Synopsis
#### SYNOPSIS

1. **Name of the Student:** Surya Pratap Singh  
2. **Name of the Constituent Institution:** Tata Memorial Centre  
3. **Enrolment No.:** LIFE09200804011  
4. **Title of Thesis:** Development of *in vivo* Raman Diagnostic methodologies for oral pre-cancers and cancers  
5. **Board of Studies:** Life Sciences
INTRODUCTION

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

Clinical examination and biopsies followed by histopathological analysis is considered as the gold standard for diagnosis and surveillance of oral cancer. However, the method has several limitations such as: inability in screening and detecting early malignancy associated changes; difficulty in recognizing subtle clinical changes in precancerous lesions or in a normal mucosa that
are indicative of neoplastic transformation; inability in distinguishing premalignant lesions from more common benign or inflammatory conditions; clinical or histological risk stratification lacks accuracy, reproducibility and requires large experience on part of the clinician [14-22]. Moreover, surveillance and biopsy of precancers is a mammoth task especially in populous countries like India. For example incidence of leukoplakia itself is up to 1% of general population. Considering these facts, it is imperative to develop a new rapid and accurate diagnostic method for early oral cancer detection.

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

Raman effect is based on inelastic scattering of photons. Unlike FT-IR, Raman spectroscopy does not suffer from water interference. However, week signals, as only very small fractions of photons (1 in 10^8) are inelastically scattered, is a major drawback associated with this technique [29]. Raman spectra of biological tissues are also often swamped by parasitic fluorescence. However, latest developments in light sources (lasers) and detectors (CCDs) have made Raman spectroscopy of biological samples like tissues and cells feasible. Further, use of near infrared excitation photons e.g. 785, 830 or 850 nm make this technique less harmful and also minimize the associated fluorescence. Most important attribute of Raman spectroscopy lies in its adaptation for *in vivo* applications. Using optical fibers laser light can be delivered to the desired site and Raman photons can also be collected. In view of above attributes Raman spectroscopy is projected as an ideal tool in pursuing biomedical applications. Raman spectroscopic differentiation of *ex vivo* normal and pathological conditions of oral, breast, cervix, colon, stomach ovarian and other forms of cancers have already been reported in the literature [30-38]. *In vivo* Raman measurements from bladder and prostate, oesophagus, skin, cervix and arteries are also reported [39-46].

**RATIONALE AND OBJECTIVES:**

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

Following are the specific objectives of the thesis -
Synopsis

1. Standardization of data acquisition and analysis methods on *ex vivo* oral tissues and correlation with histopathology and biochemical estimation.

2. To demonstrate feasibility of acquiring and classifying *in vivo* Raman spectra from buccal mucosa of normal, cancerous and pre-cancerous subjects and correlation with histopathology.

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

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

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

**Instrument details and spectral acquisition**

Spectra were acquired using HE-785 commercial Raman spectrometer (Jobin-Yvon-Horiba, France). Briefly, this system consists of a diode laser (Process Instruments) of 785 nm wavelength as excitation source, and a HE-785 spectrograph coupled with a CCD (Synapse) as dispersion and detection elements. The spectrograph is equipped with a fixed 950 gr/mm grating and spectral resolution, as specified by manufacturer, is \(\sim 4\, \text{cm}^{-1}\). Commercially available InPhotonics (Inc, Downy St. USA) probe consisting of 105 μm excitation fiber and 200 μm collection fiber (NA-
Synopsis

0.40) was used to couple excitation source and detection system. As per specifications of manufacturer of the Inphotonics probe, theoretical spot size and depth of field are 105 μm and 1mm, respectively. A XYZ precision stage along with a probe holder was assembled to record spectra of ex vivo tissues. During spectral acquisition, biopsy samples were placed on CaF2 window and mounted on XYZ translational stage. Spectra were recorded with a spacing of ~2 mm in XY direction with parameters: laser power-80 mW, integration-10 seconds and 5-accumulations. These parameters were kept constant during all measurements.

Spectral pre-processing and multivariate analysis

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

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

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

**Development of standard model and evaluation with independent test data**

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

**Correlation with band intensity and biochemical estimations**

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

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

Study on origin of spectral features in normal oral tissues

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

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

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

Spectra from contralateral normal were dominated by lipid features indicated by C=O band of esters, strong δCH₂ bend, two sharp bands in amide III region, and a sharp peak in amide I region. Predominant protein bands indicated by broad amide III, broad and shifted δCH₂, and broad amide I were observed in mean tumor spectra. These findings corroborate earlier reports of ex-vivo and in
Synopsis

*vivo* conditions [31,33,51,52]. Spectra from premalignant patches show features associated with tobacco induced hypercellularity and profile similar to tumor spectra in amide III, amide I, and δCH₂ regions.

**Development of standard models and evaluation with independent test data**

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

Influence of variability in tumor grade and differentiation status on classification was also explored. Findings suggest that it has no influence on classification with normal or precancerous conditions. Misclassifications between different groups can be primarily attributed to mucosal heterogeneity. Spectra from tumors gave best prediction efficiency (86%) followed by contralateral normal (79%) and premalignant (60%). Misclassification of tumor spectra as contralateral normal can be explained on the basis of fact that spectra were recorded at different points therefore, possibility of acquiring spectra from normal or inflammatory patches in a tumor cannot be completely ruled out. Maximum misclassification was observed between contralateral normal and premalignant spectra. This is probably due to the fact that premalignant patches in the study were from contralateral side.
Synopsis

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

Study on classification of different premalignant lesions in oral cavity

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

*Study on cancer field effects or malignancy associated changes*

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

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

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

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

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

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

**Cell line and spectral acquisition**

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

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

Multivariate analysis

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

Raman micro-spectroscopic studies

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

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

**Summary and conclusions**

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

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

**References:**

Synopsis


Synopsis


Synopsis

List of publications


Surya Pratap Singh
Synopsis

Doctoral Committee

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Name</th>
<th>Designation</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Dr. S. M. Zingde</td>
<td>Chairman</td>
<td></td>
<td>6/1/13</td>
</tr>
<tr>
<td>2.</td>
<td>Dr. C. Murali Krishna</td>
<td>Convener/Guide</td>
<td></td>
<td>6/8/13</td>
</tr>
<tr>
<td>3.</td>
<td>Dr. G. Krishnamoorthy</td>
<td>Member</td>
<td></td>
<td>7/1/13</td>
</tr>
<tr>
<td>4.</td>
<td>Dr. G.B. Maru</td>
<td>Member</td>
<td></td>
<td>6/15/13</td>
</tr>
<tr>
<td>5.</td>
<td>Dr. Pankaj Chaturvedi</td>
<td>Member</td>
<td></td>
<td>6/1/12</td>
</tr>
</tbody>
</table>

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

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

Dr. R. Sarin
(Director)
ACTREC

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