oral pre-malignant lesions showing very high orangeophilia, parakeratotic and anucleate cells, cells with minor dyskeratosis and mild increase in nuclear:cytoplasmic ratio.

In the HV smears, normal proportions of cells from stratum intermedium, stratum superficial, and few cells from stratum basale were observed. A very small percentage of parakeratotic and

orangeophilic cells were detected in few cases. As in HV, HT smears showed higher numbers of orangeophilic, and occasional parakeratotic cells and anucleate squames (Figure 3.9b). In OPL smears, very high frequency of orangeophilic cells, numerous parakeratotic and anucleate squames were observed for most cases. In some cases, mild changes in the nuclear features and nuclear: cytoplasmic ratios and dyskeratosis-like features were also observed. These subtle changes may be indicative of early dysplasia.

Higher parakeratotic cells and orthokeratinized cells indicate higher mitotic index in the lower layers of the epithelium. Thus, higher parakeratinization and orthokeratinization observed in OPL may be associated with epithelial dysplasia and hyperplasia[404,405].

3.3.2.2 Spectral Features:

The average spectra for the four groups are shown in Figure 3.10a-c. Prominent spectral features include bands at 830, 857, 1008, 1090, 1216, bands in the region of 1260-1340, 1452 and 1660 cm^{-1}. Thus, contributions from phenylalanine, DNA-phosphate backbone related features, amide III, CH_2 twisting in proteins and lipids, CH_2 bending in protein and lipids and amide I features from proteins were observed[329,330]. Intensity-related variations at most of these peaks and shifts in the amide III region were observed across the groups. Broad amide I was observed with increase in the pathological status, from healthy non-tobacco to oral premalignant conditions.

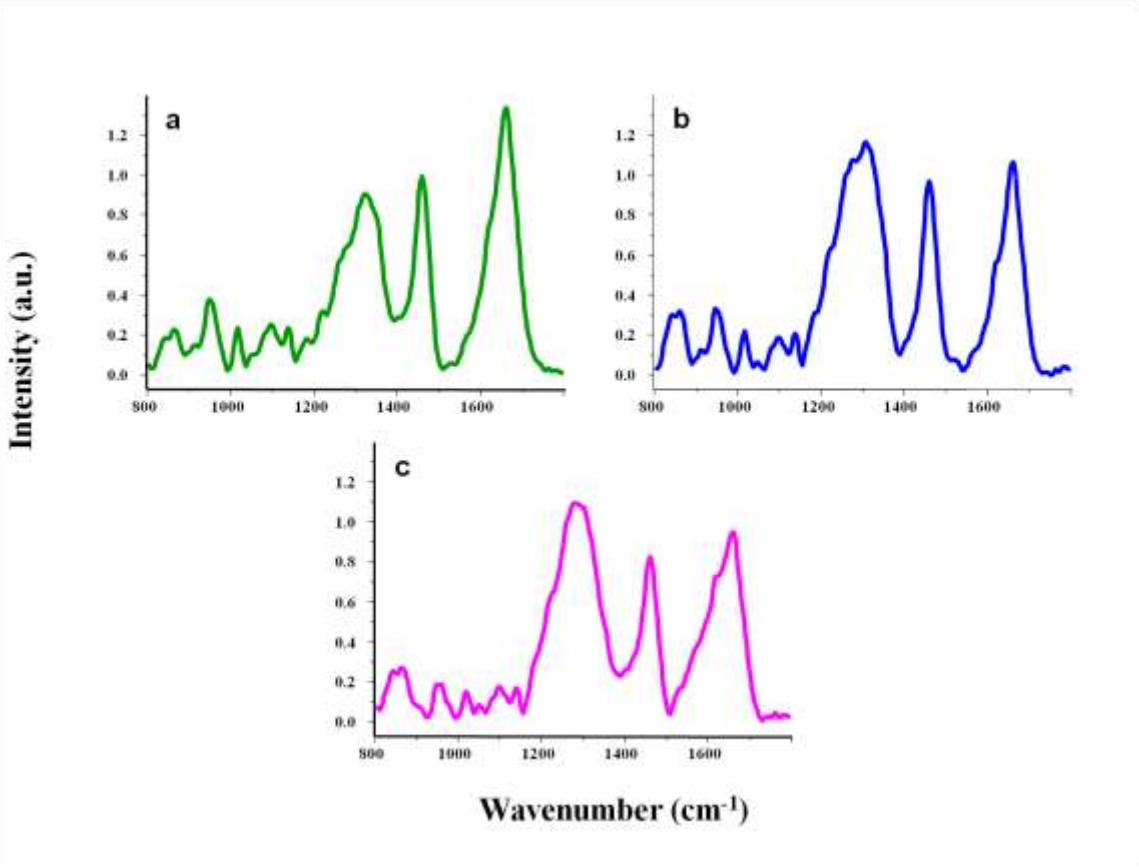


Figure 3.10. Mean Raman spectra exfoliated cells from, a) Healthy volunteers (HV), b) Habit-controls (HT), c) Pre-malignant (OPL) subjects, for spectral comparisons

3.3.2.3 Multivariate analysis

Data analysis was carried out in 2 steps: spectra-wise and patient-wise approach using PCA and PC-LDA, followed by LOOCV.

3.3.2.3.1 Spectra-wise analysis

In the first step, spectra-wise approach was adopted. First, all spectra were used as input for PCA to examine trends in the data. Scores of factor 1 and 3 were explored for classification. Factor

loadings for factor 1 and 3 are shown in Figure 3.11a-b. The scatter plot in Figure 3.11c indicates three slightly overlapping clusters for HV, HT and OPL.

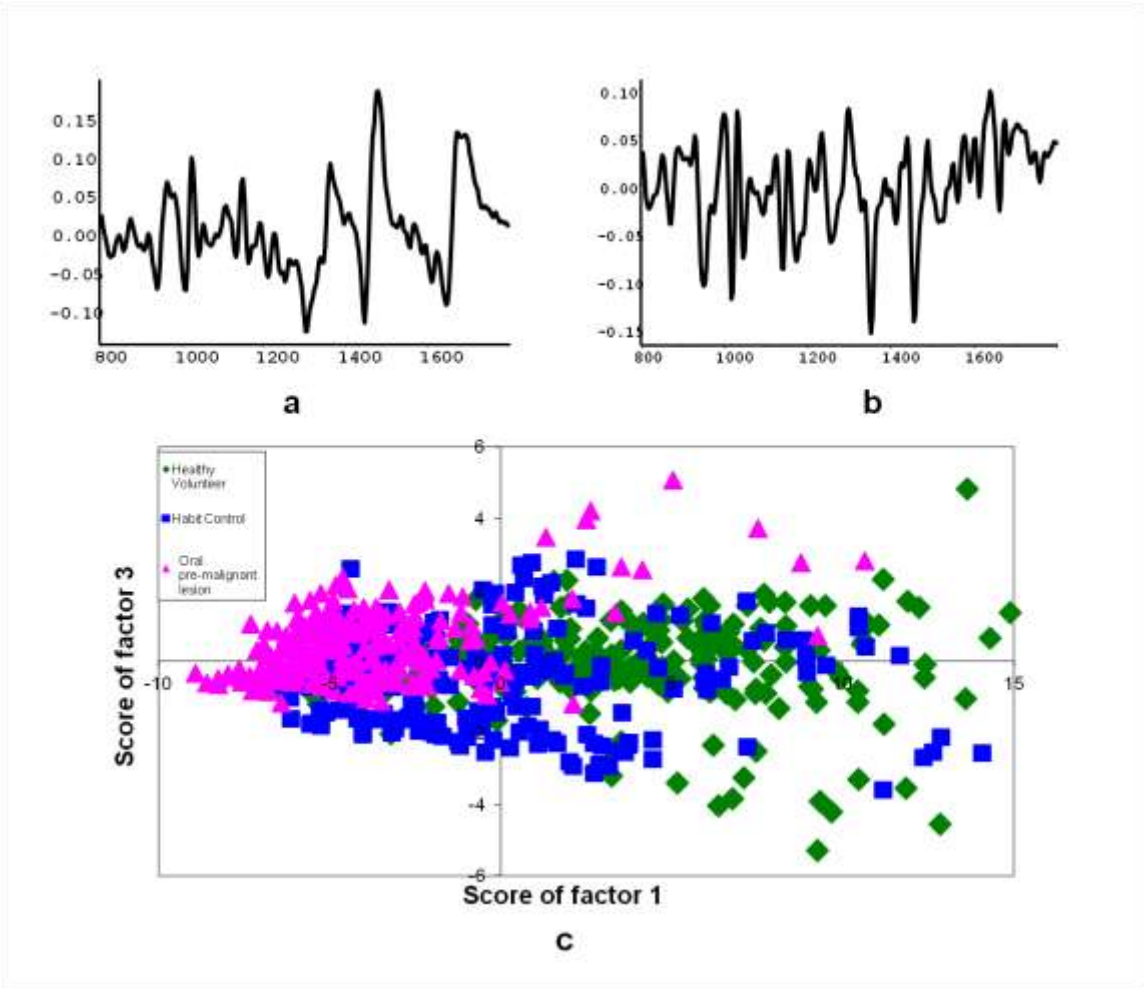


Figure 3.11. PCA for exploring differences between exfoliated cells from healthy volunteers, habit controls, premalignant subjects- Patient-wise approach, a) Loadings of factor 2 b) Loadings of factor 3, c) Scatter plots

As PCA is not a classification tool but is used for data compression and visualization to indicate trends in the data, PC-LDA was employed to explore classification between the groups. Seven factors (Figure 3.12a) were used for the analysis which accounted for ~75% classifications.

Scores of factor 2 and 3 were employed for obtaining scatter plot, as shown in Figure 3.12b. As seen in the scatter plot, better classification was observed between the three groups.

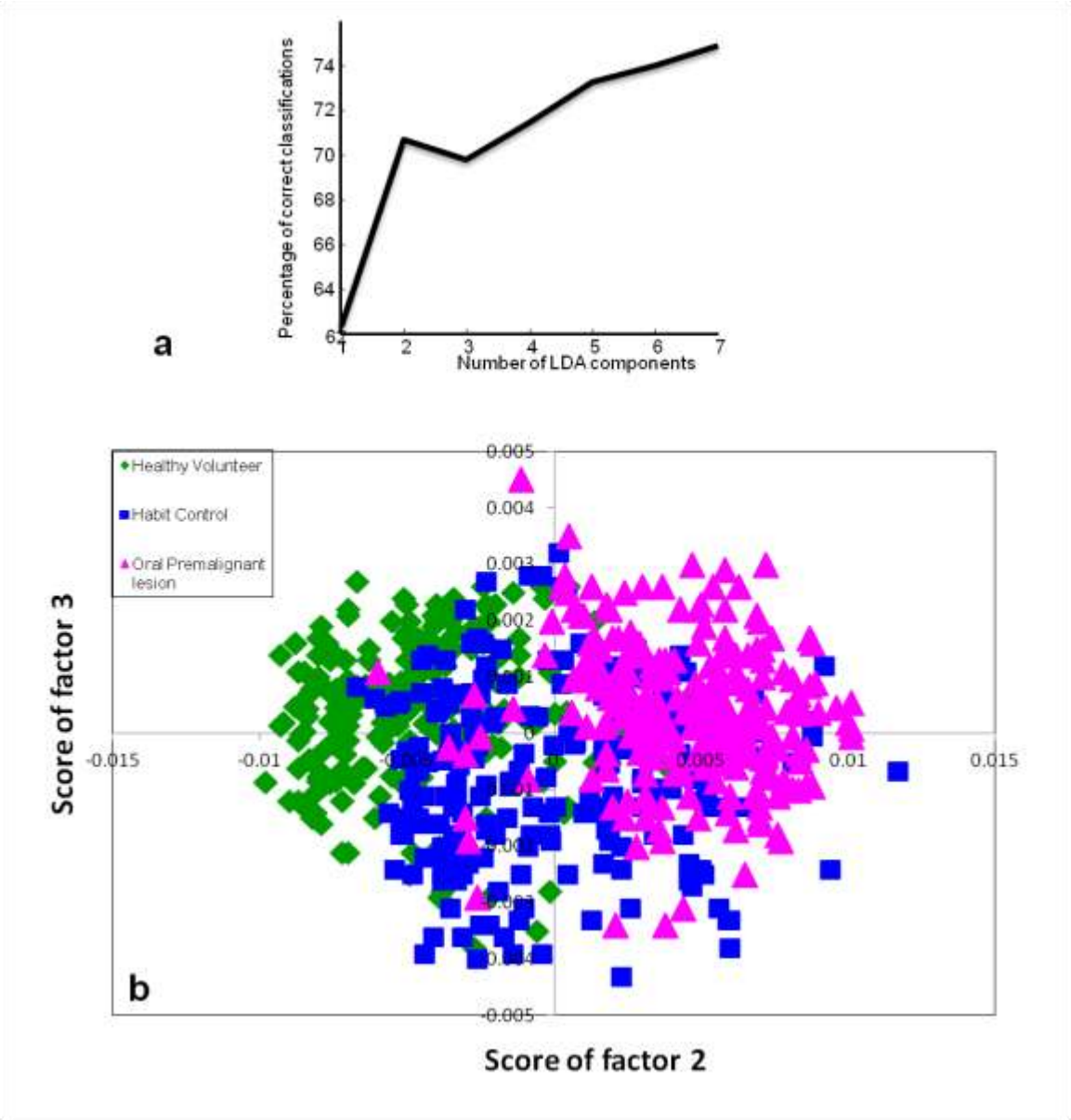


Figure 3.12. PC-LDA for exploring differences between exfoliated cells from healthy volunteers, habit controls, premalignant subjects- spectra-wise approach, a) Scree plot b) Scatter plot

As seen in the LOOCV confusion matrix (Table 3.5), 174/218 HV spectra were correctly classified while 35 and 9 misclassified with HT and OPL, respectively. In case of HT, 137/213 spectra were correctly classified. Major misclassifications were observed with both HV and OPL groups. For OPL, 179/234 spectra were correctly predicted while major misclassifications (49) were with HT. In case of HT, normal proportions of intermediate and superficial cells along with orangeophilia were present in most cases. As previously explained, tobacco exposure may not be uniform and therefore presence of normal cells in these smears may lead to overlap with HV group. HT smears also displayed parakeratotic cells, cells with increased nuclear size but maintained N: C ratios indicating mild atypia due to chronic irritation. Such features were also characteristic of the OPL group, hence the overlap between HT and OPL can be understood.

	Healthy volunteer	Habit control	Premalignant	Total
Healthy volunteer	174	35	9	218
Habit control	38	137	38	213
Premalignant	6	49	179	234

Table 3.5. PC-LDA for exploring differences between exfoliated cells from healthy volunteers, habit controls and pre-malignant subjects using spectra-wise approach- confusion matrix of LOOCV results

3.3.2.3.2 Patient-wise analysis

The second approach- patient-wise approach, wherein average spectrum of each sample was used for data analysis was consequently explored. Average spectra were calculated to yield a true

representative of the sample and also to circumvent the intra-sample heterogeneity. These average spectra were then subjected to multivariate analysis PCA and PC-LDA. PCA was carried out using 10 factors, scores of factor 1 and 3 were used to obtain scatter plot. Factor loadings are shown in Figure 3.13a-b. As seen in Figure 3.13c, slightly distinct clusters for HV, HT and OPL were observed.

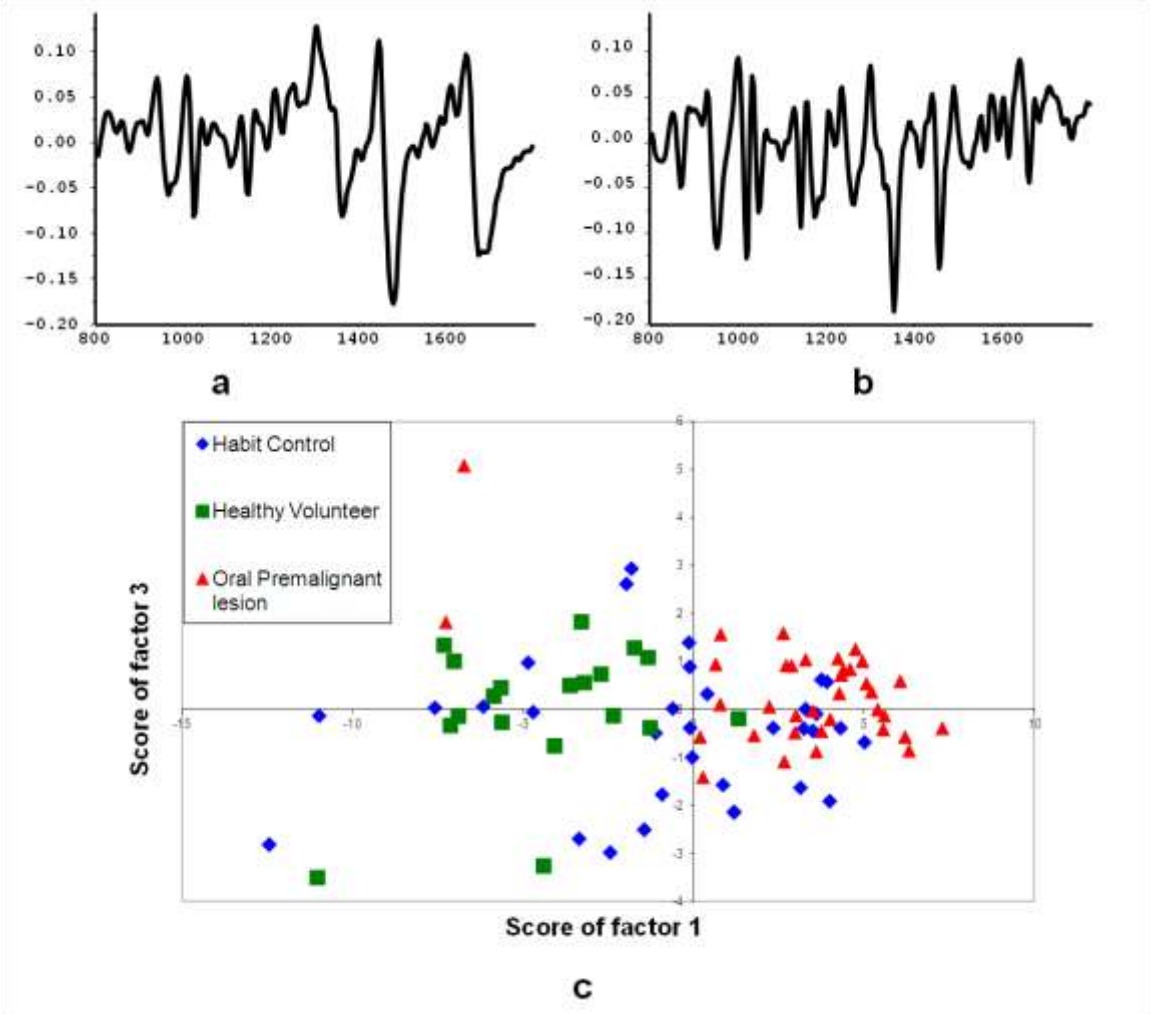


Figure 3.13. PCA for exploring differences between exfoliated cells from healthy volunteers, habit controls, premalignant subjects- Patient-wise approach, a) Loadings of factor 1, b) Loadings of factor 3, c) Scatter plot

PC-LDA was subsequently carried out to build standard models, using 7 factors (Figure 3.14a). The scatter plot shown in Figure 3.14b indicates almost exclusive clusters for the 3 groups.

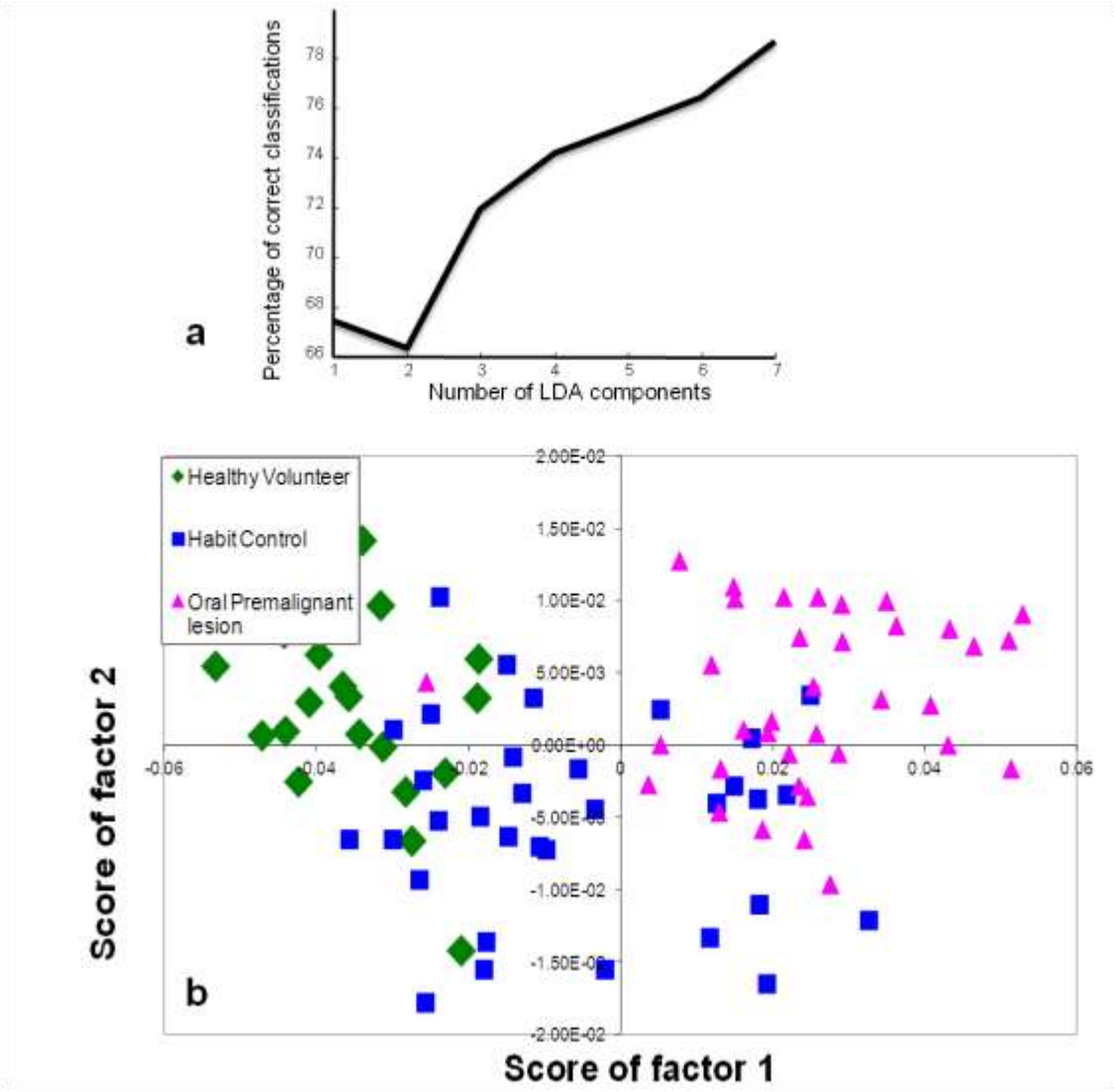


Figure 3.14. PC-LDA for exploring differences between exfoliated cells from healthy volunteers, habit controls, premalignant subjects- Patient-wise approach, a) Scree plot b) Scatter plot

LOOCV findings demonstrate that 14/20 HV, 20/33 HT and 25/36 cases were correctly predicted (Table 3.6). For HV, 6 cases misclassified with HT, for HT, 7 and 6 misclassified with HV and OPL, respectively. While for OPL, 2 and 9 misclassified with HV and HT, respectively.

	Healthy volunteer	Habit control	Premalignant	Total
Healthy volunteer	14	6	0	20
Habit control	7	20	6	33
Premalignant	2	9	25	36

Table 3.6. PC-LDA for exploring differences between exfoliated cells from healthy volunteers, habit controls and pre-malignant subjects using patient-wise approach-confusion matrix of LOOCV

In case of OPL and HT, both site of lesion and tobacco usage, respectively along with their respective contralateral side were included in the study. Thus, misclassifications of HT with HC can be understood because of cells acquired from contralateral mucosa (non-tobacco usage site) of HT subjects with less or no tobacco exposure was observed. For OPL, major misclassifications (n=9) were observed with HT, which can be explained on the basis of the fact that pre-malignant lesions may be heterogenous at microscopic level, thus matching of apparently normal albeit tobacco-exposed mucosal areas with HT cannot be ruled out. Further, OPLs included in the present study were clinically proven leukoplakia and TPK cases having long-term tobacco habits and not histopathologically-confirmed dysplastic lesions. These samples may have been benign keratotic lesions that also demonstrated hyperkeratosis in the

cytological smears. Thus, the overlap between HT and OPL can be explained. Use of biopsy-confirmed dysplastic conditions may lead to better classification between the two groups.

The present study demonstrates the potential of Raman spectroscopy in differentiating healthy and tumor cells, and also a tendency for classification of cells from premalignant conditions. Several cell-based Raman spectroscopic studies using either cell pellets, cell suspensions and thin-cell layers have been carried out. Recent Raman spectroscopic studies have demonstrated difference between normal and CIN cytology using ThinPrep approach (an improvised exfoliated cell sample collection and processing method), while another study using monolayer of cells has discriminated normal, dysplastic and SCC cell lines[280,384]. Thus, Raman mapping of oral cell smears using automated scanning may be an ideal approach for early oral cancer diagnosis.

Studies on cell-pellets have differentiated HPV+ and HPV- cell lines; normal and abnormal cervical exfoliated samples and identified HPV infection in cervical cancer patients. Even a single cell population could be identified in mixed cancer cell populations using this approach[398]. One of the major restraints for this approach is cell sample heterogeneity, as Raman probing involves accumulating information from several stacks of cells. Multiple spectra have to be recorded from cell pellets and then tested against robust standard models. If even a single spectrum is classified as abnormal, as per standard histopathological guidelines, the sample is treated as abnormal. For this diagnosis, pure standard models are required for normal and abnormal groups. Because of heterogeneity in malignant and pre-malignant lesions, it is difficult to obtain pure samples for spectral acquisition. An alternative to this point-spectra approach is the patient-wise approach, where all spectra from a sample are averaged to yield a representative spectrum which is subsequently employed for data analysis. It is known that different sub-

cellular regions of a cell, different cells within a single population and different cell types can be clearly distinguished due to the inherent sensitivity of the Raman spectroscopy. Thus, by calculating average spectra, contributions from all cellular features of the pellet are included and intra-sample variation is thereby minimized. Several previous studies have also employed a similar approach: single cell studies have compared the averaged Raman spectra of many single cells from each sample, while some pellet based studies compared spatially averaged spectra, obtained from many cells at once from a pellet[406-409]. Thus, the patient-wise approach may be a more practical and faster approach in low-resource settings. In this study, the classification rates for identification of premalignant conditions were not very satisfactory, even after cytological correlation analysis. The classification may be improved by using single-cell approaches as against pellet-based studies in light of the cellular heterogeneity, by using high-speed biomolecular imaging approaches like CARS and SRS[410,411].

Classification may also be improved by using better collection devices. Collection devices like Cytobrush are suitable for collection of superficial and intermediate layers. Detection of dysplasia in these layers is plausible as certain degree of nuclear abnormality is also present in the surface layers. This nuclear abnormality reflects the disturbance in cellular maturation in the whole thickness of the epithelium. However, the early changes occurring in the basal layers may not be detected. In the present study, rare presence of basal and parabasal cells was encountered. Thus, early cytological changes could not be detected. Better sampling devices may circumvent this limitation. Further, sampling of keratotic lesions does not yield cells from lower layers. Several studies have indicated need for better sampling devices from such areas or proceed for a biopsy-confirmation.

3.4 Summary

EC is a promising approach for screening of cancers. However, it is more useful for cervix cancer screening as in these cancers the target tissue is anatomically discrete. Screening of wide area of oral mucosa for early microscopic changes is difficult. Alternatively, EC can be employed for evaluation of suspicious lesions or suspected abnormal mucosal areas. After a positive result, transepithelial sampling and mucosal biopsy may be advised for confirmatory purposes. Further, as morphological features may not be sufficient for oral cancer diagnosis, EC can be coupled to Raman spectroscopy. Thus, the utility of oral exfoliative cytology and Raman spectroscopy for oral cancer and pre-cancer was explored in this study. Cytological analysis of same samples was also carried out using Pap staining for parallel confirmation of cellularity and morphological features of exfoliated cell specimens.

- In the proof-of-concept study, Raman spectral differences were explored between exfoliated cells from healthy subjects with and without tobacco habits and contralateral normal and tumor sites in oral cancer patients. In view of blood contamination in tumor samples, all exfoliated specimens were subjected to RBC lysis treatment to remove blood influence. Data analysis was carried out using both spectra-wise and patient-wise approaches. Major misclassifications were observed within healthy (with and without tobacco habits) and cancer (contralateral and tumor) subjects. Major misclassifications between contralateral and tumor groups may be attributed to dysplastic changes in the contralateral mucosa of cancer patients, as also observed during cytological findings.

Thus, early alterations like malignancy associated changes (MAC) in contralateral sites could be also be detected in this study.

- In the subsequent study, differences between cells from healthy subjects (with and without tobacco habits) and premalignant subjects were explored. Results indicate feasibility of classification between healthy without tobacco and premalignant groups, while healthy tobacco group misclassified with both groups. The major misclassifications between premalignant and healthy tobacco group and vice versa could be attributed to increased keratinization of the mucosa due to chronic tobacco exposure. More effective sampling strategies and approaches, especially for the keratotic pre-cancer lesions may improve the classification efficiency and aid in diagnosis of early cancer changes.

Chapter 4

Serum Raman spectroscopy for sub-site classification in oral cancers

Introduction

Oral squamous cell carcinomas (SCC) form a significant proportion (~95%) of the cancers of oral cavity[41]. In South and South-Asian countries, buccal mucosa followed by tongue are the most commonly affected sites for oral cancer[412]. Although histopathologically cancers at these subsites are SCC, they are characterized by differing clinical manifestations. Several clinico-epidemiological studies suggest that the biological characteristics of buccal mucosa and tongue cancers differ: buccal mucosa and tongue cancers demonstrate varying prognosis, aggressiveness, metastasis to lymph nodes and overall survival. Different biological markers like size of tumor, oncogene mutation expression, apoptotic signals have been detected for buccal and tongue tumors on qualitative and quantitative estimation[412-414]. Studies on amplification pattern of oncogenes suggest that different molecular pathways are involved in head and neck SCCs of different sub-sites and localizations[413]. Another study suggests that p53 immuno-expression does not vary according to SCC sub-site in oral cavity[415]. Extensive proteomic studies have indicated that different proteins and functional pathways may be involved in these two cancers. Only few proteins such as alpha B-crystallin and heat shock protein (HSP)-27, mitochondrial ribosomal protein (MRP-L13) showed similar trends of down-regulation and up-regulation, respectively in buccal mucosa and tongue cancers[416,417]. These differential over-expressed proteins can be classified into several categories based on their functions, including glycolytic enzymes, HSPs, detoxification and anti-oxidant proteins, cytoskeletal proteins, proteins involved in mitochondrial and intracellular signaling pathways, and tumor antigens.

Serum Raman spectroscopy based-classification of oral cancers has been demonstrated in the previous chapters. In this chapter, the feasibility of classifying cancer at subsites buccal mucosa and tongue was explored to understand if tumorigenesis at different subsites can be identified. In the first section, potential of resonance Raman spectroscopy in this classification was investigated. In the second section, the classification between buccal mucosa and tongue cancers was explored using conventional Raman employing 785 nm excitation. Analysis was first carried out in 2-model system between buccal mucosa and tongue cancer followed by a more robust 3-model analysis including normal, buccal mucosa and tongue cancers.

4.1. Exploring resonance Raman for classification of buccal mucosa and tongue cancers

The difference between buccal mucosa and tongue cancers was first investigated using resonance Raman employing 532 nm excitation to explore major beta-carotene related changes in the two cancers.

4.1.1. Materials and Methods

4.1.1.1. Subject details

A total of 54 oral cancer patients with tumors at subsites tongue (n=40) and buccal mucosa (n=14) were included in this study. As described previously, blood samples were collected after overnight fasting and seeking an informed written consent from these patients. Healthy subjects (n=16) were also incorporated in this study.

4.1.1.2. Sample collection

As described in section 2.1.1.2, samples were processed to obtain serum and stored at -80 º C.

4.1.1.3. Raman spectral acquisition

Raman spectra were acquired using a commercial Raman microscope WITec alpha300RS (WITec GmbH, Ulm, Germany) as elaborated in section 2.1.1.3.

4.1.1.4. Data pre-processing and analysis

As discussed in section 2.1.1.4, spectra were pre-processed and subjected to PC-LDA followed by LOOCV. Average spectra were also computed for spectral comparisons.

4.1.2. Results

4.1.2.1. Spectral features

Contributions of proteins, DNA, β carotene and several amino acids were observed in the almost similar mean spectra of both groups, as shown in Figure 4.1. The β-carotene specific bands were observed in both groups with higher relative intensity in tongue cancer mean spectrum. Overall, intensity related variations relating to increased DNA and proteins while decreased β carotene were observed in buccal mucosa samples.

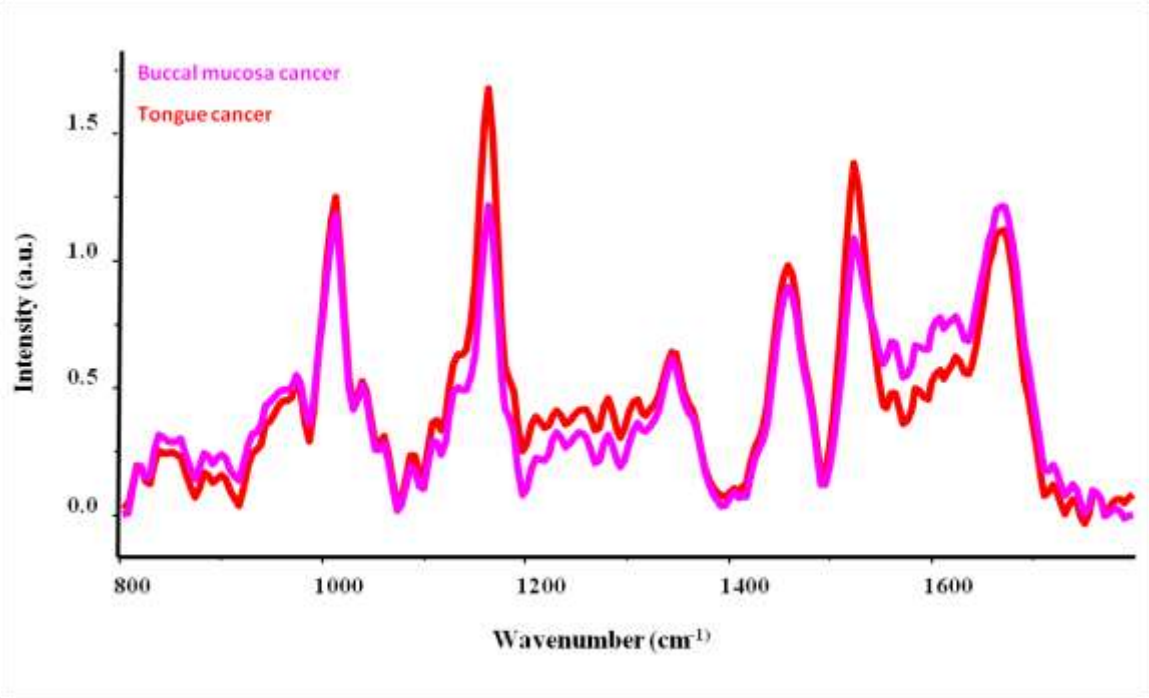


Figure 4.1 Mean Raman spectra of serum from buccal mucosa and tongue cancer subjects for spectral comparisons in resonance Raman study

4.1.2.2. Multivariate analysis

The differences between serum from buccal mucosa and tongue cancer patients was explored by employing PC-LDA followed by LOOCV using patient-wise approach. Both, 2-model and 3-model systems were investigated.

a) 2-model system

PC-LDA was carried out using spectra of buccal mucosa and tongue cancer groups using the patient-wise approach. In the PC-LDA for patient-wise analysis, 4 factors contributing 75% correct classifications were used (Figure 4.2a). Scores of factors 1 and 2 used in the scatter plot shown in Figure 4.2b indicate two overlapping clusters for buccal mucosa and tongue cancer groups.

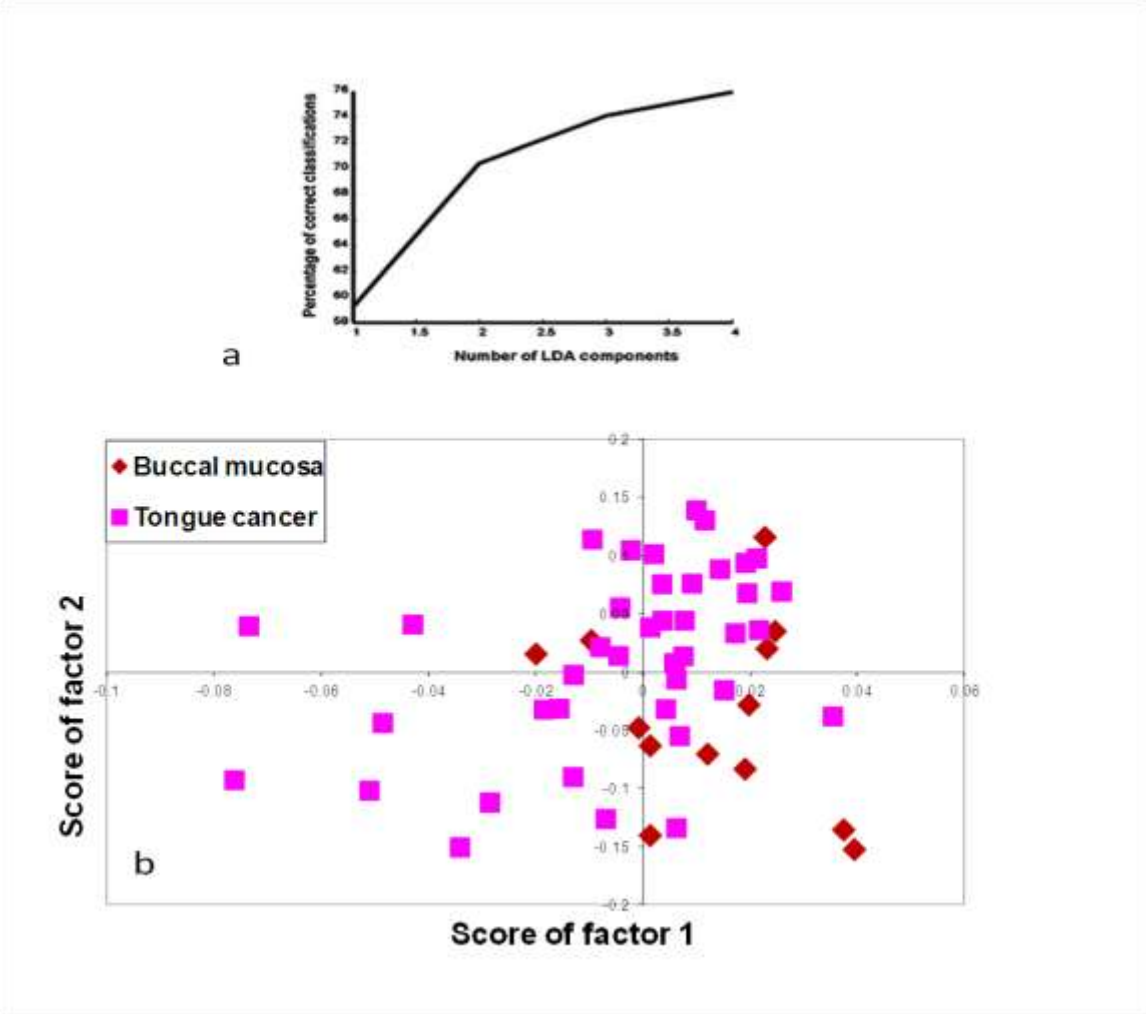


Figure 4.2 PC-LDA for exploring differences between serum from buccal mucosa and tongue cancer subjects in resonance Raman study, a) Scree plot, b) Scatter plot

A leave-one- patient-out cross validated classification efficiency of 64%, and 73% was achieved for buccal mucosa and tongue cancer, respectively (Table 4.1). Thus, preliminary studies suggest feasibility of classifying these two cancers.

	Buccal mucosa cancer	Tongue cancer	Total	% efficiency
Buccal mucosa cancer	9	5	14	64
Tongue cancer	11	29	40	73

Table 4.1. PC-LDA for exploring differences between serum from buccal mucosa and tongue cancer subjects- confusion matrix of LOOCV results in resonance Raman study

b) 3-model system

A 3-model system of normal, buccal mucosa cancer and tongue cancer groups was also developed for a more robust testing of the hypothesis. Three factors accounting for 60% correct classifications were used (Figure 4.3a). Scores of factors 1 and 2 were used to explore classification between the groups and as seen in Figure 4.3b, three minimally overlapping clusters for the two cancers and the normal group were observed.

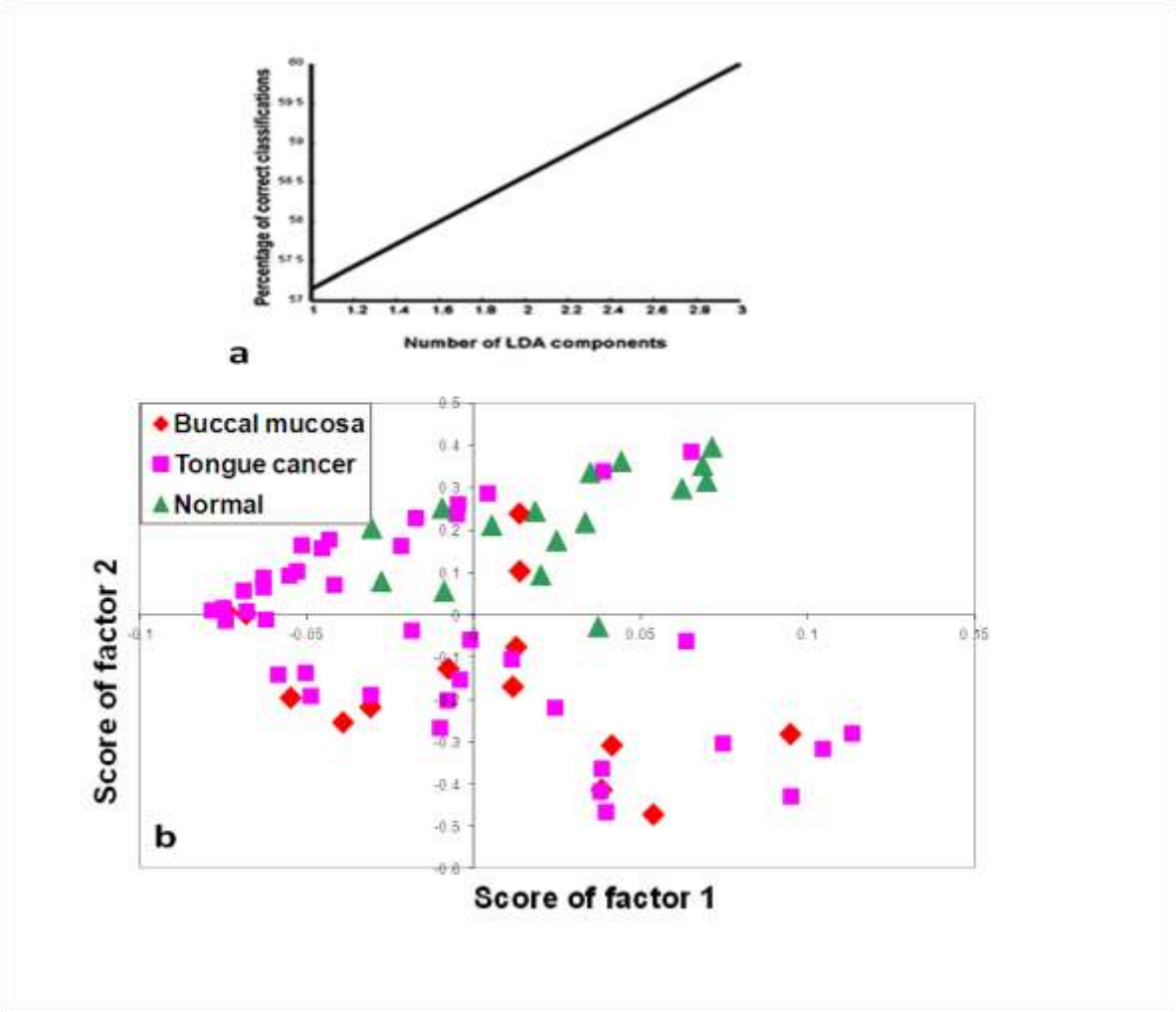


Figure 4.3. PC-LDA for exploring differences between serum from normal, buccal mucosa and tongue cancer serum in resonance Raman study, a) Scree plot, b) Scatter plot

As detailed in Table 4.2, LOOCV showed correct predictions for 6/14 cases for buccal mucosa cancer, 18/40 for tongue cancer and 12/16 for normal group.

	Buccal mucosa cancer	Tongue cancer	Normal	Total	% efficiency
Buccal mucosa cancer	6	6	2	14	42
Tongue cancer	15	18	7	40	45
Normal	1	3	12	16	75

Table 4.2. PC-LDA for exploring differences between serum from buccal mucosa cancer, tongue cancer and normal subjects- confusion matrix of LOOCV results in resonance Raman study

Major misclassifications of tongue cancer were seen with buccal mucosa group and vice versa. These results confirm the observed similarities between these groups in the average spectral analysis, where mainly minor intensity-related variations were observed between the groups.

4.2. Exploring subsite differences in buccal mucosa and tongue cancers using conventional Raman

In the previous resonance Raman studies, differences between buccal mucosa and tongue cancer serum were observed. In this section, spectral differences between these groups were explored using conventional Raman employing 785 nm excitation- the most widely used laser for biomedical applications.

4.2.1. Materials and Methods

4.2.1.1. Subject details

Oral cancer patients with either buccal mucosa (n=62) or tongue (n=58) cancer and normal (n=126) subjects were included in this study, as described in section 2.3.1.1.

4.2.1.2. Sample collection

Blood samples were collected. Serum was separated and stored at -80 º C as described in section 2.1.1.2.

4.2.1.3. Raman spectral acquisition

Spectra were acquired by placing 30 µl volume on calcium fluoride (CaF_2) window and spectra were recorded using fiber-optic Raman microprobe (Horiba-Jobin-Yvon, France), as described in section 2.2.1.3. Spectral acquisition details were- λ_{ex}: 785 nm, objective: 40X, laser power: 30 mW, integration time: 15 s and average: 4.

4.2.1.4. Spectral pre-processing and data analysis

The acquired Raman spectra were pre-processed as detailed in section 2.1.1.4. Multivariate analysis was carried out using supervised PC-LDA followed by LOOCV using patient-wise approach. Average spectra were also computed as described in section 2.1.1.4.

4.2.2. Results and discussion

4.2.2.1. Spectral features

As shown in Figure 4.4, major spectral features were observed at amide I (around 1660 cm^{-1}), amide III (1260 cm^{-1}), δCH_2 deformation (around 1450 cm^{-1}), amino acids like Phe (1008 cm^{-1}), Trp (1560 cm^{-1}, 1622 cm^{-1}), DNA (1080 cm^{-1}, 1340 cm^{-1}) and carotenoids (1162 cm^{-1}, 1530 cm^{-1}). Only minor spectral variations and shifts were observed between the two cancer groups. Major spectral features indicate increased protein -amide III (1242 cm^{-1}), amide I (1646 cm^{-1}, 1681 cm^{-1}) and decreased carotenoids (1168 cm^{-1}, 1529 cm^{-1}) in the buccal mucosa spectra. Thus overall, minor spectral variations in the peaks corresponding to proteins: Phe, amide III, amide I, CH_2 bending, and DNA- DNA backbone and bases were observed between buccal mucosa and tongue cancer.

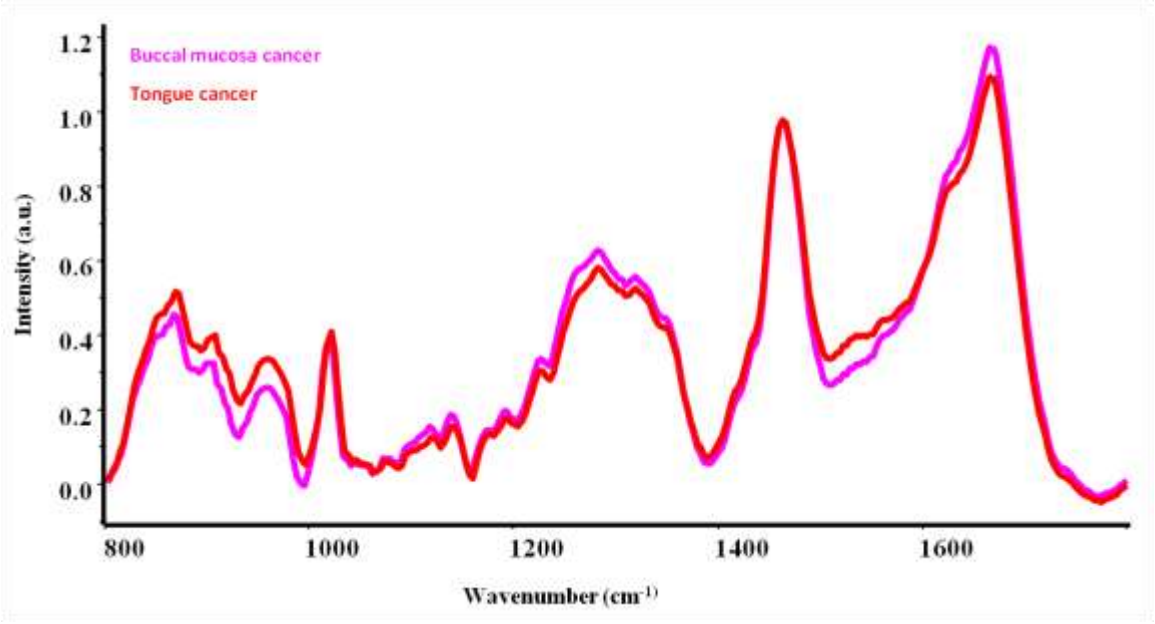


Figure 4.4. Mean Raman spectra of serum from buccal mucosa and tongue cancer subjects for spectral comparisons in conventional Raman study

4.2.2.2. Multivariate analysis

Data analysis was carried out in both 2-model and 3-model systems using PC-LDA followed by LOOCV and patient-wise approach.

a) 2-model system

PC-LDA in a 2-model system of 62 buccal mucosa and 58 tongue cancer spectra was carried out using 6 factors accounting for ~76% efficiency (Figure 4.5a). Scores of factor 1 and 3 were explored were classification. As shown in Figure 4.5b, two overlapping clusters for buccal mucosa and tongue cancer were observed.

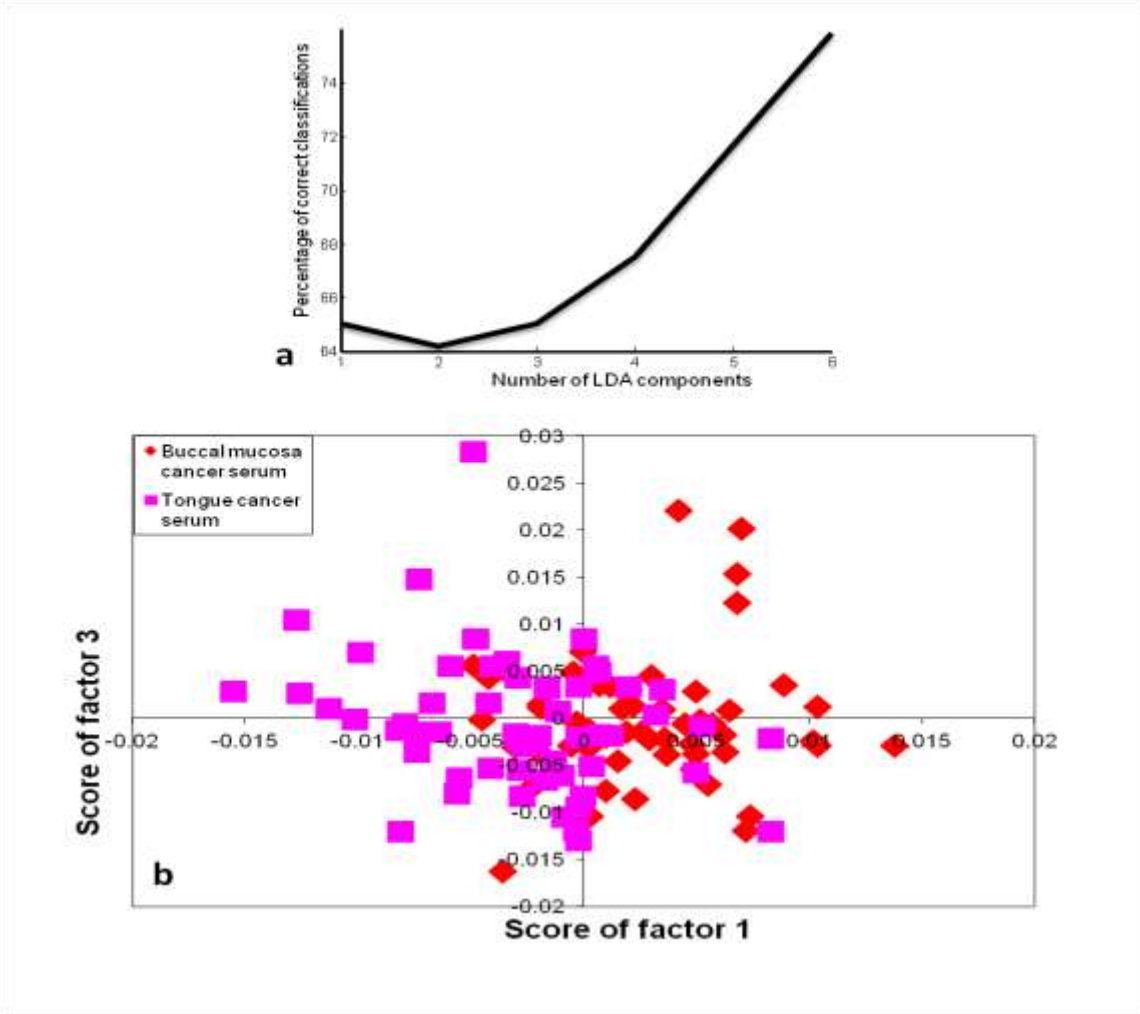


Figure 4.5. PC-LDA for exploring differences between serum from buccal mucosa and tongue cancer serum in conventional Raman study, a) Scree plot, b) Scatter plot

LOOCV findings exhibit correct classifications for 43/62 buccal mucosa and 42/58 tongue cancer (Table 4.3). Overall classification efficiencies of 69.3% and 72.4% were observed for buccal mucosa and tongue cancer, respectively.

	Buccal mucosa cancer	Tongue cancer	Total	% efficiency
Buccal mucosa cancer	43	19	62	69.3
Tongue cancer	16	42	58	72.4

Table 4.3. PC-LDA for exploring differences between serum from buccal mucosa and tongue cancer subjects- confusion matrix of LOOCV results in conventional Raman study

b) 3-model system

A 3-model system for normal, buccal mucosa cancer and tongue cancer groups was also developed for a more robust assessment of the data. PC-LDA for 62 buccal mucosa, 58 tongue cancer and 126 normal subjects was carried out using 5 factors accounting for 75% correct classifications (Fig 4.6a). Scores of factor 2 and 3 were employed to explore classification; scatter plot indicates several misclassifications between the cancer groups, and slight separation of normal group (Fig 4.6b).

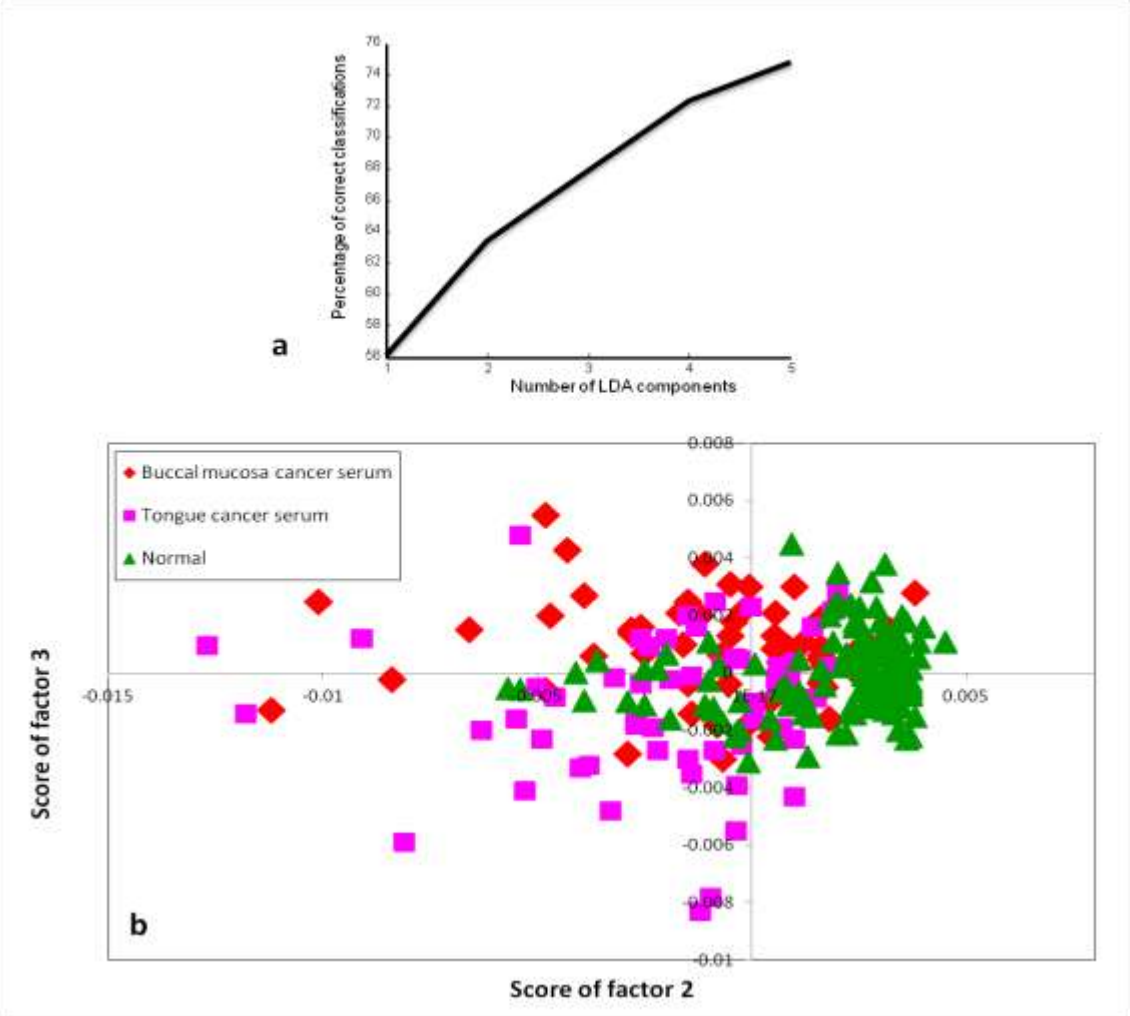


Figure 4.6. PC-LDA for exploring differences between serum from normal, buccal mucosa and tongue cancer serum in conventional Raman study, a) Scree plot, b) Scatter plot

PC-LDA followed by LOOCV yields 38/62 (61%), 31/58 (53%) and 105/126 (83%) correct classifications for buccal mucosa cancer, tongue cancer and normal groups, respectively (Table 4.4).

	Buccal mucosa cancer	Tongue cancer	Normal	Total	% efficiency
Buccal mucosa cancer	38	14	10	62	61.2
Tongue cancer	22	31	5	58	53.4
Normal	11	10	105	126	83.3

Table 4.4. PC-LDA for exploring differences between serum from buccal mucosa cancer, tongue cancer and normal subjects - confusion matrix of LOOCV results in conventional Raman study

Major misclassifications were observed between tongue and buccal mucosa cancer group and vice versa. These results confirm the observed similarities between these groups encountered in the average spectral analysis and are similar to the findings of the resonance Raman studies.

4.3. Summary

Buccal and tongue cancers are known to differ in terms of prognosis, metastasis to lymph nodes, aggressiveness and overall survival. These characteristic biological differences may be responsible for classification between buccal and tongue SCC. It is known that there is differential expression of molecules like p16, p21 and Cyclin D1 during oral cancers, with expression dependent on site of development of SCC. In the present study, Raman spectral differences in serum of buccal mucosa and tongue cancer patients were explored.

- In the resonance Raman study, an average classification efficiency ~70% was observed in the 2-model systems. In the 3-model system, normal could be distinctly classified while major overlap was observed between serum from the two cancer groups.
- In the conventional Raman study, similar classification efficiency for distinguishing buccal mucosa and tongue cancer serum was observed in both 2-model and 3-model systems.

Chapter 5

In vivo Raman spectroscopy of sub-site classification and oral cancer diagnosis

Introduction

Current diagnosis of oral cancers includes visual inspection followed by biopsy and histopathology of suspicious lesions. This procedure suffers from major disadvantages. While visual examination has shown to be useful only for high-risk population that comprises of chronic tobacco and alcohol users[6], biopsy is an invasive procedure. Therefore in most cases, patients and even clinicians are reluctant to perform biopsy for suspicious lesions. Further, in the oral cavity, multiple abnormal areas may be synchronously present. In such circumstances, selection of area of biopsy becomes difficult for the clinician, as taking multiple biopsies is associated with practical and ethical limitations. Overall, the procedure of biopsy is usually associated with patient non-compliance, subjectivity, and possible sampling errors. Histopathological assessment by a trained pathologist is known to be affected by pathologist experience, errors like fatigue and inconsistency in interpretation especially in early/premalignant stages[418,419]. Although blood based markers are extensively being investigated both by molecular and spectroscopic approaches, they rely on systemic changes and not on in situ changes. Further, there is no single established biomarker for oral cancer or pre-cancer diagnosis in clinics. Thus, there is a need for alternative approaches to enable non-invasive and objective in situ diagnosis of oral cancer and pre-cancers.

In view of the inherent accessibility of oral cavity, development of effective, non-invasive diagnostic modalities can help in early diagnosis of pre-cancers and cancers. In vivo spectroscopic methods can be employed as diagnostic adjuncts to the existing methods. Optical methods such as optical coherence tomography (OCT)[420], high-resolution micro endoscopy

(HRME)[421], elastic scattering spectroscopy (ESS)[157], fluorescence[422] and Raman spectroscopy (RS)[19,23] have shown potential in classifying healthy and malignant tissues. Additionally, RS[21], fluorescence spectroscopy[143,144] and OCT[423] have shown potential in discriminating normal, premalignant and malignant conditions. RS has also shown to detect age-related physiological conditions and even subtle cancer-field effects (CFE)[275,424]. However, previously undertaken studies established the potential of RS for only buccal mucosa subsite. Specific sites of oral cavity cancer include the lip, floor of the mouth, oral tongue (anterior two-thirds of the tongue), lower alveolar ridge, upper alveolar ridge, retromolar trigone (retromolar gingiva), hard palate, and buccal mucosa. The anatomical location of cancer occurrence may also serve as a prognostic indicator. In Western countries, tongue (27.5%), floor of mouth (24.5%) and lip (19%) account for ~70% of all cancers while in the Indian sub-continent, buccal mucosa and tongue along with lip are the most commonly affected subsites[425,426.427] Studies on amplification pattern of oncogenes suggest that different molecular pathways are involved in head and neck SCCs of different sub-sites and localizations[413]. This difference in site-distribution of oral cancers can be attributed to the variation in carcinogen exposure of the native populations[428].

The oral cavity is composed of distinct anatomic sites[429]. It is known that sub-sites in the oral cavity have distinct origins, and consequently distinct anatomical and molecular features[430,431]. Previous fluorescence[432] and reflectance spectroscopy[433,434] studies have demonstrated definite spectral contrasts between different sub-sites in healthy populations. Three major RS studies in the finger-print and high wavenumber region have also demonstrated differences arising due to epithelial and sub-epithelial structures, submucosa and degree of keratinization in subsites of healthy subjects and suggested clustering of sites based on anatomical and spectral

similarities[20,22,272]. While the first in vivo study by Guze et al suggested that spectra for different oral sites within the same ethnic group are significantly different and clearly separable, the consequent study by Bergholt et al divided major sub-sites into 3 different clusters based on their histological and spectroscopic characteristics. The 3 groups included- a) buccal mucosa, inner lip and soft palate, b) dorsa, ventral tongue and floor of mouth, c) gingiva and hard palate. Krishna et al proposed division of sub-sites into 4 major anatomical clusters- a) outer lip, and lip vermillion, b) buccal mucosa c) hard palate d) dorsal, lateral and ventral tongue and soft palate. Further, the authors also suggest the use of anatomy-matched algorithms to increase discrimination between healthy and abnormal conditions[276]. Thus, there exist ambiguities in the anatomical classification of the subsites, especially buccal mucosa, tongue and lip sub-sites. Further, it is known that incidence of oral cancer is age-related[428]. In the first section of this chapter, spectral contrast between the anatomical sites buccal mucosa, lip and tongue were investigated first in healthy as well different pathological conditions. As enough subject accrual could not be obtained for all subsites, the study was carried out on 3 major sites for oral cancer development- buccal mucosa, lip and tongue. In the suceeding section, the effect of this spectral contrast on efficiency of diagnostic algorithms was explored. In the final section, detection of age-associated physiological changes in the healthy subjects was explored at the three sub-sites. The effect of these physiological changes on the diagnostic algorithm was also subsequently evaluated.

5.1. Exploring anatomical differences in healthy, contralateral, premalignant and tumor conditions

Several studies have shown that different subsites in the oral cavity may classify based on their anatomic and thus spectroscopic differences. Because of similar and dissimilar epithelial and sub-epithelial structures, Bergholt et al[272] and Guze et al[20] proposed grouping of movable labial mucosa (lip) and buccal mucosa and tongue as a distinct cluster, while another study[22] did not include inner lip as one of the subsites. To corroborate the grouping of buccal mucosa and lip in a single cluster and tongue as an independent cluster, the anatomical differences between these 3 subsites were investigated in healthy, contralateral, premalignant and malignant conditions.

5.1.1. Materials and Methods:

5.1.1.1. Sample Details

Oral cancer patients visiting the OPD of TMH, Mumbai were recruited for this study. Eighty-five (85) subjects with tobacco-associated histopathologically verified SCC lesions in the oral cavity were recruited. Tumors were encountered mainly at buccal mucosa, lip, retro-molar trigone (RMT) and tongue subsites in these patients. Spectra were acquired from these lesions. From these patients, spectra were also acquired from different contralateral normal sites in the oral cavity namely, buccal mucosa, lip, tongue, floor of mouth (FOM), RMT, gingival and tongue. Premalignant patches like leukoplakia and condition like oral sub-mucous fibrosis were also observed in oral cavity of some patients (n=40) at sites buccal mucosa, lip and tongue. These premalignant lesion/conditions were also included in the study and spectra were acquired from these sites. Although no histopathological verification was done for these premalignant sites, such lesions found in the oral cavity of oral cancer patients are usually dysplastic. Multiple spectra (2-3) were recorded from each contralateral normal, premalignant or malignant site.

These sites were selected based on clinical presentation and diagnosed by an experienced senior oral pathologist. Thus, spectral acquisition was based only on clinical diagnosis. As the overall aim of the study was to achieve non-invasive or in vivo diagnosis of oral cancers, no biopsies were taken from any subjects. Histo-pathological data was available only for the 'tumor' subclass from oral cancer patients. Information about tobacco usage, age, sex and tumor grade of all subjects was obtained from the Electronic Medical Record (EMR) system of Tata Memorial Hospital, Mumbai, India.

A total of 72 subjects from Advanced Center for Treatment, Research, Education in Cancer (ACTREC), Mumbai were recruited as healthy volunteers (HV). Subjects recruited for the study comprised of both genders between the age group 21-60 years. Like in the case of patients, spectra were acquired from 7 different sites, buccal mucosa, lip, tongue, FOM, RMT, gingival and tongue in the oral cavity. Subjects with no current/past tobacco or alcohol habits and no history of malignancy were considered as HV. To avoid any differences because of the mouth environment, subjects were required to wash their mouth with distilled water before spectral acquisitions. A total of 1128 HV spectra, 1107 contralateral spectra, 106 premalignant and 277 tumor spectra were acquired from different subsites of the recruited subjects. All subjects were recruited only after obtaining an informed and written consent. The study was approved by IRB. Subject accrual is summarized in Table 5.1.

Sr. No.	Category	No. of subjects	Tobacco habits	Age range	Spectra acquired
1.	Healthy	72	No	21-60 years	1128
2.	Oral cancer	85			
	-Tumor	85	Yes	27-82 years	277
	-Contralateral	85	Yes	27-82 years	1107
	-Premalignant	40	Yes	27-71 years	106

Table 5.1. Subject accrual summary for in vivo studies

5.1.1.2. Raman spectroscopy

Spectra were recorded with an HE-785 commercial Raman spectrometer (Horiba-Jobin-Yvon, France)[275]. This system consists of a diode laser (Process Instruments, USA) of 785 nm wavelength as the excitation source, a high efficiency spectrograph with fixed 950 gr/mm grating coupled with a CCD (Synapse). The instrument has no movable parts and the spectral resolution as specified by the manufacturer is 4 $cm^{-1.}$ The commercially available In Photonics (Inc, Downy St. USA) probe consisting of a 105 µm excitation fiber and a 200 µm collection fiber (NA-0.40) was used to couple the excitation source and the detection system. The photographic representation of instrument employed in the in vivo study is presented in Figure 5.1. The estimated spot size and depth of the field as per the manufacturer's specifications are 105 µm and 1 mm, respectively. The working distance of the probe is 5 mm and therefore, a detachable spacer of length 5 mm was attached at the tip of the probe to maintain focus during all measurements. Prior to each measurement, these spacers were disinfected by CIDEX (Johnson

and Johnson, Mumbai, India) solution to avoid inter-subject contamination. Spectral acquisition parameters were: λ_{ex}–785 nm, laser power–80 mW, spectra were integrated for 3 seconds and averaged over 3 accumulations.

Figure 5.1. Photographic representation of instrument employed for in vivo applications (Inset- probe placement in oral cancer patient)

As spectra were recorded directly on patients, certain important logistic factors were considered before the study was initiated. Disinfection of the stainless steel spacers was meticulously carried out to ensure patient safety. Acquisition time was patient and clinician friendly to obviate movements during spectral acquisition and obtain good quality spectra. The laser power at the sampling point employed in the present study (0.08 W) was well within the maximum permissible exposure allowed on skin (0.3 W/cm^2) for a 785-nm continuous wave laser based on

the American National Standards Institute (ANSI) regulation. To prevent variability from physician to physician in terms of probe placement and effect of varying pressures applied to probe during acquisition, spectral acquisition was carried out at pre- designated sites and spectra were acquired by a single clinician for all subjects. Spectra were acquired from specific sites - from the buccal mucosa, spectra were acquired from mucosa opposing teeth positions of second pre-molar, first molar and second molar. In case of lip, labial mucosa opposing the two central incisors was selected for spectral acquisition. For tongue, spectra were acquired from lateral borders in the anterior part of the tongue. In case of hard palate, point above the hard palate-soft junction was used for spectra recording. Floor of mouth points included those adjacent to the first and second premolars. For retromolar trigone, mucosal area posterior to the last molar was employed for spectral acquisition. Gingival area on the premolars was used as spectra recording sites. Photographs representing the spectra acquisition scheme are shown in Figure 5.2.

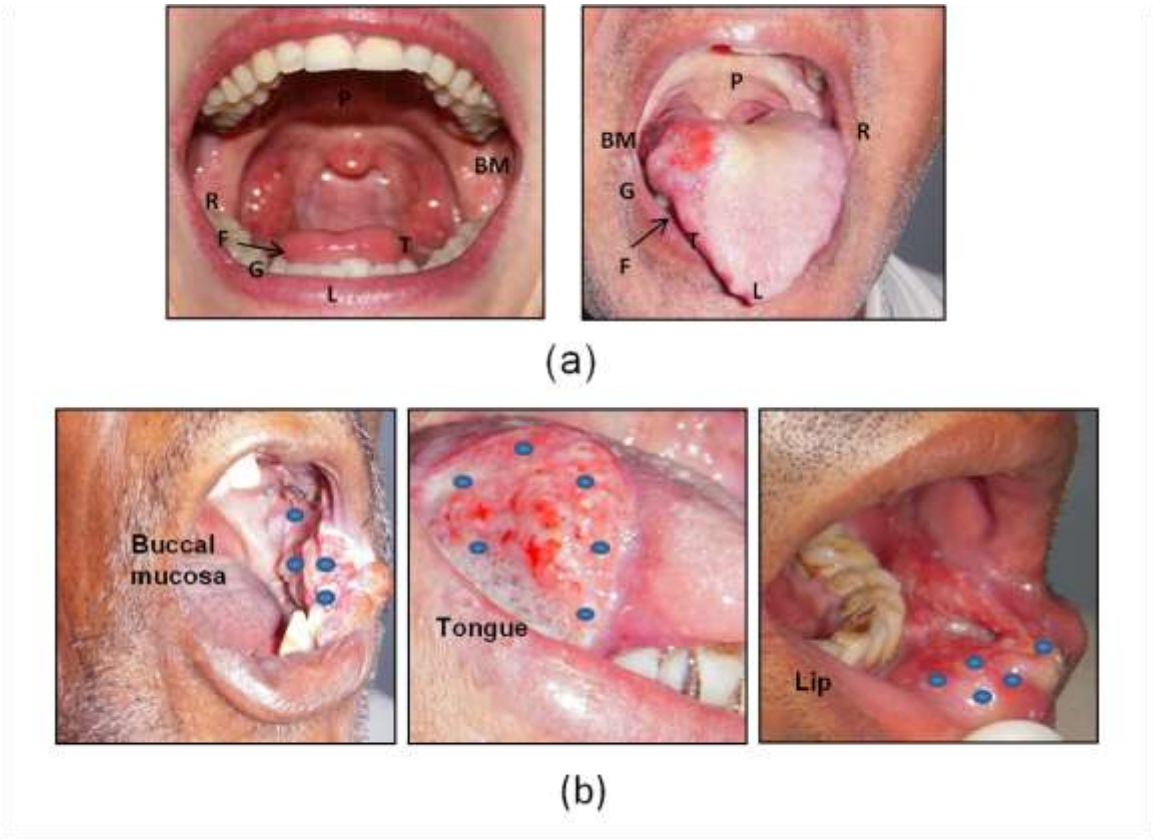


Figure 5.2. Photographic representation of spectra acquisition scheme for in vivo study, a) Spectral acquisition patterns in healthy and contralateral mucosa, b) Spectra acquisition from tumors at different subsites.

5.1.1.3. Spectral pre-processing and data analysis

Spectra were corrected for CCD response followed by the subtraction of background signals from optical elements and substrate as described in section 2.1.1.4. First-derivatized, interpolated in 1200-1800 cm^{-1} and normalized spectra were used as input for multivariate analysis PC-LDA followed by LOOCV. Average spectra were also computed for spectral comparisons across the groups. Background-corrected spectra were baseline corrected prior to derivatization by fitting a

fifth order polynomial function. These baseline-corrected, smoothed (Savitzky-Golay, 3) and normalized spectra were used for spectral comparisons.

5.1.2. **Results and discussion**

5.1.2.1. **Spectral features**

Representative mean spectra from all groups- healthy, contralateral normal, premalignant and tumor for buccal mucosa, lip and tongue have been shown in Figure 5.3 a-c. Spectral features of healthy buccal mucosa show higher lipid dominance (two sharp features in the amide III region, strong δCH_2 bend and sharp peak in the amide I region and an ester band at around 1742 cm^{-1}) while lip shows both lipid and protein features. Tongue spectra are characterized by higher protein features (features like broad amide III, amide I along with CH_2 deformation and CH_2, CH_3 wagging (collagen assignment) at 1343 cm^{-1}). With increase in severity of pathological condition, an increase in protein-related features like broadening of amide III and amide I, shifted amide III and shifted δCH_2 bend were observed for all groups (Figure a(ii-iv), b(ii-iv) and c(ii-iv)). DNA and heme-related features were also observed at 1340, 1480 and 1560 cm^{-1} in these groups. Thus, spectra from contralateral, premalignant and tumor had increased contribution from proteins, DNA and heme. These findings corroborate our previous ex vivo and in vivo findings[19,21,23,275,424].

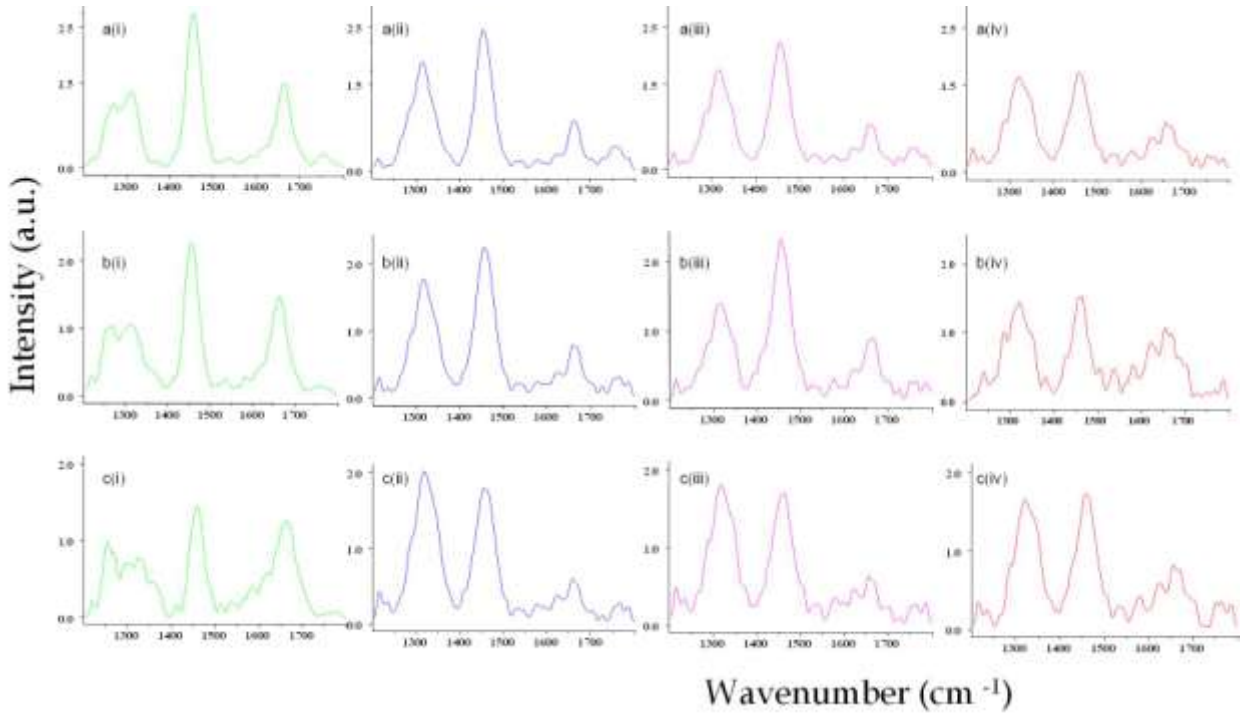


Figure 5.3. Mean in vivo Raman spectra from subsites a) Buccal mucosa, b) Lip, c) Tongue in (i) healthy (green), (ii) contralateral (blue), (iii) premalignant (pink) and (iv) tumor (red) conditions, for spectral comparisons

Spectral features were also investigated in depth in individual conditions. In healthy subsites, healthy buccal mucosa demonstrates two sharp features in the amide III region, strong δCH_2 bend and sharp peak in the amide I region and an ester band at around 1742 cm^{-1}. Broad features in amide III, amide I and δCH_2 deformation in the lip spectra indicate significant contributions from proteins with lipids. In the healthy tongue spectrum, features around 1240 cm^{-1}, broad amide III and amide I regions, CH_2, CH_3 wagging (collagen assignment) at 1343 cm^{-1} indicate a predominantly protein-dominated spectrum, which is expected owing to the muscular nature of this subsite.

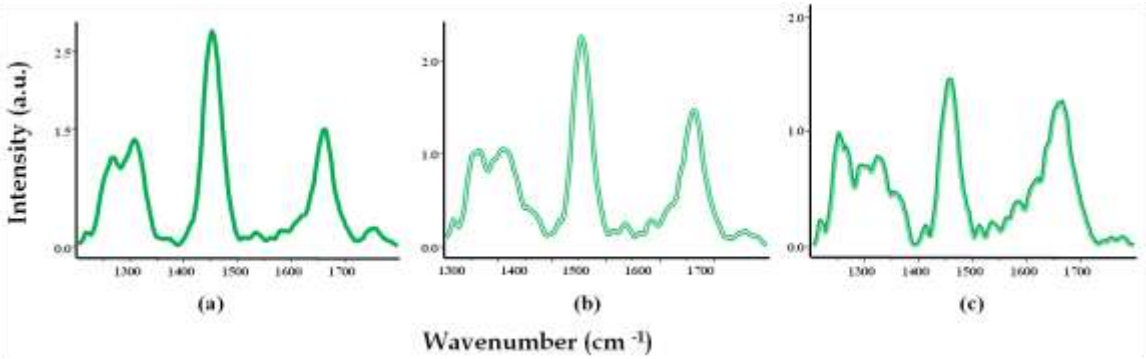


Figure 5.4. Mean in vivo Raman spectra from healthy sites, a) buccal mucosa, b) lip and c) tongue, for spectral comparisons

Spectra acquired from buccal mucosa, lip and tongue subsites on the contralateral sites in oral cancer patients are shown in Figure 5.5. Like in healthy, sharp characteristic features were observed for buccal mucosa at 1306, 1445, 1655 and 1744 cm^{-1}. In case of lip, broader features were observed at amide III, amide I and δCH_2 deformation. In case of tongue, broad amide III, collagen features at 1343, features at 1440 and 1450, and 1648 and 1660 cm^{-1} were observed.

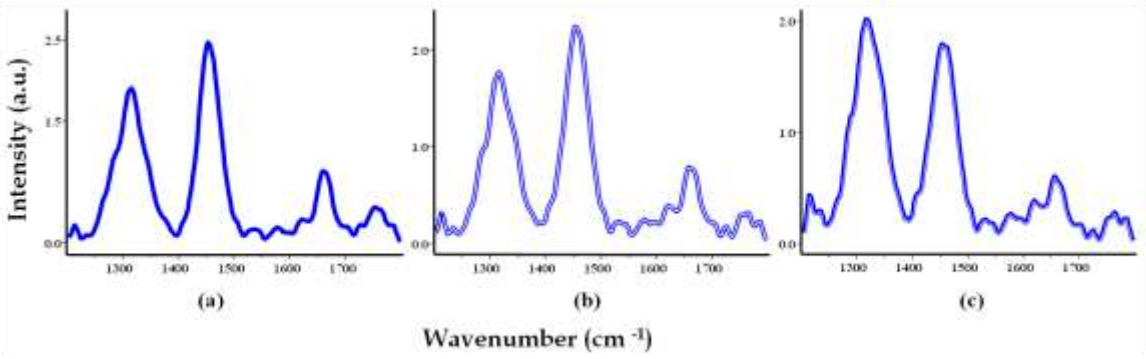


Figure 5.5. Mean in vivo Raman spectra from contralateral sites, a) buccal mucosa, b) lip and c) tongue, for spectral comparisons

Spectra from premalignant lesions from buccal mucosa, lip and tongue are shown in Figure 5.6. Similar spectral profiles were observed for all groups, with major features at 1306, 1448, 1650

and 1750 $cm^{-1.}$ Minor spectral shifts were observed for lip and tongue at amide III and only tongue in CH_2 deformation region.

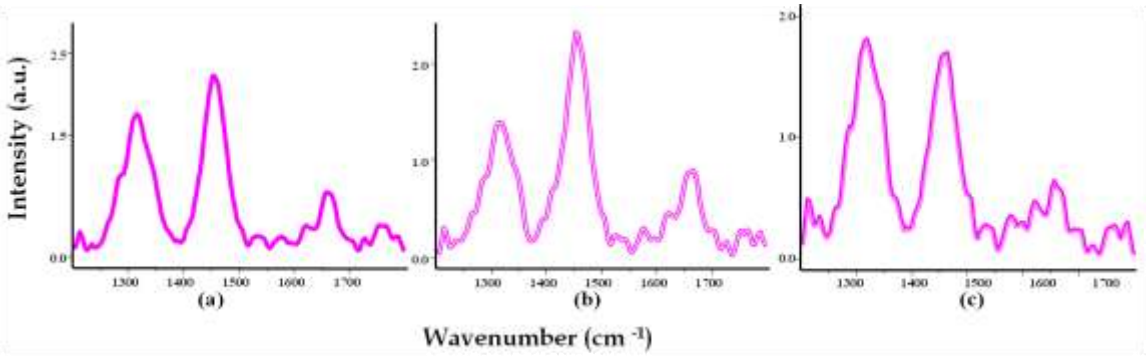


Figure 5.6. Mean in vivo Raman spectra from premalignant lesions at sites, a) buccal mucosa, b) lip and c) tongue, for spectral comparisons

Spectra acquired from oral squamous cell carcinoma present on buccal mucosa, lip or tongue are shown in Figure 5.7. Major spectral features observed in all spectra were observed at 1310, 1450, 1615, 1648 cm^{-1}, representing amide III, CH2 deformation, C=C in protein, amide I. Almost same spectral profiles with minor intensity variations were observed, indicating tumors at these subsites may be difficult to classify.

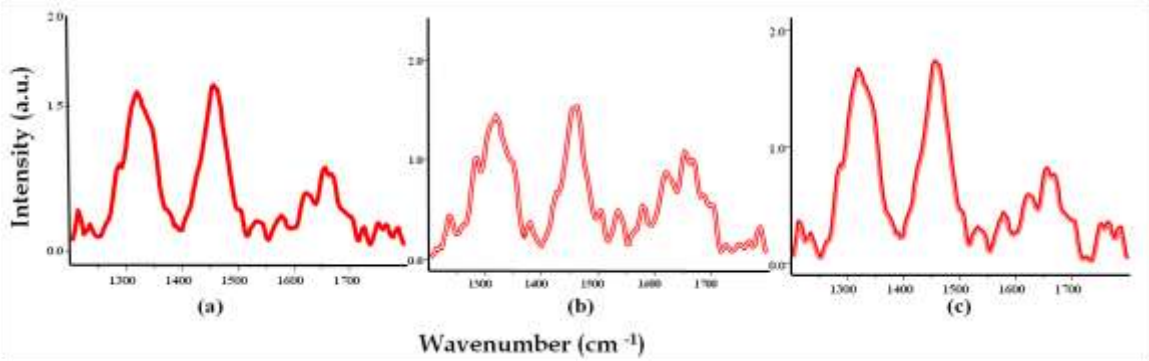


Figure 5.7. Mean in vivo Raman spectra from tumors at sites, a) buccal mucosa, b) lip and c) tongue, for spectral comparisons

5.1.2.2. Multivariate analysis

The contribution of anatomical differences at buccal mucosa, lip and tongue were explored in healthy, contralateral, premalignant and tumor conditions using PC-LDA followed by LOOCV. For PC-LDA, factors with minimum over-fitting and maximum classification efficiency were selected for analysis.

a) Anatomical differences in healthy sub-sites

To investigate spectroscopically evident anatomical differences between these subsites, spectra acquired from buccal mucosa, lip and tongue of 72 healthy volunteers were subjected to PC-LDA. 3 factors accounting for ~72% variance were employed for PC-LDA (Figure 5.8a). Scatter plot indicates two distinct clusters for buccal mucosa and tongue, while lip misclassifies with both subsites, as shown in Figure 5.8b.

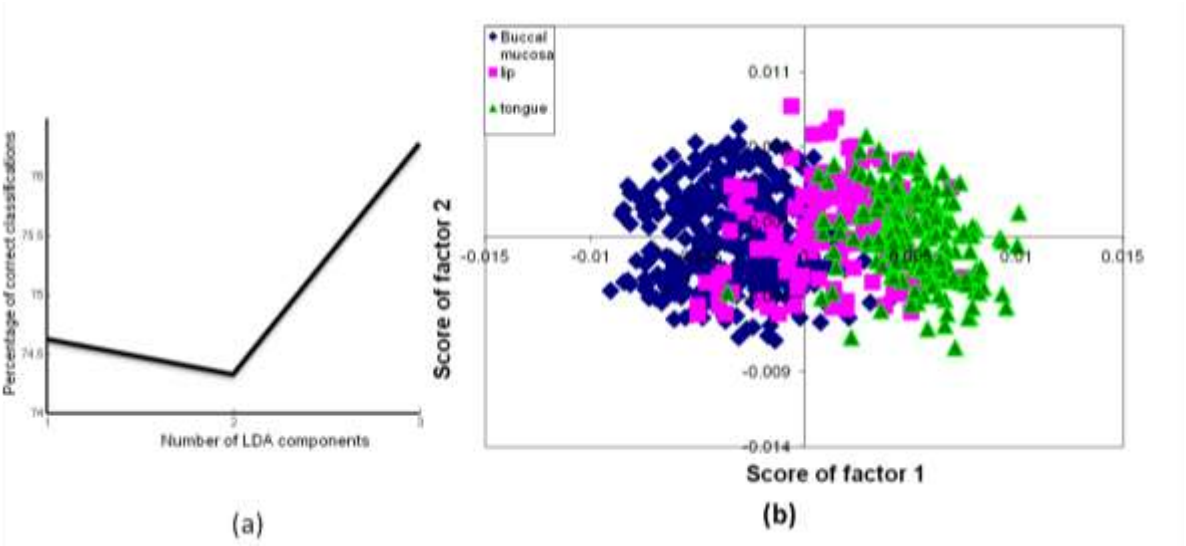


Figure 5.8. PC-LDA for exploring subsite differences in vivo between buccal mucosa, lip and tongue in healthy subjects, a) Scree plot, b) Scatter plot

LOOCV findings shown in Table 5.2 indicate that of the 353 buccal mucosa spectra, 298 were correctly identified while 52 and 3 misclassified with lip and tongue, respectively. For lip, 62/137 spectra (45.2%) were correctly classified, while 31 and 44 misclassified with buccal mucosa and tongue, respectively. In case of tongue, 143/176 spectra were correctly classified, 32 misclassified as lip. Thus, buccal mucosa and tongue are classified as almost distinct subsites, major overlap of both was observed with the intermediate subsite lip.

	Buccal mucosa	Lip	Tongue	Total
Buccal mucosa	298	52	3	353
Lip	31	62	44	137
Tongue	1	32	143	176

Table 5.2. PC-LDA for exploring subsite differences in vivo between buccal mucosa, lip and tongue in healthy subjects- confusion matrix of LOOCV results

b) Anatomical differences in contralateral sub-sites

After observing differences in healthy subsites, spectra acquired from buccal mucosa, lip and tongue subsites on the contralateral area were subjected to PC-LDA. PC-LDA was carried out using 2 factors accounting for 66% variance (Figure 5.9a). PC-LDA scatter plot shows three overlapping clusters. (Figure 5.9b)

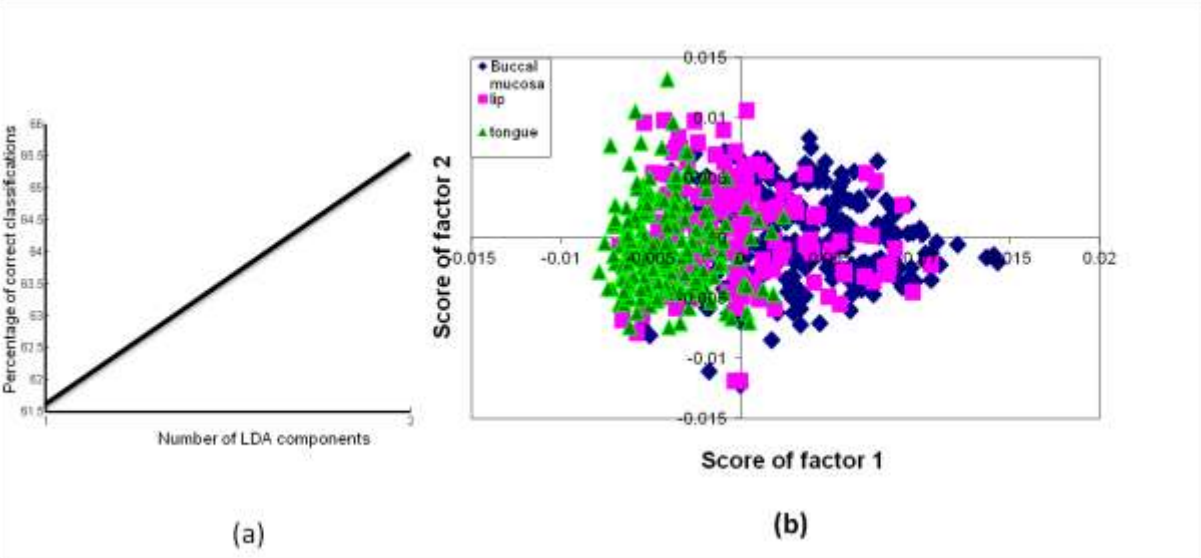


Figure 5.9. PC-LDA for exploring subsite differences in vivo between buccal mucosa, lip and tongue in contralateral sites of oral cancer patients, a) Scree plot, b) Scatter plot

Out of 231 buccal mucosa spectra, 161 were correctly classified, 53 and 17 misclassified as lip and tongue, respectively. 73/164 lip spectra were correctly classified (~44%), while 41 and 50 misclassified as buccal mucosa and tongue. For tongue, 149/191 spectra were correctly predicted as tongue, 38 and 4 misclassified as lip and buccal mucosa (Table 5.3). Thus, like healthy subsites, contralateral buccal mucosa and tongue are almost distinct subsites, overlap of both is majorly observed with lip.

	Buccal mucosa	Lip	Tongue	Total
Buccal mucosa	161	53	17	231
Lip	41	73	50	164
Tongue	4	38	149	191

Table 5.3. PC-LDA for exploring subsite differences in vivo between buccal mucosa, lip and tongue in contralateral sites of oral cancer patients - confusion matrix of LOOCV results

c) Anatomical differences in premalignant sub-sites

Subsites with presence of premalignant conditions leukoplakia and oral sub-mucous fibrosis on buccal mucosa and leukoplakia on lip and tongue were employed for this data analysis. 3 factors were employed for PC-LDA, covering ~77% classification efficiency (Figure 5.10a). Less number of spectra, with respect to buccal mucosa, were present in the lip and tongue groups. Scatter plot shown in Figure 5.10b using scores of factor 1 and 2 display 3 slightly overlapping clusters.

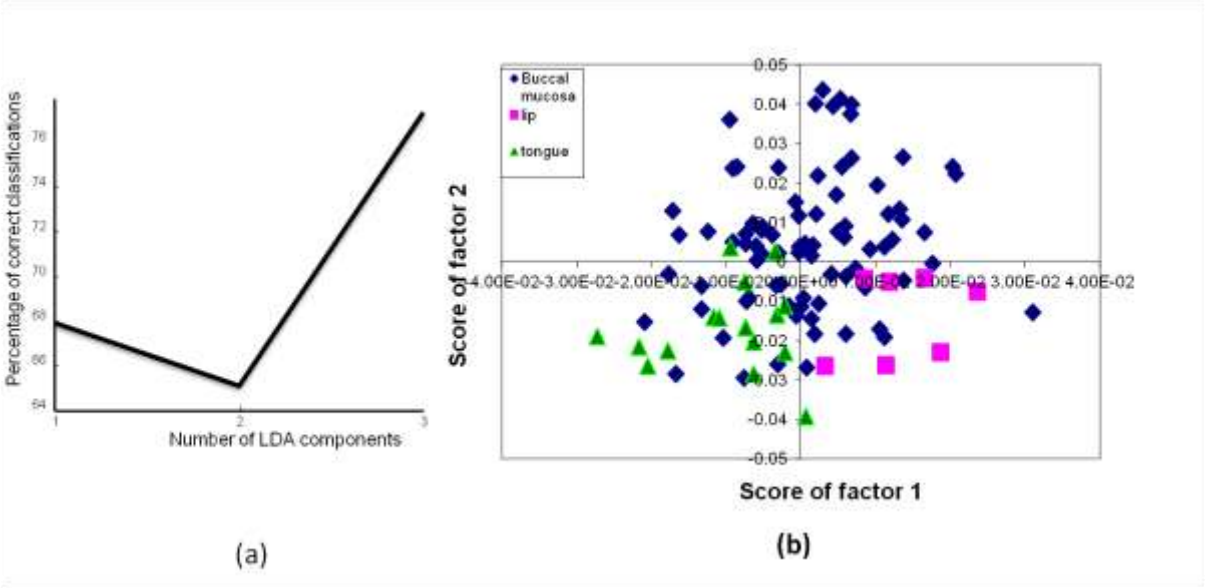


Figure 5.10. PC-LDA for exploring subsite differences in vivo between buccal mucosa, lip and tongue in premalignant sites of oral cancer patients, a) Scree plot, b) Scatter plot

As shown in LOOCV findings in Table 5.4, 58/82 (70%) buccal mucosa spectra were correctly identified while 9 and 15 misclassified as lip and tongue. The 7 spectra for lip mostly misclassified with buccal mucosa (4) and tongue (1), 2 were correctly predicted as lip. For tongue, 15/17 were correctly identified as tongue leukoplakia. In the premalignant lesion/condition found on these subsites, buccal mucosa and tongue could be distinctly identified. However, due to less number of spectra in lip and tongue groups, no definite conclusions could be derived.

	Buccal mucosa	Lip	Tongue	Total
Buccal mucosa	58	9	15	82
Lip	4	2	1	7
Tongue	1	1	15	17

Table 5.4. PC-LDA for exploring subsite differences in vivo between buccal mucosa, lip and tongue in premalignant sites of oral cancer patients - confusion matrix of LOOCV results

d) Anatomical differences in malignant sub-sites

Spectra acquired from tumor at buccal mucosa, lip and tongue were further explored for classification between tumor-site using 2 factors accounting for ~52% correct classifications (Scree plot shown in Figure 5.11a). No definite cluster for any groups was identified in the scatter plot shown in Figure 5.11b.

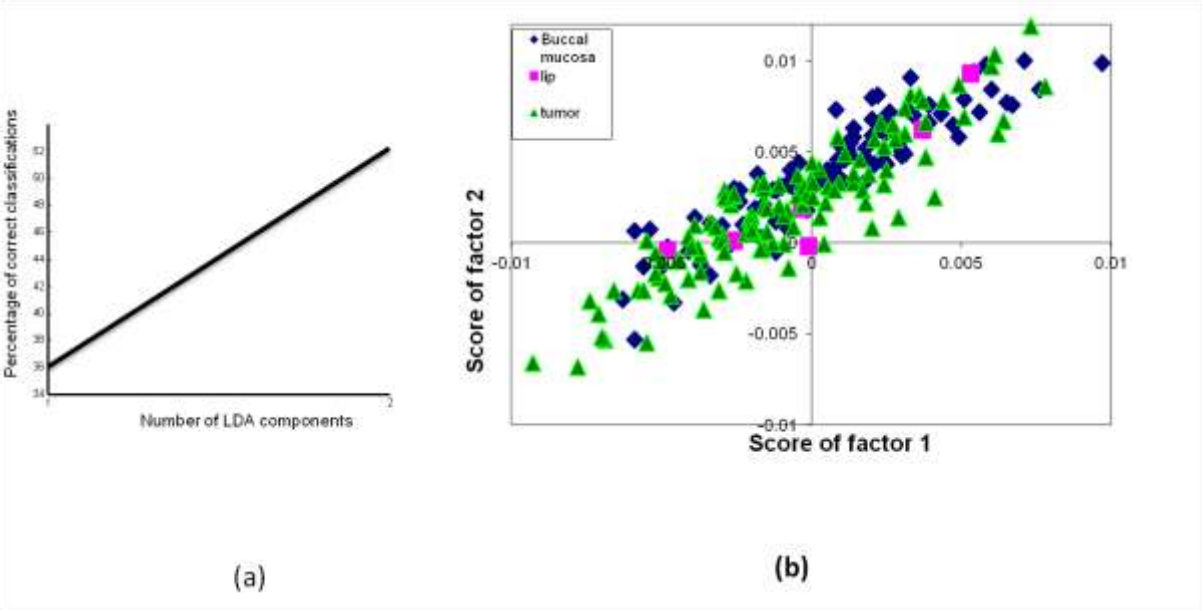


Figure 5.11. PC-LDA for exploring subsite differences in vivo between buccal mucosa, lip and tongue in malignant sites of oral cancer patients, a) Scree plot, b) Scatter plot

Of the 120 buccal mucosa tumor spectra, only 49 were correctly predicted. For 9 lip tumor spectra, 1 was correctly identified while for 124 tongue tumor spectra, 80 were correctly classified as tongue (Table 5.5). In tumors, loss of characteristic anatomical features is encountered as tumor mass is the major contributor to the Raman signal.

	Buccal mucosa	Lip	Tongue	Total
Buccal mucosa	49	22	49	120
Lip	5	1	3	9
Tongue	39	5	80	124

Table 5.5. PC-LDA for exploring subsite differences in vivo between buccal mucosa, lip and tongue in malignant sites of oral cancer patients - confusion matrix of LOOCV results

In the first section that investigated spectral contrasts at the 3 anatomical sites, it was observed that RS- based discrimination between buccal mucosa, lip and tongue reduced sequentially from healthy to the most severe pathological condition (tumor), with total loss of anatomical influence in tumor conditions. It is known that OSCC at different sites are histopathologically a squamous cell carcinoma, and therefore the site of cancer cannot be detected based on histopathology of the tumor specimens[435]. Due to loss of information from the underlying architectural arrangement and supposedly seeming similarity in squamous cell carcinoma tissue irrespective of subsite, tumor subsites could not be classified. The similarity of lip subsite with both buccal mucosa and tongue could be explained due to the fundamental anatomical characteristics of this subsite. Although both buccal mucosa and lip are covered with the same lining mucosa, the underlying connective tissue and submucosa are different for these subsites. The similarity in buccal and lip could be preliminary attributed to the presence of a thick, stratified, squamous, non-keratinizing mucosa at both sites. This could be the basis of what has been previously reported by Bergholt et al and Guze et al. However, below the labial lining mucosa, dense fibrous connective tissues and

the submucosa consisting of collagen and elastin fibers interspersed with fat and small mixed glands, orbicularis oris muscle is located. In case of buccal mucosa, beyond the thick non-keratinized epithelium, dense fibro-elastic tissue from lamina propria penetrates the fatty elastic submucosa consisting prominently of loose connective tissue, large blood vessels, nerves, salivary glands and adipose tissue. In contrast, tongue is lined by a specialized mucosa below which thin, papillated lamina propria connects with the underlying compact masses of skeletal muscle fibers. Thus, due to the overlying epithelium and small similarities in lamina propria and submucosa, buccal mucosa and lip misclassify with each other. Due to presence of underlying skeletal muscle fibers of orbicularis oris, lip misclassifies with tongue, majorly composed of skeletal muscle fibers[436-438]. These anatomical attributes give lip an intermediate position in subsite classification. However, in lieu of observed similarities and disparities, it is pertinent to explore the effect of this inter-anatomical variability on Raman based oral cancer diagnostics.

5.2. Raman spectroscopy based diagnostics for oral cancers at different subsites

To enable accurate tissue diagnosis and characterization, the influence of the anatomical variations on normal versus abnormal classification using RS was explored in this section. RS of healthy and pathological-contralateral normal, premalignant and malignant conditions was undertaken for both, pooled subsites and individual subsites. In the first step, spectra from all subsites (buccal mucosa, lip and tongue) were clustered and efficiency of classification for - healthy, contralateral normal, premalignant and malignant classification was explored using PC-LDA. In the second step, efficiency of RS in classifying healthy versus pathology at individual subsites-buccal mucosa, lip and tongue was undertaken. Data analysis was carried out using PC-LDA followed by LOOCV.

5.2.1. Materials and Methods:

5.2.1.1. Sample Details

The same subjects employed in the previous section - Healthy volunteers (n=72) and oral cancer patients (n=85) were also included in the study. Subject accrual is elaborated in section 5.1.1.1.

5.2.1.2. Raman spectroscopy

Spectra were acquired from healthy volunteers at different subsites, mainly buccal mucosa, lip and tongue as described in the previous section. From oral cancer patients, spectra were acquired from site of tumor, contralateral normal subsites and premalignant lesions (if any), as described in section.5.1.1.2. Photographic representation of spectra acquisition is provided in Figure 5.12. Spectra were acquired using a fiber-optic probe coupled HE-785 Raman instrument.

Figure 5.12. Photographic representation of spectra acquisition scheme from healthy and pathological conditions at a particular subsite- buccal mucosa for in vivo study

5.2.1.3. Spectral preprocessing and data analysis

As described in section, spectra were corrected for CCD response, background signals, first derivatized, interpolated in 1200-1800 cm^{-1} range, normalized and subjected to PC-LDA followed by LOOCV as described in section 2.1.1.3.

5.2.2. Results and discussion

5.2.2.1. Spectral features

As previously shown in Figure 5.2, healthy spectra were characterized by higher lipid features. With increase in severity of pathological condition, an increase in protein-related features was observed for all subsites. DNA and heme-related features were also observed with increasing pathology in contralateral, premalignant and tumor. The change from lipid to protein and DNA can be explained on the basis of changes in tissue architecture and morphology, increased angiogenesis and cellular proliferation in pre-malignant and malignant conditions. These tentative assignments were made based on available literature[329,330].

Spectra differences were also evaluated in detail at individual subsite. Spectral features from healthy, contralateral normal, premalignant and malignant conditions (Figure 5.13a-d) acquired from buccal mucosa demonstrate sequential decrease in lipid and increase in protein features. Heme-related features that signify increased presence of blood were observed in tumor spectra.

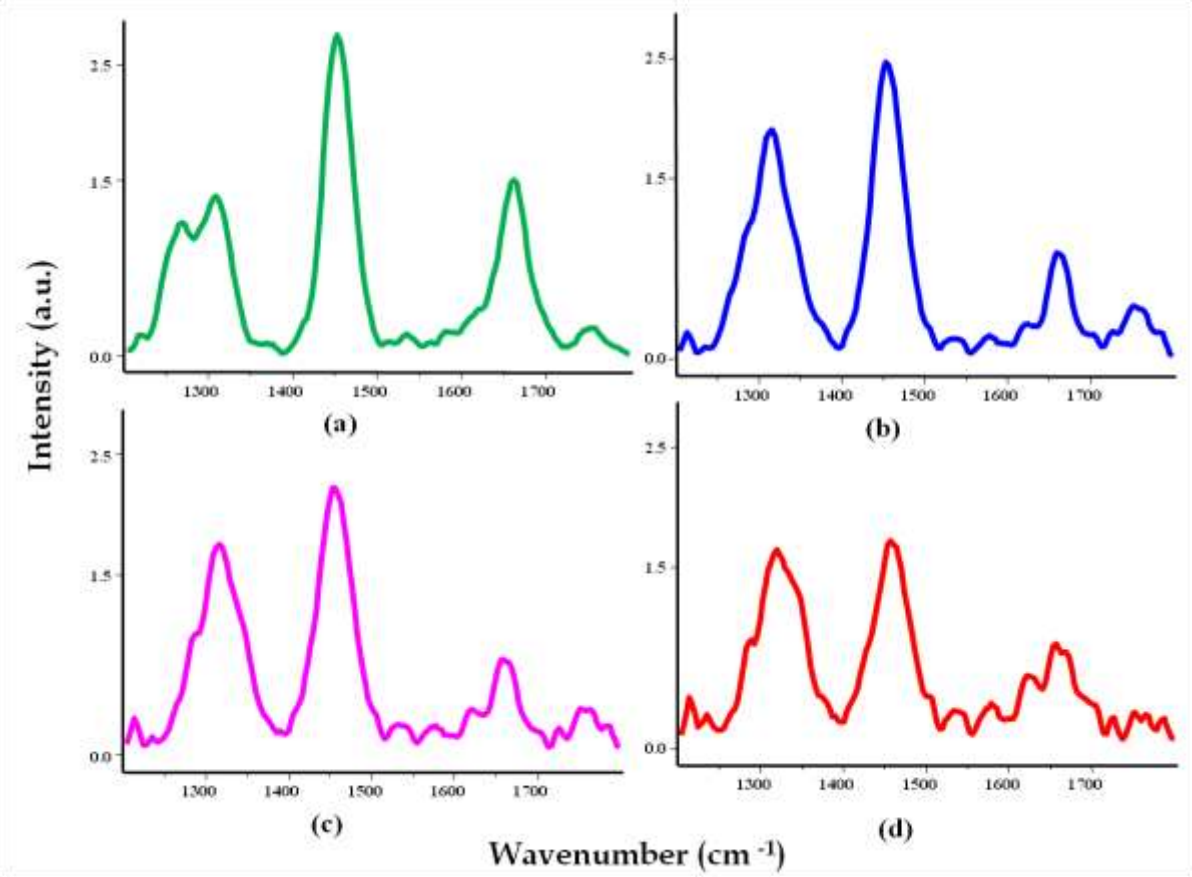


Figure 5.13. Mean in vivo Raman spectra from a) healthy, b) contralateral, c) premalignant and d) tumor conditions at buccal mucosa subsite for spectral comparisons

Average spectra from healthy, contralateral normal, premalignant and malignant conditions acquired from lip are shown in Figure 5.14a-d. Like buccal mucosa, a decrease in lipid-like features are observed in pathological conditions. Spectra from contralateral and premalignant

sites show very similar attributes. Broad amide III, amide I and heme-related features were observed in tumor spectra.

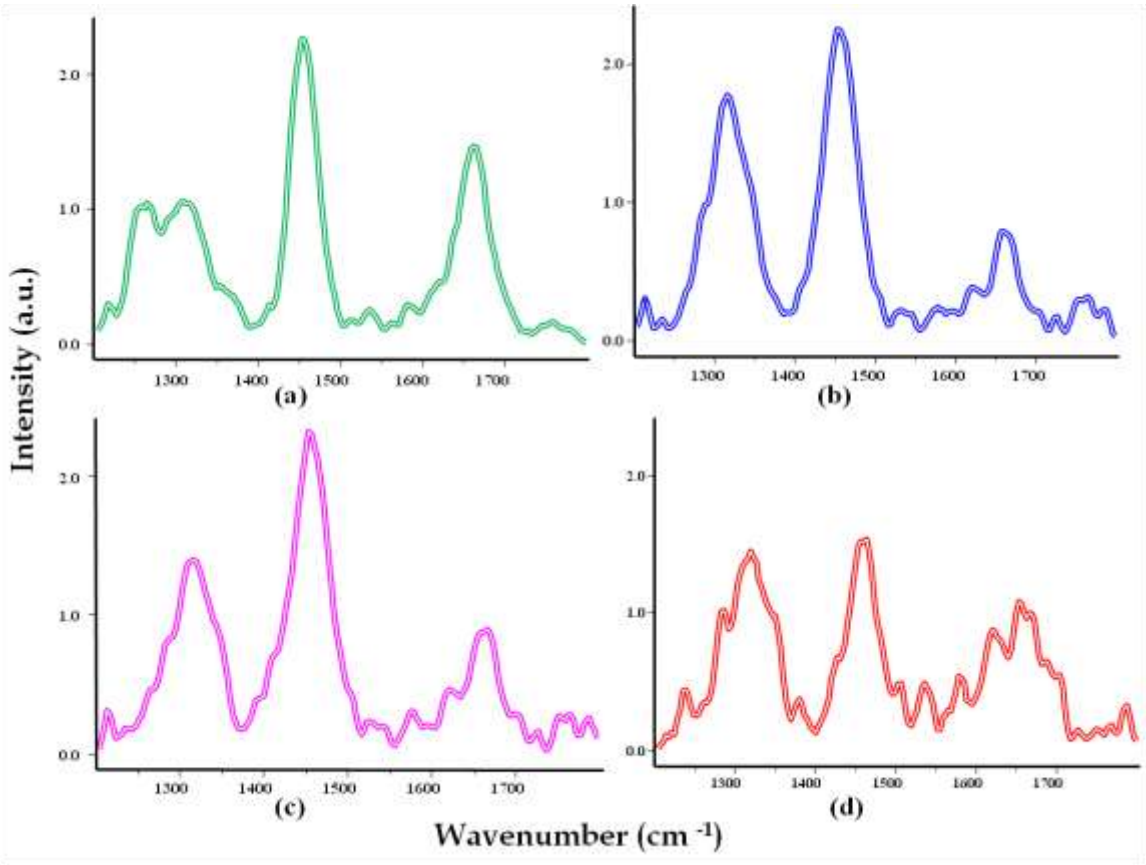


Figure 5.14. Mean in vivo Raman spectra from a) healthy, b) contralateral, c) premalignant and d) tumor conditions at lip subsite for spectral comparisons

Average spectra for healthy, contralateral normal, premalignant and malignant for tongue subsite are shown in Figure 5.15a-d. Spectra indicate changes in protein and lipid content in the healthy and pathological conditions. Spectral shifts and intensity-variations were observed at amide III, I and CH_2 deformation with increasing pathological condition.

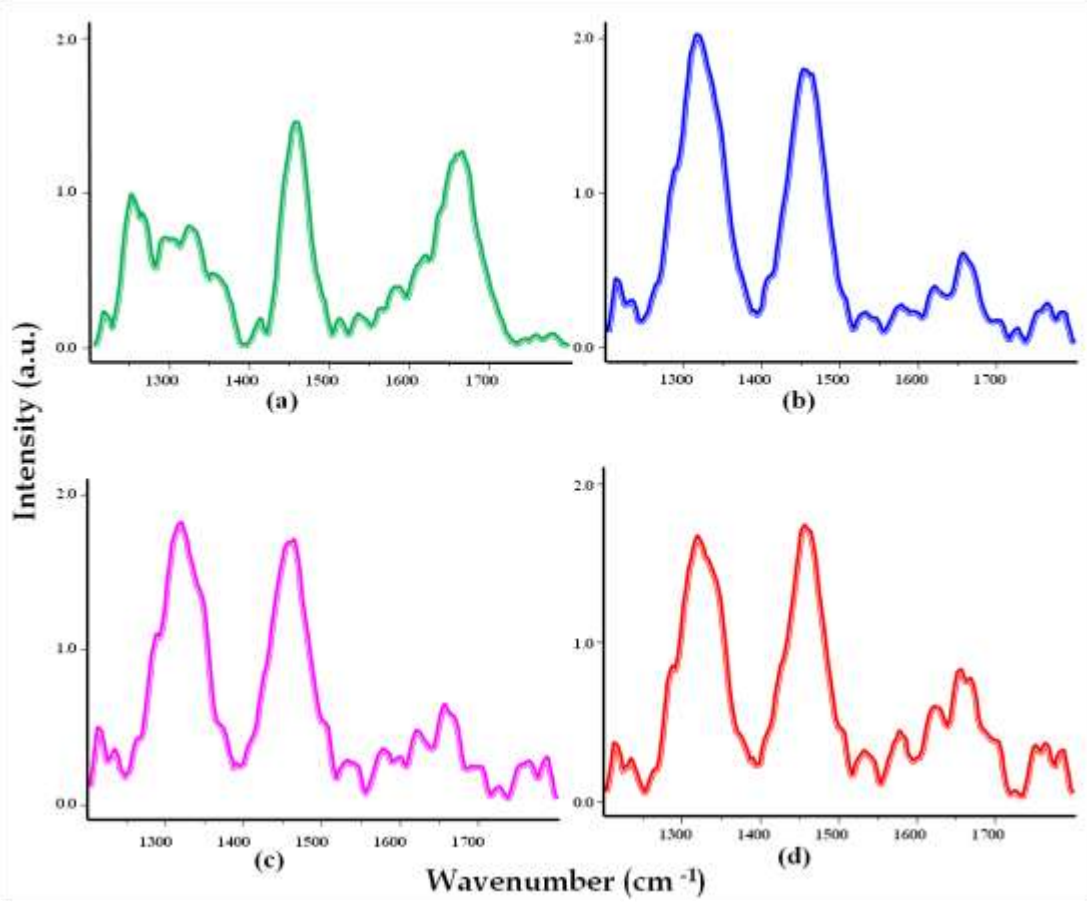


Figure 5.15. Mean in vivo Raman spectra from a) healthy, b) contralateral, c) premalignant and d) tumor conditions at tongue subsite for spectral comparisons

5.2.2.2. Multivariate analysis

Multivariate analysis was carried out to explore efficacy of discriminating healthy and pathological conditions at pooled subsites and each individual subsite.

a) Raman spectroscopic differences between healthy and pathology at pooled subsites: buccal mucosa, lip and tongue

Spectra from buccal mucosa, lip and tongue were grouped together as per their pathological status: healthy, contralateral normal, premalignant and malignant and PC-LDA was carried out using 3 factors (76% efficiency of classification- Figure 5.16a). Scatter plot shown in Figure 5.16b indicate distinct groups for healthy and pathological conditions, while overlap is observed between all the pathological conditions.

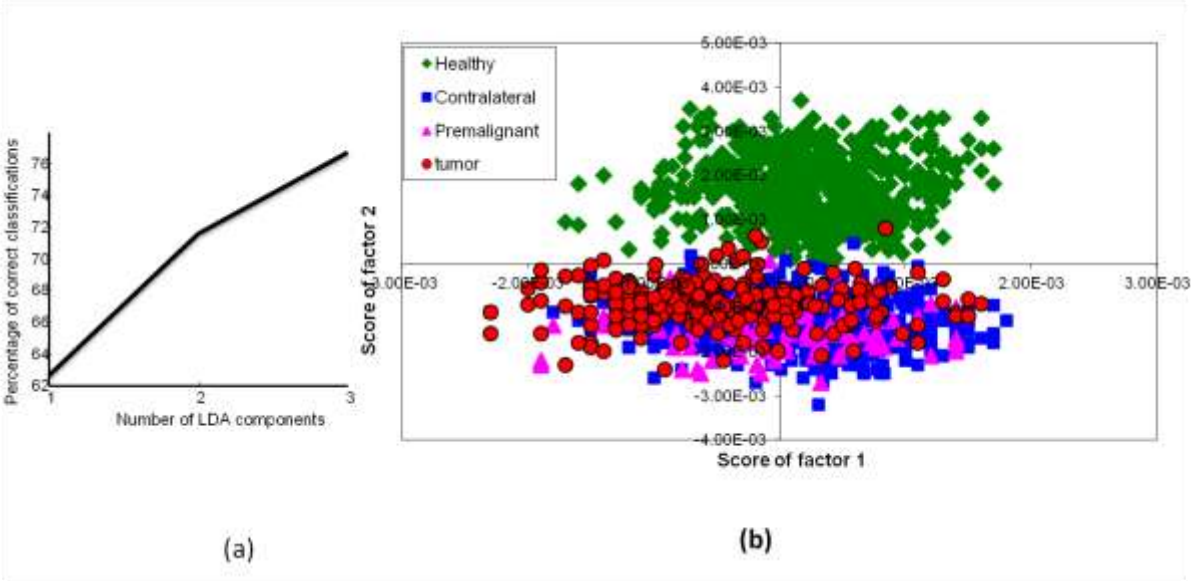


Figure 5.16. PC-LDA for exploring healthy versus pathological differences in vivo in healthy, contralateral, premalignant and tumor conditions of all subsites (buccal mucosa, lip and tongue), a) Scree plot, b) Scatter plot

LOOCV findings shown in Table 5.6 indicate that out of 666 healthy buccal, lip and tongue spectra, 652 were correctly predicted, 14 spectra misclassified with tumor. For contralateral, 352/580 spectra were correctly classified, 164 and 64 misclassified as premalignant and tumor. 32/106 premalignant were correctly predicted, 48 and 26 misclassified with contralateral normal and tumor. Out of 251 tumor spectra, 192 were correctly classified while 25 and 34 were misclassified as contralateral and premalignant, respectively.

	Healthy	Contralateral	Premalignant	Tumor	Total
Healthy	652	0	0	14	666
Contralateral	0	352	164	64	580
Premalignant	0	48	32	26	106
Tumor	0	25	34	192	251

Table 5.6. PC-LDA for exploring healthy versus pathological differences in vivo in healthy, contralateral, premalignant and tumor conditions of all subsites (buccal mucosa, lip and tongue)- confusion matrix of LOOCV results

Thus, overall classification rates were identified at 98.6%, 57%, 44% and 69% for healthy, contralateral normal, premalignant and malignant, respectively.

b) Raman spectroscopic differences between healthy and pathology at individual subsites: buccal mucosa

These spectra were subjected to PC-LDA using 5 factors. The scree and scatter plot are shown in Figure 5.17. A distinct cluster for healthy spectra and three slightly overlapping clusters were observed for contralateral, premalignant and tumor conditions.

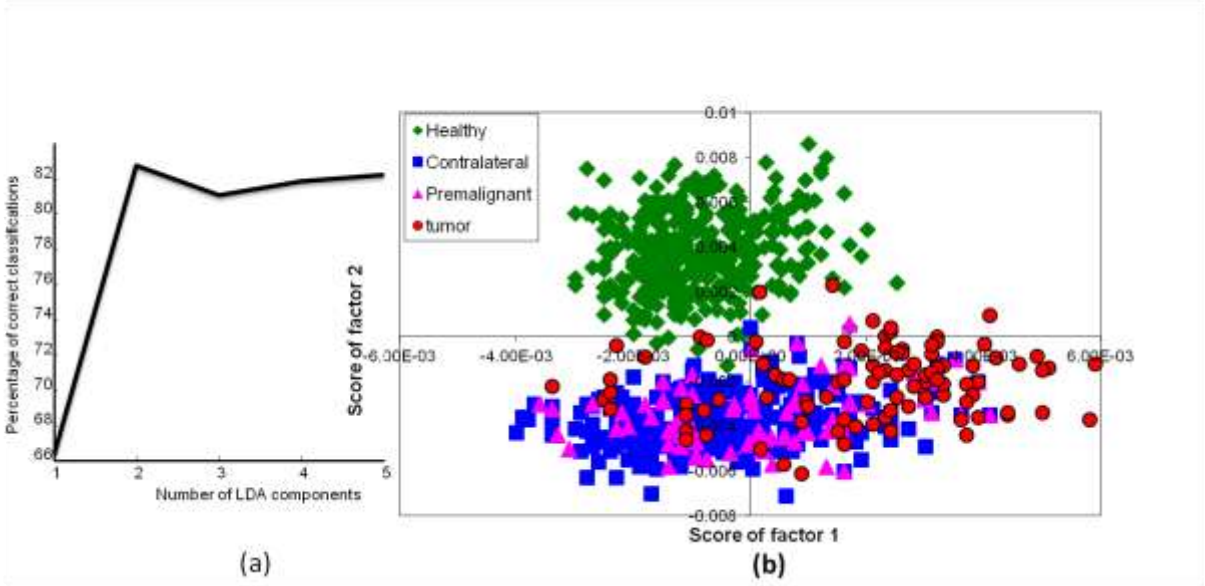


Figure 5.17. PC-LDA for exploring healthy versus pathological differences in vivo in healthy, contralateral, premalignant and tumor conditions of buccal mucosa, a) Scree plot, b) Scatter plot

LOOCV findings shown in Table 5.7 indicate that of the 353 healthy spectra, 349 were correctly classified, 162/233 spectra contralateral were correctly classified, 46/82 premalignant and 85/118 tumor were correctly predicted.

	Healthy	Contralateral	Premalignant	Tumor	Total
Healthy	349	3	0	1	353
Contralateral	0	162	64	7	233
Premalignant	0	21	46	15	82
Tumor	0	2	31	85	118

Table 5.7. PC-LDA for exploring healthy versus pathological differences in vivo in healthy, contralateral, premalignant and tumor conditions of buccal mucosa subsite- confusion matrix of LOOCV results

Less misclassification between the pathological groups- contralateral normal, premalignant and malignant was observed. Overall classification efficiency for healthy, contralateral normal, premalignant and malignant was found to be 99%, 70%, 56% and 72%.

c) Raman spectroscopic differences between healthy and pathology at individual subsites: lip

In the present study, only few lip tumor (3) and leukoplakia (2) cases could be recruited. 2 factors accounting for 79% efficiency were used for PC-LDA (Figure 5.18a). Scatter plot using scores of factors 1 and 2 is shown in Figure 5.18b. Two distinct clusters for healthy and pathology are observed, while a tendency of classification in the pathological conditions was also observed.

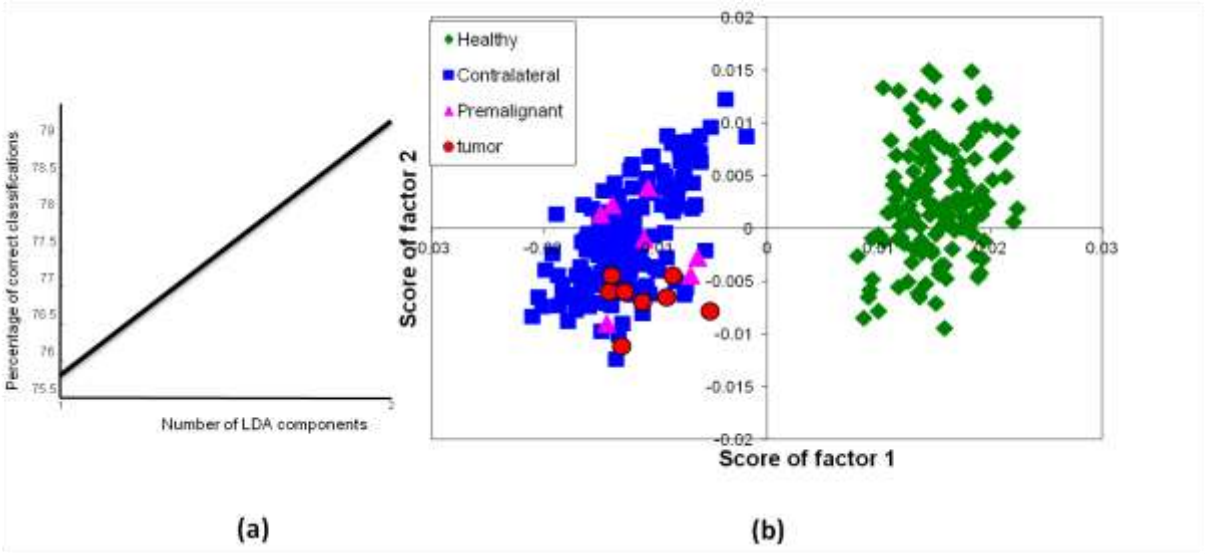


Figure 5.18. PC-LDA for exploring healthy versus pathological differences in vivo in healthy, contralateral, premalignant and tumor conditions of lip, a) Scree plot, b) Scatter plot

LOOCV findings shown in Table 5.8 indicate that 131/137 healthy, 105/163 contralateral, 2/7 premalignant and 5/11 tumor spectra were correctly predicted.

	Healthy	Contralateral	Premalignant	Tumor	Total
Healthy	131	0	0	6	137
Contralateral	0	105	52	6	163
Premalignant	0	4	2	1	7
Tumor	0	0	6	5	11

Table 5.8. PC-LDA for exploring healthy versus pathological differences in vivo in healthy, contralateral, premalignant and tumor conditions of lip subsite- confusion matrix of LOOCV results

Overall classification efficiency for healthy, contralateral normal, premalignant and malignant was found to be 96%, 64%, 29% and 45%. The findings will have to be validated after inclusion of additional lip tumor and lip leukoplakia cases.

d) Raman spectroscopic differences between healthy and pathology at individual subsites: tongue

In the next step, spectra acquired were subjected to PC-LDA using 2 factors accounting for ~75 % efficiency, as shown in Figure 5.19a. Scatter plot using scores of factors 1 and 2 indicate two separate clusters for healthy and pathology (Figure 5.19b). A minor tendency of classification was also observed in the pathological groups.

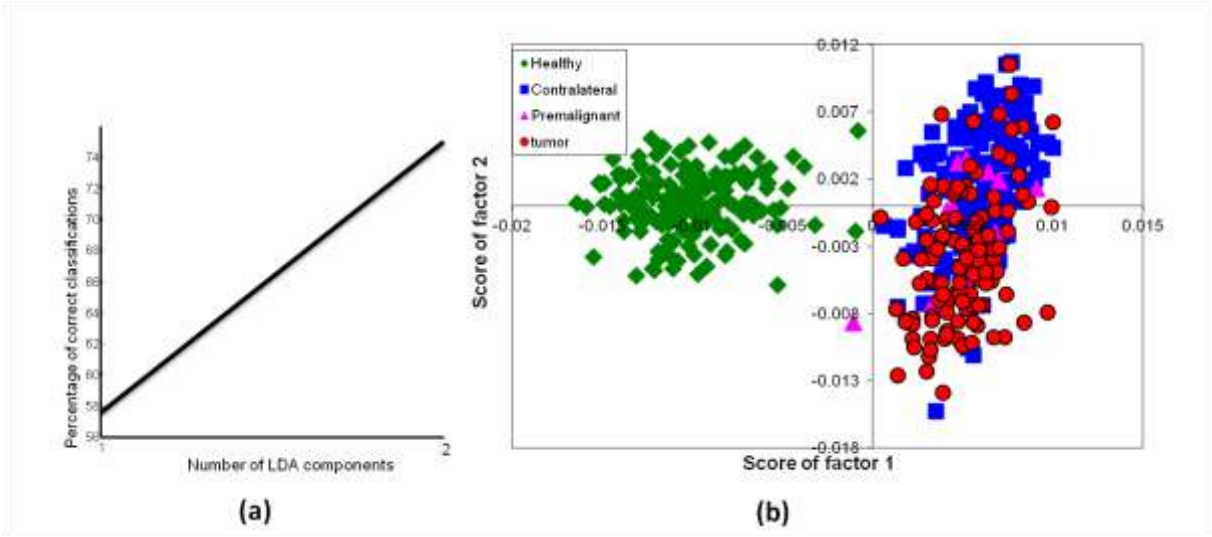


Figure 5.19. PC-LDA for exploring healthy versus pathological differences in vivo in healthy, contralateral, premalignant and tumor conditions of tongue, a) Scree plot, b) Scatter plot

LOOCV findings indicate 173/176 healthy spectra, 119/184 contralateral, 2/17 premalignant and 79/122 tumor spectra were correctly classified (Table 5.9).

	Healthy	Contralateral	Premalignant	Tumor	Total
Healthy	173	1	1	1	176
Contralateral	0	119	31	34	184
Premalignant	0	7	2	8	17
Tumor	0	29	14	79	122

Table 5.9. PC-LDA for exploring healthy versus pathological differences in vivo in healthy, contralateral, premalignant and tumor conditions of tongue subsite- confusion matrix of LOOCV results

Higher misclassification between contralateral and tumor were encountered at tongue subsite. Less number of premalignant spectra could not lead to proper conclusions for this pathological condition. Thus, overall classification rates were identified at 98%, 65%, 11% and 65% for healthy, contralateral normal, premalignant and malignant, respectively. Higher numbers of premalignant subjects will have to be recruited to evaluate the efficacy of RS in identifying premalignant lesions on tongue.

In this section, effect of this apparent anatomical variability on healthy versus pathological classification was assessed in clustered and independent sites. It was observed that PC-LDA of individual sub-sites yielded enhanced overall outcomes. Misclassifications observed between the

pathological conditions contralateral normal, premalignant and malignant could be explained due to the inherent heterogeneity of the premalignant lesions and tumor. Furthermore, for all tumor spectral acquisitions, the advancing front of the tumor was also probed along with the center of the malignancy. Several normal sites probed during the advancing front measurements cannot be ruled out[439]. Early MAC/CFE changes in the contralateral mucosa of the patients because of chronic tobacco habits may also explain the misclassification between contralateral and tumor. In case of tongue, increased misclassification between contralateral and tumor were observed. Contralateral tongue sites may not be an ideal normal for comparison with tumors. Internal infiltration of tumor to the contralateral site without obvious external clinical presentation may also be a putative reason for increased misclassification. The misclassifications of contralateral with tumor were observed mostly for advanced stages of disease, when the tumor reaches or crosses tongue midline. As premalignant lesions were probed in the same malignant subjects' oral cavity, misclassifications between the contralateral normal areas and premalignant conditions can also be understood. Further, like malignant sites, premalignant sites may also be characterized by dysplastic regions. Thus, misclassifications between these sites can also be understood.

It is noteworthy that in spite of the slightly higher classification rate observed for the individual subsites, the pooled subsites analysis also yielded overall classification rates of 98.6%, 57%, 44% and 69% for healthy, contralateral normal, premalignant and malignant, respectively. Thus, healthy could be predicted as healthy with 98% efficiency. Further as no abnormal/pathological spectra misclassified with the healthy group, abnormal group could be identified with a sensitivity of 100%. Therefore, the pooled subsites standard model differentiated between

healthy and abnormal conditions with 100% sensitivity and 98% specificity. A standard model inclusive of all subsites may be a more desirable and practical approach for preliminary screening applications. Although anatomical differences between subsites were observed in the present study, no major bearing on healthy versus pathological classification was observed. Thus, a single standard model to differentiate between healthy and abnormal conditions can be employed as a useful adjunct during oral cancer screening. The abnormality of a given sample can be further scrutinized by routine clinical methods like biopsy followed by histopathology.

5.3. Exploring age-related changes in healthy subjects

Like many other cancers, risk of developing oral cancers increases with age. Possible reasons include increased oxidative stress and reduced anti-oxidant levels in the oral cavity and persistent low-grade inflammation with increasing age[440-445]. Age-associated changes in oral mucosa and sub-mucosa such as smoothening, drying and thinning of epithelium, hypocellularity and collagen maturation in lamina propria and prominent vascular changes are reported in the literature. Other changes include gradual epithelial atrophy, salivary gland hypo-function, decreased keratinization, loss of elasticity and change in surface characteristics. Other changes of the oral epithelium such as decrease of the prominence of rete ridges, decreased cell density, decreased mitotic activity and cell proliferation and a slowdown in tissue regeneration and healing rates have also been observed[443,445-454]. To investigate pathogenesis in detail, normal changes within physiological limits and the transformation of physiology into pathology should be demarcated. Adaptive alterations of oral mucosa induced by external stimuli decide the biochemical and morphological architecture of oral cavity. Oxidative-antioxidative elements, lifestyle-related practices, level of micro and macronutrients are some of the contributing factors for numerous oral pathologies including premalignant and malignant changes. Clinical inspection and palpation have been the gold standard techniques in the diagnosis of superficial and deeply located oral mucosal pathologies; however the amount of mucosal wear and insults caused by aging-related processes cannot be monitored by clinical inspection or bimanual palpation of mucosal surfaces. The clinical experts and scientist are keen on acquiring information about the clinically significant, invisible mucosal changes in the quantifiable or readable forms. Owing to sensitivity to biochemical changes, RS can be used to study changes in

biophysical parameters which can predict complex biochemical changes within tissues. In this section, potential of RS to record clinically invisible changes induced by age-related physiological changes and differentiating these from pathological changes was explored. Although efficacy of RS-based methodologies in oral cancer management have been demonstrated in above sections, it is necessary to evaluate the same under complex conditions. Physiological variations (in age groups ranging from 20 yrs to 60 yrs) along with pathosis like contralateral mucosa, premalignant lesions and oral squamous cell carcinomas were analyzed using Raman spectroscopy. In the first step, aging-related physiological changes were explored at buccal mucosa, lip and tongue. In the next step, effect of these physiological changes on normal versus pathological classification was evaluated.

5.3.1. Materials and Methods:

5.3.1.1. Sample Details

The same cohort of study patients analyzed in the first and second sections- Healthy volunteers (n=72) and oral cancer patients (n=85) were included in this study. Subject accrual is elaborated in section 5.1.1.1. In the healthy volunteers, subjects were divided into 4 age groups- 21-30 years, 31-40 years, 41-50 years and 51-60 years.

5.3.1.2. Raman spectroscopy

Spectra were acquired from healthy volunteers at subsites buccal mucosa, lip and tongue as described in section 5.1.1.2. For buccal mucosa, 347 spectra were acquired from mucosa opposing teeth positions of second pre-molar, first molar and second molar. These spectra were

divided into 4 age groups- 21-30 years, 31-40 years, 41-50 years and 51-60 years. For lip, 137 spectra were acquired from labial mucosa opposing the position of the two central incisors teeth. For tongue, 172 spectra were acquired from the right and left lateral borders of the tongue, as they are the most commonly affected sites for cancer development. On an average, 2-3 spectra were recorded from each subject.

Spectra acquired from contralateral, premalignant and tumor of oral cancer patients (n=85) were also included in the study. For buccal mucosa, 233 contralateral, 82 premalignant and 118 tumor spectra were used. In case of lip, 163 contralateral, 7 premalignant and 11 tumor spectra were employed. For tongue, 184 contralateral, 17 premalignant and 120 tumor spectra were used.

Spectra were acquired using a fiber-optic probe coupled HE-785 Raman instrument as described in section 5.1.1.2.

5.3.1.3. Spectral preprocessing and data analysis

Spectra were corrected for CCD response, background signals, first derivatized, interpolated in 1200-1800 cm^{-1} range, normalized and subjected to PC-LDA followed by LOOCV as described in section 5.1.1.3.

5.3.2. Results and discussion

5.3.2.1. Spectral features

Spectral features in the age groups were investigated at individual subsites-buccal mucosa, lip and tongue. Average spectra for each age group recorded at buccal mucosa shown in Figure 5.20 show lipid rich features, indicated by C=O band of esters, strong δCH_2 bend, two sharp features

in amide III region and sharp amide I. Subtle differences in the amide I, amide III, δCH_2 bend and C=O ester band regions were observed in the different age group, suggesting a tendency of decrease in the overall protein content and increase in lipid content with increasing age groups.

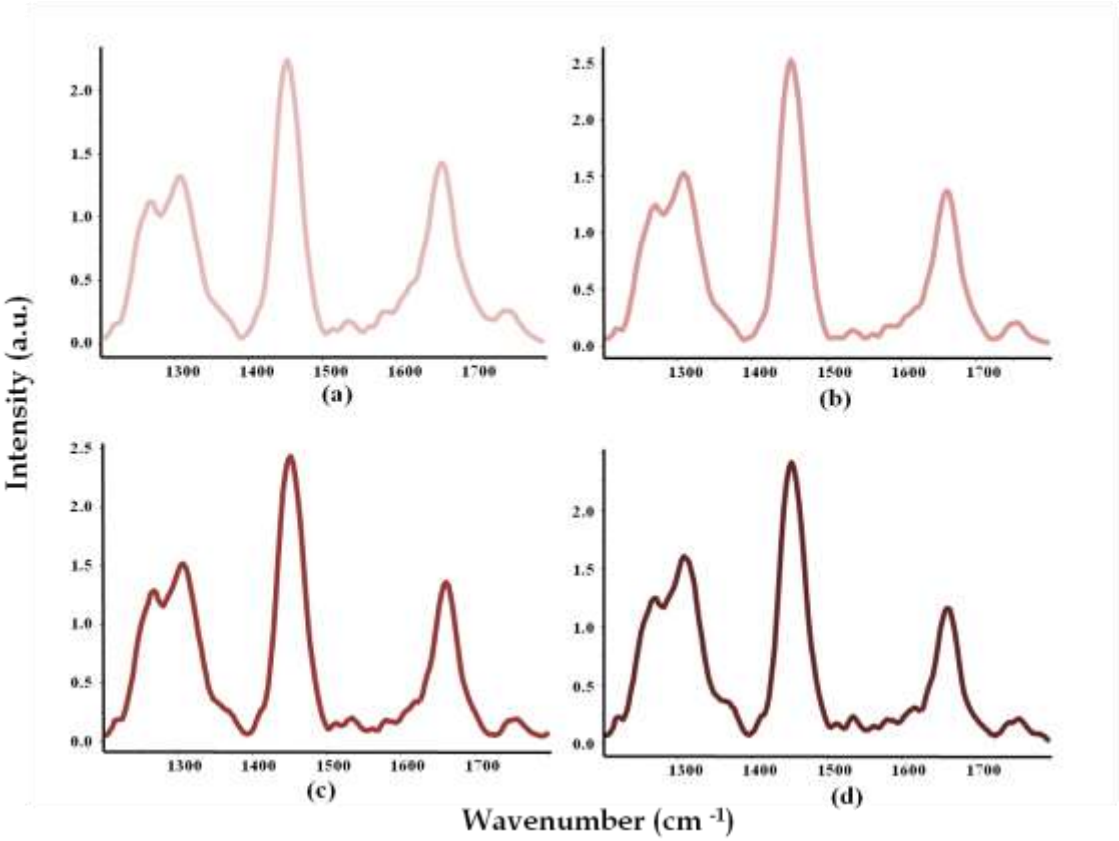


Figure 5.20. Mean in vivo Raman spectra of buccal mucosa of healthy volunteers belonging to age groups (a) 21-30 years; (b) 31-40 years; (c) 41-50 years; (d) 51-60 years, for spectral comparisons

Average spectra for lip shown in Figure 5.21 show two sharp features in the amide III region, strong δCH_2 bend and sharp peak in the amide I region and an ester band at around 1742 cm^{-1} along with broader amide III and amide I regions, indicating both lipid and protein contributions

in the spectra. Similar age-related trends like buccal mucosa- a decrease in overall protein and slight increase in lipid content were observed with increasing age groups.

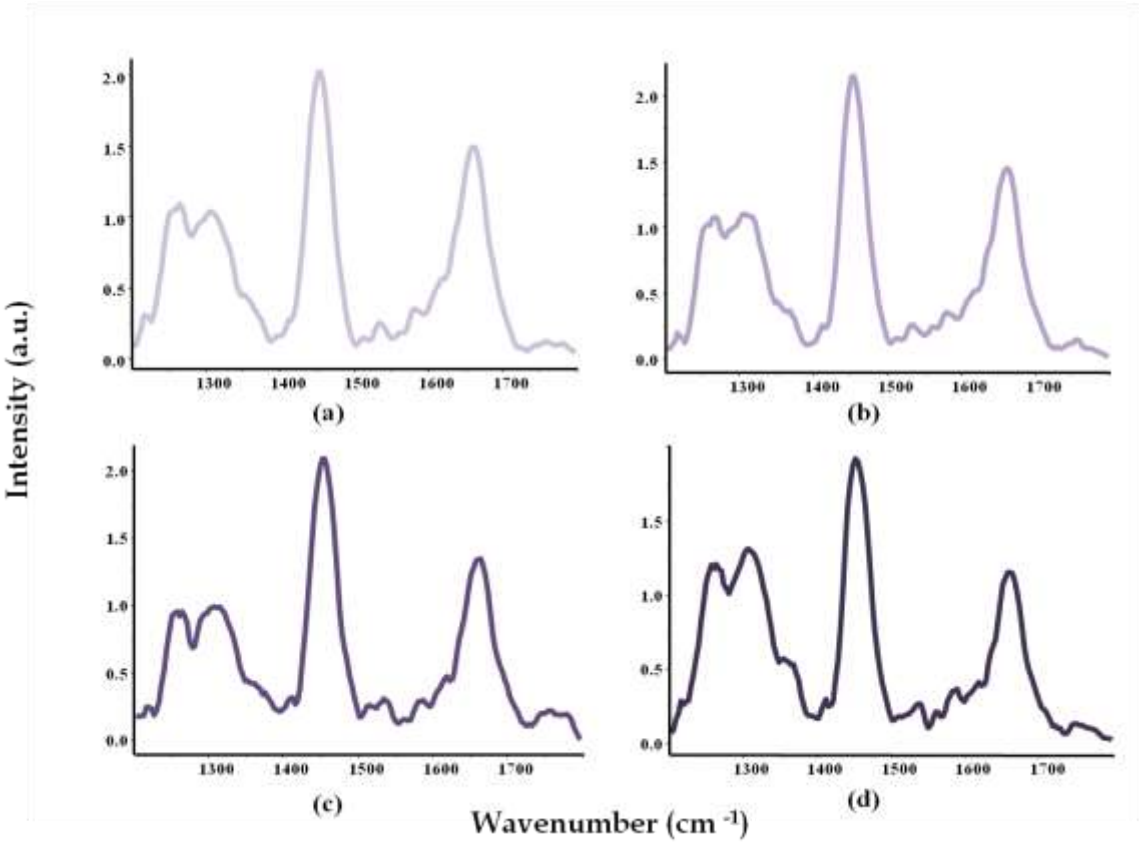


Figure 5.21. Mean in vivo Raman spectra of lip of healthy volunteers belonging to age groups (a) 21-30 years; (b) 31-40 years; (c) 41-50 years; (d) 51-60 years, for spectral comparisons

Average spectra for tongue are shown in Figure 5.22. Features like broad amide III, amide I along with CH_2 deformation and CH_2, CH_3 wagging (collagen assignment) at 1343 cm^{-1} were prominently observed in all groups. Features indicate higher protein-related features in the late age group (51-60 years), as compared to the early age groups.

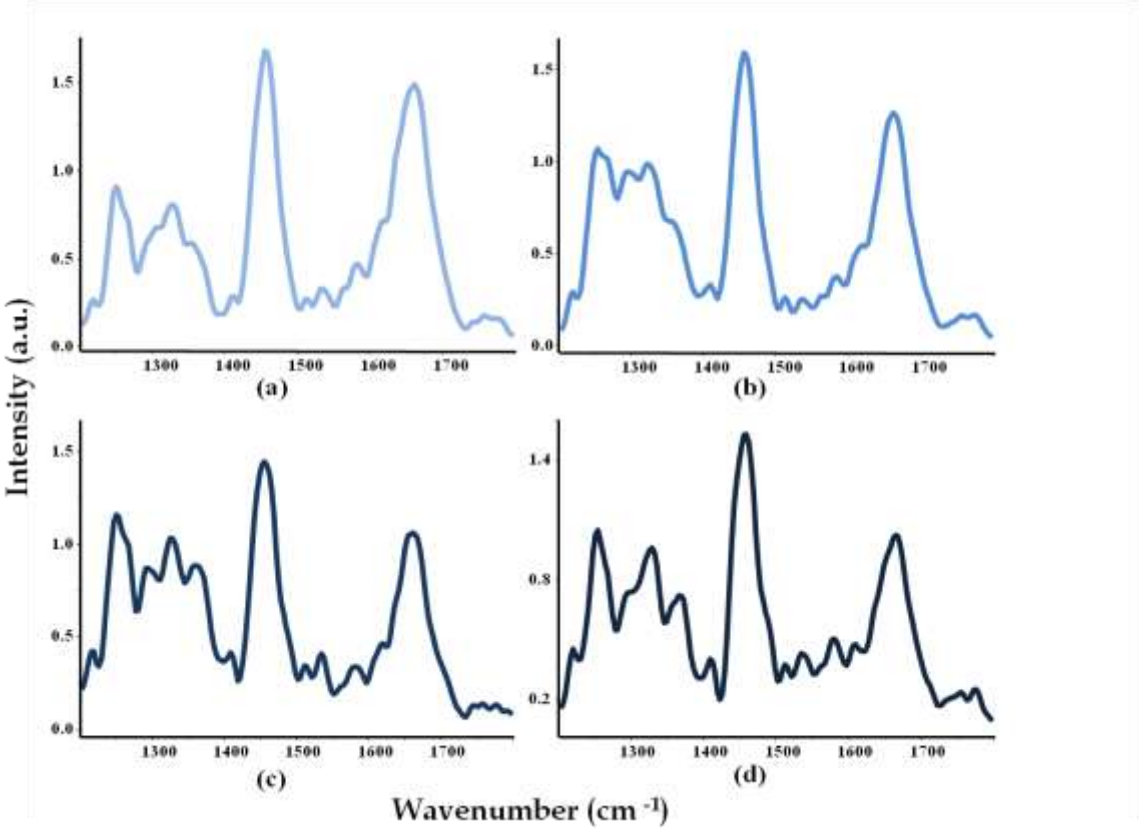


Figure 5.22. Mean in vivo Raman spectra of tongue of healthy volunteers belonging to age groups (a) 21-30 years; (b) 31-40 years; (c) 41-50 years; (d) 51-60 years, for spectral comparisons

Spectral differences between healthy age groups and pathological conditions were also explored. While the late age groups are associated with decreased protein and higher lipid at all subsites, pathological conditions are associated with increased protein, heme and DNA-related features.

5.3.2.2. Multivariate analysis

Multivariate analysis was carried out using PC-LDA followed by LOOCV. First, feasibility of detection of age-related physiological changes was explored in individual subsites. Subsequently, effect of aging on healthy versus pathological classification was explored at individual subsites

5.3.2.2.1.Detecting physiological changes associated with aging of oral mucosa

a) Buccal mucosa

Age-related changes like thinning of epithelium, atrophy of salivary glands, decreased keratinization and cell proliferation are reported for buccal mucosa[438,445,449,453]. These changes were explored by PC-LDA using 8 factors accounting for 54% correct classifications (Figure 5.23a). Scores of factors 1 and 2 were explored for classification. Four consequently overlapping clusters were observed for the four different groups in the PC-LDA scatter plot (Figure 5.23b).

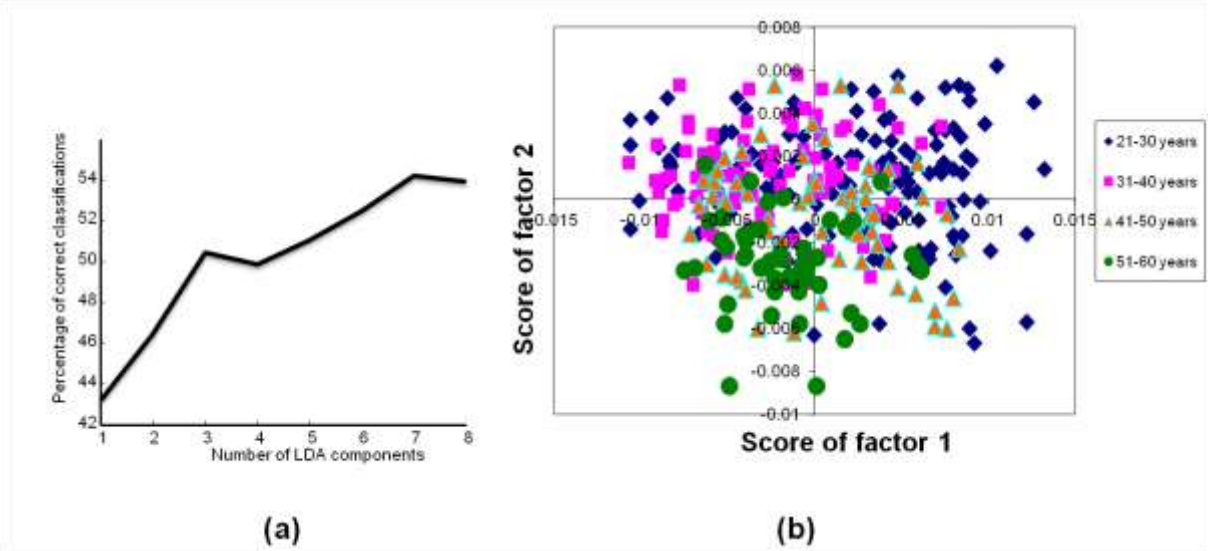


Figure 5.23. PC-LDA for exploring age wise differences in healthy volunteers at buccal mucosa in the age groups- 21-30 years, 31-40 years, 41-50 years, 51-60 years, a) Scree plot; b) Scatter plot

LOOCV findings (Table 5.10) demonstrate that 79/139 spectra of 21-30 group were correctly classified, while 28, 18, 14 spectra misclassified with 31-40, 41-50 and 51-60 groups, respectively. For 31-40 age group, 50 were correctly classified while 14, 17 and 16 misclassified

with other age groups. In case of 41-50 group, 20/65 spectra were correctly classified while 15, 14 and 16 misclassified with other groups. For the last age group, 28/46 were correctly classified, while 9, 6, 3 spectra misclassified with consecutively decreasing age groups.

	21-30 years	31-40 years	41-50 years	51-60 years	Total
21-30 years	79	28	18	14	139
31-40 years	14	50	17	16	97
41-50 years	15	14	20	16	65
51-60 years	3	6	9	28	46

Table 5.10. PC-LDA for exploring age wise differences in healthy volunteers at buccal mucosa in the age groups- 21-30 years, 31-40 years, 41-50 years, 51-60 years- confusion matrix of LOOCV results

Thus, early and late groups are mostly distinct, with high overlap between the median and late median groups and misclassifications mostly in the contiguous age groups i.e. early with median, median with late median. This could be attributed to the fact that the chronological and biological age of tissues of human body may vary. The chronological age is considered from birth onwards while biological age represents wear and tear of cellular components caused by internal and external insulting stimuli. Accumulations of multiple such injuries at cellular level lead to enhanced biological aging of tissue and this varies from individual to individual depending on the exposure. These trends of classification and misclassification could also be attributed to slow process of aging.

b) Lip

Lip or the overlying labial mucosa is susceptible to ageing effects such as atrophy of minor salivary glands and changes in distensibility and moisture content of the lower labial mucosa with age[455,456]. To evaluate these changes, PC-LDA was carried out using 3 factors accounting for ~47% correct classifications; the scree plot is shown in Figure 5.24a. PC-LDA scatter plot was plotted using scores of factors 1 and 3. As shown in Figure 5.24b, a tendency of classification was observed between early (21-30) and late (51-60 age groups) while major overlap between the middle age groups was observed.

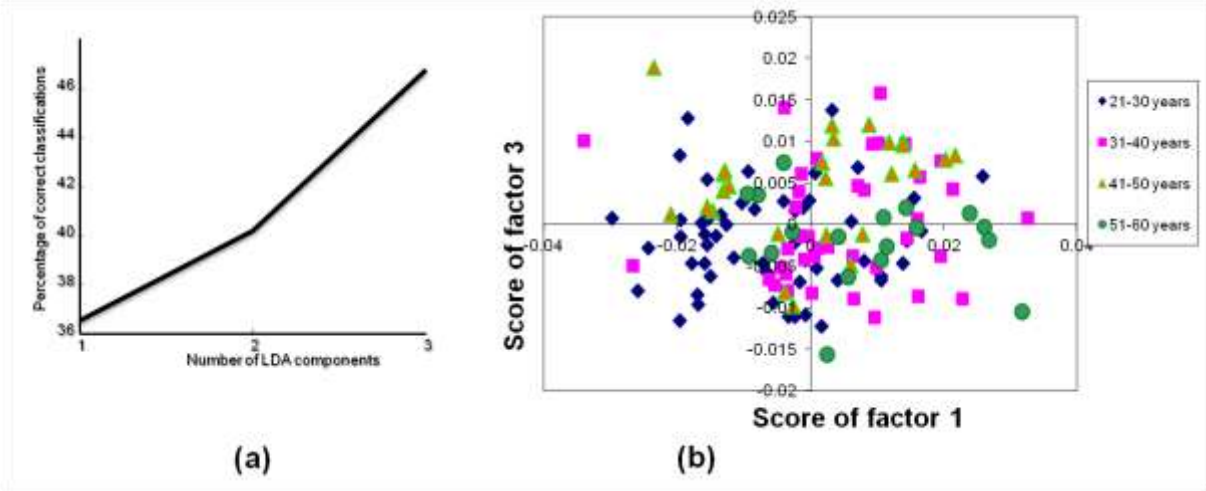


Figure 5.24. PC-LDA for exploring age wise differences in healthy volunteers at lip in the age groups- 21-30 years, 31-40 years, 41-50 years, 51-60 years, a) Scree plot, b) Scatter plot

LOOCV findings indicate 31/54 correct classifications for 21-30 age group. While only 2/39 correct classifications were observed for 31-40 group, 14/26 correct classifications were observed for 41-50 age group. For 51-60 age group, 8/18 spectra were correctly classified. Thus, the LOOCV findings indicate no definite trends in classification (Table 5.11).

	21-30 years	31-40 years	41-50 years	51-60 years	Total
21-30 years	31	1	11	11	54
31-40 years	11	2	14	12	39
41-50 years	8	2	14	2	26
51-60 years	4	2	4	8	18

Table 5.11. PC-LDA for exploring age wise differences in healthy volunteers at lip in the age groups- 21-30 years, 31-40 years, 41-50 years, 51-60 years- confusion matrix of LOOCV results

Although changes in oral mucosa (that includes labial mucosa) with increasing age are reported in literature, no clear classification trends were observed in this study. Like in buccal mucosa, the classifications and misclassifications may be explained due to differences in the chronological and biological age, as well as the slow transitional nature of aging.

c) Tongue

Several age-associated physiological changes in tongue have been reported. From changes in muscular tension, increase in venous blood vessels, atrophy of filiform and fungiform papillae, to epithelial atrophy and overall simplification of structure have already been documented[446,457-459]. To evaluate spectroscopic detection of these changes, PC-LDA was carried out using 4 factors accounting for ~46% variance (Figure 5.25a). Scores of factor 1 and 2 were explored for classification between the groups. As shown in Figure 5.25b, no definite trends of classification were observed.

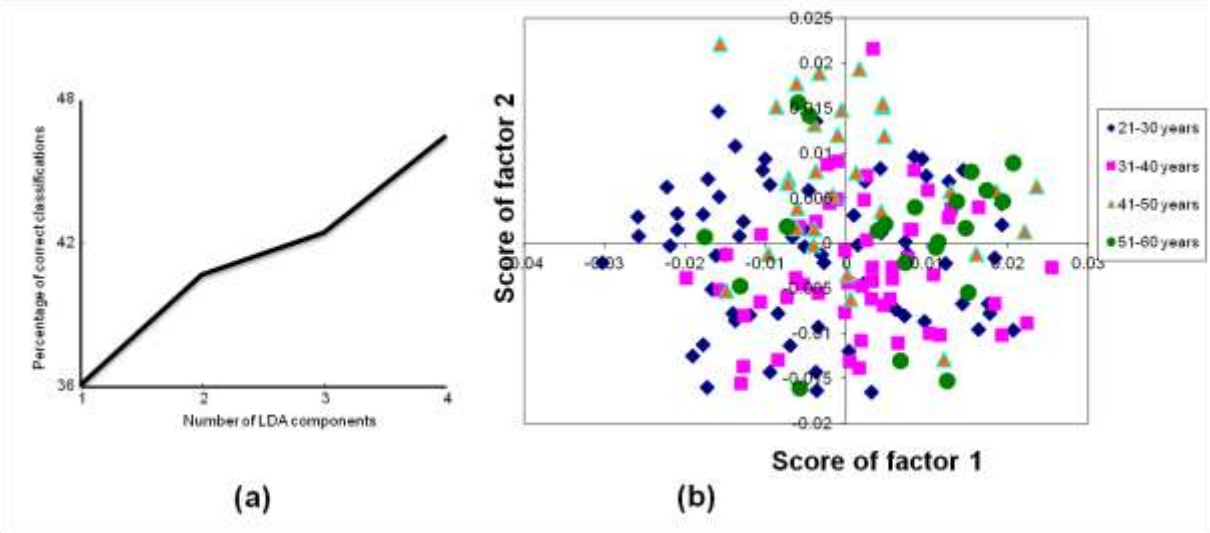


Figure 5.25. PC-LDA for exploring age wise differences in healthy volunteers at tongue in the age groups- 21-30 years, 31-40 years, 41-50 years, 51-60 years, a) Scree plot, b) Scatter plot

Upon LOOCV, 25/66 spectra from 21-30 age group were correctly classified. For 31-40 group, 22/54 were correctly classified while 11/31 spectra showed correct classification for 41-50 age group. In case of 51-60, 12/21 (>50%) were correctly classified (Table 5.12).

	21-30 years	31-40 years	41-50 years	51-60 years	Total
21-30 years	25	15	14	12	66
31-40 years	13	22	6	13	54
41-50 years	8	6	11	6	31
51-60 years	4	2	3	12	21

Table 5.12. PC-LDA for exploring age wise differences in healthy volunteers at tongue in the age groups- 21-30 years, 31-40 years, 41-50 years, 51-60 years- confusion matrix of LOOCV results

As in the PC-LDA scatter plot, no definite classification trends were observed. This could be attributed to the site of spectral probing, as major age-related changes are reported with respect to decrease in size and number of papillae present on the dorsal surface of the tongue. In this study, spectra were acquired from the lateral surface of the tongue. Thus, no age-related changes could be detected.

5.3.2.2.2. Effect of physiological aging on classifier models

The feasibility of Raman spectroscopic detection of aging-associated changes between the early and late age groups was observed. In the next step, the potential of Raman spectroscopy in distinctly identifying age-related physiological and cancer-associated pathological changes was explored at each of the three subsites- buccal mucosa, lip and tongue.

a) Buccal mucosa

The feasibility of differentiating age-related physiological changes and cancer-related pathological changes were first explored at buccal mucosa. Three-hundred forty seven healthy spectra belonging to different age groups (139 to 21-30 group, 97 to 31-40 group, 65 to 41-50 group and 46 to 51-60 age groups) and 433 pathological spectra (233-contralateral, 82-premalignant and 118-tumor spectra) were subjected to multivariate analysis. PC-LDA was carried out using 7 factors accounting for ~60% correct classifications. Scores of factor 1 and 2 were employed for classifications. As shown in the PC-LDA scatter plot shown in Figure 5.26, two distinct clusters for healthy and pathological conditions were observed. Although no trends in age-related classification were observed, feasibility of classification was seen for contralateral, premalignant and tumor conditions.

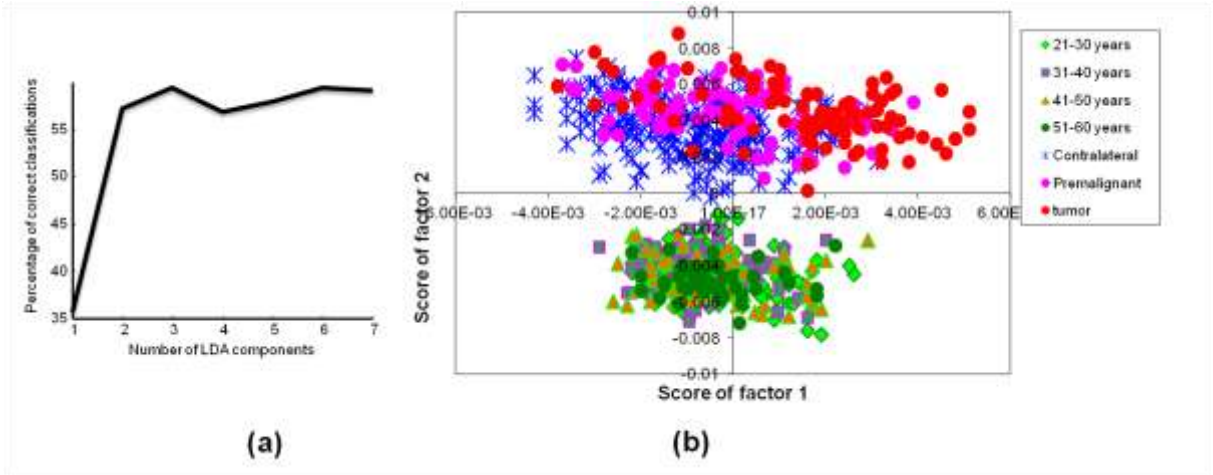


Figure 5.26. PC-LDA for exploring effect of age-wise differences on healthy (21-30 years, 31-40 years, 41-50 years and 51-60 years) versus pathological conditions (contralateral mucosa, premalignant and tumor) at buccal mucosa, a) Scree plot, b) Scatter plot

LOOCV findings demonstrate distinct classification between healthy and pathological groups, as shown in Table 5.13. While major overlap was observed between healthy age-groups as in previous section, only 1 misclassification of healthy was observed with contralateral group. Similarly, pathological groups showed expected classifications with several misclassifications amongst themselves. Only 1 contralateral and 1 tumor spectra misclassified with the late age group 51-60 years.

	21-30 years	31-40 years	41-50 years	51-60 years	Contralateral	Premalignant	Tumor	Total
21-30 years	77	26	20	15	1	0	0	139
31-40 years	20	44	12	21	0	0	0	97
41-50 years	28	13	8	16	0	0	0	65
51-60 years	7	6	5	28	0	0	0	46
Contralateral	0	0	0	1	161	61	10	233
Premalignant	0	0	0	0	18	46	18	82
Tumor	0	0	0	1	3	26	88	118

Table 5.13. PC-LDA for exploring effect of age-wise differences on healthy (21-30 years, 31-40 years, 41-50 years and 51-60 years) versus pathological conditions (contralateral mucosa, premalignant and tumor) at buccal mucosa- confusion matrix of LOOCV results

The healthy group (different physiological age groups) is distinct from the pathological group. Thus, high sensitivity of Raman spectroscopy in differentiating normal and abnormal conditions is observed. More importantly, the aging related changes do not seem to have any bearing on classification of normal from abnormal conditions.

b) Lip

Detection of physiological and pathological changes was evaluated next on lip subsite. A total of 137 spectra were analyzed along with 181 pathological spectra (163-contralateral, 7-premalignant and 11-tumor spectra). PC-LDA was carried out using 2 factors accounting for ~54% correct classifications. Scatter plot was plotted using scores of factor 1 and 2. As shown in Figure 5.27, two distinct groups corresponding to healthy and pathology were obtained. Although no age-related trends could be observed in the scatter plot, a gradient classification was observed for pathological changes.

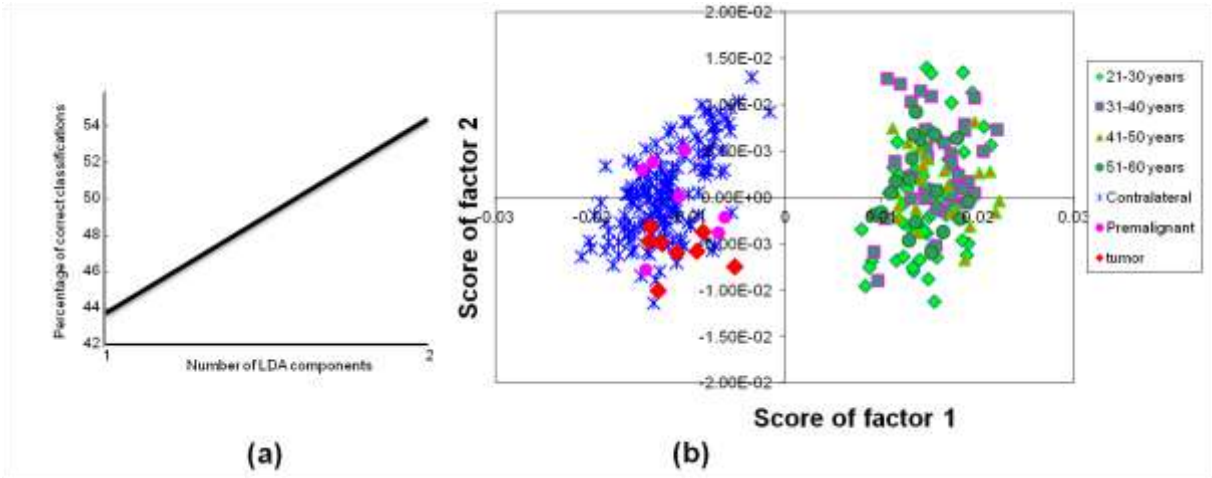


Figure 5.27. PC-LDA for exploring effect of age-wise differences on healthy (21-30 years, 31-40 years, 41-50 years and 51-60 years) versus pathological conditions (contralateral mucosa, premalignant and tumor) at lip, a) Scree plot, b) Scatter plot

Upon LOOCV, most healthy groups classified as healthy while 3 spectra from 21-30 and 1 spectra from 31-40 groups misclassified as tumor, respectively (Table 5.14).

	21-30 years	31-40 years	41-50 years	51-60 years	Contralateral	Premalignant	Tumor	Total
21-30 years	28	17	3	3	0	0	3	54
31-40 years	9	18	6	5	0	0	1	39
41-50 years	12	9	3	2	0	0	0	26
51-60 years	8	6	1	3	0	0	0	18
Contralateral	0	0	0	0	105	52	6	163
Premalignant	0	0	0	0	4	2	1	7
Tumor	0	0	0	0	0	6	5	11

Table 5.14. PC-LDA for exploring effect of age-wise differences on healthy (21-30 years, 31-40 years, 41-50 years and 51-60 years) versus pathological conditions (contralateral mucosa, premalignant and tumor) at lip- confusion matrix of LOOCV results

In case of pathological conditions, no spectra misclassified with the healthy groups. Thus, RS distinctly classifies healthy and pathological groups.

c) Tongue

In the final step, effect of aging on classification with pathological changes was confirmed on tongue subsite. One hundred seventy-two healthy spectra and 321 pathological spectra (172-contralateral, 17-premalignant and 120-tumor spectra) were subjected to PC-LDA. Classification was explored using 7 factors accounting for 60% variance. Scores of factor 1 and 2 were employed for classification. Scatter plot shown in Figure 5.28 demonstrates different clusters for healthy and pathological groups, while major overlap within both healthy and pathological conditions.

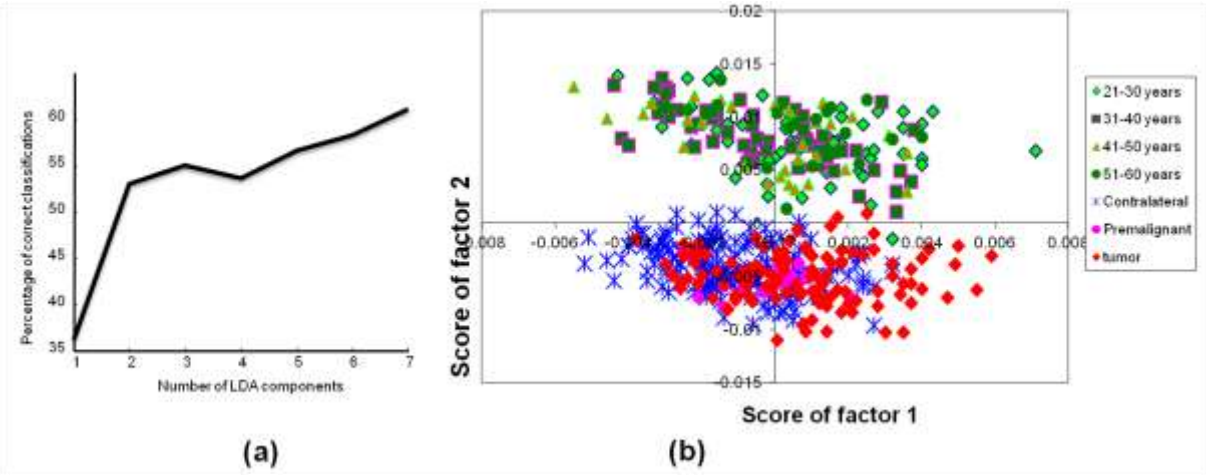


Figure 5.28. PC-LDA for exploring effect of age-wise differences on healthy (21-30 years, 31-40 years, 41-50 years and 51-60 years) versus pathological conditions (contralateral mucosa, premalignant and tumor) at tongue, a) Scree plot, b) Scatter plot

LOOCV findings indicate correct prediction as healthy for most spectra from healthy groups. Only 1 spectra each from 21-30, 31-40 and 51-60 misclassified as premalignant while 1 spectra from 41-50 misclassified as tumor. For the pathological groups, most spectra were correctly classified as pathology; however 2 spectra from tumor each misclassified as 21-30 and 51-60 groups (Table 5.15). Thus, age-related changes do not influence healthy and pathological classification even at tongue subsite.

	21-30 years	31-40 years	41-50 years	51-60 years	Contralateral	Premalignant	Tumor	Total
21-30 years	27	14	14	10	0	1	0	66
31-40 years	12	24	10	7	0	1	0	54
41-50 years	8	4	10	8	0	0	1	31
51-60 years	6	3	6	5	0	1	0	21
Contralateral	0	0	0	0	124	32	28	184
Premalignant	0	0	0	0	2	8	7	17
Tumor	2	0	0	2	20	24	74	120

Table 5.15. PC-LDA for exploring effect of age-wise differences on healthy (21-30 years, 31-40 years, 41-50 years and 51-60 years) versus pathological conditions (contralateral mucosa, premalignant and tumor) at tongue- confusion matrix of LOOCV results

5.4. Summary

Oral cancers suffer from poor disease-free survival rates due to delayed diagnosis. Non-invasive, rapid, objective approaches as adjuncts to visual inspection can help in better management of oral cancers. Optical spectroscopies like OCT, fluorescence and Raman spectroscopy have shown potential in identification of oral premalignant and malignant conditions. RS has also shown potential in detection of early cancer changes like cancer-field-effects at buccal mucosa sub-site. Anatomic differences between oral sub-sites have been observed using fluorescence, reflectance and RS studies in healthy subjects.

In this study, anatomical differences between buccal mucosa, lip and tongue were investigated in healthy, contralateral, premalignant and cancer conditions. The possible influence of anatomical differences on healthy versus pathological classification was also explored. Aging-related physiological changes were also explored.

- BM and tongue classified as distinct subsites in healthy and contralateral conditions while lip misclassified with both subsites. Decrease in subsite classification with increasing pathology was observed, with total loss of subsite influence in tumor sites.

- Both, pooled subsites model and individual subsite diagnostic models were investigated for better healthy versus pathological classification. Although diagnostic algorithm for individual subsites gave slightly higher classification efficiencies, the 'pooled subsites' standard model could satisfactorily differentiate between healthy and pathological conditions. Such a model may serve as a more useful and practical preliminary screening tool.

- Aging-associated changes were also evaluated at these three subsites. Although minor physiological changes were detectable especially at buccal mucosa, these changes did not influence healthy versus pathological classification at any of the three subsites.

Chapter 6

Conclusions and Future perspectives

Oral cancers suffer from a low 5-year survival rate, even lower than occult cancers like breast and cervix cancers. Better diagnostic approaches may improve overall outcomes of oral cancers. In this study, non-invasive and less-invasive approaches for early diagnosis and screening of oral cancers were studied. In this chapter, the conclusions to the study are presented in the first section while future directions are presented in the second section.

6.1. Conclusions

The major findings and their conclusions are summarized below.

6.1.1. Serum for oral cancer diagnosis

Screening of cancers using peripheral blood samples may be a practical and feasible approach. This approach is associated with multiple advantages like accessibility, low invasive procedure, low cost, and multiple sampling[363]. Further, samples can be transported from primary screening centers to a centralized facility for analysis. Therefore, oral cancer detection using serum Raman spectroscopy was explored. Two pilot studies using resonance and conventional Raman indicated the feasibility of differentiating normal and oral cancer serum samples with ~78% efficiency. In the subsequent study, premalignant (n=47) and disease control (n=35) subjects and larger cohorts of normal (n=126) and oral cancer (n=120) subjects were included. Spectral features indicated differences in the protein, amino acids, β-carotene, lipids and DNA between the groups. Premalignant and disease control subjects could be differentiated from normal with reasonable accuracy. The analysis on a large sample size and independent test data evaluation demonstrates discrimination between normal and oral cancer groups with an average efficiency of ~86%.

Screening-related applications were explored in the next step. The normal versus abnormal screening model yielded sensitivity of 64% and specificity of ~80%. Although the sensitivity and specificity rates are not very high, they are comparable to established screening techniques like fecal-occult-blood test (FOBT), Pap smear and mammography[31-33]. As the current screening approaches (visual inspection followed by biopsy and histopathology) are useful mainly for high-risk populations, oral cancer screening using serum RS may serve as a useful adjunct and help in improving low disease-free survival rates associated with the oral cancers.

Recurrence is a major problem in oral cancer management; low disease free-survival rates are attributed to recurrence and delays in diagnosis. The current exploratory study investigated the feasibility of identifying recurrence-prone patients using before and after surgery serum samples. Findings indicate possibility of identifying recurrence using after serum surgery samples. Higher DNA and protein features were observed in recurrence spectra. Although RS cannot predict localization of recurrent tumor, it can serve as a preliminary test. On the basis of these results, regular imaging modalities like magnetic resonance imaging (MRI) followed by more comprehensive adjuvant treatment decisions and stringent follow-ups may improve overall outcome of the disease.

Serum Raman spectroscopy could identify differences between buccal mucosa and tongue serum with ~70% efficiency using both resonance Raman (532 nm) and conventional Raman (employing 785 nm excitation) studies.

6.1.2. Exfoliated cells

For oral cancers, visual inspection is the currently available method for screening. This method may be useful only in screening of mainly high-risk cases. Exfoliative cytology is a more practical tool for mass screening and also monitoring of high-risk populations.

The feasibility of differentiating healthy and tumor cells was first investigated in the proof-of-concept approach. RBC lysis treatment to remove influence of blood was carried out for all samples, and spectral differences were explored between healthy subjects (with and without tobacco habits) and contralateral normal and tumor sites in oral cancer patients. Increase in DNA content and changes in protein secondary structure were observed with increasing pathology. Both spectra-wise and patient-wise approaches were employed for data analysis. Findings on 70 specimens demonstrate potential in discriminating healthy and cancer cells with high efficiency. A major overlap (~25%) between contralateral and tumor sites in oral cancer patients could be attributed to detection of malignancy associated changes in the clinically normal contralateral mucosa.

In the subsequent study, differences between cells from healthy subjects (n=40, with and without tobacco habits) and subjects with pre-cancer lesion/condition (n=27) were explored to enable early diagnosis. Spectral differences were mainly observed for DNA and protein-related features. Results indicate feasibility of classification between healthy without tobacco and pre-cancer groups, while healthy tobacco group misclassified with both groups. The major misclassifications between pre-cancer and healthy tobacco group and vice versa could be attributed to increased keratinization of the mucosa due to chronic tobacco exposure. Further,

obtaining good cellular yields from keratotic lesions like leukoplakia may be difficult. Improvements in sampling strategies may enhance overall outcomes.

6.1.3. In vivo

Oral cancers suffer from poor disease-free survival rates due to delayed diagnosis. Current diagnosis includes visual inspection followed by biopsy and histopathology of suspicious lesions. This procedure suffers from several disadvantages. Non-invasive, rapid, objective approaches as adjuncts to visual inspection can help in better management of oral cancers. In the present study, anatomical differences between buccal mucosa, movable mucosa of the lip and tongue-the most commonly affected subsites in Indian subcontinent were investigated in healthy, contralateral, premalignant and cancer conditions; effect of anatomical variability on oral cancer diagnostic algorithms and detection of physiological changes due to aging at these three subsites were explored on n=72 healthy volunteers and n=85 oral cancer patients. From patients, spectra were acquired from contralateral normal, premalignant and tumor sites. Spectral analysis indicated higher lipid content in most healthy conditions while higher DNA, heme and protein-related features in severe pathological conditions.

In the first analysis for investigating spectral contrasts at the 3 anatomical sites, it was observed that RS- based discrimination between buccal mucosa, lip and tongue reduced sequentially from healthy to the most severe pathological condition (tumor), with total loss of anatomical influence in tumor conditions. For healthy and contralateral conditions, buccal mucosa and tongue were classified as distinct entities while lip misclassified with both subsites. The similarity of lip subsite with both buccal mucosa and tongue can be explained due to the fundamental anatomical

characteristics of this subsite. In the next step, effect of this anatomical variability on healthy versus pathological classification was assessed in clustered and independent site-wise diagnostic algorithms. Although diagnostic algorithm for individual subsites gave slightly higher classification efficiencies, the 'pooled subsites' standard model could satisfactorily differentiate between healthy and pathological conditions. Such a model may serve as a more useful and practical preliminary screening tool. In the aging studies, although minor physiological changes were detectable especially at buccal mucosa, these changes did not influence healthy versus pathological classification at any subsite.

6.2. Future perspectives

The ultimate goal of these studies is to provide an objective, real-time screening and diagnostic adjunct for oral cancer diagnosis. Reasonable classification efficiencies for classifying healthy and cancer using serum, exfoliated cells and in vivo approach were obtained in the present thesis. Following appropriate modifications and large scale validation studies, Raman spectroscopy can prospectively help in oral cancer screening and diagnosis.

6.2.1. Serum

Serum Raman spectroscopy-based screening of oral cancers can serve as a useful adjunct to conventional screening using the normal versus abnormal model after more rigorous validation studies using various population cohorts, several disease controls and more premalignant cases. Other approaches like Raman mapping, and sample concentration methods like DCDR may also improve the overall sensitivity rates of serum Raman spectroscopy-based screening. Prospectively, after preliminary screening using the RS results from serum, patients with a

positive result could be referred for further confirmatory investigations like histopathological diagnosis. For early detection of recurrence, a large-scale validation study with a huge sample size can help in establishing Raman spectral markers for recurrence, which could further be confirmed by biological assays, prospectively leading to implementation of this method in clinics.

6.2.2. Exfoliated cells

More effective sampling strategies and approaches, especially for the keratotic pre-cancer lesions may improve the classification efficiency. Better sampling approaches on large pre-cancer and cancer cohorts can confirm the utility of this method for monitoring of suspicious lesions and early detection of malignancy associated changes in patients. After obtaining pure models of oral cancer, pre-cancer and normal groups, the spectra-wise approach on the pellet can be employed for future applications in low resource settings. If even a single spectrum is classified as abnormal, as per standard histopathological guidelines, the sample would be treated as abnormal and subjected to further confirmatory procedures like biopsy followed by histopathology. Alternatively, Raman mapping of single cells in oral smears using automated scanning (like CARS and SRS) may aid in better identification of abnormal cells and building of robust standard models that may facilitate early oral cancer diagnosis.

6.2.3. In vivo

More number of premalignant and malignant subjects are needed to confirm the efficacy of pooled subsites model for early cancer diagnosis. In the present study, healthy could be identified

with ~98% efficiency in the single diagnostic pooled subsites model. Thus the model demonstrates high specificity and sensitivity for discriminating normal and pathological conditions. As the main aim during preliminary screening is to distinguish between normal and all other pathological conditions, this model, after prospective evaluation with higher premalignant and malignant subjects, could be employed for these applications. After preliminary screening using this model, sites predicted as abnormal/pathology can be tested for determining type of pathology using a more accurate subsite-specific model.

Studies in this thesis have demonstrated potential of Raman spectroscopy in oral cancer diagnosis using less-invasive and non-invasive approaches. Large-scale clinical and field trials are important for prospective clinical applications of the methodology. Validation on multiple-cohorts after inclusion of proper controls needs to be undertaken in future to obtain adequate data for development and testing of models to develop diagnostic algorithms. Additionally, improvements in data algorithms to enable development of rapid, unambiguous, objective and user friendly interfaces for clinical applications is also mandatory. Translation of this approach to clinics may help in better preliminary oral cancer screening and may lead to better overall outcomes.

References

References

1 Warnakulasuriya, S. Global epidemiology of oral and oropharyngeal cancer. *Oral oncology* **45**, 309-316 (2009).
2 Coelho, K. R. Challenges of the oral cancer burden in India. *Journal of cancer epidemiology* **2012** (2012).
3 Khandekar, S., Bagdey, P. & Tiwari, R. Oral cancer and some epidemiological factors: A hospital based study. *Indian J Community Med* **31**, 157-159 (2006).
4 Prasad, L. Burden of oral cancer: An Indian scenario. *Journal of Orofacial Sciences* **6**, 77 (2014).
5 Khan, Z. U. An Overview of Oral Cancer in Indian Subcontinent and Recommendations to Decrease its Incidence. (2012).
6 Sankaranarayanan, R. *et al.* Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial. *The Lancet* **365**, 1927-1933 (2005).
7 Lane, P. M. *et al.* Simple device for the direct visualization of oral-cavity tissue fluorescence. *Journal of biomedical optics* **11**, 024006-024006-024007 (2006).
8 Messadi, D. V. Diagnostic aids for detection of oral precancerous conditions. *International journal of oral science* **5**, 59-65 (2013).
9 Mehrotra, R. & Gupta, D. K. Exciting new advances in oral cancer diagnosis: avenues to early detection. *Head & neck oncology* **3**, 1-9 (2011).
10 Liu, C. J. *et al.* Increase of microRNA miR-31 level in plasma could be a potential marker of oral cancer. *Oral diseases* **16**, 360-364 (2010).
11 Sidransky, D. Emerging molecular markers of cancer. *Nature Reviews Cancer* **2**, 210-219 (2002).
12 Sawant, S. S., Zingde, S. M. & Vaidya, M. M. Cytokeratin fragments in the serum: their utility for the management of oral cancer. *Oral oncology* **44**, 722-732 (2008).
13 Malati, T. Tumour markers: an overview. *Indian Journal of Clinical Biochemistry* **22**, 17-31 (2007).
14 Bro, R. *et al.* Forecasting individual breast cancer risk using plasma metabolomics and biocontours. *Metabolomics*, 1-5 (2015).
15 Swinson, B., Jerjes, W., El-Maaytah, M., Norris, P. & Hopper, C. Optical techniques in diagnosis of head and neck malignancy. *Oral Oncology* **42**, 221-228 (2006).
16 Ellis, D. I., Cowcher, D. P., Ashton, L., O'Hagan, S. & Goodacre, R. Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool. *The Analyst* **138**, 3871-3884 (2013).
17 Fenn, M. B. *et al.* Raman spectroscopy for clinical oncology. *Advances in Optical Technologies* **2011** (2011).
18 Evers, D., Hendriks, B., Lucassen, G. & Ruers, T. Optical spectroscopy: current advances and future applications in cancer diagnostics and therapy. *Future oncology* **8**, 307-320 (2012).
19 Malini, R. *et al.* Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: a Raman spectroscopy study. *Biopolymers* **81**, 179-193 (2006).
20 Guze, K. *et al.* Parameters defining the potential applicability of Raman spectroscopy as a diagnostic tool for oral disease. *Journal of biomedical optics* **14**, 014016-014016-014019 (2009).
21 Singh, S., Deshmukh, A., Chaturvedi, P. & Krishna, C. M. In vivo Raman spectroscopic identification of premalignant lesions in oral buccal mucosa. *Journal of biomedical optics* **17**, 1050021-1050029 (2012).
22 Krishna, H., Majumder, S. K., Chaturvedi, P. & Gupta, P. K. Anatomical variability of in vivo Raman spectra of normal oral cavity and its effect on oral tissue classification. *Biomedical Spectroscopy and Imaging* **2**, 199-217 (2013).
23 Singh, S., Deshmukh, A., Chaturvedi, P. & Krishna, C. M. in *SPIE BiOS*. 82190K-82190K-82196 (International Society for Optics and Photonics).
24 Pichardo-Molina, J. *et al.* Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients. *Lasers in medical science* **22**, 229-236 (2007).
25 Harris, A. T. *et al.* Potential for Raman spectroscopy to provide cancer screening using a peripheral blood sample. *Head & neck oncology* **1**, 1-8 (2009).
26 González-Solís, J. L. *et al.* Cervical cancer detection based on serum sample Raman spectroscopy. *Lasers in medical science* **29**, 979-985 (2014).
27 Feng, S. *et al.* Nasopharyngeal cancer detection based on blood plasma surface-enhanced Raman spectroscopy and multivariate analysis. *Biosensors and Bioelectronics* **25**, 2414-2419 (2010).
28 Owens, G. L. *et al.* Vibrational biospectroscopy coupled with multivariate analysis extracts potentially diagnostic features in blood plasma/serum of ovarian cancer patients. *Journal of biophotonics* **7**, 200-209 (2014).
29 Lin, D. *et al.* Colorectal cancer detection by gold nanoparticle based surface-enhanced Raman spectroscopy of blood serum and statistical analysis. *Optics express* **19**, 13565-13577 (2011).
30 Papamarkakis, K. *et al.* Cytopathology by optical methods: spectral cytopathology of the oral mucosa. *Laboratory Investigation* **90**, 589-598 (2010).

References

31 Nanda, K. *et al.* Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Annals of internal medicine* **132**, 810-819 (2000).

32 Niv, Y. & Sperber, A. D. Sensitivity, specificity, and predictive value of fecal occult blood testing (Hemoccult II) for colorectal neoplasia in symptomatic patients: a prospective study with total colonoscopy. *The American journal of gastroenterology* **90**, 1974-1977 (1995).

33 Ferrini, R., Mannino, E., Ramsdell, E. & Hill, L. Screening mammography for breast cancer: American College of Preventive Medicine practice policy statement. *American journal of preventive medicine* **12**, 340-341 (1996).

34 Rubina, S., Amita, M., Bharat, R. & Krishna, C. M. Raman spectroscopic study on classification of cervical cell specimens. *Vibrational Spectroscopy* **68**, 115-121 (2013).

35 Naghavi, M. *et al.* Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **385**, 117-171 (2015).

36 Organization, W. H. Global status report on noncommunicable diseases. Geneva: 2010. *World Health Organization.[acesso em 2012 Fev 28]. Disponível em: http://www. who. int/nmh/publications/ncd_report2010/en* (2014).

37 Organization, W. H. (2008).

38 Hunter, D. J. & Reddy, K. S. Noncommunicable diseases. *New England Journal of Medicine* **369**, 1336-1343 (2013).

39 Organization, W. H. Scaling up action against noncommunicable diseases: How much will it cost? (2011).

40 Werning, J. W. *Oral cancer: diagnosis, management, and rehabilitation*. (Thieme, 2011).

41 Pereira, M. C., Oliveira, D. T., Landman, G. & Kowalski, L. P. Histologic subtypes of oral squamous cell carcinoma: prognostic relevance. *Journal (Canadian Dental Association)*, 339-344 (2007).

42 Williams, H. Molecular pathogenesis of oral squamous carcinoma. *Molecular Pathology* **53**, 165 (2000).

43 Albo, D., Granick, M. S., Jhala, N., Atkinson, B. & Solomon, M. P. The relationship of angiogenesis to biological activity in human squamous cell carcinomas of the head and neck. *Annals of plastic surgery* **32**, 588-594 (1994).

44 Neville, B. W. *Oral and maxillofacial pathology*. (Elsevier Brasil, 2009).

45 Ho, P.-S. *et al.* Malignant transformation of oral potentially malignant disorders in males: a retrospective cohort study. *BMC cancer* **9**, 260 (2009).

46 Guillaud, M., Zhang, L., Poh, C., Rosin, M. P. & MacAulay, C. Potential use of quantitative tissue phenotype to predict malignant risk for oral premalignant lesions. *Cancer research* **68**, 3099-3107 (2008).

47 Nair, M. K. & Sankaranarayanan, R. Epidemiologic leads to cancer control in India. *Cancer Causes and Control* **2**, 263-265 (1991).

48 Shafer, W. G. Initial mismanagement and delay in diagnosis of oral cancer. *The Journal of the American Dental Association* **90**, 1262-1264 (1975).

49 de Camargo Cancela, M. *et al.* Oral cavity cancer in developed and in developing countries: Population-based incidence. *Head & neck* **32**, 357-367 (2010).

50 John, R., Sung, H. & Max, W. Economic cost of tobacco use in India, 2004. *Tobacco control* **18**, 138-143 (2009).

51 Imam, S. Z. *et al.* Use of smokeless tobacco among groups of Pakistani medical students–a cross sectional study. *BMC Public Health* **7**, 231 (2007).

52 Kuper, H., Boffetta, P. & Adami, H. O. Tobacco use and cancer causation: association by tumour type. *Journal of internal medicine* **252**, 206-224 (2002).

53 Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International journal of cancer* **127**, 2893-2917 (2010).

54 Elango, J. K., Gangadharan, P., Sumithra, S. & Kuriakose, M. Trends of head and neck cancers in urban and rural India. *Asian Pacific journal of cancer prevention* **7**, 108 (2006).

55 Health, N. I. o. Detecting Oral Cancer: A Guide for Health Care Professionals. (2008).

56 Napier, S. S. & Speight, P. M. Natural history of potentially malignant oral lesions and conditions: an overview of the literature. *Journal of oral pathology & medicine* **37**, 1-10 (2008).

57 van der Waal, I. Potentially malignant disorders of the oral and oropharyngeal mucosa; present concepts of management. *Oral oncology* **46**, 423-425 (2010).

58 Greene, F. L. *AJCC cancer staging manual*. Vol. 1 (Springer Science & Business Media, 2002).

59 Reis, L., Hankey, B. & Miller, B. Cancer statistics review 1973-1988. *Bethesda, Md, National Cancer Institute, NIH publication 91-2789, III* **39** (1991).

60 Po Wing Yuen, A. *et al.* Prognostic factors of clinically stage I and II oral tongue carcinoma—A comparative study of stage, thickness, shape, growth pattern, invasive front malignancy grading, martinez-gimeno score, and pathologic features. *Head & neck* **24**, 513-520 (2002).

61 Clara, S. Histological Typing of Cancer and Precancer of the Oral Mucosa.

62 Neville, B. W. & Day, T. A. Oral cancer and precancerous lesions. *CA: a cancer journal for clinicians* **52**, 195-215 (2002).

63 Marshall, J. R. *et al.* Smoking, alcohol, dentition and diet in the epidemiology of oral cancer. *European Journal of Cancer Part B: Oral Oncology* **28**, 9-15 (1992).

64 Johnson, N. Tobacco use and oral cancer: a global perspective. *Journal of Dental Education* **65**, 328-339 (2001).

65 Humans, I. W. G. o. t. E. o. C. R. t., Organization, W. H. & Cancer, I. A. f. R. o. *Tobacco Smoke and Involuntary Smoking: This Publication Represents the Views and Expert Opinions of an IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, which Met in Lyon, 11-18 June 2002*. (Iarc, 2004).

66 Carbone, D. Smoking and cancer. *The American journal of medicine* **93**, S13-S17 (1992).

67 Pednekar, M. S., Gupta, P. C., Yeole, B. B. & Hébert, J. R. Association of tobacco habits, including bidi smoking, with overall and site-specific cancer incidence: results from the Mumbai cohort study. *Cancer Causes & Control* **22**, 859-868 (2011).

68 Mashberg, A., Boffetta, P., Winkelman, R. & Garfinkel, L. Tobacco smoking, alcohol drinking, and cancer of the oral cavity and oropharynx among US veterans. *Cancer* **72**, 1369-1375 (1993).

69 Andre, K., Schraub, S., Mercier, M. & Bontemps, P. Role of alcohol and tobacco in the aetiology of head and neck cancer: a case-control study in the Doubs region of France. *European Journal of Cancer Part B: Oral Oncology* **31**, 301-309 (1995).

70 Rahman, M., Sakamoto, J. & Fukui, T. Bidi smoking and oral cancer: A meta-analysis. *International journal of cancer* **106**, 600-604 (2003).

71 Thoma, K. H. Stomatitis nicotina and its effect on the palate. *American Journal of Orthodontics and Oral Surgery* **27**, C38-C47 (1941).

72 Kramer, I., Lucas, R., Pindborg, J. & Sobin, L. WHO Collaborating Centre for Oral Precancerous Lesions. *Definition of leukoplakia and related lesions: an aid to studies on oral precancer, Oral Surg Oral Med Oral Pathol* **46**, 518-539 (1978).

73 Van der Eb, M. *et al.* Reverse smoking as a risk factor for palatal cancer: A cross-sectional study in rural andhra pradesh, india. *International journal of cancer* **54**, 754-758 (1993).

74 Stoykewych, A., DeBrouwere, R. & Curran, J. Reverse smoking and its effects on the hard palate: a case report. *Journal (Canadian Dental Association)* **58**, 215-216 (1992).

75 Siddiqi, K. *et al.* Global burden of disease due to smokeless tobacco consumption in adults: analysis of data from 113 countries. *BMC medicine* **13**, 194 (2015).

76 Khan, Z., Tönnies, J. & Müller, S. Smokeless tobacco and oral cancer in South Asia: a systematic review with meta-analysis. *Journal of cancer epidemiology* **2014** (2014).

77 Humans, I. W. G. o. t. E. o. C. R. t., Organization, W. H. & Cancer, I. A. f. R. o. *Smokeless tobacco and some tobacco-specific N-nitrosamines*. Vol. 89 (World Health Organization, 2007).

78 Warnakulasuriya, S. Smokeless tobacco and oral cancer. *Oral diseases* **10**, 1-4 (2004).

79 Idris, A., Ahmed, H. & Malik, M. Toombak dipping and cancer of the oral cavity in the Sudan: A case-control study. *International Journal of Cancer* **63**, 477-480 (1995).

80 Warnakulasuriya, S., Trivedy, C. & Peters, T. J. Areca nut use: an independent risk factor for oral cancer: The health problem is under-recognised. *BMJ: British Medical Journal* **324**, 799 (2002).

81 Jovanovic, A., Schulten, E. A., Kostense, P. J., Snow, G. B. & Waal, I. Tobacco and alcohol related to the anatomical site of oral squamous cell carcinoma. *Journal of oral pathology & medicine* **22**, 459-462 (1993).

82 Blot, W. J. *et al.* Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer research* **48**, 3282-3287 (1988).

83 Lewin, F. *et al.* Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck. *Cancer* **82**, 1367-1375 (1998).

84 Moreno-Lopez, L. *et al.* Risk of oral cancer associated with tobacco smoking, alcohol consumption and oral hygiene: a case-control study in Madrid, Spain. *Oral oncology* **36**, 170-174 (2000).

85 Ko, Y. C. *et al.* Betel quid chewing, cigarette smoking and alcohol consumption related to oral cancer in Taiwan. *Journal of oral pathology & medicine* **24**, 450-453 (1995).

86 Sugerman, P. & Shillitoe, E. The high risk human papillomaviruses and oral cancer: evidence for and against a causal relationship. *Oral diseases* **3**, 130-147 (1997).

87 Gupta, K. & Metgud, R. Evidences suggesting involvement of viruses in oral squamous cell carcinoma. *Pathology research international* **2013** (2013).

88 Park, N.-H., Li, S.-L., Xie, J.-F. & Cherrick, H. M. In vitro and animal studies of the role of viruses in oral carcinogenesis. *European Journal of Cancer Part B: Oral Oncology* **28**, 145-152 (1992).

89 Sand, L. & Jalouli, J. Viruses and oral cancer. Is there a link? *Microbes and infection* **16**, 371-378 (2014).

90 Maden, C. *et al.* Human papillomaviruses, herpes simplex viruses, and the risk of oral cancer in men. *American journal of epidemiology* **135**, 1093-1102 (1992).

91 Ankathil, R., Mathew, A., Joseph, F. & Nair, M. Is oral cancer susceptibility inherited? Report of five oral cancer families. *European Journal of Cancer Part B: Oral Oncology* **32**, 63-67 (1996).

92 Pogoda, J. M. & Preston-Martin, S. Solar radiation, lip protection, and lip cancer risk in Los Angeles County women (California, United States). *Cancer Causes & Control* **7**, 458-463 (1996).

93 Winn, D. M. Diet and nutrition in the etiology of oral cancer. *The American journal of clinical nutrition* **61**, 437S-445S (1995).

94 Winn, D. M. *et al.* Diet in the etiology of oral and pharyngeal cancer among women from the southern United States. *Cancer research* **44**, 1216-1222 (1984).

95 Potter, J. D. & Steinmetz, K. Vegetables, fruit and phytoestrogens as preventive agents. *IARC scientific publications*, 61-90 (1995).

96 Block, G., Patterson, B. & Subar, A. Fruit, vegetables, and cancer prevention: a review of the epidemiological evidence. *Nutrition and cancer* **18**, 1-29 (1992).

97 Shafer, W. & Waldron, C. A. A clinical and histopathologic study of oral leukoplakia. *Surgery, gynecology & obstetrics* **112**, 411 (1961).

98 Van der Waal, I. Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management. *Oral oncology* **45**, 317-323 (2009).

99 Bouquot, J. E. & Gorlin, R. J. Leukoplakia, lichen planus, and other oral keratoses in 23,616 white Americans over the age of 35 years. *Oral surgery, oral medicine, oral pathology* **61**, 373-381 (1986).

100 Waldron, C. A. & Shafer, W. G. Leukoplakia revisited. A clinicopathologic study 3256 oral leukoplakias. *Cancer* **36**, 1386-1392 (1975).

101 Warnakulasuriya, S., Johnson, N. & Van der Waal, I. Nomenclature and classification of potentially malignant disorders of the oral mucosa. *Journal of oral pathology & medicine* **36**, 575-580 (2007).

102 Carnelio, S., Rodrigues, G. S., Shenoy, R. & Fernandes, D. A brief review of common oral premalignant lesions with emphasis on their management and cancer prevention. *Indian Journal of Surgery* **73**, 256-261 (2011).

103 Silverman, S. & Rozen, R. D. Observations on the clinical characteristics and natural history of oral leukoplakia. *The Journal of the American Dental Association* **76**, 772-777 (1968).

104 Hogewind, W. F., van der Kwast, W. A. & van der Waal, I. Oral leukoplakia, with emphasis on malignant transformation: a follow-up study of 46 patients. *Journal of Cranio-Maxillofacial Surgery* **17**, 128-133 (1989).

105 Hansen, L. S., Olson, J. A. & Silverman, S. Proliferative verrucous leukoplakia: a long-term study of thirty patients. *Oral surgery, oral medicine, oral pathology* **60**, 285-298 (1985).

106 Silverman, S. & Gorsky, M. Proliferative verrucous leukoplakia: a follow-up study of 54 cases. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology* **84**, 154-157 (1997).

107 Cabay, R. J., Morton, T. H. & Epstein, J. B. Proliferative verrucous leukoplakia and its progression to oral carcinoma: a review of the literature. *Journal of oral pathology & medicine* **36**, 255-261 (2007).

108 Reichart, P. A. & Philipsen, H. P. Oral erythroplakia—a review. *Oral oncology* **41**, 551-561 (2005).

109 Hashibe, M. *et al.* Chewing tobacco, alcohol, and the risk of erythroplakia. *Cancer Epidemiology Biomarkers & Prevention* **9**, 639-645 (2000).

110 Shafer, W. G. & Waldron, C. A. Erythroplakia of the oral cavity. *Cancer* **36**, 1021-1028 (1975).

111 Arakeri, G. & Brennan, P. A. Oral submucous fibrosis: an overview of the aetiology, pathogenesis, classification, and principles of management. *British Journal of Oral and Maxillofacial Surgery* **51**, 587-593 (2013).

112 Mahomed, F. Oral submucous fibrosis-a potentially malignant condition of growing concern: clinical review. *South African Dental Journal* **67**, 562, 564-565 (2012).

113 Yardimci, G., Kutlubay, Z., Engin, B. & Tuzun, Y. Precancerous lesions of oral mucosa. *World Journal of Clinical Cases: WJCC* **2**, 866 (2014).

114 Tilakaratne, W., Klinikowski, M., Saku, T., Peters, T. & Warnakulasuriya, S. Oral submucous fibrosis: review on aetiology and pathogenesis. *Oral oncology* **42**, 561-568 (2006).

115 Jaitley, S., Agarwal, P. & Upadhyay, R. Role of oral exfoliative cytology in predicting premalignant potential of oral submucous fibrosis: A short study. *Journal of cancer research and therapeutics* **11**, 471 (2015).

116 Smith, J. F., Mincer, H. A., Hopkins, K. P. & Bell, J. Snuff-dipper's lesion: a cytological and pathological study in a large population. *Archives of Otolaryngology* **92**, 450-456 (1970).

117 Grady, D. *et al.* Oral mucosal lesions found in smokeless tobacco users. *The Journal of the American Dental Association* **121**, 117-123 (1990).

118 Ismail, S. B., Kumar, S. K. & Zain, R. B. Oral lichen planus and lichenoid reactions: etiopathogenesis, diagnosis, management and malignant transformation. *Journal of oral science* **49**, 89-106 (2007).

119 Al-Hashimi, I. *et al.* Oral lichen planus and oral lichenoid lesions: diagnostic and therapeutic considerations. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology* **103**, S25. e21-S25. e12 (2007).

120 Mehrotra, R. *Oral Cytology: A Concise Guide*. (Springer Science & Business Media, 2012).

121 Sciubba, J. J. & Group, U. C. O. S. Improving detection of precancerous and cancerous oral lesions: computer-assisted analysis of the oral brush biopsy. *The Journal of the American Dental Association* **130**, 1445-1457 (1999).

122 Scheifele, C., Schmidt-Westhausen, A.-M., Dietrich, T. & Reichart, P. A. The sensitivity and specificity of the OralCDx technique: evaluation of 103 cases. *Oral oncology* **40**, 824-828 (2004).

123 Handschel, J. *et al.* Additional use of DNA-image cytometry improves the assessment of resection margins. *Journal of oral pathology & medicine* **36**, 472-475 (2007).

124 Graveland, A. P. *et al.* Molecular screening of oral precancer. *Oral oncology* **49**, 1129-1135 (2013).

125 Bremmer, J. F. *et al.* Screening for oral precancer with noninvasive genetic cytology. *Cancer Prevention Research* **2**, 128-133 (2009).

126 Gupta, A., Singh, M., Ibrahim, R. & Mehrotra, R. Utility of toluidine blue and oral brush biopsy in oral precancerous lesions and squamous cell carcinoma. *Acta Cytol* **51**, 788-794 (2007).

127 Martin, I., Kerawala, C. & Reed, M. The application of toluidine blue as a diagnostic adjunct in the detection of epithelial dysplasia. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology* **85**, 444-446 (1998).

128 Epstein, J. B. & Güneri, P. The adjunctive role of toluidine blue in detection of oral premalignant and malignant lesions. *Current opinion in otolaryngology & head and neck surgery* **17**, 79-87 (2009).

129 Rahman, F. *et al.* A study to evaluate the efficacy of toluidine blue and cytology in detecting oral cancer and dysplastic lesions. *Quintessence international (Berlin, Germany: 1985)* **43**, 51-59 (2012).

130 Patel, B. P., Shah, S. V., Shukla, S. N., Shah, P. M. & Patel, P. S. Clinical significance of MMP-2 and MMP-9 in patients with oral cancer. *Head & neck* **29**, 564-572 (2007).

131 Li, Y. *et al.* Salivary transcriptome diagnostics for oral cancer detection. *Clinical Cancer Research* **10**, 8442-8450 (2004).

132 Brinkmann, O. & Wong, D. Salivary transcriptome biomarkers in oral squamous cell cancer detection. *Advances in clinical chemistry* **55**, 21-34 (2010).

133 Lallemant, B. *et al.* Clinical relevance of nine transcriptional molecular markers for the diagnosis of head and neck squamous cell carcinoma in tissue and saliva rinse. *BMC cancer* **9**, 370 (2009).

134 Wong, D. T. Salivary diagnostics powered by nanotechnologies, proteomics and genomics. *The Journal of the American Dental Association* **137**, 313-321 (2006).

135 Ramanujam, N. Fluorescence spectroscopy of neoplastic and non-neoplastic tissues. *Neoplasia* **2**, 89-117 (2000).

136 Schantz, S. P. *et al.* In vivo native cellular fluorescence and histological characteristics of head and neck cancer. *Clinical Cancer Research* **4**, 1177-1182 (1998).

137 Harris, D. M. & Werkhaven, J. Endogenous porphyrin fluorescence in tumors. *Lasers in surgery and medicine* **7**, 467-472 (1987).

138 Onizawa, K. *et al.* Analysis of fluorescence in oral squamous cell carcinoma. *Oral oncology* **38**, 343-348 (2002).

139 Onizawa, K., Okamura, N., Saginoya, H. & Yoshida, H. Characterization of autofluorescence in oral squamous cell carcinoma. *Oral oncology* **39**, 150-156 (2003).

140 Onizawa, K., Saginoya, H., Furuya, Y., Yoshida, H. & Fukuda, H. Usefulness of fluorescence photography for diagnosis of oral cancer. *International Journal of Oral & Maxillofacial Surgery* **28**, 206-210 (1999).

141 Chen, C. T. *et al.* Light-induced fluorescence spectroscopy: a potential diagnostic tool for oral neoplasia. *Proceedings of the National Science Council, Republic of China. Part B, Life sciences* **20**, 123-130 (1996).

142 Wang, C. Y. *et al.* Autofluorescence spectroscopy for in vivo diagnosis of DMBA-induced hamster buccal pouch pre-cancers and cancers. *Journal of oral pathology & medicine* **32**, 18-24 (2003).

143 Gillenwater, A. *et al.* Noninvasive diagnosis of oral neoplasia based on fluorescence spectroscopy and native tissue autofluorescence. *Archives of Otolaryngology–Head & Neck Surgery* **124**, 1251-1258 (1998).

144 Shaiju, S. N. *et al.* Habits with killer instincts: in vivo analysis on the severity of oral mucosal alterations using autofluorescence spectroscopy. *Journal of biomedical optics* **16**, 087006-087006-087013 (2011).

145 Farwell, D. G. *et al.* Time-resolved fluorescence spectroscopy as a diagnostic technique of oral carcinoma: validation in the hamster buccal pouch model. *Archives of Otolaryngology–Head & Neck Surgery* **136**, 126-133 (2010).

146 Francisco, A. L. N. *et al.* Fluorescence spectroscopy for the detection of potentially malignant disorders and squamous cell carcinoma of the oral cavity. *Photodiagnosis and photodynamic therapy* **11**, 82-90 (2014).

147 Francisco, A. L. N. *et al.* Analysis of surgical margins in oral cancer using in situ fluorescence spectroscopy. *Oral oncology* **50**, 593-599 (2014).

148 Yuvaraj, M. *et al.* Fluorescence spectroscopic characterization of salivary metabolites of oral cancer patients. *Journal of Photochemistry and Photobiology B: Biology* **130**, 153-160 (2014).

149 Bellisola, G. & Sorio, C. Infrared spectroscopy and microscopy in cancer research and diagnosis. *American journal of cancer research* **2**, 1 (2012).

150 Schultz, C. P., Liu, K.-Z., Kerr, P. D. & Mantsch, H. H. In situ infrared histopathology of keratinization in human oral/oropharyngeal squamous cell carcinoma. *Oncology research* **10**, 277-286 (1997).

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

152 Wu, J. G. *et al.* Distinguishing malignant from normal oral tissues using FTIR fiber-optic techniques. *Biopolymers* **62**, 185-192 (2001).

153 Banerjee, S. *et al.* Fourier-transform-infrared-spectroscopy based spectral-biomarker selection towards optimum diagnostic differentiation of oral leukoplakia and cancer. *Analytical and bioanalytical chemistry* **407**, 7935-7943 (2015).

154 Müller, M. G. *et al.* Spectroscopic detection and evaluation of morphologic and biochemical changes in early human oral carcinoma. *Cancer* **97**, 1681-1692 (2003).

155 Jerjes, W., Swinson, B., Pickard, D., Thomas, G. & Hopper, C. Detection of cervical intranodal metastasis in oral cancer using elastic scattering spectroscopy. *Oral oncology* **40**, 673-678 (2004).

156 Jerjes, W., Swinson, B., Johnson, K., Thomas, G. & Hopper, C. Assessment of bony resection margins in oral cancer using elastic scattering spectroscopy: a study on archival material. *Archives of oral biology* **50**, 361-366 (2005).

157 Sharwani, A. *et al.* Assessment of oral premalignancy using elastic scattering spectroscopy. *Oral oncology* **42**, 343-349 (2006).

158 De Veld, D. C. *et al.* Autofluorescence and diffuse reflectance spectroscopy for oral oncology. *Lasers in surgery and medicine* **36**, 356-364 (2005).

159 Subhash, N. *et al.* Oral cancer detection using diffuse reflectance spectral ratio R540⁄R575 of oxygenated hemoglobin bands. *Journal of biomedical optics* **11**, 014018-014018-014016 (2006).

160 Jayanthi, J. *et al.* Diffuse reflectance spectroscopy: diagnostic accuracy of a non-invasive screening technique for early detection of malignant changes in the oral cavity. *BMJ open* **1**, e000071 (2011).

161 Stephen, M. M. *et al.* Diagnostic accuracy of diffuse reflectance imaging for early detection of pre-malignant and malignant changes in the oral cavity: a feasibility study. *BMC cancer* **13**, 278 (2013).

162 Chen, Y.-W., Liaw, Y.-K., Hung, H.-R., Chung, P.-C. & Tseng, S.-H. in *Asia Communications and Photonics Conference*. AF4I. 3 (Optical Society of America).

163 Yu, B., Shah, A., Nagarajan, V. K. & Ferris, D. G. Diffuse reflectance spectroscopy of epithelial tissue with a smart fiber-optic probe. *Biomedical optics express* **5**, 675-689 (2014).

164 Raman, C. V. & Krishnan, K. S. A new type of secondary radiation. *Nature* **121**, 501-502 (1928).

165 Müller, U. M. Diem: Introduction to modern vibrational spectroscopy, J. Wiley, New York, Chichester, 1993, ISBN 0-471-59584-5, 285 Seiten, Preis:£ 49.50. *Berichte der Bunsengesellschaft für physikalische Chemie* **98**, 1347-1348 (1994).

166 Pelletier, M. J. *Analytical applications of Raman spectroscopy*. (Wiley-Blackwell, 1999).

167 Das, R. S. & Agrawal, Y. Raman spectroscopy: Recent advancements, techniques and applications. *Vibrational Spectroscopy* **57**, 163-176 (2011).

168 Ferraro, J. R. *Introductory raman spectroscopy*. (Academic press, 2003).

169 Kneipp, K., Kneipp, H., Itzkan, I., Dasari, R. R. & Feld, M. S. Surface-enhanced Raman scattering and biophysics. *Journal of Physics: Condensed Matter* **14**, R597 (2002).

170 Xie, W. & Schlücker, S. Medical applications of surface-enhanced Raman scattering. *Physical Chemistry Chemical Physics* **15**, 5329-5344 (2013).

171 Dingari, N. C., Horowitz, G. L., Kang, J. W., Dasari, R. R. & Barman, I. Raman spectroscopy provides a powerful diagnostic tool for accurate determination of albumin glycation. *PLoS One* **7**, e32406 (2012).

172 Barman, I. *et al.* Raman spectroscopy-based sensitive and specific detection of glycated hemoglobin. *Analytical chemistry* **84**, 2474-2482 (2012).

173 Filik, J. & Stone, N. Drop coating deposition Raman spectroscopy of protein mixtures. *The Analyst* **132**, 544-550 (2007).

174 Tolles, W. M., Nibler, J., McDonald, J. & Harvey, A. A review of the theory and application of coherent anti-Stokes Raman spectroscopy (CARS). *Applied Spectroscopy* **31**, 253-271 (1977).

175 Kudelski, A. Analytical applications of Raman spectroscopy. *Talanta* **76**, 1-8 (2008).

176 White, J. C. in *Tunable lasers* 115-207 (Springer, 1987).

177 Yakovlev, V. V. *et al.* Stimulated Raman scattering: old physics, new applications. *Journal of modern optics* **56**, 1970-1973 (2009).

178 Eesley, G. L. *Coherent Raman Spectroscopy*. (Elsevier, 2013).

179 Matousek, P. *et al.* Subsurface probing in diffusely scattering media using spatially offset Raman spectroscopy. *Applied spectroscopy* **59**, 393-400 (2005).

180 Stone, N. *et al.* Surface enhanced spatially offset Raman spectroscopic (SESORS) imaging–the next dimension. *Chemical Science* **2**, 776-780 (2011).

181 Sharma, B., Ma, K., Glucksberg, M. R. & Van Duyne, R. P. Seeing through bone with surface-enhanced spatially offset Raman spectroscopy. *Journal of the American Chemical Society* **135**, 17290-17293 (2013).

182 Choquette, S. J., Etz, E. S., Hurst, W. S., Blackburn, D. H. & Leigh, S. D. Relative intensity correction of Raman spectrometers: NIST SRMs 2241 through 2243 for 785 nm, 532 nm, and 488 nm/514.5 nm excitation. *Applied spectroscopy* **61**, 117-129 (2007).

183 do Lago, C. L., Juliano, V. F. & Kascheres, C. Applying moving median digital filter to mass spectrometry and potentiometric titration. *Analytica chimica acta* **310**, 281-288 (1995).

184 Ehrentreich, F. & Sümmchen, L. Spike removal and denoising of Raman spectra by wavelet transform methods. *Analytical chemistry* **73**, 4364-4373 (2001).

185 Friedrichs, M. S. A model-free algorithm for the removal of baseline artifacts. *Journal of Biomolecular NMR* **5**, 147-153 (1995).

186 Zhao, J., Lui, H., McLean, D. I. & Zeng, H. Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy. *Applied spectroscopy* **61**, 1225-1232 (2007).

187 Lieber, C. A. & Mahadevan-Jansen, A. Automated method for subtraction of fluorescence from biological Raman spectra. *Applied spectroscopy* **57**, 1363-1367 (2003).

188 Mosier-Boss, P., Lieberman, S. & Newbery, R. Fluorescence rejection in Raman spectroscopy by shifted-spectra, edge detection, and FFT filtering techniques. *Applied spectroscopy* **49**, 630-638 (1995).

189 Zhang, D. & Ben-Amotz, D. Enhanced chemical classification of Raman images in the presence of strong fluorescence interference. *Applied Spectroscopy* **54**, 1379-1383 (2000).

190 Hu, Y. *et al.* A background elimination method based on wavelet transform for Raman spectra. *Chemometrics and Intelligent Laboratory Systems* **85**, 94-101 (2007).

191 Zhang, Z.-M. *et al.* An intelligent background-correction algorithm for highly fluorescent samples in Raman spectroscopy. *Journal of Raman Spectroscopy* **41**, 659-669 (2010).

192 Schulze, G. *et al.* Investigation of selected baseline removal techniques as candidates for automated implementation. *Applied spectroscopy* **59**, 545-574 (2005).

193 Loethen, Y. L., Zhang, D., Favors, R. N., Basiaga, S. B. & Ben-Amotz, D. Second-derivative variance minimization method for automated spectral subtraction. *Applied spectroscopy* **58**, 272-278 (2004).

194 Savitzky, A. & Golay, M. J. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* **36**, 1627-1639 (1964).

195 Kastanos, E., Kyriakides, A., Hadjigeorgiou, K. & Pitris, C. A novel method for bacterial UTI diagnosis using Raman spectroscopy. *International Journal of Spectroscopy* **2012** (2012).

196 Bocklitz, T., Walter, A., Hartmann, K., Rösch, P. & Popp, J. How to pre-process Raman spectra for reliable and stable models? *Analytica chimica acta* **704**, 47-56 (2011).

197 Randolph, T. W. Scale-based normalization of spectral data. *Cancer Biomark* **2**, 135-144 (2006).

198 Baker, F. B. & Hubert, L. J. Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association* **70**, 31-38 (1975).

199 Schott, J. R. Principles of Multivariate Analysis: A User's Perspective. *Journal of the American Statistical Association* **97** (2002).

200 Varmuza, K. & Filzmoser, P. *Introduction to multivariate statistical analysis in chemometrics*. (CRC press, 2009).

201 Polat, K. & Güneş, S. Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer. *Expert Systems with Applications* **34**, 214-221 (2008).

202 Wang, L. & Mizaikoff, B. Application of multivariate data-analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy. *Analytical and bioanalytical chemistry* **391**, 1641-1654 (2008).

203 Chatfied, C. & Collins, A. J. *Introduction to multivariate analysis*. (Springer, 2013).

204 Schervish, M. J. A review of multivariate analysis. *Statistical Science*, 396-413 (1987).

205 Barker, M. & Rayens, W. Partial least squares for discrimination. *Journal of chemometrics* **17**, 166-173 (2003).

206 Vapnik, V. N. & Vapnik, V. *Statistical learning theory*. Vol. 1 (Wiley New York, 1998).

207 Gautam, R., Vanga, S., Ariese, F. & Umapathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Techniques and Instrumentation* **2**, 1-38 (2015).

208 Lepot, L., De Wael, K., Gason, F. & Gilbert, B. Application of Raman spectroscopy to forensic fibre cases. *Science & Justice* **48**, 109-117 (2008).

209 Mogilevsky, G., Borland, L., Brickhouse, M. & Fountain III, A. W. Raman spectroscopy for homeland security applications. *International Journal of Spectroscopy* **2012** (2012).

210 Gantner, E. & Steinert, D. Applications of laser Raman spectrometry in process control, using optical fibers. *Fresenius' Journal of Analytical Chemistry* **338**, 2-8 (1990).

211 Schmitt, M. & Popp, J. Raman spectroscopy at the beginning of the twenty-first century. *Journal of Raman Spectroscopy* **37**, 20-28 (2006).

212 Herrero, A. M. Raman spectroscopy a promising technique for quality assessment of meat and fish: A review. *Food Chemistry* **107**, 1642-1651 (2008).

213 Ashton, L., Lau, K., Winder, C. L. & Goodacre, R. Raman spectroscopy: lighting up the future of microbial identification. *Future microbiology* **6**, 991-997 (2011).

214 Virkler, K. & Lednev, I. K. Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene. *Forensic Science International* **188**, 1-17 (2009).

215 Motz, J. T. *et al.* In vivo Raman spectral pathology of human atherosclerosis and vulnerable plaque. *Journal of Biomedical Optics* **11**, 021003-021003-021009 (2006).

216 De Jong, B. *et al.* Raman spectroscopic detection of changes in molecular composition of bladder muscle tissue caused by outlet obstruction. *Vibrational spectroscopy* **32**, 57-65 (2003).

217 Maher, J. R., Takahata, M., Awad, H. A. & Berger, A. J. Raman spectroscopy detects deterioration in biomechanical properties of bone in a glucocorticoid-treated mouse model of rheumatoid arthritis. *Journal of biomedical optics* **16**, 087012-087012-087016 (2011).

218 Chen, T., Kozloff, K. M., Goldstein, S. A. & Morris, M. D. in *Biomedical Optics 2004.* 85-92 (International Society for Optics and Photonics).

219 Hanlon, E. *et al.* Prospects for in vivo Raman spectroscopy. *Physics in medicine and biology* **45**, R1 (2000).

220 Nijssen, A., Koljenović, S., Schut, T. C. B., Caspers, P. J. & Puppels, G. J. Towards oncological application of Raman spectroscopy. *Journal of biophotonics* **2**, 29-36 (2009).

221 Choo-Smith, L. P. *et al.* Medical applications of Raman spectroscopy: from proof of principle to clinical implementation. *Biopolymers* **67**, 1-9 (2002).

222 Başar, G. *et al.* Investigation of preeclampsia using raman spectroscopy. *Journal of Spectroscopy* **27**, 239-252 (2012).

223 Neugebauer, U. *et al.* Fast differentiation of SIRS and sepsis from blood plasma of ICU patients using Raman spectroscopy. *Journal of biophotonics* **7**, 232-240 (2014).

224 Saade, J., Pacheco, M. T. T. & Rodrigues, M. R. Identification of hepatitis C in human blood serum by near-infrared Raman spectroscopy. *Journal of Spectroscopy* **22**, 387-395 (2008).

225 Schipper, H. M. *et al.* Spectroscopy of human plasma for diagnosis of idiopathic Parkinson's disease. (2008).

226 Carmona, P. *et al.* Discrimination analysis of blood plasma associated with Alzheimer's disease using vibrational spectroscopy. *Journal of Alzheimer's Disease* **34**, 911-920 (2013).

227 Ryzhikova, E. *et al.* Raman spectroscopy of blood serum for Alzheimer's disease diagnostics: specificity relative to other types of dementia. *Journal of biophotonics* **9999** (2014).

228 Rehman, A. *et al.* Dengue blood analysis by Raman spectroscopy. *Laser Physics* **22**, 1085-1089 (2012).

229 Saleem, M., Bilal, M., Anwar, S., Rehman, A. & Ahmed, M. Optical diagnosis of dengue virus infection in human blood serum using Raman spectroscopy. *Laser Physics Letters* **10**, 035602 (2013).

230 Torres Filho, I. P., Terner, J. & Scheider, L. Resonance Raman spectroscopy of human sickle cell hemoglobin from transgenic mice. *The FASEB Journal* **23**, 768.762 (2009).

231 Wood, B. R. & McNaughton, D. Resonance raman spectroscopy of erythrocytes. *Handbook of Vibrational Spectroscopy* (2008).

232 Bueno Filho, A. C., Silveira Jr, L., Yanai, A. L. S. A. & Fernandes, A. B. Raman spectroscopy for a rapid diagnosis of sickle cell disease in human blood samples: a preliminary study. *Lasers in medical science* **30**, 247-253 (2015).

233 Sahu, A., Dalal, K., Naglot, S., Aggarwal, P. & Krishna, C. M. Serum based diagnosis of asthma using Raman spectroscopy: an early phase pilot study. **8(11)**, e78921 (2013).

234 Li, X. & Jin, H. in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE.* 3221-3224 (IEEE).

235 Bhattacharjee, T. *et al.* Raman spectroscopy of serum: A study on 'pre'and 'post'breast adenocarcinoma resection in rat models. *Journal of biophotonics* **9999** (2014).

236 Li, X., Yang, T. & Li, S. Discrimination of serum Raman spectroscopy between normal and colorectal cancer using selected parameters and regression-discriminant analysis. *Applied optics* **51**, 5038-5043 (2012).

237 Wang, G. *et al.* Detection of the potential pancreatic cancer marker MUC4 in serum using surface-enhanced Raman scattering. *Analytical chemistry* **83**, 2554-2561 (2011).

238 Diem, M. *et al.* Applications of infrared and Raman microspectroscopy of cells and tissue in medical diagnostics: Present status and future promises. *Journal of Spectroscopy* **27**, 463-496 (2012).

239 Vargis, E., Tang, Y.-W., Khabele, D. & Mahadevan-Jansen, A. Near-infrared Raman microspectroscopy detects high-risk human papillomaviruses. *Translational oncology* **5**, 172-179 (2012).

240 Haka, A. S. *et al.* In vivo Margin Assessment during Partial Mastectomy Breast Surgery Using Raman Spectroscopy [? Q1: Running head: Raman Margin Assessment at Partial Mastectomy. Short title OK? Q1]. *Cancer Research* **66**, 3317-3322 (2006).

241 Bitar, R. *et al.* in *Biomedical Optics (BiOS) 2008.* 68530O-68530O-68537 (International Society for Optics and Photonics).

242 Lieber, C. A., Majumder, S. K., Ellis, D. L., Billheimer, D. D. & Mahadevan-Jansen, A. In vivo nonmelanoma skin cancer diagnosis using Raman microspectroscopy. *Lasers in surgery and medicine* **40**, 461-467 (2008).

243 Kirsch, M., Schackert, G., Salzer, R. & Krafft, C. Raman spectroscopic imaging for in vivo detection of cerebral brain metastases. *Analytical and bioanalytical chemistry* **398**, 1707-1713 (2010).

244 Beljebbar, A., Dukic, S., Amharref, N. & Manfait, M. Ex vivo and in vivo diagnosis of C6 glioblastoma development by Raman spectroscopy coupled to a microprobe. *Anal Bioanal Chem* **398**, 477-487, doi:10.1007/s00216-010-3910-6 (2010).

245 Bergholt, M. S. *et al.* Raman endoscopy for in vivo differentiation between benign and malignant ulcers in the stomach. *The Analyst* **135**, 3162-3168 (2010).

246 Bergholt, M. S. *et al.* In vivo diagnosis of gastric cancer using Raman endoscopy and ant colony optimization techniques. *International journal of cancer* **128**, 2673-2680 (2011).

247 Bergholt, M. S. *et al.* Fiber-optic Raman spectroscopy probes gastric carcinogenesis in vivo at endoscopy. *J Biophotonics* **6**, 49-59 (2013).

248 Huang, Z. *et al.* In vivo early diagnosis of gastric dysplasia using narrow-band image-guided Raman endoscopy. *Journal of biomedical optics* **15**, 037017-037017-037015 (2010).

249 Huang, Z. *et al.* In vivo detection of epithelial neoplasia in the stomach using image-guided Raman endoscopy. *Biosensors and Bioelectronics* **26**, 383-389 (2010).

250 Duraipandian, S. *et al.* Real-time Raman spectroscopy for in vivo, online gastric cancer diagnosis during clinical endoscopic examination. *Journal of biomedical optics* **17**, 081418-081418 (2012).

251 Day, J. *et al.* A miniature confocal Raman probe for endoscopic use. *Physics in medicine and biology* **54**, 7077 (2009).

252 Short, M. A. *et al.* Development and preliminary results of an endoscopic Raman probe for potential in vivo diagnosis of lung cancers. *Optics letters* **33**, 711-713 (2008).

253 Robichaux-Viehoever, A. *et al.* Characterization of Raman spectra measured in vivo for the detection of cervical dysplasia. *Applied spectroscopy* **61**, 986-993 (2007).

254 Kanter, E. M. *et al.* Application of Raman spectroscopy for cervical dysplasia diagnosis. *Journal of biophotonics* **2**, 81 (2009).

255 Kanter, E. M. *et al.* Multiclass discrimination of cervical precancers using Raman spectroscopy. *Journal of Raman Spectroscopy* **40**, 205-211 (2009).

256 Richards-Kortuml, R. Rapid Communication Development of a Fiber Optic Probe to Measure NIR Raman Spectra of Cervical Tissue In Vivo. *Photochemistry and photobiology* **68**, 427-431 (1998).

257 Kanter, E. M., Majumder, S., Kanter, G. J., Woeste, E. M. & Mahadevan-Jansen, A. Effect of hormonal variation on Raman spectra for cervical disease detection. *American journal of obstetrics and gynecology* **200**, 512. e511-512. e515 (2009).

258 Duraipandian, S. *et al.* Non-invasive analysis of hormonal variations and effect of postmenopausal Vagifem treatment on women using in vivo high wavenumber confocal Raman spectroscopy. *The Analyst* **138**, 4120-4128 (2013).

259 Duraipandian, S. *et al.* Near-infrared-excited confocal Raman spectroscopy advances in vivo diagnosis of cervical precancer. *Journal of biomedical optics* **18**, 067007-067007 (2013).

260 Buckley, K. *et al.* Measurement of abnormal bone composition in vivo using noninvasive Raman spectroscopy. *IBMS BoneKEy* **11** (2014).

261 Shaikh, R. *et al.* In vivo Raman spectroscopy of human uterine cervix: exploring the utility of vagina as an internal control. *Journal of biomedical optics* **19**, 087001-087001 (2014).

262 Bergholt, M. S. *et al.* Characterizing Variability of In Vivo Raman Spectroscopic Properties of Different Anatomical Sites of Normal Colorectal Tissue towards Cancer Diagnosis at Colonoscopy. *Analytical chemistry* **87**, 960-966 (2014).

263 Bergholt, M. S. *et al.* Simultaneous fingerprint and high-wavenumber fiber-optic Raman spectroscopy enhances real-time in vivo diagnosis of adenomatous polyps during colonoscopy. *Journal of biophotonics* **9999** (2015).

264 Jermyn, M. *et al.* Intraoperative brain cancer detection with Raman spectroscopy in humans. *Science translational medicine* **7**, 274ra219-274ra219 (2015).

265 Desroches, J. *et al.* Characterization of a Raman spectroscopy probe system for intraoperative brain tissue classification. *Biomed Opt Express* **6**, 2380-2397, doi:10.1364/BOE.6.002380 (2015).

266 Bakker Schut, T. *et al.* In vivo detection of dysplastic tissue by Raman spectroscopy. *Analytical chemistry* **72**, 6010-6018 (2000).

267 Venkatakrishna, K. *et al.* Optical pathology of oral tissue: a Raman spectroscopy diagnostic method. *Current science* **80**, 665-669 (2001).

268 Krishna, C. M. *et al.* Micro-Raman spectroscopy for optical pathology of oral squamous cell carcinoma. *Applied spectroscopy* **58**, 1128-1135 (2004).

269 Oliveira, A. P., Bitar, R. A., Silveira Jr, L., Zângaro, R. A. & Martin, A. A. Near-infrared Raman spectroscopy for oral carcinoma diagnosis. *Photomedicine and Laser Therapy* **24**, 348-353 (2006).

270 Hu, Y., Jiang, T. & Zhao, Z. in *Intelligent Networks and Intelligent Systems, 2008. ICINIS'08. First International Conference on.* 633-636 (IEEE).

271 Sunder, N., Rao, N., Kartha, V., Ullas, G. & Kurien, J. Laser raman spectroscopy: A novel diagnostic tool for oral cancer. *Journal of Orofacial Sciences* **3**, 15 (2011).

272 Bergholt, M. S., Zheng, W. & Huang, Z. Characterizing variability in in vivo Raman spectroscopic properties of different anatomical sites of normal tissue in the oral cavity. *Journal of Raman Spectroscopy* **43**, 255-262 (2012).

273 Singh, S., Deshmukh, A., Chaturvedi, P. & Krishna, C. M. Raman spectroscopy in head and neck cancers: Toward oncological applications. *Journal of cancer research and therapeutics* **8**, 126 (2012).

274 Deshmukh, A., Singh, S., Chaturvedi, P. & Krishna, C. M. Raman spectroscopy of normal oral buccal mucosa tissues: study on intact and incised biopsies. *Journal of biomedical optics* **16**, 127004-12700410 (2011).

275 Singh, S., Sahu, A., Deshmukh, A., Chaturvedi, P. & Krishna, C. M. In vivo Raman spectroscopy of oral buccal mucosa: a study on malignancy associated changes (MAC)/cancer field effects (CFE). *The Analyst* **138**, 4175-4182 (2013).

276 Krishna, H., Majumder, S. K., Chaturvedi, P., Sidramesh, M. & Gupta, P. K. In vivo Raman spectroscopy for detection of oral neoplasia: A pilot clinical study. *Journal of biophotonics* **7**, 690-702 (2014).

277 Singh, S. & Krishna, C. M. Raman spectroscopic studies of oral cancers: correlation of spectral and biochemical markers. *Analytical Methods* **6**, 8613-8620 (2014).

278 Behl, I. *et al.* Raman mapping of oral buccal mucosa: a spectral histopathology approach. *Journal of biomedical optics* **19**, 126005-126005 (2014).

279 Cals, F. L. *et al.* Investigation of the potential of Raman spectroscopy for oral cancer detection in surgical margins. *Laboratory Investigation* (2015).

280 Carvalho, L. F. C. *et al.* Raman micro-spectroscopy for rapid screening of oral squamous cell carcinoma. *Experimental and molecular pathology* **98**, 502-509 (2015).

281 Carvalho, L. F. C. *et al.* in *SPIE Biophotonics South America.* 953125-953125-953127 (International Society for Optics and Photonics).

282 Kumar, P., Bhattacharjee, T., Ingle, A., Maru, G. & Krishna, C. M. Raman spectroscopy of experimental oral carcinogenesis: Study on sequential cancer progression in hamster buccal pouch model. *Technol Cancer Res Treat* **In Press** (2015).

283 Kho, K. W., Malini, O., Shen, Z. X. & Soo, K. C. in *Biomedical Optics 2005.* 84-91 (International Society for Optics and Photonics).

284 Rekha, P. *et al.* in *Photonics (ICP), 2013 IEEE 4th International Conference on.* 135-137 (IEEE).

285 Elumalai, B., Prakasarao, A., Ganesan, B., Dornadula, K. & Ganesan, S. Raman spectroscopic characterization of urine of normal and oral cancer subjects. *Journal of Raman Spectroscopy* **46**, 84-93 (2015).

286 Gadducci, A., Cosio, S., Carpi, A., Nicolini, A. & Genazzani, A. R. Serum tumor markers in the management of ovarian, endometrial and cervical cancer. *Biomedicine & Pharmacotherapy* **58**, 24-38 (2004).

287 Sturgeon, C. M. *et al.* National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers. *Clinical chemistry* **54**, e11-e79 (2008).

288 Meany, D. L., Sokoll, L. J. & Chan, D. W. Early detection of cancer: immunoassays for plasma tumor markers. *Expert opinion on medical diagnostics* **3**, 597-605 (2009).

289 Bhatt, A. N., Mathur, R., Farooque, A., Verma, A. & Dwarakanath, B. Cancer biomarkers-Current perspectives. (2010).

290 Menon, U. *et al.* Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *The lancet oncology* **10**, 327-340 (2009).

291 Daniele, B., Bencivenga, A., Megna, A. S. & Tinessa, V. α-fetoprotein and ultrasonography screening for hepatocellular carcinoma. *Gastroenterology* **127**, S108-S112 (2004).

292 Harris, R. & Lohr, K. N. Screening for prostate cancer: an update of the evidence for the US Preventive Services Task Force. *Annals of internal medicine* **137**, 917-929 (2002).

293 Roschewski, M. *et al.* Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study. *The Lancet Oncology* **16**, 541-549 (2015).

294 Hou, L. *et al.* Blood Telomere Length Attrition and Cancer Development in the Normative Aging Study Cohort. *EBioMedicine* (2015).

295 Dhakar, N., Astekar, M., Jain, M., Saawarn, S. & Saawarn, N. Total sialic acid, total protein and total sugar levels in serum and saliva of oral squamous cell carcinoma patients: A case control study. *Dental research journal* **10**, 343 (2013).

296 Singh, S., Ramesh, V., Premalatha, B., Prashad, K. V. & Ramadoss, K. Alterations in serum lipid profile patterns in oral cancer. *Journal of natural science, biology, and medicine* **4**, 374 (2013).

297 Schwarzenbach, H., Hoon, D. S. & Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nature Reviews Cancer* **11**, 426-437 (2011).

298 Gormally, E. *et al.* Amount of DNA in plasma and cancer risk: a prospective study. *International journal of cancer* **111**, 746-749 (2004).

299 Aravindh, L. *et al.* Estimation of plasma antioxidants beta carotene, vitamin c and vitamin e levels in patients with osmf and oral cancer-indian population. (2012).

300 Toniolo, P. *et al.* Serum carotenoids and breast cancer. *American Journal of Epidemiology* **153**, 1142-1147 (2001).

301 Lai, H.-S., Lee, J.-C., Lee, P.-H., Wang, S.-T. & Chen, W.-J. in *Seminars in cancer biology.* 267-276 (Elsevier).

302 Al-Salhi, M., Masilamani, V., Vijmasi, T., Al-Nachawati, H. & VijayaRaghavan, A. Lung cancer detection by native fluorescence spectra of body fluids—a preliminary study. *Journal of fluorescence* **21**, 637-645 (2011).

303 Madhuri, S. *et al.* Native Fluorescence Spectroscopy of Blood Plasma in the Characterization of Oral Malignancy¶. *Photochemistry and photobiology* **78**, 197-204 (2003).

304 Kalaivani, R. *et al.* Fluorescence spectra of blood components for breast cancer diagnosis. *Photomedicine and Laser Surgery* **26**, 251-256 (2008).

305 Burns, D. H. *et al.* Near-infrared spectroscopy of blood plasma for diagnosis of sporadic Alzheimer's disease. *Journal of Alzheimer's disease: JAD* **17**, 391-397 (2008).

306 Gajjar, K. *et al.* Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer. *The Analyst* **138**, 3917-3926 (2013).

307 Nahavandi, M., Nichols, J., Hassan, M., Gandjbakhche, A. & Kato, G. Near-infrared spectra absorbance of blood from sickle cell patients and normal individuals. *Hematology* **14**, 46-48 (2009).

308 Erukhimovitch, V., Talyshinsky, M., Souprun, Y. & Huleihel, M. FTIR spectroscopy examination of leukemia patients plasma. *Vibrational Spectroscopy* **40**, 40-46 (2006).

309 Ahmed, S. S., Santosh, W., Kumar, S. & Christlet, T. H. T. Neural network algorithm for the early detection of Parkinson's disease from blood plasma by FTIR micro-spectroscopy. *Vibrational Spectroscopy* **53**, 181-188 (2010).

310 Backhaus, J. *et al.* Diagnosis of breast cancer with infrared spectroscopy from serum samples. *Vibrational Spectroscopy* **52**, 173-177 (2010).

311 Peto, R., Doll, R., Buckley, J. D. & Sporn, M. Can dietary beta-carotene materially reduce human cancer rates? (1981).

312 Wald, N., Thompson, S., Densem, J., Boreham, J. & Bailey, A. Serum beta-carotene and subsequent risk of cancer: results from the BUPA Study. *British Journal of Cancer* **57**, 428 (1988).

313 Yuan, J.-M., Ross, R. K., Chu, X.-D., Gao, Y.-T. & Mimi, C. Y. Prediagnostic levels of serum β-cryptoxanthin and retinol predict smoking-related lung cancer risk in Shanghai, China. *Cancer Epidemiology Biomarkers & Prevention* **10**, 767-773 (2001).

314 Ray, G. & Husain, S. A. Oxidants, antioxidants and carcinogenesis. *Indian journal of experimental biology* **40**, 1213-1232 (2002).

315 Borek, C. Dietary antioxidants and human cancer. *Integrative cancer therapies* **3**, 333-341 (2004).

316 Chippagiri, P., Ali, I. & Banavar, S. R. Role of Serum ß-Carotene in the Diagnosis and Prevention of Oral Squamous Cell Carcinoma–A Case Control Study. *Journal of clinical and diagnostic research: JCDR* **8**, ZC01 (2014).

317 Aggarwal, A., Shetti, A., Keluskar, V. & Bagewadi, A. Estimation of serum beta carotene levels in patients with oral submucous fibrosis in India. *Journal of oral science* **53**, 427-431 (2011).

318 Tasumi, M., Inagaki, F. & Miyazawa, T. Resonance Raman spectra of β-carotene and its molecular structure in the excited electronic state. *Chemical Physics Letters* **22**, 515-518 (1973).

319 Inagaki, F., Tasumi, M. & Miyazawa, T. Excitation profile of the resonance Raman effect of β-carotene. *Journal of Molecular Spectroscopy* **50**, 286-303 (1974).

320 Darvin, M. *et al.* Determination of beta carotene and lycopene concentrations in human skin using resonance Raman spectroscopy. *Laser Phys* **15**, 295-299 (2005).

321 Sufra, S., Dellepiane, G., Masetti, G. & Zerbi, G. Resonance Raman spectrum of β-carotene. *Journal of Raman Spectroscopy* **6**, 267-272 (1977).

322 Nijssen, A. *et al.* Discriminating basal cell carcinoma from perilesional skin using high wave-number Raman spectroscopy. *Journal of biomedical optics* **12**, 034004-034004-034007 (2007).

323 Koljenović, S., Schut, T. C. B., Kros, J. M., van den Berge, H. J. & Puppels, G. J. Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy. *Laboratory Investigation* **82**, 1265-1277 (2002).

324 Crow, P. *et al.* The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines. *British journal of cancer* **92**, 2166-2170 (2005).

325 Harvey, T. *et al.* Factors influencing the discrimination and classification of prostate cancer cell lines by FTIR microspectroscopy. *The Analyst* **134**, 1083-1091 (2009).

326 Kelly, J. G. *et al.* Biospectroscopy to metabolically profile biomolecular structure: a multistage approach linking computational analysis with biomarkers. *Journal of proteome research* **10**, 1437-1448 (2011).

327 Ghanate, A., Kothiwale, S., Singh, S., Bertrand, D. & Krishna, C. M. Comparative evaluation of spectroscopic models using different multivariate statistical tools in a multicancer scenario. *Journal of Biomedical Optics* **16**, 025003-025003-025009 (2011).

328 PALAN, P. R., ROMNEY, S. L., MIKHAIL, M., BASU, J. & VERMUND, S. H. Decreased plasma β-carotene levels in women with uterine cervical dysplasias and cancer. *Journal of the National Cancer Institute* **80**, 454-455 (1988).

329 Parker, F. S. *Applications of infrared, Raman, and resonance Raman spectroscopy in biochemistry*. (Springer Science & Business Media, 1983).

330 Movasaghi, Z., Rehman, S. & Rehman, I. U. Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews* **42**, 493-541 (2007).

331 Sahu, A., Sawant, S., Mamgain, H. & Krishna, C. M. Raman spectroscopy of serum: an exploratory study for detection of oral cancers. *The Analyst* **138**, 4161-4174, doi:10.1039/c3an00308f (2013).

332 Sahu, A., Sawant, S., Talathi-Desai, S. & Murali Krishna, C. Raman spectroscopy of serum: A study on oral cancers. *Biomedical Spectroscopy and Imaging* **4**, 171-187 (2015).

333 Singh, A. P., Kumar, N., Raju, M. S., Singh, N. N. & Nagendrareddy, S. G. Estimation of serum β2-microglobulin in potentially malignant disorders and squamous cell carcinoma of the oral cavity: A clinicopathological study. *Dental research journal* **11**, 109 (2014).

334 Kamath, V., Satelur, K. & Komali, Y. Biochemical markers in oral submucous fibrosis: A review and update. *Dental research journal* **10**, 576 (2013).

335 Pereira, T., Shetty, S. & Pereira, S. Estimation of serum lactate dehydrogenase level in patients with oral premalignant lesions/conditions and oral squamous cell carcinoma: A clinicopathological study. *Journal of cancer research and therapeutics* **11**, 78 (2015).

336 Iwamoto, F. M. *et al.* Serum YKL-40 is a marker of prognosis and disease status in high-grade gliomas. *Neuro-oncology* **13**, 1244-1251 (2011).

337 Radić, B., Vukelić, Ž. & Kalanj Bognar, S. Serum gangliosides in patients with brain tumors. *Collegium antropologicum* **32**, 171-175 (2008).

338 Arslan, M., Demir, H., Arslan, H., Gokalp, A. S. & Demir, C. Trace elements, heavy metals and other biochemical parameters in malignant glioma patients. *Asian Pac J Cancer Prev* **12**, 447-451 (2011).

339 Popescu, I. D. *et al.* Potential serum biomarkers for glioblastoma diagnostic assessed by proteomic approaches. *Proteome science* **12**, 1-14 (2014).

340 da Silva, S. D., Hier, M., Mlynarek, A., Kowalski, L. P. & Alaoui-Jamali, M. A. Recurrent oral cancer: current and emerging therapeutic approaches. *Frontiers in pharmacology* **3** (2012).

341 Kostakoglu, L. *et al.* Early detection of recurrent disease by FDG-PET/CT leads to management changes in patients with squamous cell cancer of the head and neck. *The oncologist* **18**, 1108-1117 (2013).

342 Reis, P. P. *et al.* A gene signature in histologically normal surgical margins is predictive of oral carcinoma recurrence. *BMC cancer* **11**, 437 (2011).

343 Graveland, A. P. *et al.* Loss of heterozygosity at 9p and p53 immunopositivity in surgical margins predict local relapse in head and neck squamous cell carcinoma. *International journal of cancer* **128**, 1852-1859 (2011).

344 Rosin, M. P. *et al.* 3p14 and 9p21 loss is a simple tool for predicting second oral malignancy at previously treated oral cancer sites. *Cancer research* **62**, 6447-6450 (2002).

345 Chung, C. H. *et al.* Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer cell* **5**, 489-500 (2004).

346 Lumachi, F., Basso, S. M. & Basso, U. in *Methods of Cancer Diagnosis, Therapy and Prognosis* 109-115 (Springer, 2008).

347 Han, M. *et al.* Serum acid phosphatase level and biochemical recurrence following radical prostatectomy for men with clinically localized prostate cancer. *Urology* **57**, 707-711 (2001).

348 Partin, A. W. *et al.* Prostate-Specific Antigen as a Marker of Disease Activity in Prostate Cancer: Part 2. *Oncology* **16**, 1 (2002).

349 Wallace, T. *et al.* Current approaches, challenges and future directions for monitoring treatment response in prostate cancer. *Journal of Cancer* **5**, 3 (2014).

350 Tan, E. *et al.* Diagnostic precision of carcinoembryonic antigen in the detection of recurrence of colorectal cancer. *Surgical oncology* **18**, 15-24 (2009).

351 Lai, Q. *et al.* Alpha-fetoprotein and novel tumor biomarkers as predictors of hepatocellular carcinoma recurrence after surgery: a brilliant star raises again. *International journal of hepatology* **2012** (2012).

352 Widschwendter, A. *et al.* Human papillomavirus DNA in sera of cervical cancer patients as tumor marker. *Cancer letters* **202**, 231-239 (2003).

353 Hao, T. *et al.* Circulating cell-free DNA in serum as a biomarker for diagnosis and prognostic prediction of colorectal cancer. *British journal of cancer* **111**, 1482-1489 (2014).

354 Behne, T. & Copur, M. S. Biomarkers for hepatocellular carcinoma. *International journal of hepatology* **2012** (2012).

355 Shotelersuk, K. *et al.* Epstein-Barr virus DNA in serum/plasma as a tumor marker for nasopharyngeal cancer. *Clinical Cancer Research* **6**, 1046-1051 (2000).

356 Ahn, S. M. *et al.* Saliva and Plasma Quantitative Polymerase Chain Reaction–Based Detection and Surveillance of Human Papillomavirus–Related Head and Neck Cancer. *JAMA Otolaryngology–Head & Neck Surgery* **140**, 846-854 (2014).

357 Chen, L.-M. *et al.* External validation of a multiplex urinary protein panel for the detection of bladder cancer in a multicenter cohort. *Cancer Epidemiology Biomarkers & Prevention* **23**, 1804-1812 (2014).

358 Long, Q. *et al.* Protein-coding and microRNA biomarkers of recurrence of prostate cancer following radical prostatectomy. *The American journal of pathology* **179**, 46-54 (2011).

359 Chatterjee, M. *et al.* Tumor autoantibodies as biomarkers for predicting ovarian cancer recurrence. *Cancer biomarkers: section A of Disease markers* **11** (2012).

360 Braakhuis, B., Brakenhoff, R. & Leemans, C. R. Treatment choice for locally advanced head and neck cancers on the basis of risk factors: biological risk factors. *Annals of Oncology* **23**, x173-x177 (2012).

361 van Oijen, M. G. & Slootweg, P. J. Oral field cancerization: carcinogen-induced independent events or micrometastatic deposits? *Cancer Epidemiology Biomarkers & Prevention* **9**, 249-256 (2000).

362 Wang, Y. *et al.* Analysis of diferentially expressed protein from primary and recurrent ovarian cancer serum. *Asian Pacific journal of tropical medicine* **5**, 573-576 (2012).

363 Good, D. M. *et al.* Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future. *Journal of proteome research* **6**, 4549-4555 (2007).

364 Diniz-Freitas, M., García-García, A., Crespo-Abelleira, A., Martins-Carneiro, J. & Gándara-Rey, J. Applications of exfoliative cytology in the diagnosis of oral cancer. *Medicina oral: órgano oficial de la Sociedad Española de Medicina Oral y de la Academia Iberoamericana de Patología y Medicina Bucal* **9**, 355 (2004).

365 PAPANICOLAOU, G. N. A survey of the actualities and potentialities of exfoliative cytology in cancer diagnosis. *Annals of internal medicine* **31**, 661-674 (1949).

366 Papanicolaou, G. N. & Traut, H. F. Diagnosis of uterine cancer by the vaginal smear. *New York* **46** (1943).

367 Weinmann, J. The keratinization of the human oral mucosa. *Journal of Dental Research* **19**, 57-71 (1940).

368 D.E. Ziskin, P. K., I. Kitley. Epithelial smears from oral mucosa. *J Dent Res.* **20**, 386-387 (1941).

369 Papanicolaou, G. N. Atlas o. *Ex [oliative Cytology*, 13-21 (1954).

370 Montgomery, P. W. & Von Haam, E. A study of the exfoliative cytology in patients with carcinoma of the oral mucosa. *Journal of dental research* **30**, 308-313 (1951).

371 Sandler, H. & Stahl, S. S. Exfoliative cytology as a diagnostic aid in the detection of oral neoplasms. *Journal of oral surgery* **16**, 414 (1958).

372 Mehrotra, R., Hullmann, M., Smeets, R., Reichert, T. & Driemel, O. Oral cytology revisited. *Journal of Oral Pathology & Medicine* **38**, 161-166 (2009).

373 Proia, N. K., Paszkiewicz, G. M., Nasca, M. A. S., Franke, G. E. & Pauly, J. L. Smoking and smokeless tobacco-associated human buccal cell mutations and their association with oral cancer—a review. *Cancer Epidemiology Biomarkers & Prevention* **15**, 1061-1077 (2006).

374 Mehrotra, R. The role of cytology in oral lesions: A review of recent improvements. *Diagnostic cytopathology* **40**, 73-83 (2012).

375 Remmerbach, T. W. *et al.* Cytologic and DNA-cytometric early diagnosis of oral cancer. *Analytical Cellular Pathology* **22**, 211-221 (2001).

376 Maraki, D., Becker, J. & Boecking, A. Cytologic and DNA-cytometric very early diagnosis of oral cancer. *Journal of oral pathology & medicine* **33**, 398-404 (2004).

377 Remmerbach, T. W. *et al.* Toward a multimodal cell analysis of brush biopsies for the early detection of oral cancer. *Cancer Cytopathology* **117**, 228-235 (2009).

378 Remmerbach, T. W., Weidenbach, H., Hemprich, A. & Böcking, A. Earliest detection of oral cancer using non-invasive brush biopsy including DNA-image-cytometry: report on four cases. *Analytical Cellular Pathology* **25**, 159-166 (2003).

379 Remmerbach, T. W. *et al.* Diagnostic value of nucleolar organizer regions (AgNORs) in brush biopsies of suspicious lesions of the oral cavity. *Analytical cellular pathology* **25**, 139-146 (2003).

380 Wong, P., Wong, R. K., Caputo, T. A., Godwin, T. A. & Rigas, B. Infrared spectroscopy of exfoliated human cervical cells: evidence of extensive structural changes during carcinogenesis. *Proceedings of the National Academy of Sciences* **88**, 10988-10992 (1991).

381 Fung, M. F. K. *et al.* Comparison of Fourier-transform infrared spectroscopic screening of exfoliated cervical cells with standard Papanicolaou screening. *Gynecologic oncology* **66**, 10-15 (1997).

382 Wood, B. R. *et al.* FTIR microspectroscopic study of cell types and potential confounding variables in screening for cervical malignancies. *Biospectroscopy* **4**, 75-91 (1998).

383 Romeo, M. J., Wood, B. R., Quinn, M. A. & McNaughton, D. Removal of blood components from cervical smears: implications for cancer diagnosis using FTIR spectroscopy. *Biopolymers* **72**, 69-76 (2003).

384 Bonnier, F. *et al.* Processing ThinPrep cervical cytological samples for Raman spectroscopic analysis. *Analytical Methods* **6**, 7831-7841 (2014).

385 Solomon, D. The 1988 Bethesda system for reporting cervical/vaginal cytologic diagnoses: developed and approved at a National Cancer Institute workshop, Bethesda MD, USA, December 12-13, 1988. *Journal of reproductive medicine* **34**, 779-785 (1989).

386 Luff, R. D. The Bethesda system for reporting cervical/vaginal cytologic diagnoses: report of the 1991 Bethesda Workshop. *Human pathology* **23**, 719-721 (1992).

387 Koss, L. G. & Durfee, G. R. Diagnostic Cytology and its Histopathologic Bases. *The American Journal of the Medical Sciences* **242**, 790 (1961).

388 MacDonald, G. in *Satellite Symposium on Epithelial Dysplasia and Borderline Cancer of the Head and Neck: Controversies and Future Directions. Joint BSOMP, BSOM, BAHNO Meeting, Oxford.*

389 Boon, M. E. & Suurmeijer, A. J. *The pap smear*. (CRC Press, 1996).

390 Hoda, R. S. & Hoda, S. A. *Fundamentals of Pap test cytology*. (Springer Science & Business Media, 2007).

391 Stone, N., Kendall, C., Shepherd, N., Crow, P. & Barr, H. Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. *Journal of Raman spectroscopy* **33**, 564-573 (2002).

392 Offenbacher, A. R., Chen, J. & Barry, B. A. Perturbations of aromatic amino acids are associated with iron cluster assembly in ribonucleotide reductase. *Journal of the American Chemical Society* **133**, 6978-6988 (2011).

393 Kengne-Momo, R. *et al.* Protein interactions investigated by the Raman spectroscopy for biosensor applications. *International Journal of Spectroscopy* **2012** (2012).

394 Hu, Q. & Hou, H. *Tobacco Smoke Exposure Biomarkers*. (CRC Press, 2015).

395 Lin-Vien, D., Colthup, N. B., Fateley, W. G. & Grasselli, J. G. *The handbook of infrared and Raman characteristic frequencies of organic molecules*. (Elsevier, 1991).

396 Gunasekaran, S., Sailatha, E., Seshadri, S. & Kumaresan, S. FTIR, FT Raman spectra and molecular structural confirmation of isoniazid. *Indian J Pure Appl Phys* **47**, 12-18 (2009).

397 Matthews, Q., Jirasek, A., Lum, J., Duan, X. & Brolo, A. G. Variability in Raman spectra of single human tumor cells cultured in vitro: correlation with cell cycle and culture confluency. *Applied spectroscopy* **64**, 871-887 (2010).

398 Krishna, C. M. *et al.* Micro-Raman spectroscopy of mixed cancer cell populations. *Vibrational Spectroscopy* **38**, 95-100 (2005).

399 Slaughter, D. P., Southwick, H. W. & Smejkal, W. "Field cancerization" in oral stratified squamous epithelium. Clinical implications of multicentric origin. *Cancer* **6**, 963-968 (1953).

400 Khandelwal, S. & Solomon, M. C. Cytomorphological analysis of keratinocytes in oral smears from tobacco users and oral squamous cell carcinoma lesions—A histochemical approach. *International journal of oral science* **2**, 45 (2010).

401 Babshet, M., Nandimath, K., Pervatikar, S. & Naikmasur, V. Efficacy of oral brush cytology in the evaluation of the oral premalignant and malignant lesions. *Journal of cytology/Indian Academy of Cytologists* **28**, 165 (2011).

402 Gray, W. & Kocjan, G. *Diagnostic Cytopathology*. (Churchill Livingstone/Elsevier, 2010).

403 Mignogna, M. D., Fedele, S. & Russo, L. L. Dysplasia/neoplasia surveillance in oral lichen planus patients: a description of clinical criteria adopted at a single centre and their impact on prognosis. *Oral oncology* **42**, 819-824 (2006).

404 Sahiar, B., Daftary, D. & Mehta, F. S. Cytological and histological keratinization studies in leukoplakias of the mouth. *Journal of Oral Pathology & Medicine* **4**, 19-26 (1975).

405 GUJRAL, I., SINGH, P., SHARMA, S. & GANGANE, N. A CYTOLOGICAL STUDY OF LEUKOPLASTIC LESIONS IN ORAL CAVITY.

406 Verrier, S., Notingher, I., Polak, J. & Hench, L. In situ monitoring of cell death using Raman microspectroscopy. *Biopolymers* **74**, 157-162 (2004).

407 Notingher, I., Verrier, S., Haque, S., Polak, J. & Hench, L. Spectroscopic study of human lung epithelial cells (A549) in culture: living cells versus dead cells. *Biopolymers* **72**, 230-240 (2003).

408 Kunapareddy, N., Freyer, J. P. & Mourant, J. R. Raman spectroscopic characterization of necrotic cell death. *Journal of biomedical optics* **13**, 054002-054002-054009 (2008).

409 Short, K. W., Carpenter, S., Freyer, J. P. & Mourant, J. R. Raman spectroscopy detects biochemical changes due to proliferation in mammalian cell cultures. *Biophysical journal* **88**, 4274-4288 (2005).

410 Fu, D. *et al.* Imaging the intracellular distribution of tyrosine kinase inhibitors in living cells with quantitative hyperspectral stimulated Raman scattering. *Nature chemistry* **6**, 614-622 (2014).

411 Bonn, M., Muller, M., Rinia, H. A. & Burger, K. N. Imaging of chemical and physical state of individual cellular lipid droplets using multiplex CARS microscopy. *Journal of Raman Spectroscopy* **40**, 763-769 (2009).

412 Sathyan, K. *et al.* Carcinoma of tongue and the buccal mucosa represent different biological subentities of the oral carcinoma. *Journal of cancer research and clinical oncology* **132**, 601-609 (2006).

413 Freier, K. *et al.* Tissue microarray analysis reveals site-specific prevalence of oncogene amplifications in head and neck squamous cell carcinoma. *Cancer research* **63**, 1179-1182 (2003).

414 Schwartz, J., Gu, X., Kittles, R., Baptiste, A. & Shklar, G. Experimental oral carcinoma of the tongue and buccal mucosa: possible biologic markers linked to cancers at two anatomic sites. *Oral oncology* **36**, 225-235 (2000).

415 Sá, C. T., Fonseca, L. M., Cardoso, S. V., Aguiar, M. C. F. & Carmo, M. A. V. d. p53 immunoexpression in oral squamous cell carcinomas from different anatomical sites: a comparative study. *Int J Morphol* **24**, 231-238 (2006).

416 He, Q. Y., Chen, J., Kung, H. F., Yuen, A. P. W. & Chiu, J. F. Identification of tumor-associated proteins in oral tongue squamous cell carcinoma by proteomics. *Proteomics* **4**, 271-278 (2004).

417 Chen, J., He, Q. Y., Yuen, A. P. W. & Chiu, J. F. Proteomics of buccal squamous cell carcinoma: the involvement of multiple pathways in tumorigenesis. *Proteomics* **4**, 2465-2475 (2004).

418 Poh, C. F. *et al.* Biopsy and histopathologic diagnosis of oral premalignant and malignant lesions. *Journal of the Canadian Dental Association* **74**, 283-288 (2008).

419 Karabulut, A. *et al.* Observer variability in the histologic assessment of oral premalignant lesions. *Journal of oral pathology & medicine* **24**, 198-200 (1995).

420 DeCoro, M. & Wilder-Smith, P. Potential of optical coherence tomography for early diagnosis of oral malignancies. (2010).

421 Muldoon, T. J. *et al.* Noninvasive imaging of oral neoplasia with a high-resolution fiber-optic microendoscope. *Head & neck* **34**, 305-312 (2012).

422 Chaturvedi, P., Majumder, S. K., Krishna, H., Muttagi, S. & Gupta, P. K. Fluorescence spectroscopy for noninvasive early diagnosis of oral mucosal malignant and potentially malignant lesions. *Journal of cancer research and therapeutics* **6**, 497 (2010).

423 Lee, C.-K. *et al.* Diagnosis of oral precancer with optical coherence tomography. *Biomedical optics express* **3**, 1632-1646 (2012).

424 Sahu, A. *et al.* Raman spectroscopy of oral buccal mucosa: a study on age-related physiological changes and tobacco-related pathological changes. *Technology in cancer research & treatment* **11**, 529-541 (2012).

425 Pollock, R. E. & Doroshow, J. H. *UICC manual of clinical oncology*. (Wiley-liss, 2004).

426 Carnelio, S. & Rodrigues, G. Oral cancer at a glance. *The Internet Journal of Dental Science* **1** (2004).

427 Kirita, T. & Omura, K. *Oral Cancer: Diagnosis and Therapy*. (Springer, 2015).

428 Burket, L. W., Greenberg, M. S., Glick, M. & Ship, J. A. *Burket's oral medicine*. (PMPH-USA, 2008).

429 Boyd Gillespie, M. (Wiley Online Library, 2004).

430 German, R. Z. & Palmer, J. B. Anatomy and development of oral cavity and pharynx. *GI Motility online* (2006).

431 Winning, T. A. & Townsend, G. C. Oral mucosal embryology and histology. *Clinics in dermatology* **18**, 499-511 (2000).

432 de Veld, D. C. *et al.* Autofluorescence characteristics of healthy oral mucosa at different anatomical sites. *Lasers in surgery and medicine* **32**, 367-376 (2003).

433 McGee, S. *et al.* Anatomy-based algorithms for detecting oral cancer using reflectance and fluorescence spectroscopy. *Annals of Otology, Rhinology & Laryngology* **119**, 817-826 (2010).

434 McGee, S. *et al.* Model-based spectroscopic analysis of the oral cavity: impact of anatomy. *Journal of biomedical optics* **13**, 064034-064034-064015 (2008).

435 Baldwin, K. M. Wheater\'s Review of Histology and Basic Pathology. (2010).

436 Kumar, G. *Orban's Oral Histology & Embryology*. (Elsevier Health Sciences, 2014).

437 Nanci, A. *Ten Cate's Oral Histology-Pageburst on VitalSource: Development, Structure, and Function*. (Elsevier Health Sciences, 2007).

438 Squier, C. & Finkelstein, M. Oral mucosa. *Ten Cate's oral histology: development, structure, and function. 6th ed. St. Louis: Mosby*, 369-372 (2003).

439 Guze, K. *et al.* Pilot study: Raman spectroscopy in differentiating premalignant and malignant oral lesions from normal mucosa and benign lesions in humans. *Head & neck* **37**, 511-517 (2015).

440 Hershkovich, O., Shafat, I. & Nagler, R. M. Age-related changes in salivary antioxidant profile: possible implications for oral cancer. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **62**, 361-366 (2007).

441 Denison, M. E., Soto, L. & Reyes-Ortiz, C. A. The association between population aging and oral and pharynx cancer incidence in the American continent. *Colombia Médica* **42**, 191-198 (2011).

442 Baric, J. M., Alman, J. E., Feldman, R. S. & Chauncey, H. H. Influence of cigarette, pipe, and cigar smoking, removable partial dentures, and age on oral leukoplakia. *Oral Surgery, Oral Medicine, Oral Pathology* **54**, 424-429 (1982).

443 Pinzón, T. M. & Gaitán, C. L. [Aging and the oral cavity]. *Practica odontologica* **10**, 33-36 (1989).

444 Bäck, M., Hlawaty, H., Labat, C., Michel, J.-B. & Brink, C. The oral cavity and age: a site of chronic inflammation. *PLoS One* **2**, e1351 (2007).

445 Mckenna, G. & Burke, F. M. Age-related oral changes. *Dent Update* **37**, 519-523 (2010).

446 Scott, J., Valentine, J., St Hill, C. & Balasooriya, B. A quantitative histological analysis of the effects of age and sex on human lingual epithelium. *Journal de biologie buccale* **11**, 303-315 (1983).

447 Eid, R. A., Sawair, F., Landini, G. & Saku, T. Age and the architecture of oral mucosa. *Age* **34**, 651-658 (2012).

448 Shklar, G. The Effects Of Aging Upon Oral Mucosa1. *Journal of Investigative Dermatology* **47**, 115-120 (1966).

449 Hill, M. W., Squier, C. A., Squier, C. & Hill, M. *The Effect of Aging in Oral Mucosa and Skin*. (CRC Press, 1994).

450 Hill, M. Influence of age on the morphology and transit time of murine stratified squamous epithelia. *Archives of oral biology* **33**, 221-229 (1988).

451 Barakat, N. J., Toto, P. D. & Choukas, N. C. Aging and cell renewal of oral epithelium. *Journal of periodontology* **40**, 599 (1969).

452 Karring, T. & Löe, H. The effect of age on mitotic activity in rat oral epithelium. *Journal of periodontal research* **8**, 164-170 (1973).

453 Squier, C. A. & Kremer, M. J. Biology of oral mucosa and esophagus. *Journal of the National Cancer Institute. Monographs*, 7-15 (2000).

454 Cate-Ten, A. (Mosby-Year Book, Saint Louis, 1998).

455 Nakagawa, K., Sakurai, K., UEDA-KODAIRA, Y. & Ueda, T. Age-related changes in elastic properties and moisture content of lower labial mucosa. *Journal of oral rehabilitation* **38**, 235-241 (2011).

456 SYRJÄNEN, S. Age-related changes in structure of labial minor salivary glands. *Age and ageing* **13**, 159-165 (1984).

457 Miles, A. 'Sans teeth': changes in oral tissues with advancing age. *Proceedings of the Royal Society of Medicine* **65**, 801 (1972).

458 Bottomley, W. Physiology of the oral mucosa. *Otolaryngologic clinics of North America* **12**, 15 (1979).

459 Pickett, H. G., Appleby, R. G. & Osborn, M. O. Changes in the denture supporting tissues associated with the aging process. *The Journal of prosthetic dentistry* **27**, 257-262 (1972).

Prints and re-prints