

**A STUDY OF BREAST CANCER PROGRESSION  
IN RODENT MODELS USING RAMAN  
SPECTROSCOPY**

**By**

**TANMOY BHATTACHARJEE**

**LIFE09200904015**

**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**



**December, 2015**

# HOMI BHABHA NATIONAL INSTITUTE

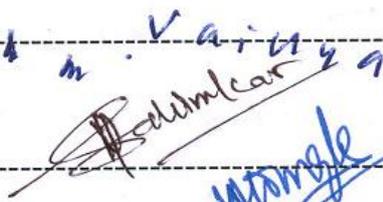
## Recommendations of the Viva Voce Board

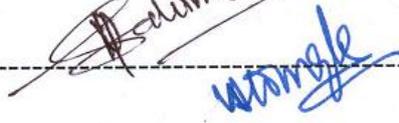
As members of the Viva Voce Board, we certify that we have read the dissertation prepared by Tanmoy Bhattacharjee entitled "A study of breast cancer progression in rodent models using Raman spectroscopy" and recommend that it may be accepted as fulfilling the dissertation requirements for the Degree of Doctor of Philosophy.

-----  
Chairperson: --Prof. R. Sarin  Date: 7/12/15

-----  
Convener: -- Dr. C. Murali Krishna  Date: 7.12.15

-----  
Member: -- Dr. M. Vaidya

-----  
Member: -- Dr. M. Mahimkar  Date: 07.12.15

-----  
Technical advisor: -- Dr. A. Ingle  Date: 7/12/2015  
07-12-15

-----  
Invitee: -- Dr. S. Zingde  Date: 7/12/15

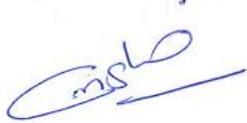
-----  
External Examiner: -- Prof. A. Roy  Date: 07/12/2015

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

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

Date: 07/12/2015

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 source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Navi Mumbai,

Tanmoy Bhattacharjee

Date:

## **DECLARATION**

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

Navi Mumbai,

August 2015.

Tanmoy Bhattacharjee

# List of Publications arising from the thesis

## Journal

1. “*Ex vivo* Raman spectroscopic study of breast metastatic lesions in lungs in animal models”,  
**T. Bhattacharjee**, S. Tawde, R. Hudlikar, M. Mahimkar, G. Maru, A. Ingle, C. Murali Krishna, J. Biomed. Opt., 2015, under revision.
2. “Transcutaneous *in vivo* Raman spectroscopy of breast tumors and pretumors”,  
**T. Bhattacharjee**, G. Maru, A. Ingle, C. Murali Krishna, Journal of Raman Spectroscopy, 2015, DOI 10.1002/jrs.4739.
3. “Raman spectroscopy of urine: Diagnosis of breast cancer in animal model”,  
**T. Bhattacharjee**, A. Khan, A. Ingle, G. Maru, C. Murali Krishna, Analyst, 2015, 140, 456-466 (selected as a ‘Hot Article’ by Analyst, Royal Society of Chemistry).
4. “Raman spectroscopy of serum: A study on ‘pre’ and ‘post’ breast adenocarcinoma resection in rat models”,  
**T. Bhattacharjee**, A. Khan, P. Kumar, A. Ingle, G. Maru, C. Murali Krishna, J. Biophotonics, 2015, 8 (7), 575-583.
5. “Swiss Bare mice: A suitable model for transcutaneous *in vivo* Raman spectroscopic studies of breast cancer”,  
**T. Bhattacharjee**, Piyush Kumar, G. Maru, A. Ingle, C. Murali Krishna, Lasers. Med. Sci., 2014, 29 (1), 325 -333.
6. “Transcutaneous *in vivo* Raman spectroscopic studies in mouse model: Evaluation of changes in breast associated with pregnancy and lactation”,

**T. Bhattacharjee**, G. Maru, A. Ingle, C. Murali Krishna, J. Biomed. Opt., 2013, 18(4), 047004.

7. “Transcutaneous *in vivo* Raman spectroscopy: Detection of age-related changes in mouse breast”,

**T. Bhattacharjee**, G. Maru, A. Ingle, C. Murali Krishna, Vibrational Spectroscopy, 2013, 67, 80–86.

### **Books:**

1. “Raman spectroscopy and its applications in breast cancer management”,

**T. Bhattacharjee**, C. Murali Krishna, 2015, (synopsis accepted for **SPIE eBook** series, Full draft under revision)

### **Conferences:**

1. 31st Annual convention of Indian Association for Cancer Research & International symposium, 26-29<sup>th</sup> January, 2012, Navi Mumbai
2. International Conference of Raman spectroscopy, 12-17<sup>th</sup> August, 2012, Bangalore
3. Global Cancer Genomics Consortium, 19-20<sup>th</sup> November, 2012, Navi Mumbai

### **Others**

1. “Raman spectroscopy of experimental oral carcinogenesis: Study on sequential cancer progression in hamster buccal pouch model”,

P. Kumar, **T. Bhattacharjee**, A. Ingle, G.B. Maru, C. Murali Krishna, Technologies in Cancer Research and Therapeutics, 2015, accepted.

Tanmoy Bhattacharjee

# **Dedication**

I dedicate this thesis to my parents,

Karabi Bhattacharjee and Tapobrata Bhattacharjee

## **ACKNOWLEDGEMENTS**

I would like to thank my guide, Dr. C. Murali Krishna for providing me a research platform and key inputs, the Director, Dr. S. Chiplunkar for her support, former Director and current Chairperson of Doctoral Committee Dr. R. Sarin for his counsel, former chairperson of Doctoral Committee and Deputy Director Dr. S.M. Zingde for her help and advice, Dr. A. Ingle for invaluable suggestions with animal models, Dr. G. Maru for his ideas and Dr. M. Vaidya for his support. I also thank ACTREC for funding.

I am also thankful to the staff of Laboratory animal facility and Imaging facility, my labmates Piyush, Pramod, Aditi, Sneha, Nikhila, Mahesh, Sarika, Surya, Arti and especially Aarif. I am grateful to my best friends in ACTREC and ACTREC PhD batch 2009, who made my stay in ACTREC awesome. I would also like to thank my all my seniors, specifically Dilip and Gaurav, and all my juniors. Finally, and most importantly, I would like to acknowledge the indispensable support of my parents, Karabi and Tapobrata Bhattacharjee.

## Contents

SYNOPSIS.....	10
LIST OF FIGURES .....	25
LIST OF TABLES.....	29
CHAPTER 1 INTRODUCTION.....	31
CHAPTER 2 DEVELOPMENT OF A RODENT MODEL OF BREAST NEOPLASMS SUITABLE FOR CHARACTERIZATION BY RAMAN SPECTROSCOPY ..	71
CHAPTER 3 TO STUDY DEVELOPMENT OF BREAST NEOPLASMS INDUCED BY CHEMICAL CARCINOGEN USING RAMAN SPECTROSCOPY IN TARGET ORGANS AND BODY FLUIDS .....	134
CHAPTER 4 STUDY RAMAN SPECTRAL SIGNATURES OF EXPERIMENTAL LUNG METASTASIS FROM BREAST CANCER CELL LINE.....	219
CHAPTER 5 THESIS SUMMARY.....	238
REFERENCES .....	242
Reprints .....	256

## **SYNOPSIS**



1. **Name of the Student:** Tanmoy Bhattacharjee
2. **Name of the Constituent Institution:** Tata Memorial Centre (ACTREC).
3. **Enrolment no.:** LIFE09200904015
4. **Title of the Thesis:** A study of breast cancer progression in rodent models using Raman spectroscopy
5. **Board of Studies :** Life Sciences

## **Introduction**

Breast cancer is the most frequently diagnosed as well as the most fatal cancer amongst women worldwide, with estimated 16 million new cases and 5 million cancer deaths in 2012 (1). Breast cancer is a heterogeneous disease encompassing several entities with distinct morphological features and clinical outcomes; the most common being invasive ductal and lobular carcinomas. (2). The etiology of this cancer is emerging and several factors such as ageing, BRCA mutations, detection of atypical hyperplasia, familial history of cancer, etc. are known risk factors (3). All the currently available techniques for breast cancer screening - mammography, magnetic resonance imaging (MRI), breast self examination (BSE), clinical breast examination (CBE), positron emission tomography (PET), ultrasonography, molecular breast imaging and thermography suffer from low sensitivity, low specificity and/ or poor cost effectiveness (4). The gold standard for diagnosis of breast cancer remains biopsy followed by histopathology, although it has several drawbacks like low sensitivity, low specificity, inter-observer variation, tedious procedures and long output times (5). Modalities used for treatment of this cancer include combination of surgery, radiation, chemotherapy, hormonal therapy and targeted therapy (eg. Herceptin) (3).

It has been repeatedly observed that treatment outcomes depend on stage of diagnosis and that prognosis worsens with late detection (6). Thus, it becomes imperative to detect breast cancer early. Several genomic and proteomic studies have been undertaken to identify biomarkers for early detection of breast cancer (7). It has been proposed that such holistic approaches have better chance at early detection than single or few markers. Holistic information regarding a sample can also be obtained using optical spectroscopic techniques like diffuse reflectance spectroscopy, infrared spectroscopy and Raman spectroscopy. Raman spectroscopy is based on inelastic scattering of light. The shift in the frequency of the

inelastically scattered photon compared to the incident photon is specific to the molecular vibration causing the scattering. Thus, the Raman scattered light can provide detailed information regarding the chemical composition of a sample. Since progression of breast cancer involves massive biochemical changes, this technique may be ideal for studying carcinogenesis. Raman spectroscopy has shown promising results in the diagnosis of cervix, lung, gastrointestinal, oral, skin, colon and several other cancers (8) including breast cancers (9-11). Identification of atypical ductal hyperplasia, surgical margins; detection of HER-2, EGFR, lymph node abnormality, microcalcification, etc. using Raman spectroscopy has been demonstrated (12-19). These studies have relied on histopathology and/ or immunohistochemistry (IHC) as gold standard for pathological (atypia)/ biomarker (EGF, ErbB2, IGF-1) status. However, as mentioned earlier, histopathology/IHC themselves suffer from low sensitivity, specificity and inter-observer variations. Thus, it would be ideal if biochemical changes that lead to the clinical event of cancer can be identified using Raman spectroscopy. To this end, animals can be treated with a carcinogen and their mammary glands can be spectroscopically scanned at regular time intervals post carcinogen treatment till tumor appearance. In such sequential follow-up studies, spectra from regions of mammary glands that actually had a clinical outcome of cancer can be analyzed. This may provide better insights into the process of carcinogenesis and help identify spectral features for predicting tumor development.

**Aim and objective:**

To sequentially follow chemical carcinogen/cell line induced breast neoplasm in rodents and identify spectral signatures of precancerous, malignant and metastatic stages.

Following are the specific objectives of the thesis:

1. To develop a rodent model of breast neoplasms suitable for characterization by Raman spectroscopy.
2. To study development of breast neoplasms induced by chemical carcinogen using Raman spectroscopy in target organs and body fluids.
3. To study Raman spectral signatures of experimental lung metastasis from breast cancer cell line.

**Objective 1: To develop a rodent model of breast neoplasms suitable for characterization by Raman spectroscopy**

*1. Exploring rodent strains suitable for transcutaneous in vivo spectroscopy:* To study progression of breast neoplasm, a model suitable for non invasive spectroscopy (that does not require animal sacrifice) and induction of breast neoplasm is imperative. Several rodent strains - black haired C57 (n=2), brown haired C3H/J (n=2), Swiss albino mice (n=2), hairless Swiss bare (Swiss B) mice (n=2), and albino Sprague-Dawley (SD) rats (n=2) were screened to meet the above criteria. Spectra were acquired transcutaneously from left inguinal mammary glands of these strains using a Raman spectrometer consisting of a 785 nm wavelength diode laser (PI-ECL-785-300-FC, Process Instruments), a high efficiency spectrograph (HE-785, Jobin-Yvon-Horiba, France) coupled to a CCD (CCD-1024X256-BIDD-SYN, Synapse) using Commercial RamanProbe (RPS 785/ 12-5, In Photonics Inc, Downy St. USA). The spectra were preprocessed by correcting for CCD response, subtracting background and polynomial fitting before calculating mean spectra. Comparison of mean transcutaneous spectra with ex-vivo spectra and spectra from literature showed Swiss Bare mice (SB)/ shaved Sprague Dawley (SD) rats to be best suited for transcutaneous

spectroscopy. *These observations suggest that Swiss B and shaved SD rats can be used for transcutaneous in vivo Raman spectroscopy.*

**2. Classification of transcutaneous breast spectra from transcutaneous spectra of other anatomical sites:** In order to evaluate the possibility of distinguishing *in vivo* breast spectra from other anatomical sites, transcutaneous spectra were acquired from scalp, cheek, chest, leg shin, spine, leg thigh, neck, and tail of Swiss B mice (n=10) and compared with transcutaneous breast spectra of Swiss B mice (n=10). The spectra were preprocessed and subjected to Principal Component Analysis (PCA) and Principal Component – Linear Discriminant Analysis (PC-LDA). PCA as well as PC-LDA could *distinguish breast spectra from other anatomical sites*. This further confirms feasibility of *in vivo* spectroscopy (**T. Bhattacharjee, et.al. Lasers in Medical Sciences, 2014**).

**3. Non-invasive follow up of physiological processes using Raman spectroscopy:** The next step was to test the possibility of non-invasively studying biological processes using transcutaneous *in vivo* spectroscopy. Two physiological processes were followed *in vivo*

1. Pregnancy and lactation: Female Swiss B mice were allowed to mate. Spectra were acquired from left inguinal breast of mice a day after successful mating (n=6); ascertained by the development of vaginal plugs. Spectra were also acquired from the same site when the mice were visibly pregnant (n=6) and a day after delivering pups, that is during lactation (n=6). Changes in lipid, protein and DNA profiles were observed in the spectra of pregnant and lactating mice breast compared to control. PC-LDA could classify control, pregnant and lactating transcutaneous spectra (**T. Bhattacharjee, et.al. Journal of Biomedical Optics, 2013**).

2. Ageing: transcutaneous spectra were acquired from left inguinal breast of 2 (n=5), 4-6 (n=4), 10-12 (n=5) and 13-15 (n=5) months old Swiss B mice. Increase in lipids and decrease in DNA were observed in spectra of older mice compared to younger mice. PC-LDA could classify spectra of 2 and 4-6 months old mice distinctly, but overlap was observed between spectra of 10-12 and 13-15 months old mice (**T. Bhattacharjee, et.al. Vibrational Spectroscopy, 2013**). It is known that cell proliferation in breast is high during menarche (2 months old) and reproductive phase (4-6 months), but drops during perimenopause and menopause (10 – 15 months), which may explain the overlap observed. The results corroborate with the Pike's theory of 'breast ageing' (20).

These results suggest possibility of studying physiological processes non invasively. Despite change from normal during these conditions, they could be classified with high efficiency from transcutaneous frank adenocarcinoma spectra of tumor bearing Swiss B mice (n=2). *Thus, pregnancy, lactation and ageing may not act as confounding factors in the detection of breast cancer.*

**4. Model suitable for breast carcinogenesis and in vivo spectroscopy:** The previous section explores a model suitable for transcutaneous spectroscopy. This section explores strains and protocols suitable for induction of breast cancer from the point of view of *in vivo* spectroscopy. Rats or mice were treated on 47<sup>th</sup> day after birth using several different protocols - intragastric administration (gavage) of DMBA (1mg/mouse) in Swiss Bare mice, subcutaneous injection of MNU (0.1mg/mouse) in Swiss Bare mice, intragastric administration of 7,12-Dimethylbenz(a)anthracene (DMBA, 65mg/kg) in Sprague Dawley rats, subcutaneous injection of MNU (1mg/rat) in Sprague Dawley rats, subcutaneous dusting of DMBA (1mg/rat) in Sprague Dawley rats, injection of DMBA (1mg/rat) into nipple of

Sprague Dawley rats. Tumor incidence was low in Swiss Bare mice. Of a total of 25 gavage treated rats, 16 rats developed adenocarcinoma. Although the incidence is high, transcutaneous spectra needs to be acquired from all 12 rat mammary glands, since the site of tumor appearance is difficult to predict. Acquiring spectra from 12 glands is time consuming and hence impractical. Hence, this model is also unsuitable for transcutaneous *in vivo* Raman spectroscopy study of breast cancer progression. Although dusting resulted in site specific - high tumor incidence, injury and surgical wound healing at the site of incision can influence Raman spectra. To avoid this, subcutaneous injection was applied. Out of 28 treated rats, 21 rats developed fibroadenoma at the site of injection. Using nipple injection, 6 out of 10 rats developed adenocarcinoma at the site of injection. However, the time of tumor appearance ranges between 9 -24 weeks post treatment. Further standardization of this model may help study of breast adenocarcinoma progression using *in vivo* Raman spectroscopy.

**Objective 2: To study development of breast neoplasms induced by chemical carcinogen using Raman spectroscopy in target organs and body fluids**

1. ***Study development of fibroadenoma using in vivo Raman spectroscopy:*** Of 28 rats injected subcutaneously with DMBA, 21 rats developed benign breast neoplasm (fibroadenoma), while 7 rats did not develop any abnormality even 30 weeks post carcinogen treatment. Of 21 that developed neoplasms, 7 developed neoplasm approximately 18 weeks post carcinogen treatment. Spectra acquired 3, 8-10 and 12-14 weeks post carcinogen treatment from these 7 rats (labelled pre fibroadenoma or PF hence forth), along with spectra of corresponding weeks from control (C), rats that did not develop abnormality despite carcinogen treatment (no fibroadenoma or NF) and frank fibroadenoma (F) were analyzed. Control spectral features suggest lipid dominance whereas fibroadenoma spectra show dominance of proteins. PCA shows PF and F spectra

clustering away from controls. PC-LDA validation using independent test data (24 C, 16 PF and 17 NF)) shows similar results. Thus, results suggest possibility of distinguishing pretumor spectra from controls (**T. Bhattacharjee, et.al. Journal of Raman Spectroscopy, 2014 submitted**).

2. ***Study development of breast adenocarcinoma using in vivo Raman spectroscopy:*** Based on fibroadenoma progression study, transcutaneous *in vivo* study of malignant breast neoplasm (adenocarcinoma) was also carried out. Site specific adenocarcinoma was induced by injection of carcinogen in the nipple. A pilot study of adenocarcinoma progression (n=15, 5 controls, 6 pretumors, 4 no tumors despite treatment and 6 frank tumors) was carried out. Although adenocarcinoma could be distinguished from control, preadenocarcinoma could not be distinguished from control using the current model. Probable reasons are wide range time of tumor appearance (9<sup>th</sup> -24<sup>th</sup> week post carcinogen treatment) and tumor incidence is ~ 50%. Further standardization of protocol such as a) stereotactic mechanism for injecting carcinogen at precise location and depth b) exploration of vehicles to deliver high concentration with low amount of fluid, are required for large scale studies.

3. ***Study development of breast adenocarcinoma using urine based Raman spectroscopy:*** As described above, transcutaneous *in vivo* study of adenocarcinoma progression is limited by unsuitability of model. To circumvent this, urine based Raman spectroscopy was used to study pre-adenocarcinoma condition using the well-established protocol for inducing breast adenocarcinoma in SD rats - intragastric carcinogen administration of DMBA. A total of 42 SD rats were used (test =25, control =17). 20 rats developed breast tumors (histopathologically confirmed adenocarcinoma of breast) approximately six months post carcinogen treatment. None of the control rats developed breast tumors. Rats were

restrained; airlifted and voided urine was collected in sterile petridishes. Urine was then transferred to sterile eppendorf tubes using a micropipette. Spectra from urine divided into six groups were acquired: a) Unprocessed control urine (n=9): urine samples of control rats were thawed and spectra were acquired from these samples b) Unprocessed tumor urine (n=9): urine samples of tumor bearing rats were thawed and spectra were acquired from these samples c) Concentrated control urine (n=8): urine samples of control were thawed, dehydrated in vacuum using Speed Vac and rehydrated with 40ul normal saline before spectra acquisition d) Concentrated tumor urine (n=7): urine samples of tumor bearing were thawed and dehydrated in vacuum using Speed Vac e) Concentrated TT urine (n=4): urine samples were collected 5 months post carcinogen (DMBA) treatment from visibly and palpably normal rats. These rats were palpated every two weeks after urine collection. Approximately 1 month post urine collection (~ 6 months post carcinogen treatment), these rats developed breast tumors. Biopsy followed by histopathology confirmed the tumors to be adenocarcinoma. The urine samples collected from these rats were labelled 'Tumors Treated' and will hence forth be referred to as 'TT'. f) Concentrated NTT urine (n=4): urine samples were collected 5 months post carcinogen (DMBA) treatment from visibly and palpably normal rats. However, these rats failed to develop tumor even 8 months post carcinogen treatment. Urine samples from these rats were labeled as 'No Tumors Treated' and henceforth are referred as 'NTT'. After passive thawing/rehydrating, samples were subjected to Raman spectroscopy by placing 40 µl sample on calcium fluoride (CaF<sub>2</sub>) window and spectra were recorded using Raman microprobe. Difference in the urea and creatinine bands is observed in spectra of different groups. The pilot study suggests that pre adenocarcinoma urine can be distinguished from control with ~ 75% efficiency (**T. Bhattacharjee, A. Khan, et.al. Analyst, 2014**). At present, the success rate of urine

collection is low. Before undertaking a major study, urine collection needs to be standardized.

4. ***Study development of breast adenocarcinoma using serum based Raman spectroscopy:***

Serum based Raman spectroscopic study of pre adenocarcinoma in rats (n=21; control serum 6, tumor serum 6, pre adenocarcinoma serum 4, no adenocarcinoma despite treatment serum 5) was also carried out. Results suggest that pre adenocarcinoma serum can be distinguished from control (Fig 4). For a major study, methodology for multiple blood collection to increase the efficiency of collection, and others factors as mentioned for urine studies need to be standardized.

Study of control (n=5) serum pre (n=5) and post (n=6) adenocarcinoma tumor surgery was also carried out to understand changes after tumor removal was also carried out (**T. Bhattacharjee, et.al. Journal of Biophotonics, 2014**). Results suggest ‘post surgery’ spectra can be distinguished from ‘pre surgery’ spectra.

**5. Objective 3: Study Raman spectral signatures of experimental lung metastasis from breast cancer cell line**

Accurate diagnosis of breast metastasis is vital to guide an appropriate systemic therapy and achieve better prognosis. Therefore, feasibility of distinguishing metastatic lung lesions from primary lung and breast tumors was explored. Tumors from Mouse Mammary Tumor Virus (MMTV) - induced spontaneous tumorigenesis model, C3H Jax mouse were harvested after sacrificing the mouse by cervical dislocation and used to acquire spectra of primary breast tumor (n=4). Lung adenoma was induced by intraperitoneal injection of benzo[a]pyrene (B[a]P) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) once a week for 8 weeks in AJ mice [32]. Mice were sacrificed after 28 weeks by cervical dislocation and excised

lungs were used to acquire spectra of primary lung tumor (n=4). Spectra were also acquired from normal breast (n=5) and normal lung tissues (n=6). Breast metastasis in lung was induced by intravenous injection of C3H Jax tumor single cell suspension ( $4 \times 10^6$  cells) into new 8 weeks old C3H Jax mice. After 3 weeks, the mice were sacrificed by cervical dislocation, lungs harvested (n=8) and used for spectroscopy. Distinct cluster of breast metastasis afflicted lung was observed. This suggests that Raman spectroscopy may identify breast metastasis correctly from normal lung, lung tumor, normal breasts and breast tumor (T. Bhattacharjee, et.al. *Journal of Biomedical Optics* 2014, submitted).

## 6. Summary:

The aim of the study was to identify spectral signatures/ patterns that may help distinguish pre-breast neoplasm condition from controls. Results of the study suggest that 1) Raman spectra can be used to discriminate pregnancy, lactation and ageing associated changes from each other as well as frank breast tumors; 2) pre- neoplasm (fibroadenoma) can be distinguished from controls using transcutaneous *in vivo* Raman spectroscopy; 3) although pre adenocarcinoma could not be distinguished from control using transcutaneous spectroscopy due to problems with the rodent model used, the condition could be distinguished from control using urine-based Raman spectroscopy. These studies suggest potential of breast cancer screening and early diagnosis using Raman spectroscopy; 4) Feasibility of discriminating metastatic lung lesions from primary breast and lung tumors using Raman spectroscopy was also shown. This may help guide therapy and aid breast cancer management. The feasibility of *in vivo* transcutaneous spectra acquisition, detection of early changes, and use of body fluids such as urine for progression of the disease suggest that Raman spectroscopy has potential which needs further validation in defined study designs for use in the clinics.

## References

1. "Globocan 2012," [http://globocan.iarc.fr/Pages/fact\\_sheets\\_cancer.aspx](http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx) (February, 2015).
2. P. P. Rosen et al., *Rosen's breast pathology*, Fourth edition. ed.
3. C. B. Matsen, and L. A. Neumayer, "Breast cancer: a review for the general surgeon," *JAMA surgery* **148**(10), 971-979 (2013).
4. A. B. Nover et al., "Modern breast cancer detection: a technological review," *International journal of biomedical imaging* **2009**(902326 (2009).
5. S. M. Ismail et al., "Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia," *BMJ* **298**(6675), 707-710 (1989).
6. N. A. Howlander N, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds), "SEER Cancer Statistics Review, 1975-2011," *National Cancer Institute. Bethesda, MD* (2013).
7. J. A. Shaw et al., "Genomic analysis of circulating cell-free DNA infers breast cancer dormancy," *Genome research* **22**(2), 220-231 (2012).
8. D. Evers et al., "Optical spectroscopy: current advances and future applications in cancer diagnostics and therapy," *Future oncology* **8**(3), 307-320 (2012).
9. M. Chowdary et al., "Biochemical correlation of Raman spectra of normal, benign and malignant breast tissues: a spectral deconvolution study," *Biopolymers* **91**(7), 539-546 (2009).
10. A. S. Haka et al., "Diagnosing breast cancer by using Raman spectroscopy," *Proceedings of the National Academy of Sciences of the United States of America* **102**(35), 12371-12376 (2005).
11. J. Pichardo-Molina et al., "Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients," *Lasers in medical science* **22**(4), 229-236 (2007).
12. I. Barman et al., "Application of Raman spectroscopy to identify microcalcifications and underlying breast lesions at stereotactic core needle biopsy," *Cancer research* **73**(11), 3206-3215 (2013).
13. X. Bi et al., "Evaluating HER2 amplification status and acquired drug resistance in breast cancer cells using Raman spectroscopy," *Journal of biomedical optics* **19**(2), 025001 (2014).
14. J. Horsnell et al., "Raman spectroscopy--a new method for the intra-operative assessment of axillary lymph nodes," *The Analyst* **135**(12), 3042-3047 (2010).
15. B. H. Jun et al., "Multifunctional silver-embedded magnetic nanoparticles as SERS nanoprobe and their applications," *Small* **6**(1), 119-125 (2010).
16. M. D. Keller et al., "Development of a spatially offset Raman spectroscopy probe for breast tumor surgical margin evaluation," *Journal of biomedical optics* **16**(7), 077006-077006-077008 (2011).
17. A. Saha et al., "Raman spectroscopy: a real-time tool for identifying microcalcifications during stereotactic breast core needle biopsies," *Biomedical optics express* **2**(10), 2792-2803 (2011).
18. L. Xiao et al., "Imaging of epidermal growth factor receptor on single breast cancer cells using surface-enhanced Raman spectroscopy," *Analytica chimica acta* **843**(73-82 (2014).
19. C. Zheng et al., "The use of Au@SiO<sub>2</sub> shell-isolated nanoparticle-enhanced Raman spectroscopy for human breast cancer detection," *Analytical and bioanalytical chemistry* **406**(22), 5425-5432 (2014).
20. M. C. Pike et al., "'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer," (1983).

## **Publications in Refereed Journal:**

### **a. Published**

1. **T. Bhattacharjee**, A. Khan, A. Ingle, G. Maru, C. Murali Krishna, Raman spectroscopy of urine: Diagnosis of breast cancer in animal model, *Analyst* (2015), 140, 456-466
2. **T. Bhattacharjee**, A. Khan, P. Kumar, A. Ingle, G. Maru, C. Murali Krishna, Raman spectroscopy of serum: A study on 'pre' and 'post' breast adenocarcinoma resection in rat models, *J. Biophotonics* (2014), DOI10.1002/jbio.201400040
3. **T. Bhattacharjee**, Piyush Kumar, G. Maru, A. Ingle, C. Murali Krishna, Swiss Bare mice: A suitable model for transcutaneous *in vivo* Raman spectroscopic studies of Breast cancer, *Lasers. Med. Sci.*, (2014), 29 (1), 325 - 333
4. **T. Bhattacharjee**, G. Maru, A. Ingle, C. Murali Krishna, Transcutaneous *in vivo* Raman spectroscopic studies in mouse model: Evaluation of changes in Breast associated with Pregnancy and Lactation, *J. Biomed. Opt.* (2013), 18(4), 047004
5. **T. Bhattacharjee**, G. Maru, A. Ingle, C. Murali Krishna, Transcutaneous *in vivo* Raman spectroscopy: Detection of Age-related changes in Mouse Breast, *Vibrational Spectroscopy* (2013), 67, 80–86

**b. Accepted**

**c. Communicated**

1. **T. Bhattacharjee**, G. Maru, A. Ingle, C. Murali Krishna, Transcutaneous *in vivo* Raman spectroscopy of breast tumor and pre tumor, *Journal of Raman Spectroscopy* (2014)
2. **T. Bhattacharjee**, S. Tawde, R. Hudlikar, M. Mahimkar, G. Maru, A. Ingle, C. Murali Krishna, Raman spectroscopic study of breast cancer metastasis in lungs, *JBO* (2014)

**d. Other Publications**

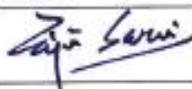
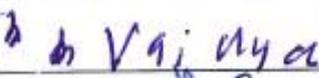
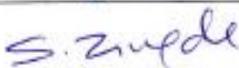
1. P. Kumar, **T. Bhattacharjee**, A. Ingle, G.B. Maru, C. Murali Krishna, Raman spectroscopy vis-a-vis histopathology of hamster buccal pouch carcinogenesis: a study on early and late stages, *Journal of Innovative Health Sciences*, 2015 (submitted).

2. P. Kumar, T. Bhattacharjee, A. Ingle, G.B. Maru, C. Murali Krishna, Raman spectroscopy of experimental oral carcinogenesis: Study on sequential cancer progression in hamster buccal pouch model, Technologies in Cancer Research and Therapeutics, 2015 (submitted).

**Signature of Student:**

**Date:**

**Doctoral Committee**

Sr. No.	Name	Designation	Signature	Date
1	Dr. R. Sarin	Chairperson		20-2-15
2	Dr. C. M. Krishna	Convener (Guide)		20-3-15
3	Dr. G. B. Maru	Co-Guide		20-03-2015
4	Dr. M. M. Vaidya	Member		20-03-15
5	Dr. A. D. Ingle	Technical Advisor		20-03-15
6	Dr. S. M. Zingde	Invitee		16-3-15

Forwarded Through:



Dr. S.V. Chiplunkar  
Director, ACTREC  
Chairperson,  
Academic & training Program, ACTREC

Dr. S. V. Chiplunkar  
Director  
Advanced Centre for Treatment, Research &  
Education in Cancer (ACTREC)  
Tata Memorial Centre  
Kharghar, Navi Mumbai 410210.



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

**PROF. K. S. SHARMA**  
DIRECTOR (ACADEMICS)  
TATA MEMORIAL CENTRE,  
PAREL, MUMBAI

## **LIST OF FIGURES**

<b>Figure</b>	<b>Page no.</b>
Figure 1.1 Anatomy of the breast .....	34
Figure 1.2 Jablonksi's diagram for IR absorption, elastic and inelastic scattering and fluorescence .....	51
Figure 1.3 Raman spectrometer schematic .....	53
Figure 1.4 Overview of study methodology and rationale .....	68
Figure 2.1 Illustration of fiber-optic based Raman spectrometer employed in the study.....	74
Figure 2.2 Choosing a suitable strain for <i>in vivo</i> Raman studies - spectra .....	78
Figure 2.3 Classification of anatomical sites using <i>in vivo</i> spectroscopy - illustration of the anatomical sites investigated in the study .....	81
Figure 2.4 Classification of anatomical sites using transcutaneous <i>in vivo</i> Raman spectroscopy - mean and standard deviation of transcutaneously acquired spectra .....	85
Figure 2.5 Classification of anatomical sites using transcutaneous <i>in vivo</i> Raman spectroscopy - PC-LDA scatter plot .....	87
Figure 2.6 Classification of breast tumors using transcutaneous <i>in vivo</i> Raman spectroscopy - mean spectra and PC-LDA scatter plot.....	93
Figure 2.7 Pregnancy and lactation transcutaneous <i>in vivo</i> Raman spectroscopy study - mean <i>in vivo</i> Raman spectra of breast from a) non pregnant b) pregnant c) lactating d) tumor bearing mice.....	99
Figure 2.8 Pregnancy and lactation transcutaneous <i>in vivo</i> Raman spectroscopy study - difference spectrum.....	101
Figure 2.9 Pregnancy and lactation transcutaneous <i>in vivo</i> Raman spectroscopy study - PC-LDA to explore differences in mouse breast of non pregnant, pregnant and lactating mice.	103
Figure 2.10 Pregnancy and lactation transcutaneous <i>in vivo</i> Raman spectroscopy study - PC-LDA to explore differences between breast spectra of tumor bearing, non pregnant, pregnant and lactating mice .....	108
Figure 2.11 Aging transcutaneous <i>in vivo</i> Raman spectroscopy study - mean <i>in vivo</i> Raman spectra of mouse breast from a) 2 months (menarche), b) 4-6 months (mid-reproductive), c) 10-12 months (perimenopause) d) 13-15 months (menopause) old mice and e) frank breast tumors .....	115
Figure 2.12 Aging transcutaneous <i>in vivo</i> Raman spectroscopy study - Difference spectra.	118

Figure 2.13 Aging transcutaneous <i>in vivo</i> Raman spectroscopy study - PC-LDA for age-wise differences in breast of 2 months (menarche), 4-6 months(mid-reproductive), 10-12 months (perimenopause) and 13-15 months (menopause) old mice .....	120
Figure 2.14 Aging transcutaneous <i>in vivo</i> Raman spectroscopy study - PC-LDA to explore differences between frank breast tumors and age-related changes in breast .....	125
Figure 3.1 Scheme of spectra acquisition for transcutaneous <i>in vivo</i> Raman spectroscopic study of fibroadenoma progression.....	138
Figure 3.2 Sites of spectra acquisition for transcutaneous <i>in vivo</i> Raman spectroscopic study of fibroadenoma progression .....	141
Figure 3.3 Fibroadenoma progression study - mean spectra .....	145
Figure 3.4 Transcutaneous <i>in vivo</i> Raman spectroscopic study of fibroadenoma progression - difference spectra .....	147
Figure 3.5 Transcutaneous <i>in vivo</i> Raman spectroscopic study of fibroadenoma progression - PCA.....	151
Figure 3.6 Transcutaneous <i>in vivo</i> Raman spectroscopic study of fibroadenoma progression - PC-LDA model scatter plot of C <sub>3</sub> , C <sub>8-10</sub> , C <sub>12-14</sub> , C <sub>20</sub> and F .....	154
Figure 3.7 Breast adenocarcinoma progression study - PCA of frank adenocarcinoma, pre adenocarcinoma and control spectra. ....	172
Figure 3.8 Methodology: Urine based pre adenocarcinoma study using Raman spectroscopy .....	176
Figure 3.9 Illustration of Raman microprobe system .....	178
Figure 3.10 Urine based pre adenocarcinoma study - mean and difference spectra of control and tumor bearing rat urine.....	180
Figure 3.11 Urine based pre adenocarcinoma study a) Mean spectra of concentrated urine interpolated in 600-1800cm <sup>-1</sup> region from i) NTT and ii) TT rats, b) i) NTT – control, ii) TT – control, and iii) TT - NTT difference concentrated urine spectra. Change in urea intensity is again observed between the groups .....	182
Figure 3.12 Urine based pre adenocarcinoma study - PCA of unprocessed urine from control and tumor bearing rats .....	185
Figure 3.13 Urine based pre adenocarcinoma study - PC-LDA of unprocessed urine from control and tumor bearing rats .....	187
Figure 3.14 Urine based pre adenocarcinoma study - PCA of concentrated urine from control and tumor bearing rats .....	189

Figure 3.15 Urine based pre adenocarcinoma study - PC-LDA of concentrated urine from control and tumor bearing rats .....	191
Figure 3.16 Urine based pre adenocarcinoma study - PCA of concentrated urine from control, NTT, TT and tumor bearing rats.....	196
Figure 3.17 Urine based pre adenocarcinoma study - PC-LDA of concentrated urine from control, NTT, TT and tumor bearing rats .....	198
Figure 3.18 Serum based pre adenocarcinoma study - PC-LDA of C, NTT, TT and tumor serum samples.....	204
Figure 3.19 Pre and post surgery serum based Raman spectroscopy study - protocol.....	207
Figure 3.20 Pre and post surgery serum based Raman spectroscopy study - mean and difference spectra.....	209
Figure 3.21 Pre and post surgery serum based Raman spectroscopy study - PCA of control, pre-surgical and post-surgical groups .....	212
Figure 3.22 Pre and post surgery serum based Raman spectroscopy study - PC-LDA of control, pre-surgical and post-surgical groups.....	214
Figure 4.1 Discrimination of primary lung from metastatic breast lesions - mean spectra interpolated .....	225
Figure 4.2 Discrimination of primary lung from metastatic breast lesions - difference spectra .....	228
Figure 4.3 Discrimination of primary lung from metastatic breast lesions - PCA .....	230
Figure 4.4 Discrimination of primary lung from metastatic breast lesions - PC-LDA.....	232

## **LIST OF TABLES**

<b>Table</b>	<b>Page no.</b>
Table 2.1 Classification of anatomical sites using transcutaneous <i>in vivo</i> Raman spectroscopy - PC-LDA of anatomical sites.....	89
Table 2.2 Classification of breast tumors using transcutaneous <i>in vivo</i> Raman spectroscopy - PC-LDA of breast normal and tumor spectra .....	95
Table 2.3 Pregnancy and lactation transcutaneous <i>in vivo</i> Raman spectroscopy study - confusion matrix for PC-LDA of non pregnant, pregnant and lactating mouse breast .....	105
Table 2.4 Pregnancy and lactation transcutaneous <i>in vivo</i> Raman spectroscopy study - confusion matrix for PC-LDA of non pregnant, pregnant, lactating mouse breast and frank breast tumors .....	110
Table 2.5 Aging transcutaneous <i>in vivo</i> Raman spectroscopy study - PC-LDA of 2 months, 4-6 months, 10-12 months and months 13-15 old mice .....	122
Table 2.6 Aging <i>in vivo</i> study - PC-LDA of age groups compared with frank breast tumor	127
Table 2.7 Summary of carcinogenesis protocols and outcome.....	132
Table 3.1 Transcutaneous <i>in vivo</i> Raman spectroscopic study of fibroadenoma progression - PC-LDA confusion matrix of C3, C8-10, C12-14, C20 and F .....	157
Table 3.2 Transcutaneous <i>in vivo</i> Raman spectroscopic study of fibroadenoma progression - evaluation of PC-LDA model with control test data set .....	160
Table 3.3 Transcutaneous <i>in vivo</i> Raman spectroscopic study of fibroadenoma progression - evaluation of PC-LDA model with NF test data set .....	163
Table 3.4 Transcutaneous <i>in vivo</i> Raman spectroscopic study of fibroadenoma progression - evaluation of PC-LDA model with PF test data set .....	166
Table 3.5 Urine based pre adenocarcinoma study - PC-LDA confusion matrix for LOOCV a) unprocessed urine, b) concentrated urine.....	193
Table 3.6 Urine based pre adenocarcinoma study PC-LDA confusion matrix for LOOCV of control, NTT, TT and tumor .....	200
Table 3.7 : Pre and post surgery serum based Raman spectroscopy study - confusion matrix for PC-LDA LOOCV of control, pre-surgical and post-surgical groups.....	216
Table 4.1 Discrimination of primary lung from metastatic breast lesions - PC-LDA Confusion matrix for a) model building, b) LOOCV, and c) LOOCV - sensitivity and specificity, and d) Independent test prediction.....	235

# **CHAPTER 1 INTRODUCTION**

Breast cancer is the most fatal cancer amongst women worldwide. According to Globocan 2012, estimated number of breast cancer cases and deaths due to breast cancer worldwide were 1677000 and 522000, respectively. Although the number of cases in more developed countries and less developed countries are similar, the numbers of deaths in developed countries are fewer – probably due to accessible therapeutic interventions. In the United States of America, European Union, India and China, estimated cancer cases were 233000, 367000, 145000 and 187000, respectively whereas estimated cancer deaths were 44000, 92000, 70000 and 48000, respectively (1). Section I of this chapter covers important aspects of breast cancer such as etiology, screening and diagnostic techniques, treatment modalities and current problems associated with breast cancer management. Section II describes Raman spectroscopy, a prospective technique that may help solve some of the shortcoming currently associated with breast cancer management, it's principles, instrumentation, data analysis methods, general applications and literature pertaining to breast cancer management. Section III details the aims and objectives of this thesis.

## **I. Breast cancer**

### **Breast anatomy**

Breast (Figure 1.1) is described in Gray's Anatomy as being composed of glandular and adipose tissue held together by a loose framework of fibres called Cooper's ligaments. Histological studies show that the lobes are composed of lobules, which consist of clusters of alveoli containing lactocytes (mammary secretory epithelial cells) that synthesize breast milk. The alveoli are connected to very small ducts that join to form larger ducts draining the lobules. These larger ducts finally merge into one milk duct for each lobe. Then, under the areola, this single milk duct widens into a lactiferous sinus before narrowing at the base of

the nipple and terminating at its orifice on the surface of the nipple. The adipose tissue of the breast is typically situated between lobes rather than within lobules (21).

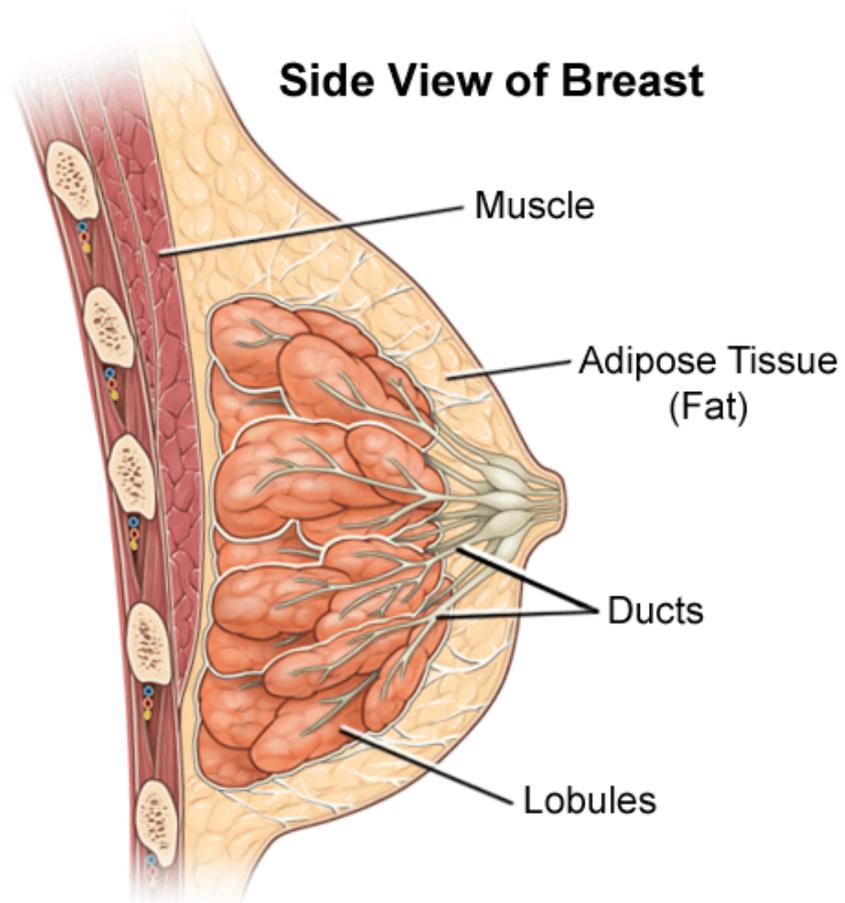


Figure 1.1 Anatomy of the breast. Figure shows different components that make up the human breast.

In mouse, there are five pairs of mammary fat pads located just below the skin, which extend from the thoracic (three pairs) to the inguinal (two pairs) regions of the animal along what is termed the milk or mammary line. Each fat pad has an exterior nipple (1–3) to which the primary epithelial duct is connected to allow the release of milk during lactation.

At birth, the mammary gland consists of the epithelial cords and the stroma, which includes connective tissue, fibroblasts, and the mammary fat pad. The parenchyma at this stage is rudimentary and consists of a small ductal tree. Each branch is composed of a single layer of epithelial cells surrounding a central lumen; the cells bordering the lumen are referred to as luminal epithelial cells. Beneath the epithelial cells, the myoepithelial cells form a basal layer that rests on basement membrane, which separates the parenchymal and stromal compartments. The myoepithelial cells are contractile and are responsible for the movement of milk out of the alveoli and down the ducts during lactation. These cells are also responsible for the secretion of basement membrane components during all developmental stages. The connective tissue stroma is thick and dense around these epithelial structures and consists of eosinophilic fibrous connective tissue and fibroblasts. The basement membrane is composed of a thin layer of proteins which lies next to the basal surface of the myoepithelial cells of the ductal structures. It consists of an organized network of proteins and proteoglycans which are locally secreted by the myoepithelial cells (22).

### **Breast physiology**

Most of the information regarding breast physiology has been derived from study of mouse and rat models. The following discussion focuses on mouse breast physiology.

### **Puberty and postpuberty**

The period of most rapid ductal growth occurs during puberty from approximately 3–6 weeks of age in the mouse. In rodents the glands do not regress during prepuberty, but maintain a small ductal tree with the terminal end bud (TEB) at tip, which initiates rapid growth at the onset of puberty. During pubertal and postpubertal growth, the TEB form epithelial and myoepithelial cells while the ducts lengthen and branch to form secondary and tertiary ducts. They ultimately fill the mammary fat pad by approximately 3 months of age. At 10–12 weeks of age, the majority of the TEBs have reached the edge of the fat pad. The development of lateral and alveolar buds is initiated in the postpubertal gland in response to the cyclic secretion of ovarian hormones with each estrous cycle. The lateral buds can form branches or cleave to form alveolar buds. Lateral buds that will form branches have a layer of cap cells at the growing tip similar to TEBs (22).

### **Pregnancy and lactation**

The peak of mammary differentiation occurs during the 19–21 days of pregnancy and culminates with formation of alveoli and a fully lactating gland at parturition. Initial pregnancy-induced mammary growth involves massive proliferation of ductal branches and formation of alveolar buds like those observed during postpubertal development. The epithelial to adipocyte ratio increases and capillaries are found within the connective tissue surrounding each individual alveolus. During the second half of pregnancy, the alveolar buds progressively cleave and differentiate into individual alveoli that will ultimately become milk-secreting lobules during lactation. By late pregnancy, the alveoli fill the majority of the fat pad, at which time the gland is capable of milk production. By day 18 of pregnancy, the alveolar epithelial cells are producing milk proteins and lipid in preparation for lactation. The epithelial cells are enlarged due to accumulation of lipid. Fully developed alveolar structures remain until the completion of lactation. The process of lactation continues for approximately

3 weeks until the pups are weaned. After weaning, the gland goes through a process of death and remodeling termed involution (22).

### **Perimenopause and menopause**

The postmenopausal breast consists of fibrous and adipose tissue with few scattered residual ducts, acini, and vessels. Atrophy of the glandular component is normal with increasing age and is characterized by loss of the glandular epithelium and increasing thickness of the basement membranes, leading to progressive obliteration of the acini (23).

### **Breast cancer pathology**

Different breast pathologies have been discussed in detail in Rosen's breast pathology (2).

Briefly, breast diseases may be grouped into

a) Inflammatory disorders

b) Infections

c) Breast neoplasms – Breast neoplasms can be broadly classified into two groups

1. Benign breast tumors: papillomas, myoepithelial and fibroepithelial neoplasms and adenosis

2. Invasive breast cancer:

Breast cancer is a heterogenous disease that includes a variety of abnormalities with distinct morphological and clinical presentations (2). The most common subtypes are ductal and lobular carcinomas that account for 40-75% of all diagnosed cases. As their names suggest, ductal carcinoma originates from ductal cells while lobular carcinoma originates from cells in breast lobes. When confined to the duct or lobe of

origin, they are referred to as ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS), respectively. On invasion of the same, they are called invasive ductal carcinoma (IDC)/ invasive lobular carcinoma (ILC). The current model of human breast cancer progression proposes a linear multi-step process which initiates as flat epithelial atypia (FEA), progresses to atypical ductal hyperplasia (ADH)/ atypical lobular hyperplasia (ALH), evolves into DCIS/ LCIS and culminates in the potentially lethal stage of IDC/ ILC (24). Other sub types of invasive carcinoma are intraepithelial carcinoma, intraductal carcinoma, invasive ductal carcinoma, tubular, papillary, medullary, squamous, mucinous, apocrine, small cell, secretory, cystic hypersecretory, adenoid, cribriform, lipid rich, glycogen rich, invasive micropapillary, invasive lobular carcinoma. Some other known malignancies include Paget's disease of nipple, sarcoma, lymphoid and haematopoietic tumors.

### **Breast cancer etiology**

The etiology of human breast cancer remains largely unknown. Risk factors associated with breast cancer can be grouped into three broad determinants: physiological factors, family history (hereditary) factors, and environmental (including lifestyle) factors (25, 26).

#### **Physiological factors**

With increasing age, breast cancer incidence also increases, doubling every 10 years until the menopause. After menopause, when the rate of increase slows dramatically. Early start in menstruation and late menopause increase the risk of breast cancer. Women who start menstruating early in life or who have a late menopause have an increased risk of developing breast cancer. Nulliparity and late age at first birth both increase the lifetime incidence of breast cancer. An early age at birth of a second child further reduces the risk of breast cancer.

### **Hereditary factors**

Up to 10% of breast cancer in Western countries is due to genetic predisposition. Breast cancer susceptibility is generally inherited as an autosomal dominant with limited penetrance. A woman's risk of breast cancer is two or more times greater if she has a first degree relative (mother, sister, or daughter) who developed the disease before the age of 50, and the younger the relative when she developed breast cancer - the greater the risk. Women with severe atypical epithelial hyperplasia have a four to five time higher risk of developing breast cancer than women who do not have any abnormal proliferative changes in their breast. Women with palpable cysts, complex fibroadenomas, duct papillomas, sclerosis adenosis, and moderate or florid epithelial hyperplasia have a slightly higher risk of breast cancer (1.5-3 times) than women without these changes, but this increase is not clinically important. Ionizing radiation also increases risk later in life, especially when exposure is during rapid breast formation.

### **Environmental factors**

Age adjusted incidence and mortality for breast cancer varies by up to a factor of five between countries. Studies of migrants from Japan to Hawaii show that the rates of breast cancer in migrants assume the rate in the host country within one or two generations, indicating that environmental factors are of greater importance than genetic factors. Obesity is associated with a twofold increase in the risk of breast cancer in postmenopausal women. Women who begin use of oral contraceptives before 20 years of age have a higher risk of breast cancer. Increased risk is also observed in the 10 years following cessation of oral contraceptive intake. Among current users of hormone replacement therapy (HRT) and those who have ceased use 1-4 years previously, the relative risk of having breast cancer diagnosed increases by a factor of 1.023 (1.011-1.036) for each year of use.

## **Breast cancer screening**

It has been observed that early diagnosis leads to better prognosis. According to Surveillance, Epidemiology, and End Results (SEER) program data (6), relative survival rate on detection of cancer when localized to the breast is as high as 98.4%. Compared to this, relative survival rate decreases to 23% if the cancer has metastasized. Thus, breast cancer screening can dramatically improve treatment efficacy and overall survival of breast cancer afflicted women.

### **Guidelines**

Several agencies have formulated screening guidelines to help detect breast cancer early. The most widely recommended screening approach in the United States has been annual mammography beginning at age 40 years. However, in November 2009, the US Preventive Services Task Force (USPSTF) issued updated breast cancer screening guidelines that recommend against routine mammography before age 50 years, but suggested biennial mammography and individualized decision to opt for screening for women aged 40 to 49 years. The USPSTF recommended against teaching breast self examination (BSE), based on studies that found that teaching BSE did not reduce breast cancer mortality but instead resulted in additional imaging procedures and biopsies. As opposed to this, the American College of Obstetricians and Gynecologists (ACOG) continues to recommend counseling patients that BSE has the potential to detect palpable breast cancer and can be performed. The USPSTF also found insufficient evidence in favor of magnetic resonance imaging (MRI) as screening technique for breast cancer. As per American Society Guidelines, women in their 20s and 30s should opt for BSE regularly and have a clinical breast exam (CBE) every 3 years. Starting at age 40, women should have a CBE every year. Women at high risk (greater than 20% lifetime risk) should get an MRI and a mammogram every year. Yearly MRI

screening is not recommended for women whose lifetime risk of breast cancer is less than 15% (27, 28).

## **Screening techniques**

### **BSE and CBE**

BSE and CBE are inexpensive and noninvasive procedures for the regular examination of breasts. Evidence supporting the effectiveness of these 2 screening methods is controversial and largely inferred. Even with appropriate training, BSE has not been found to reduce breast cancer mortality. However, with increasing improvements in treatment regimens for early localized disease, BSE and CBE, particularly among women younger than 40 years, continues to be recommended by many health care organizations. Randomized clinical trial results support combining CBE with mammography to enhance screening sensitivity, particularly in younger women in whom mammography may be less effective and in women who receive mammograms every other year as opposed to annually.

### **Mammography**

Mammography is a special type of low-dose x-ray imaging used to create detailed images of the breast. Mammography is currently the best available population-based method to detect breast cancer at an early stage, when treatment is most effective. Mammography can demonstrate microcalcifications smaller than 100  $\mu\text{m}$ ; it often reveals a lesion before it is palpable by CBE and, on average, 1-2 years before noted by BSE. An estimated 48 million mammograms are performed each year in the United States. The USPSTF estimates the benefit of mammography in women aged 50-74 years to be a 30% reduction in risk of death from breast cancer. For women aged 40-49 years, the risk of death is decreased by 17%. Of

all of the screening mammograms performed annually, approximately 90% show no evidence of cancer. On necessary further diagnostic testing, approximately 2% of all screening mammograms are shown to be abnormal and require biopsy. Among cases referred for biopsy, approximately 80% of the abnormalities are shown to be benign, and 20% are shown to be cancerous.

Although mammography remains the most cost-effective approach for breast cancer screening, the sensitivity (67.8%) and specificity (75%) are not ideal. Mammography combined with CBE slightly improves sensitivity (77.4%), with a modest reduction in specificity (72%). Mammographic sensitivity for breast cancer declines significantly with increasing breast density. Mammography uses low-dose ionizing radiation, which may be harmful to the patient. False-positive results may arise when benign microcalcifications are regarded as malignant. Tissue summation shadows may appear as local parenchymal distortion; this may erroneously be called malignant tissue. A benign, circumscribed lesion may show signs suggestive of malignancy, along with other findings, such as an irregular border and no halo sign. According to data from the Breast Cancer Detection Demonstration Project (BCDDP), the false-negative rate of mammography is approximately 8-10%. About 1-3% of women with a clinically suspicious abnormality, a negative mammogram, and a negative sonogram may still have breast cancer. Possible causes for missed breast cancers include dense parenchyma obscuring a lesion, poor positioning or technique, perception error, incorrect interpretation of a suspect finding, subtle features of malignancy, and slow growth of a lesion. A multicenter study found that on previous mammograms with missed cancers, 30% of the 115 lesions were calcifications, with 49% (17 of 35) clustered or pleomorphic. Approximately 70% were mass lesions, with 40% spiculated or irregular. For calcifications and masses, the most frequently suggested reasons for possible miss were dense breasts

(34%) and distracting lesions (44%). Some cancers (for example, mucinous carcinomas) may have well-defined borders and mammographic features suggestive of benignancy.

## **MRI**

In an effort to overcome the limitations of mammography and ultrasonography, MRI has been explored as a modality for detecting breast cancer in women at high risk and in younger women. A combination of T1, T2, and 3-D contrast-enhanced MRI techniques has been found to be highly sensitive (approximating 99% when combined with mammography and CBE to malignant changes in the breast. MRI has been demonstrated to be an important adjunct screening tool for women with *BRCA1* or *BRCA2* mutations, identifying cancers at earlier stages. However, breast MRI has limited use as a general screening tool, with a 10-fold higher cost than mammography and poor specificity (26%), resulting in significantly more false-positive reads that generate significant additional diagnostic costs and procedures. The many advantages of breast MRI over conventional breast imaging for the detection of malignancy have become apparent with increasing clinical experience. These advantages include the following: a) No ionizing radiation b) All imaging planes possible c) Capability of imaging the entire breast volume and chest wall d) Greater than 90% sensitivity to invasive carcinoma e) Detection of occult, multifocal, or residual malignancy f) Accurate size estimation for invasive carcinoma. However, the widespread use of breast MRI for the detection of breast malignancy also has many disadvantages, as follows: a) High equipment and examination costs b) Limited scanner availability c) Need for the injection of a contrast agent d) No standard technique e) Poor throughput compared with that of ultrasonography or mammography f) Large number of images g) Long learning curve for interpretation h) False-positive enhancement in some benign tissues (limited specificity) i) Variable enhancement of in situ carcinoma.

## **Ultrasonography**

Ultrasonography has become a widely available and useful adjunct to mammography in the clinical setting. Ultrasound is generally used to assist the clinical examination of a suspicious lesion detected on mammography or physical examination. As a screening device, ultrasound is limited by a number of factors, most notably by the failure to detect microcalcifications and by poor specificity (34%). Ultrasonography may be useful in detecting occult breast cancer in dense breasts.

## **Positron Emission Tomography (PET)**

PET uses the metabolic differences between normal and tumor to detect malignancies. Ability to detect small tumors is limited.

## **Thermography**

The technique attempts to identify tumor by virtue of its probable temperature difference from normal tissues. It is hypothesized that tumor will have higher temperature due to its higher metabolic rate and increased blood flow. However, no study has evaluated its efficacy.

## **Breast cancer diagnosis**

### **Biopsy**

Different techniques can be used to perform biopsy, as listed below.

**Fine needle aspiration biopsy:** Fine needle aspiration (FNA) is the least invasive method of biopsy. It involves removal of cells from suspicious area in breast using a thin needle.

**Core needle biopsy:** A needle is used to remove several cylinder-shaped samples of tissue from the suspicious area. Several insertions may be required. This technique has higher risk of “false negative” result.

**Vacuum-assisted breast biopsy:** This technique is a modification of above with advantage that single insertion is sufficient for biopsy.

**Incisional biopsy:** This involves surgical removal of tissue from suspicious area.

**Excisional biopsy:** Entire tissue from suspected area along with normal margin is excised in this form of biopsy.

### **Ductal lavage**

Cells from the milk ducts are extracted by flushing the ducts with saline solution and recovering the solution using a syringe.

### **Blood marker tests**

These tests may be used for diagnosis of breast cancer and metastasis. Some blood markers are

1. CA 15.3: used to diagnose breast and ovarian cancers
2. TRU-QUANT and CA 27.29: may indicate breast cancer
3. CA125: may signal ovarian cancer, ovarian cancer recurrence, and breast cancer recurrence
4. CEA (carcinoembryonic antigen): may indicate breast cancer metastasis.
5. Circulating tumor cells: possible indication of metastasis.

### **Breast cancer treatment**

Treatment depends on many factors; some of the important ones are type and stage of cancer and estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor (HER-2/ neu) status. Breast cancer is classified into five stages and each stage has a unique therapeutic approach. The stages are defined based on tumor size, tumor presence in lymph nodes and tumor metastasis – stage 0: cancer confined to duct or lobes (in situ carcinoma); stage I: tumors are of size 2 centimeter (cm) or less, or small clusters of tumors are found in lymph nodes; stage II: tumors larger than 2 millimeters (mm) in 1 to 3 axillary lymph nodes, or tumors larger than 2 centimeter but less than 5 cm; stage III: tumors larger than 5 cm or tumor in 9 to 10 auxiliary lymph node or collar bone lymph node or cancer has spread to chest wall; stage IV: cancer has spread to other organs of the body. Surgery, chemotherapy, radiation therapy, hormonal therapy (use of tamoxifen and aromatase inhibitors to block ER) and targeted therapy (use of herceptin to block HER-2) may be used individually or in combination depending on the stage, type on hormone receptor status. Lumpectomy with radiation or mastectomy is usually used for stage 0. For stage I and II, treatment for stage 0 is accompanied by lymph node removal followed by hormone and chemotherapy. Surgery followed by chemotherapy, hormone therapy and biological therapy or combination of all is recommended for stage III and IV (3).

Despite advancements in treatment procedures, the long-term recurrence rate is still 21.4%. Several approaches are used to avoid recurrence. Surgical margin assessment is carried out to aid complete removal of tumor, while lymph nodes are evaluated and removed to contain spread of the disease. Further, breast imaging techniques like mammography, MRI, PET and ultrasound among others are used for surveillance of curative treatments and detection of asymptomatic local recurrence early. However, none of these approaches have achieved major success in preventing/ predicting recurrence. Metastatic relapse remains incurable with

average survival less than 2 years. It is estimated that 20-30% of all breast cancers will become metastatic. PET/CT has shown promising results in detection of distant metastasis. Use of circulating tumor markers like MUC-1, CA 15-3 and CA 27.29 have been recommended for metastasis treatment monitoring along with patient history, physical examination and diagnostic imaging. However, the prognostic potential of these markers are low (3, 29).

### **Challenges in breast cancer management**

The major challenges in breast cancer management are screening and accurate diagnosis of the ailment. As described in the previous sections, treatment prognosis and overall survival depends on early detection of cancer. However, currently available screening and diagnostic techniques suffer from several disadvantages. Hence, alternative screening/ diagnostic tools are imperative for better breast cancer management. A search for better screening techniques has instigated investigation in several diverse fields such as genomics, proteomics and optical spectroscopy (30-35). the optical spectroscopic techniques have been discussed in the succeeding sections.

### **Optical spectroscopic techniques**

Optical spectroscopic techniques have an edge over others as screening tools since these techniques are rapid, objective and amenable to *in vivo* applications. Some of the most commonly used methods are fluorescence spectroscopy, diffuse reflectance spectroscopy, Fourier transform infrared spectroscopy and Raman spectroscopy. Fluorescence spectroscopy is based on spectral characteristics of specific molecules (fluorophores) in tissue after exposure to light of a specific wavelength. These fluorophores absorb the light energy and gets excited which on relaxation, light of a wavelength different from the excitation

wavelength is obtained. The fluorescence spectral characteristics depend on the concentrations of the fluorophores in the target tissue. Diffuse reflectance spectroscopy (DRS) measure the intensity of diffusely reflected light. The concentration of absorbers can be directly quantified from the reflected light spectrum, the main absorbers in soft tissues being oxygenated and deoxygenated hemoglobin and b-carotene and water, adipose tissue and collagen in the visible and near-infrared spectrum of light, respectively. This technique has shown potential in discriminating several cancers from normal. In Fourier transform infrared spectroscopy (FTIR), Infra red (IR) light promotes vibration of the covalent bonds of molecules within the sample and absorbs it. Various biomolecular components give a characteristic IR spectrum that helps measurement of complex molecular vibration modes. Raman spectroscopy is based on the phenomenon of inelastic scattering and also can provide detailed chemical information about a tissue sample. Several studies have reported ability of these techniques to distinguish skin, urinary bladder, bronchus, and gastrointestinal tract, head and neck, gynecological, breast and brain cancers from normal (8, 33, 35-42). Raman spectroscopy has several advantages – minimum sample requirements, minimal or no damage to sample, ease of sample arrangements, sensitivity to environmental changes affecting the sample, feasibility of studying different physical states such as dilute solutions, concentrated solutions as well as crystals, and sensitivity towards conformational states of molecules. Compared to IR spectroscopy, Raman is not affected by water content and hence is ideal for biological applications. As opposed dearth of fluorophores in biological samples for fluorescence spectroscopy, all biological samples are well endowed with Raman scatterers. Owing to these advantages, the applications of this technique are myriad.

## **II. Raman spectroscopy**

### **Raman spectroscopy - principle**

Excellent description of Raman spectroscopy principles can be found in several books (43-48). The same has been discussed briefly. Raman scattering can be lucidly and effectively explained by the quantum theory. The quantum theory states that energy levels are quantized. 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 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 (Figure 1.2).

Raman spectrum is a plot of intensity of scattered light against energy difference. Numerically, energy difference between the initial and final vibrational levels,  $\nu$ , or Raman shift in wavenumber can be calculated using the following equation

$$\nu = \frac{1}{\lambda_{incident}} - \frac{1}{\lambda_{scattered}}$$

where,  $\lambda_{incident}$  and  $\lambda_{scattered}$  are the wavelengths in centimetre of the incident and Raman scattered photons, respectively. Molecules have normal modes of vibration – vibrations wherein all atoms vibrate with the same frequency and pass through their equilibrium positions simultaneously. The centre of gravity does not move and the molecules do not rotate. The relative magnitude and direction of the vibrational amplitudes however may differ. A linear molecule with three atomic nuclei will have four vibrational modes – 1)

vibration of the centre nuclei along the x-axis in the XY plane (bending mode), 2) simultaneous vibration of the extreme nuclei along the y-axis in the XY plane (stretching mode), 3) vibration of centre nuclei along the y-axis in the XY plane (stretching mode), and 4) vibration of the centre nuclei along the z-axis or out-of-plane vibration in the XZ plane (bending mode). All these vibrations along with translational and rotational motions may take place simultaneously and result in seemingly erratic motions. However, the total motion is a sum total of all vibrations and other motions occurring simultaneously. In a typical vibrational Raman spectroscopy, the difference in energy between the incident and Raman scattered photon is equal to the energy of vibrations of the scattering molecules. Since the energy difference or Raman shift is unique to specific vibration of a molecule, Raman spectra can provide chemical fingerprint of the sample.

Raman spectra provide information complementary to IR spectra. IR spectroscopy is an absorption spectroscopic technique where the frequency of incident photon is same as that of the molecular vibration, resulting in absorption of energy. Absorption of energy depends on change in the dipole moment as opposed to change in polarizability due to induced dipole moment in Raman scattering. IR and Raman generally exhibit a mutual exclusion rule – vibrations that are Raman active are IR inactive and vice versa. Symmetric or in-phase vibrations generally are Raman active while asymmetric or out-of phase vibrations are IR active. But, there are few vibrations that are both Raman and IR active.

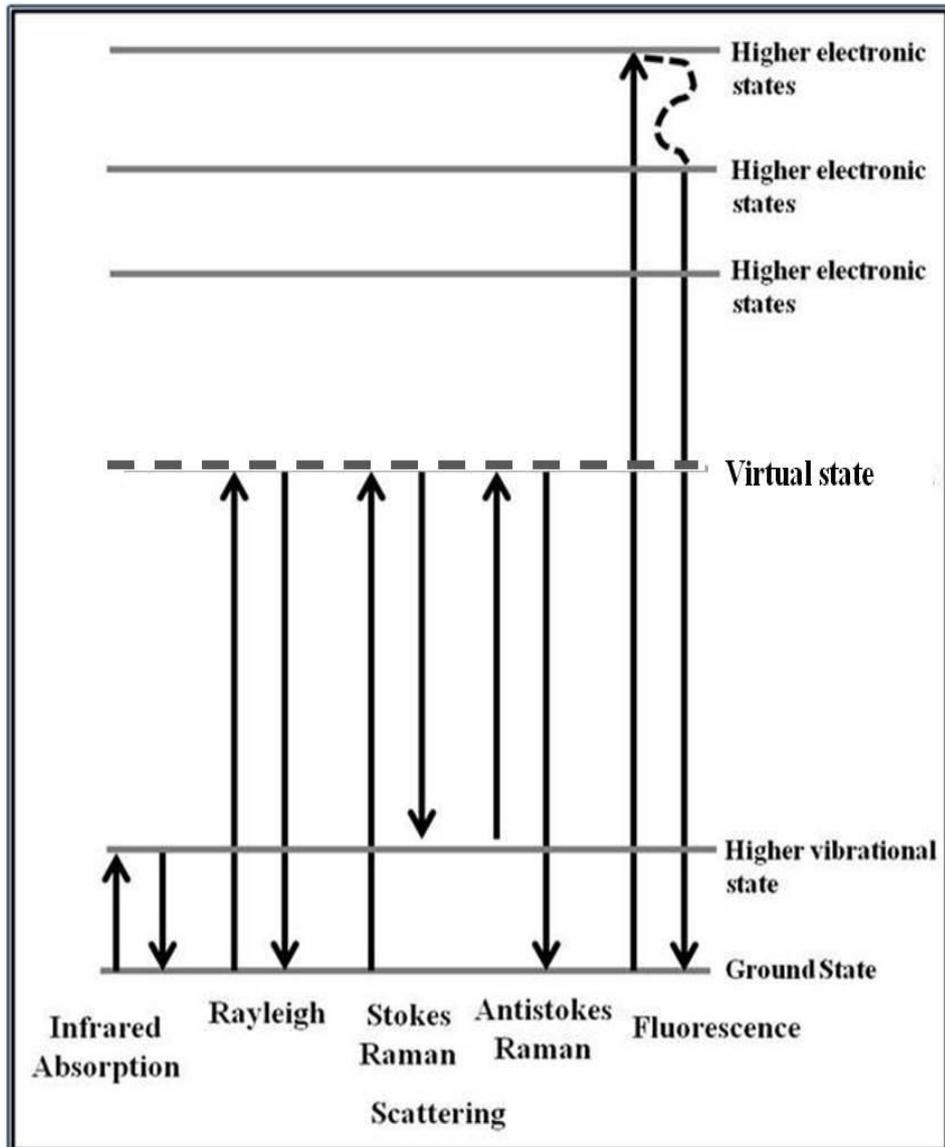


Figure 1.2 Figure shows the Jablonski's diagram for IR absorption, elastic and inelastic scattering and fluorescence

## **Raman spectroscopy – instrumentation**

This section describes some of the key components of Raman instrument. Raman instruments are of two types – interferometric and dispersive. The dispersive instruments use a fixed grating to disperse the collected signals and a CCD to measure the signals. The low laser power requirements compared to Interferometric instruments and feasibility of manufacturing compact instruments have rendered this more popular than the Interferometric type. The following brief description of dispersive Raman instrumentation is derived from the books – Analytical Instrumentation Handbook edited by Jack Cazes, Handbook of Raman spectroscopy: From research laboratory to the Process Line edited by I. Lewis and H. Edwards and Emerging Raman Application and techniques in biomedical and pharmaceutical field edited by P. Matousek and M. Morris (45, 47, 48). A typical Raman spectrometer consists of an excitation source, optical elements and a detector (Figure 1.3).

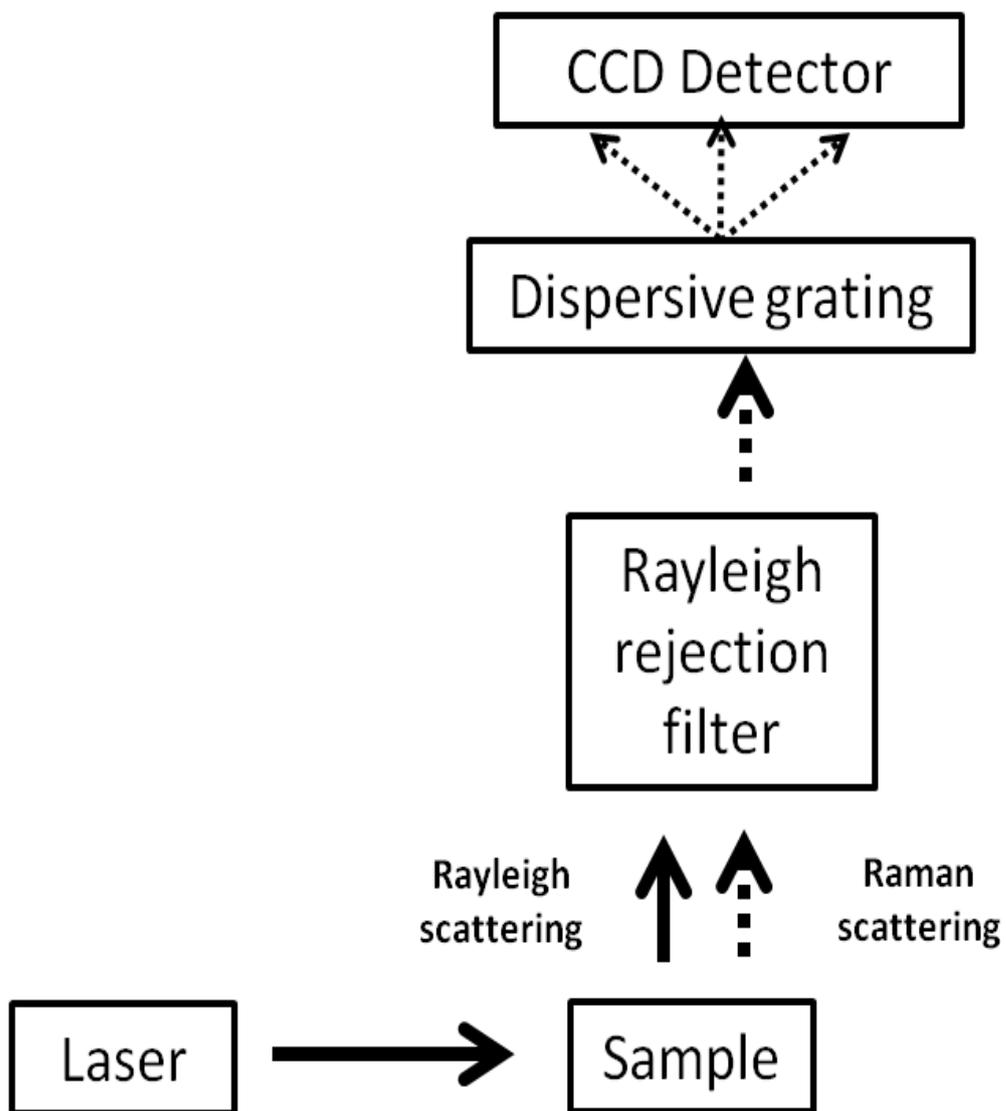


Figure 1.3 Raman spectrometer schematic. Figure illustrates components of a dispersive Raman spectrometer.

## **Excitation source**

A powerful excitation source is required to induce the weak phenomenon of Raman scattering. Lasers are the most preferred source to excite Raman signals enough for experimental purposes. Several key issues need to be considered while choosing a laser. One factor is laser stability, since output instability may result in intensity variations while wavelength instability may cause shifts in Raman spectrum. Laser power is another issue. Increasing the laser power will increase the strength of signal, but will result in corresponding increase in Rayleigh signal which might interfere with signal collection. Intense lasers may also damage the sample. Sample fluorescence, which may swamp the Raman signal, also needs consideration. Use of longer wavelengths will reduce fluorescence, but will also invariably lead to loss of signal strength. A trade off involving lower fluorescence and higher laser power may solve the issue. Early Raman spectroscopy studies employed visible range gas lasers – HeNe (632.8nm), argon (488 and 514.5nm) lasers, krypton (530.9, 568.2 and 647.1nm) or mixed argon/krypton lasers. These lasers had power output up to 1W. Use of intracavity  $\beta$ -barium in argon lasers resulted in ultraviolet (UV) and visible lines (228.9 – 457.9 nm). The UV lasers have applications in Resonance Raman spectroscopy described later. The problem with these lasers were cost effectiveness and decrease in power output with time due to progressive loss of gas pressure, electrode erosion and laser tube window deterioration. Continuous wave Nd:YAG laser (532nm) later became popular for process-oriented Raman applications due to its stability and small size. The fluorescence problem was dramatically reduced by use of Near Infrared (NIR) lasers like Nd:YAG operating at 1064nm. Initial xenon lamp discharge lasers were soon replaced by laser diode-pumped systems, which were smaller, more efficient and had low noise. However 1064nm lasers are not

compatible with CCDs. Alternatives such as 720, 785 and 830nm diode lasers also reduced fluorescence, but could be used with CCDs, increasing the signal detection limits.

### **Detector**

The benefit of using multichannel detectors like CCD is that the grating (dispersion element) can be static. Since there are no critical movable parts related to frequency measurement scale, the optical system is highly stable and reproducible. For such systems, dispersion linearization can be achieved by acquiring spectrum from a standard source and applying correction to subsequently recorded spectra. The spectral range and resolution is however fixed, depending on array size, grating and system aperture/slit. For a 1024 pixel array, spectral resolution and minimum digitization of one point are fixed at  $8\text{ cm}^{-1}$  and  $4\text{ cm}^{-1}$  respectively; if spectral acquisition range is chosen to be  $0 - 4000\text{ cm}^{-1}$ . To increase the resolution, spectral range needs to be decreased. Other option is to increase the pixel array size, such as 2048 arrays which became available in early 2000s. CCD performance depends on amount of dark current generated and chip fabrication/ implementation. The dark current can be reduced by cooling and efficient heat dissipation. By using thinned, back illuminated devices, high quantum efficiency can be achieved. Moreover, since a single wavelength illuminates a particular column of pixels, the output from each pixel can be combined to improve the signal. This process is referred to as binning. It should be noted that CCD cannot be used when longer wavelength excitation like 1064nm is used. In such cases, an interferometer combined with NIR sensitive detectors are used in tandem.

### **Sample illumination and collection**

Due to inherent weakness of the Raman signal, focusing of the incident radiation and efficient collection of the scattered signal is critical. Laser is a coherent source with very

small beam diameter ( $\sim 1\text{mm}$ ) making focusing of incident radiation easy. Several illumination and collection geometries may be used for maximum signal generation and collection. Some of the commonly used geometries are - near  $360^\circ$  backscatter or  $90^\circ/180^\circ$  straight through geometries. The collected signal consists majorly of Rayleigh scattered light. Rejection of the Rayleigh line from the scattered signal is achieved efficiently by holographic notch filters. These filters consist of multiple film layers fabricated by recording interference patterns formed by two mutually coherent lasers. These filters have high optical density, extremely narrow bandwidth, high laser damage thresholds and are free from extraneous reflection bands. The next important element is the wavelength selector. This decides the spectral range that will be recorded by the selector and the resolution of the spectrum. The major function of the wavelength selector is spatial separation of photons of different frequencies before reaching the detector. Simple selectors like interference filter generate constructive interference to transmit a fixed number of wavelength proportional to the thickness of the filters. Prism/ grating monochromators are more popular for use as wavelength selector in Raman instruments. Resolution of the instrument is defined by the wavelength region selected and dispersion of the grating within the selected region. Dispersion of the grating in turn depends on the ruling density – higher the lines per millimeter, greater is the resolution. As mentioned earlier, higher resolution means narrower spectral range covered by the grating. Commonly used ruling densities are 1000 and 2000 lines/mm. For some time now, holographically recorded diffraction gratings have completely replaced mechanically ruled gratings. Holographic grating are superior to their mechanically ruled counterpart in several aspects – they form a continuous phase diffractive structure, has very high diffraction efficiency, are cleanable and are extremely sturdy.

## **Raman spectroscopy – variations**

The Raman scattering phenomenon is weak – only 1 in  $10^6$  photons scatter inelastically. Resonance Raman spectroscopy circumvents this problem by using a laser excitation frequency that matches the electronic absorption band of a chromophore. This enhances the intensity  $10^3 - 10^6$  times (44, 49, 50). Another approach to increase the Raman signal surface enhanced Raman spectroscopy (SERS). In this technique, an enhancement material like colloidal solution/nanoparticles of metals (gold, silver) is added to the sample. Alternatively, the SERS active nanoparticle may be coated onto the sample (51) or the sample may be coated onto the nanoparticle (52). Molecules of the sample that get adsorbed on the enhancement material exhibit signal enhancement in the order of  $10^6$  (53, 54). Alternatively, non linear technique like Coherent anti Stokes Raman spectroscopy (CARS) may be used to increase signal (55-57). In this technique, all molecular bonds oscillate in phase and interfere constructively, increasing the signal by several orders of magnitude. Another variant, Spatially offset Raman spectroscopy (SORS) enables non invasive investigation of deeply buried regions of turbid samples such as subsurface cancers, concealed explosives, etc. This approach takes advantage of the diffused scattered light propagating through deep layers (58). Two important variants relevant to the thesis are described in detail as follows.

### Fiber-Optic probe based Raman spectroscopy

This technology (59, 60) has been developed for remote sensing, that is to obtain spectra from samples that are inconvenient to measure near the instrument. The technique is of special advantage to probe inaccessible organs like lung, stomach, etc. non-invasively in patients to detect cancer and other diseases. One of the major issues with fiber optic probes is the fiber spectral background (FSB) generated in the optical fibers. Use of pure silica fibers is one of the solutions. These fibers are wavelength dependent; therefore fibers have to be

manufactured with respect to the excitation wavelength. Commonly used designs are the confocal type head probe and miniaturized fiber optic probe. The former has a head consisting of filters for isolating Raman signals whereas the later has filters directly mounted on the end of optical fiber. The spectrometer specifications are dictated by the fiber probe used. Thus, the entire instrument design depends on the fiber probe.

### Raman microscopy

Raman microprobe is the combination of Raman spectrometer and a microscope. Cost-effective, bench-top microprobe spectrometers based on a single monochromator and CCD detector have found extensive applications in industrial, analytical and forensic laboratories. The major advantage is the micro sample analysis using this technique. Raman microspectroscopy wherein spectra are acquired from a thin layer of sample around the focal plane is called confocal Raman spectroscopy. In this, signals from focused layer of the sample are recorded resulting in confocal spectra acquisition. Raman imaging is the study of an area on the surface of the sample by global illumination of the area and imaging directly onto the CCD. By using liquid crystal tunable filters (LCTF), complete Raman spectrum per pixel can be acquired. In such instruments, the scattered light from a portion of sample globally illuminated passes through the LCTF before being imaged onto a pixel of the CCD. The LCTF allows spectroscopic analysis of complete Raman spectrum at each pixel allowing thousands of spectra from the sample area illuminated are acquired and analyzed using image generated by software manipulation of the data. Raman mapping is similar to Raman imaging where in spectrum from several points of the sample is obtained in a grid pattern over an area by moving the sample. However, the process is time consuming since large number of points needs to be recorded to generate an image. These techniques and their applications have been excellently reviewed recently by Stewart et.al. and Abramczyk et.al. (61, 62).

## **Raman spectroscopy – data analysis**

The Raman spectrum provides the chemical fingerprint of a sample. While the basic aim is to derive chemical information from the sample, this may not be sufficient for biomedical investigations. In biological samples, the primary aim is to discern differences between the sample and its altered states in order to easily distinguish subsequent samples. For example, it is desirable to extract discriminating parameters between a normal breast tissue and breast tumor tissue (an altered state of the normal tissue) from the spectra and help identify breast cancer in future samples. This can be achieved by two types of analysis – univariate and multivariate. In univariate analysis, single bands/peaks are considered. For example, in a study of melanoma by Martin et. al. they observed differences in distribution of polysaccharides, tyrosine and Amide I band distributions between normal from abnormal. They used these individual bands to distinguish normal and abnormal (39). Other groups have also applied similar approaches (11, 59). However, due to heterogeneity of samples and complexity involved in distinguishing several disease conditions (dysplasia, hyperplasia, carcinoma in situ, for example) from normal, many groups prefer multivariate analysis (10, 31, 63, 64). In multivariate analysis, the global profile changes are considered as opposed to individual bands. Classical chemometrics involves decomposition of sample – variable matrix to identify a small number of parameters that can describe the whole data set, the rationale being that the entire dataset is repeated measures of a small data set that define sample and spectral properties. In biomedical context, classification tools to distinguish diseased state from normal are used using the underlying principles of classical chemometrics. An in depth discussion of data analysis techniques can be found in book chapter by R. Reddy and R. Bhargava entitled ‘Chemometric methods for biomedical Raman spectroscopy and imaging’ (45).

## **Data pre-processing**

Data pre-processing is used to remove spectral contributions other than the sample and to improve the signal quality. The first step baseline correction involves removal of fluorescence signal. One of the methods for baseline correction is to fit a polynomial and subtract from the spectrum. This can be achieved by using an algorithm; many of which are available. Next, spike noise generated by the CCD detectors need to be removed. This can be achieved by median filtering, summation approach of comparing several spectra from the same sample or using transform techniques. Likelihood estimators that distinguish spikes based on prior knowledge of sample spectral shapes may also be used. Other preprocessing steps may also be carried out depending on the experiment and expected outcome. National Institute of Standards and Technology (NIST) based calibration or white light correction to compare spectra acquired from different instruments is one such step. In some experiments like *in vivo* spectroscopy, it may be advisable to record the environmental background noise and subtract the same from the acquired spectra. Spectral interpolation to analyze specific set of bands associated with the biological phenomenon under consideration may improve the final analysis. In many experiments, the change in Raman spectral profile rather than band intensities is important for a biologically relevant result. In such cases, normalization of the spectra to remove intensity related changes may be carried out.

## **Multivariate analysis**

### ***Unsupervised methods***

Principal Component Analysis (PCA) is one of the most popular unsupervised multivariate analysis methods. The analysis is called ‘unsupervised’ since no prior knowledge of the sample class or group is provided. PCA (65) explores the variation in the data, extracts

spectral features contributing to the variation (eigenvectors) and arranges them in the order of importance based on their contribution to the variation (eigenvalues) - such that eigenvectors effecting maximum variation is placed first and so on. The data set is then transformed with the highest eigenvectors forming the x and y axes (and z axis in 3D plots) of the new system, displaying variation in the data. PCA is a simple, robust and reliable analysis method. It is usually used to reveal patterns in unknown data as well as to remove noise. It can be used in conjunction with supervised techniques where PCA supplies non-redundant spectral features for further analysis. Drawbacks include incompatibility with non linear data and high computational times in case of very large data sets. Other unsupervised methods include k-means clustering and Hierarchical Clustering Analysis (HCA). In k-means clustering, the user defines number of groups expected in the data. Equivalent number of centroids is generated. Every data point is then clustered with the nearest centroid, followed by recalculation of centroids based on the achieved clusters. The centroid recalculations are carried out till there is no change in the centroid value. The clusters thus obtained after several rounds of centroid recalculations is the final result. Although, it is a good exploratory technique, the user has to provide expected number of clusters which may not be known. Moreover, error in initial point selection may skew the outcome. In HCA, samples are organized into clusters based on their relative distances. HCA can be done in two ways. In one approach, the whole data set is considered as a single object which is successively broken down into smaller groups based on similarity in distances between the points. The other approach is the reverse of the first; every sample is considered individual group and are successively grouped into larger clusters based on relative distances. This method needs no prior knowledge of groups and is an excellent exploratory method (66).

### *Supervised methods*

Supervised method involves training an algorithm or establishing a function to discriminate between two groups so as to allow classification of unknown sample. The success or failure of the training is evaluated by independent data set (validation). Several methods have been used for classification of biological samples. In Bayes's classifier method, spectral profiles of each class and intra class distribution are used to build a probabilistic model. The model then gives the probability of an unknown spectral profile belonging to a particular class. The method however requires large training data set to build an accurate model. Artificial neural networks (ANN) use several layers (input, intermediate and output) to select spectral features that will force classification into designated classes. Although efficient in handling complex distribution and large number of parameters, the method is prone to overfitting or underfitting resulting in poor outcome during validation. Support vector machines (SVM) use classifying hyperplane coupled with margins called support vectors to achieve maximum separation between classes. It may also use non linear kernels to distinguish complexly distributed classes. SVM also suffers from the problem of overfitting. Linear Discriminant Analysis (LDA) defines a linear function to effectively increase inter class variance and reduce intra class variance and classify different groups. This method has been used widely for biological data. In cases where linear function cannot distinguish the groups, non linear functions like quadratic (Quadratic Discriminant Analysis), factorial (Factorial Discriminant Analysis) and partial least square (Partial Least Square Discriminant analysis) equations may be used to classify the data. Further information on this topic can be gleaned from several sources (67-72).

## **Raman spectroscopy – applications**

### **General applications**

The technique is used in solid state physics for material characterization, determination of nanocrystals, chirality and semidiameters in nanomaterials, semiconductor impurities determination, finding crystallographic orientation, real time monitoring of polymerization reaction and for identification of the principal mineral phases or classification of rocks in geology and mineralogy. It plays an important role in pharmaceutical industry for control of quality and purity of pharmaceuticals, active substances and excipients (even through packing) and identification of adulterated pharmaceuticals. Identification of unknown or hazardous substances, trace amounts of substances in evidential materials such as paints, inks from documents, pigments, explosive particles, inflammables, drugs, illegal active ingredients, fibers, gunpowder residues, chemical and biological agents, plastics are its application in forensics and security. Non destructive investigation of paintings and archaeological remains are also performed using Raman spectroscopy.

### **Biomedical applications**

Biomedical applications include a) study of dental hard tissues, calculus, mineral components in enamel, dentin, formation of calcium fluoride on enamel; b) study of ocular fluids, development of ocular pathology, detection and tracking of ocular drugs; c) study of chemical alterations in subchondral bone, correlating changes in collagen secondary structure with aging; d) diagnosis of osteoarthritis using synovial fluid; e) detection of pathological diseases; f) glucose measurement using blood and urine; g) analysis of viral and bacterial chemical components, monitoring chemical differences occurring as a result of the growth of microorganisms, evaluate the interaction of microorganisms with active pharmaceutical agents; h) non-destructive assessment of the physical, chemical, and mechanical characteristics of load-bearing parts in arthroplastic components (that is, artificial joints); i) the characterization of atherosclerotic plaques; j) detect Alzheimer's disease brain tissues; and

k) diagnosis of cancer. With respect to cancer diagnosis, several studies have shown possibility of distinguishing skin, cervical, GI tract, lung, kidney and several other cancers including breast cancer from normal tissues (8). More detailed description of Raman biomedical applications can be found elsewhere (45). Identification of Cancer Field Effects/ Malignancy Associated Changes, pre cancer lesions, hyperplastic and dysplastic patches using Raman spectroscopy has also been reported for some cancers (73-75). The application of Raman spectroscopy in breast cancer management is elaborated in the succeeding section.

### **Raman spectroscopy – existing literature on breast cancer management**

Identification of breast cancer by Raman spectroscopy has been well documented in literature. In 1991, Alfano *et.al.* reported the difference between Raman spectra of normal and malignant breast tissues for the first time (76). Since then, several studies have demonstrated not only spectral difference between normal and malignant tissues, but also feasibility of classifying them using multivariate statistical analysis (40, 63, 77-79). The Feld group worked out the major components contributing to the spectra using basis spectra fitting (10). They considered the spectra as a linear combination of basis spectra from biochemical components. They then acquired spectra from several components such as nucleus, cytoplasm, lipids, calcium salts, etc. and fit each spectrum to the spectrum obtained from normal/ malignant breast tissue using least squares fitting. Using this method, they could easily classify normal, malignant and benign tissues. Alternatively, Krishna *et. al.* delineated the spectral components using spectral deconvolution and arrived at the same conclusion (9). Use of low cost, miniaturized Raman spectrometer and resonance Raman spectroscopy for classifying malignant tissues from normal was reported (80, 81). In depth study of lipid composition to better understand its role in carcinogenesis was carried out using Raman imaging (38, 82). In order to increase the speed of spectra acquisition and decrease the output

time, CARS was applied to breast tumor tissues. This rapidly generated Raman images clearly showed several histopathological features characteristic of ILC, intermediate and high grade IDC and fibroadenoma. These features, in turn, enabled distinction between the mentioned pathological conditions, suggesting that CARS may be applied as a rapid, objective adjunct to histopathology (83, 84).

Identification of breast cancer risk factors – premalignant conditions such as ADH, DCIS and IDC has been reported in human tissues using SERS by Xu *et. al.*(19), while classification of atypical cells – another risk factor, has been shown in animal models by Kast *et.al.* (79). Similarities between ageing process, another important risk factor and process of carcinogenesis were pointed out by Abramczyk *et .al.* (85). Nima *et.al.* demonstrated identification of a single cancer cell in a milieu of normal cells (86). It is well known that cancer begins with mutations in a few cells which selectively outgrow normal cells to form a tumor. The ability of Raman spectroscopy to identify single abnormal cell demonstrates its potential to identify malignant cells before they form a tumor. A study on biochemical changes associated with transformation of a cell line (87) was carried out by Damayanti *et.al.* Ability of the technique to detect transformation associated changes also shows its ability to detect sensitive biochemical changes that may precede tumor development. Detection of microcalcification and type of microcalcification (12, 17, 88-90) was shown by Feld *et.al.* Microcalcification, an abnormality observed even before advent of palpable tumors, can be used for early detection. Moreover, identification of type of microcalcification, that can help distinguish benign from invasive tumors, cannot be determined by radiography but can be revealed by Raman spectroscopy. Lee *et.al* have reported detection and quantification of early cancer markers – epidermal growth factor (EGF), insulin like growth factor (IGF-1) and HER-2 using SERS (91) using SERS. These early cancer markers may also help early

detection. These studies underline the potential of Raman spectroscopy for early detection of breast cancer.

Raman spectroscopic studies have also focused on surgical margin assessment, detection of prognostic markers and identification of resistance phenotype. Krishna *et. al.* have reported classification of multidrug resistant breast cancer cell line from its drug sensitive counterpart (92). The first surgical margin assessment study in human subjects was undertaken by the Feld group (93). During partial mastectomy surgery, they acquired spectra from tumor margins and determined surgical cut offs. Interestingly, they detected grossly invisible tumor which required the patient to undergo a second surgical procedure after pathologic review. Since this demonstration of sensitivity, Raman spectroscopy has been actively pursued for applications in breast cancer treatment. Detection of prognostic markers like EGFR and HER-2 (18) as well as difference between cell lines with HER-2 overexpressing plasmid and cell lines without HER-2 plasmid has been demonstrated (13). Feasibility of distinguishing drug resistant cell lines from normal (94) and possibility of monitoring drugs as well effect of photodynamic therapy has been shown (95). Surgical margin assessment rapidly using CARS and deep in tissues using SORS has also been shown (16, 96).

Raman spectroscopy has been used to explore the metastatic aspect of breast cancer. Feasibility of distinguishing cancer affected lymph nodes indicating metastasis from normal lymph nodes (14, 64, 77), identification of epithelial mesenchymal transition (EMT) phenotype (97), capability of discriminating metastatic lesions from primary tumors and possibility of identifying primary cancers (98) has been demonstrated. Lymph nodes facilitate cancer metastasis, hence are the first to be affected by cancer. Lymph nodes devoid of cancer cells indicate localized cancer which has very good prognosis. Thus, identification of metastatic lymph nodes is critical to treatment. Another important aspect is identification of a

lesion as primary or metastatic, since treatment varies for primary cancer and metastatic cancer. Moreover, treatment is also guided by the identity of primary cancer. Raman spectroscopic identification of lymph node metastasis, metastatic status of a lesion and primary cancer of a metastatic lesion may thus help guide therapy and improve prognosis.

### **III. Aim and Objectives of the thesis**

#### **Raman spectroscopy in breast cancer management – unexplored avenues**

As described in the review of literature, several groups have focused on identification of premalignant conditions using Raman spectroscopy for early detection of breast cancer. However, literature suggests that premalignant conditions are indicators for risk assessment; for example - the risk of developing breast cancer in women presenting premalignant lesions atypical ductal hyperplasia (ADH) and ductal carcinoma in situ (DCIS) are 4-5 and 8-10 times higher than those that do not exhibit such abnormalities (3). Thus, detection of a premalignant condition may not be a true predictor of prospective tumor development. The only true indicator of confirmed tumor development is the appearance of the tumor itself. Hence, the objective of this study is to find an association between spectral changes and tumor appearance.

For this, sequential follow up study of subjects from healthy condition till tumor development is imperative. Such studies are extremely difficult in human subjects, since patients in hospitals mostly present advanced stages of breast cancer. To circumvent this problem, rodent models can be used. Rodents can be treated with carcinogen and spectra can be acquired at regular intervals post carcinogen treatment till tumor appearance (Figure 1.4). The spectra acquired before tumor appearance can then be evaluated, correlated with tumor appearance and used to predict tumor occurrence in test rodents.

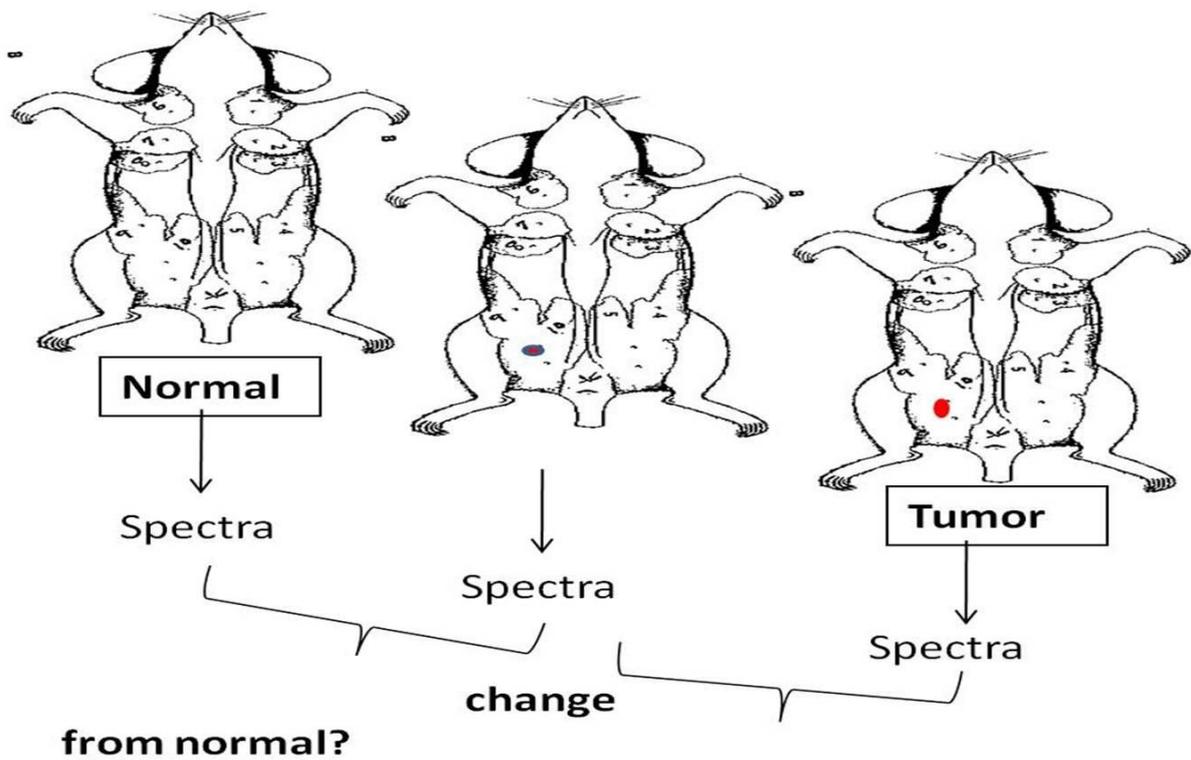


Figure 1.4 Overview of study methodology and rationale: Figure shows three rodents and the location of the breasts. The first rodent is normal with no tumor (before carcinogen treatment), in the second rodent (representing the first rodent few weeks after carcinogen treatment) a change is observed (red dot with blue border) and in the third rodent (which represents the same rodent several weeks post carcinogen treatment) a tumor is seen (red circle). Spectra can be acquired from site of tumor in breast after carcinogen treatment till tumor appearance. Spectral changes prior to and after tumor appearance can then be evaluated.

Alternatively, bio-fluid based Raman spectroscopy can also be used for early detection of cancer. As mentioned earlier, feasibility of using serum-based Raman spectroscopy for identification of breast cancer patients has been shown (11). Other bio-fluids can also be explored to achieve the same. The current study investigates feasibility of using urine-based Raman spectroscopy for breast cancer diagnosis as well as screening. Moreover, serum based Raman spectroscopy can have other applications such as treatment prognosis and drug monitoring among others. In this context, the current study evaluates differences in serum before and after breast tumor resection surgery in rats.

Another unexplored avenue is the use of fiber-optic probe based Raman spectroscopy for classification of metastatic breast tumors. Several studies have shown the inability of currently available tools to distinguish primary lung lesions from metastatic breast lesions (99). Therefore, this study investigates the possibility of using fiber-Raman to distinguish primary and metastatic cancers. Combined with low dose computed tomography (LDCT) that identifies lesions in lungs (100), but cannot distinguish as primary or metastatic – a distinction critical for therapeutic decisions, fiber-Raman may help guide therapy and improve prognosis.

**A study of breast cancer progression in rodent models using Raman spectroscopy - aim and objectives:**

The aim and specific objectives of the study are as follows:

Aim:

To sequentially follow chemical carcinogen/cell line induced breast neoplasm in rodents and identify spectral signatures of precancerous, malignant and metastatic stages.

Specific objectives:

1. To develop a rodent model of breast neoplasms suitable for characterization by Raman spectroscopy.
2. To study development of breast neoplasms induced by chemical carcinogen using Raman spectroscopy in target organs and body fluids.
3. To study Raman spectral signatures of experimental lung metastasis from breast cancer cell line.

**CHAPTER 2    DEVELOPMENT OF A  
RODENT MODEL OF BREAST NEOPLASMS  
SUITABLE FOR CHARACTERIZATION BY  
RAMAN SPECTROSCOPY**

The major objective of this study is to sequentially follow a rodent from carcinogen induction till tumor development. The study cannot be carried out in human subject due to ethical and practical consideration, hence rodent models were considered. Several models have been used to study breast carcinogenesis over years, of which rats and mice are most common (101, 102). This chapter details the exploration of a model suitable for *in vivo* spectroscopy that will enable sequential follow up of the animals and carcinogenesis protocols that will allow spectral correlation with tumor appearance. In section I, transcutaneous *in vivo* spectra were acquired from breasts of different rat and mice strains to find which of them gave the best spectra. Section II evaluates possibility of uniquely identifying breast from other anatomical sites and breast tumors using transcutaneous *in vivo* spectra. Section III describes transcutaneous follow-up of physiological conditions – pregnancy, lactation and aging. Finally, section IV investigates several carcinogenesis protocols for the one best suited for transcutaneous *in vivo* cancer progression studies.

## **I. Exploring rodent strains suitable for transcutaneous *in vivo* spectroscopy**

The first step involves screening different mouse and rats strains for one that gives the best transcutaneous *in vivo* breast spectra - spectra that are closest to spectra reported in literature and *ex vivo* spectra.

### **Material and methods**

All animal studies included in this thesis were approved by Institutional Animal Ethics Committee, ACTREC endorsed by the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA), Government of India guidelines (Project no. 19/2011 and 28/2012). All animals were housed under standard laboratory conditions, fed a diet of in-house-prepared pellets and provided with water *ad libitum*.

## **Animals**

Approximately 10 spectra were recorded transcutaneously from inguinal breast of C57 mice (n=2), Agouti mice (n=2), Swiss Albino mice (n=2), Swiss Bare (SB) mice (n=2), a hairless mutant of Swiss Albino mice (103) and Sprague-Dawley (SD) rats (n=2) for choosing the best species/strain. Skin in the inguinal breast region of SB mice was incised to expose breast and has been referred to as 'exposed breast' in the rest of the write up. Ten spectra were recorded from 'exposed breast' of SB mice (n=2). Spectra were also recorded from inguinal breast of rats after shaving.

## **Raman spectroscopy**

All spectra were recorded using the Raman spectrometer (Figure Figure 2.1) (73). This system consist of a diode laser (PI-ECL-785-300-FC, Process Instruments) of 785 nm wavelength as excitation source, a high efficiency spectrograph (HE-785, Jobin-Yvon-Horiba, France) with fixed 950 gr/mm grating coupled with a CCD (CCD-1024X256-BIDD-SYN, Synapse). The spectrograph has no movable parts and spectral resolution is  $\sim 4 \text{ cm}^{-1}$ . Commercial RamanProbe (RPS 785/ 12-5, In Photonics Inc, Downy St. USA) consisting of an excitation and a collection fiber (NA-0.40) of diameters 105 and 200  $\mu\text{m}$ , respectively, was used to couple excitation source and detection system. This probe utilizes a backscattering ( $\theta=180^\circ$ ) sampling geometry. The estimated spot size and depth of penetration as per the manufacturer's specifications is 105  $\mu\text{m}$  and 1mm, respectively. Spectral acquisition parameters were:  $\lambda_{\text{ex}} = 785 \text{ nm}$ , laser power-80 mW, spectra were integrated for 15 seconds and averaged over 3 accumulations.

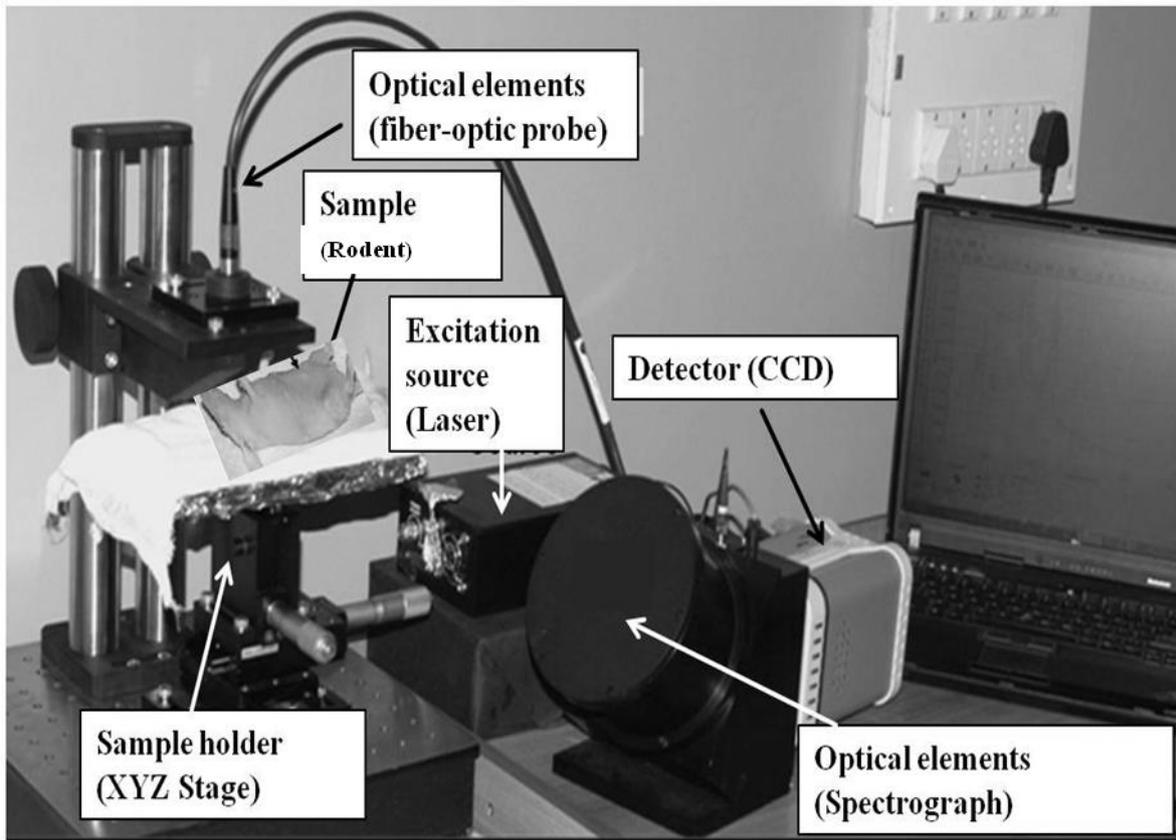


Figure 2.1 Illustration of fiber-optic based Raman spectrometer employed in the study



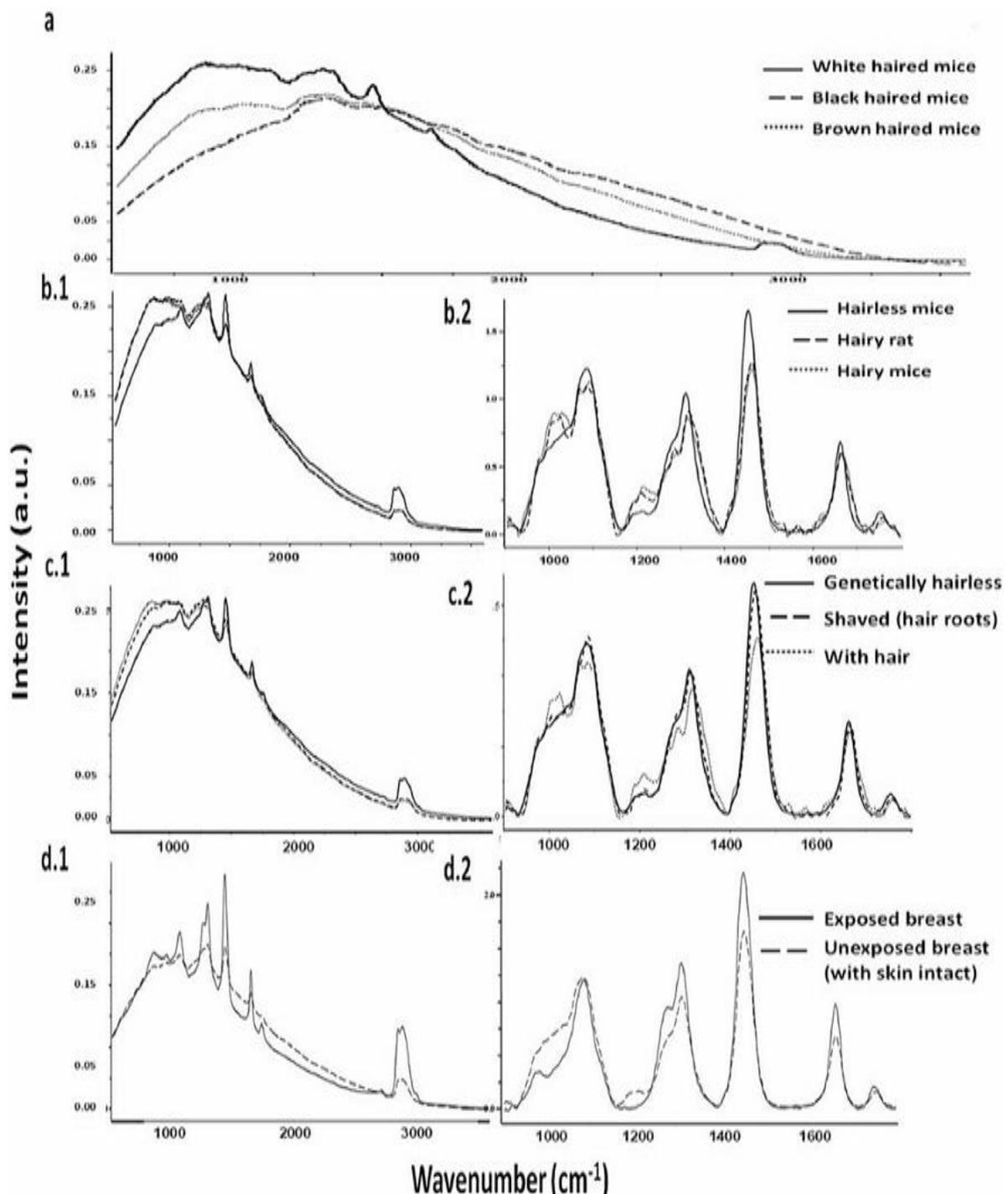
## **Data analysis**

Raw spectra, corrected for CCD and background, were baseline corrected using 5<sup>th</sup> order polynomial fit in LabSpec 4.18. Average spectra were computed from the background subtracted spectra (without derivatization) for each class and baseline corrected by fitting a fifth order polynomial function. The spectra were smoothed post averaging using LabSpec 4.18, (average method, window size 3) for representing the mean spectra. These baseline corrected, vector normalized spectra were used for spectral comparisons. Standard Deviation was also calculated to illustrate intra group variability (73).

## **Results and Discussion**

Spectra of animals with different hair colors were acquired. Typical raw spectra acquired transcutaneously from breast of black haired C57 mice, brown haired agouti mice and white haired Swiss albino mice are shown in Figure 2.2a. As seen in the figure, spectra of colored mice exhibit very high background, whereas white haired mice gave best signal to noise ratio with low background. Thus, further studies were carried out using white haired animals. Next, typical raw spectrum (Figure 2.2b.1) and corrected mean spectrum (Figure 2.2b.2) of white haired SD rat, white haired Swiss albino mice and hairless SB mice were evaluated. Amongst them, transcutaneous breast spectra of hairless SB mice has best signal to noise ratio and is closest to breast spectra reported in literature (9, 10); the major Raman bands being C=O band of esters,  $\delta\text{CH}_2$  bend, two features in amide III and sharp amide I. This suggests that hair may interfere with transcutaneous breast spectra acquisition and hence, hairless mice would be more suitable for such studies. Further, to confirm the effect of hair on spectra; hair was shaved off from the breast region of SD rats and spectra were recorded transcutaneously. Better signal to noise ratio was observed in typical raw spectrum of shaved rat compared to unshaved rodents, although hairless mice had the best signal to noise ratio

(Figure 2.2c.1). The difference between shaved rat and genetically hairless SB mice may be due to roots of hair present in shaved rat; which are absent in genetically hairless mice. Despite high background, corrected mean spectrum of breast post-shaving show spectral similarity to that of genetically hairless mice spectra, whereas hairy SD rats show several additional spectral features; which could be due to contribution of hair (Figure 2.2c.2). Further, as shown in Figure 2.2d.1 and d.2, transcutaneous breast spectrum of genetically hairless SB mice resemble exposed breast spectrum. Resemblance to *ex vivo* spectra suggests that most of the signals are originating from breast. All spectral assignments reported above and hence forth are based on available literature (44).



**Figure 2.2** Choosing a suitable strain for *in vivo* Raman studies. Figure shows different spectra obtained - a) typical transcutaneous raw spectra from inguinal breast of white, brown and black haired mice; white haired mice give the best signal to noise ratio b.1) typical transcutaneous raw spectra b.2) mean corrected spectra of hairless mice, hairy mice and hairy rats (900-1800  $\text{cm}^{-1}$ ); hairless mice spectra have the best signal to noise ratio c.1) typical transcutaneous raw spectrum and c.2) mean corrected spectrum (900-

**1800 cm<sup>-1</sup>) from breast of unshaved rats (with hair), shaved rats (with roots of hair) and genetically hairless mice (no hair/hair-roots); hairless mouse have the best signal to noise ratio d.1) typical raw spectrum and d.2) mean corrected spectrum (900-1800 cm<sup>-1</sup>) of exposed breast and transcutaneous (with skin) breast of hairless SB mice; transcutaneous breast spectrum closely resemble exposed breast spectrum**

Summing up, spectra of hairless Swiss bare mice gave best signal to noise ratio amongst their black haired, brown haired, white haired and shaved counterparts. This suggests that SB mice are better suited for transcutaneous *in vivo* Raman spectroscopy of breast.

## **II. Classification of transcutaneous breast spectra from other anatomical sites and frank tumor**

Although transcutaneous breast spectra of breast closely resembled *ex vivo* breast spectra and that reported in literature, it is important to ensure distinctness of *in vivo* breast spectra from all other anatomical sites, to avoid spectra acquisition from sites close to the breast. This section therefore evaluates the feasibility of uniquely identifying breast spectra from other anatomical sites as well as breast tumor using transcutaneous Raman spectroscopy.

### **Material and methods**

#### **Animals**

Approximately six spectra per site per animal were recorded from breast, scalp, cheek, neck, chest, thigh, shin, spine and tail of SB mice (n=10) to evaluate the uniqueness of breast spectra. The anatomical sites investigated are illustrated in

**Figure 2.3.**

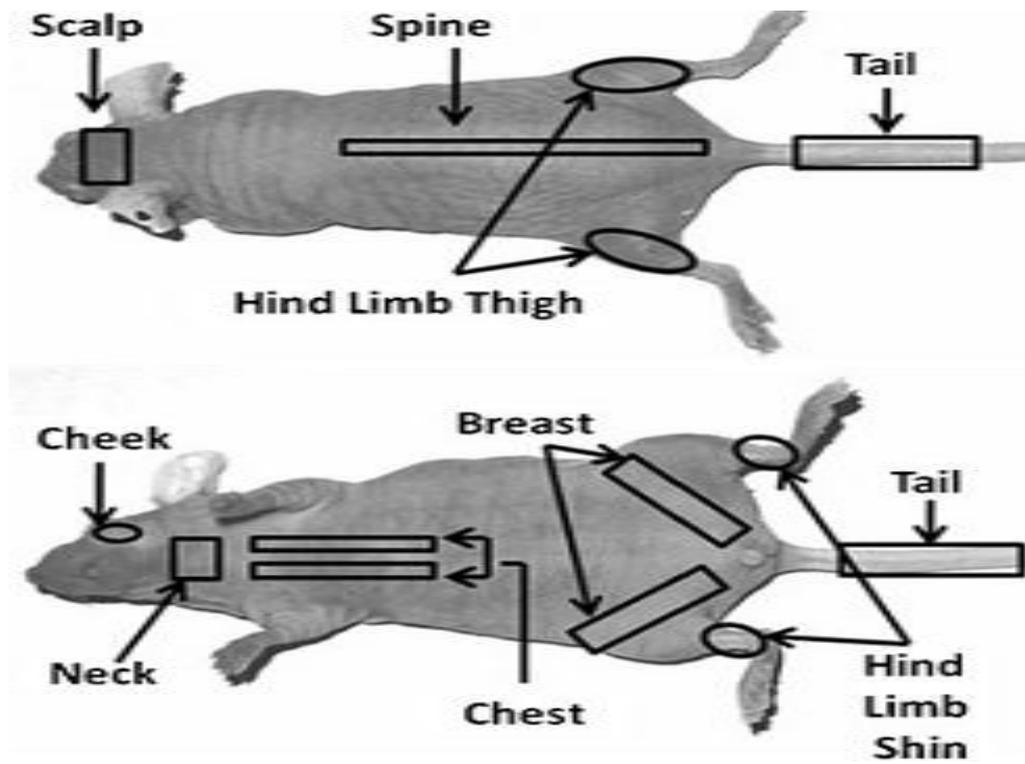


Figure 2.3 Classification of anatomical sites using *in vivo* spectroscopy. Figure illustrates the anatomical sites investigated in the study

## **Tumor transplantation**

In order to evaluate feasibility of distinguishing breast tumor transcutaneously from normal breast, tumors were transplanted subcutaneously in SB mice. ICRC mouse, a well known Mouse Mammary Tumor Virus (MMTV) - induced spontaneous breast tumorigenesis model (104), was used to extract tumors. Tumor transplantation was carried out using established protocols by an expert veterinarian. Briefly, ICRC mouse breast tumor was extracted, cut into small pieces using a scalpel, and washed in normal saline. Histopathological analysis of one such piece confirmed tumorigenesis. The rest were grafted (2 tumors per mice) subcutaneously in inguinal breast of SB mice. Spectra recorded transcutaneously from tumors transplanted in mice (n=2).

## **Raman spectroscopy**

The Raman instrumentation and spectra acquisition parameters have been described in the previous section (Section I Materials and methods).

## **Data analysis**

Spectra of different groups were preprocessed (73) by correcting for CCD response with a NIST certified SRM 2241 material and subtraction of background signals from optical elements. To remove interference of the slow moving background, first derivative of the preprocessed Raman spectra were calculated (Savitzky-Golay, window size 3), interpolated in the 900-1800  $\text{cm}^{-1}$  range (Raman fingerprint region) and vector normalized. Analysis of the preprocessed spectra was carried out using multivariate analysis tool Principal Component- Linear Discriminant Analysis (PC-LDA) implemented in MATLAB (Mathwork Inc.) based in-house software. PCA is routinely used method for data compression and visualization. It describes data variance by identifying a new set of orthogonal features,

which are called as principal components (PCs) that are linear combinations of original data variables. These PCs are calculated by identifying eigenvectors for the covariance matrix of mean-centered data. Because of their orthogonal characteristics, first few PCs are enough to represent maximum data variance. And for visual discrimination, we project each of the spectra in the newly formed co-ordinate space of these selected PCs. While PCA aims to identify features that represent variance among complete data, LDA provides data classification based on an optimized criterion which is aimed for more class separability. LDA is a method of choice when input data has higher within class variance that could lead to development of PC's which are inappropriate for visual discrimination. The classification criterion is identified using the scatter measure of within class and between class variances. LDA transformations are further identified as eigenvector matrix of this classification criterion. With the help of this LDA transform matrix, any test spectra can be classified to a class by iteratively calculating Euclidean/RMS or Mahalanobis distance of transformed test spectra and the mean of transformed input data set. In this study we have employed Mahalanobis distance for class prediction, since it handles nonlinearity well. LDA can be used in companion with PCA (PC-LDA) to further increase performance 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). 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 taking into account contribution of all factors selected for analysis. These results can also be depicted in the form of scatter plots, generated by plotting combinations of

scores of factors. Plotting different combinations of factor scores give visual understanding of classification pattern in the data. The data analysis methodology remains same throughout the thesis (67-71).

## **Results and discussion**

In the previous section (Section I), feasibility of acquiring transcutaneous breast spectra was demonstrated. The spectra show predominance of lipids. However, skin also contains lipids (105). Therefore, the transcutaneously recorded breast spectra may have (a) maximum contribution from skin with little contribution from breast (b) approximately equal contribution from skin and breast / in some ratio (c) maximum contribution from breast with minimum influence of skin. It is important to establish the origin of spectra for breast-related studies, since if (a) is true, transcutaneous breast studies cannot be carried out, while if (b) is true, sufficient information may not be acquired.

To address this, spectra were acquired transcutaneously from breast, scalp, cheek, neck, chest, thigh, shin, spine and tail of SB mice. The mean spectra and standard deviation of all sites interpolated in the range  $900\text{-}1800\text{ cm}^{-1}$  are shown in Figure 2.4. A sharp phosphate band indicative of bone is observed in shin and tail, while lower intensities of the same are observed in cheek and scalp. Predominant lipid features indicated by strong band corresponding to  $\text{CH}_2$  deformation,  $\text{CH}_2$  twisting, stretching of  $\text{C}=\text{C}$  and  $\text{C}=\text{O}$  are seen in breast; and to lesser extent in neck and spine. Muscle signatures such as amide III,  $\text{CH}_2$  twisting, deformation of  $\text{CH}_2$ ,  $\text{CH}_3$  and amide I are seen in cheek, spine, neck, chest and thigh spectra. Tail spectrum consists of collagen specific features like broad amide I band, amide III band indicative of protein backbone, and protein specific bands of  $\text{CH}_2$  and  $\text{CH}_3$  deformation (10, 106, 107).

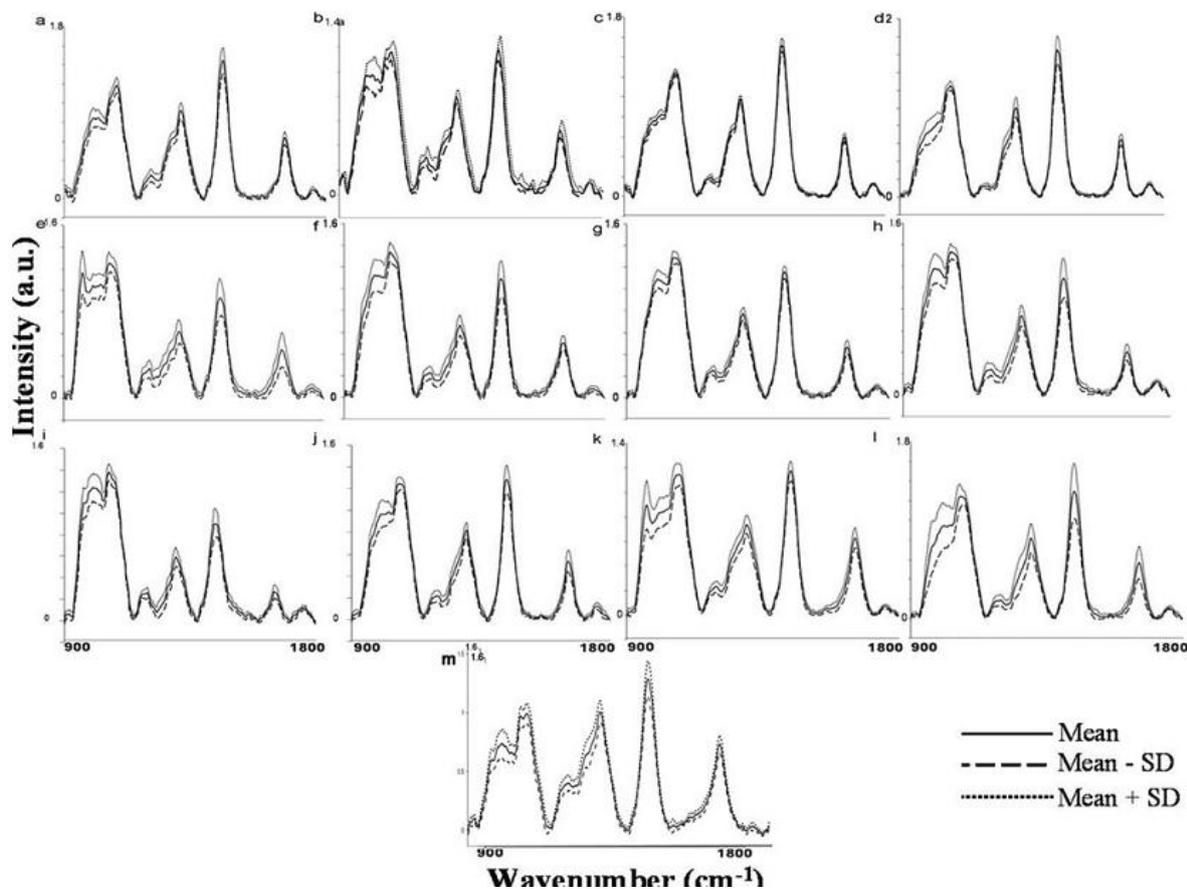


Figure 2.4 Classification of anatomical sites using transcutaneous *in vivo* Raman spectroscopy. Figure shows mean and standard deviation of transcutaneously acquired spectra (900-1800  $\text{cm}^{-1}$ ) from a) hairy mouse breast, b) hairy rat breast, c) shaved rat breast, d) hairless mouse breast, e) hairless mouse shin, f) hairless mouse cheek, g) hairless mouse chest, h) hairless mouse scalp, i) hairless mouse spine, j) hairless mouse tail, k) hairless mouse thigh, l) hairless mouse neck and m) tumor transplanted subcutaneously in hairless mouse breast region.

To explore the differences between these spectra, PC-LDA was used. The scatter plot of PC-LDA factors 1 and 2; shown in Figure 2.5 suggests classification of breast, shin, scalp and tail. Spectra of cheek, neck, chest, thigh, and spine overlap.

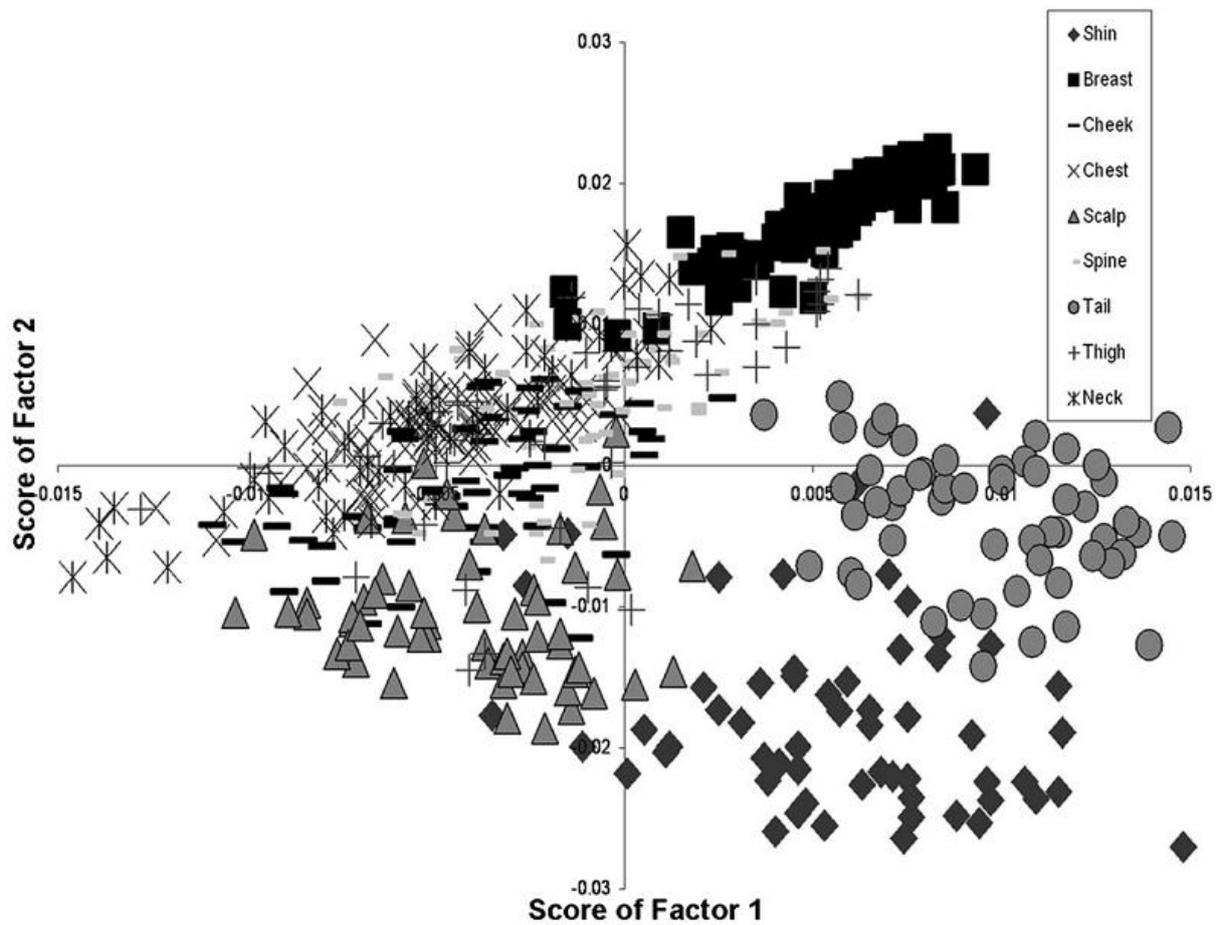


Figure 2.5 Classification of anatomical sites using transcutaneous *in vivo* Raman spectroscopy. Figure shows the PC-LDA scatter plot (factors 1 and 2) of anatomical sites' spectra from hairless SB mice. Discrete clusters are observed for breast, shin tail and scalp, while neck, spine, thigh, cheek and chest show misclassification

Confusion matrix for PC-LDA model and LOOCV is shown in Table 2.1 a and b respectively. 59 out of 63 breast spectra could be correctly identified as breast, while 4/63 spectra misclassified with spine. No breast spectra misclassified with other anatomical sites, although all anatomical sites are covered with skin. This suggests that the influence of lipids in breast spectra is not contributed by skin lipids, but by lipids in the mammary fat pad. Thus, useful information regarding breast and breast-associated physiological or pathological changes can be gleaned using transcutaneous *in vivo* Raman spectroscopy.

<b>a</b>										
Sites ( No. of spectra )	Shin	Breast	Cheek	Chest	Scalp	Spine	Tail	Thigh	Neck	Classification Efficiency (%)
Shin (58)	49	0	2	0	3	0	4	0	0	84
Breast (63)	0	59	0	0	0	4	0	0	0	94
Cheek (60)	0	0	29	5	8	5	0	10	3	48
Chest (60)	0	0	7	15	0	6	0	8	24	25
Scalp (60)	1	0	8	1	48	0	0	2	0	80
Spine (61)	0	9	9	3	1	25	0	7	7	41
Tail (60)	1	0	0	0	0	2	57	0	0	95
Thigh (66)	0	12	6	8	6	20	0	5	9	8
Neck (62)	0	3	11	8	2	15	0	4	19	31
<b>b</b>										
Shin (58)	48	0	2	0	3	0	5	0	0	83
Breast (63)	0	59	0	0	0	4	0	0	0	94
Cheek (60)	0	0	27	7	8	5	0	10	3	45
Chest (60)	0	0	7	15	0	6	0	8	24	25
Scalp (60)	1	0	8	1	48	0	0	2	0	80
Spine (61)	0	9	9	3	1	25	0	7	7	41
Tail (60)	1	0	0	0	0	2	57	0	0	95
Thigh (66)	0	12	6	8	6	20	0	5	9	8
Neck (62)	0	3	11	11	2	15	0	4	16	26

Table 2.1 Classification of anatomical sites using transcutaneous *in vivo* Raman spectroscopy. Table shows the PC-LDA confusion matrix of anatomical sites; a) model, and b) LOOCV (Diagonal elements represent true positive predictions)



Secondly, it is important to ascertain that spectra of breast are influenced by functional mammary tissue in addition to lipids in mammary fat pad. Breast consists of the mammary fat pad, connective tissue, fibroblasts and branched ductal network of mammary epithelium (22). The mammary epithelium is the functional part of the breast which forms milk glands during lactation and is the site of breast cancer. The mammary fat pad and connective tissue provide support to the functional breast tissue. Thus, the breast spectra should have contribution from mammary epithelium in addition to lipids of mammary fat pad. To address this, breast spectra were compared with spectra of neck, spine and cheek. These sites have lipids in the form of adipose deposits. If lipids are the only contributing factor, breast should misclassify with these lipid rich sites. However, as seen in Table 2.1b, breast spectra do not match neck and cheek spectra, while only 4/63 misclassify with spine. Uniqueness of breast spectra from neck, spine and cheek possibly indicates contribution of mammary tissue specific signals along with mammary fat pad to the breast spectra.

The study of anatomical sites provides further insights into the sensitivity of Raman spectroscopy to biochemical components. The sites that can be uniquely identified by transcutaneous *in vivo* Raman spectroscopy are breast, shin, tail and scalp, their classification efficiencies being 94%, 83%, 95% and 80% respectively. These sites have distinct biochemical composition; breast rich in lipids and proteins, tail rich in collagen, shin rich in phosphates, scalp rich in phosphates and proteins. These results further augment the evidence for chemical sensitivity of transcutaneous *in vivo* Raman spectroscopy.

It may be argued that comparison of skin, transcutaneous breast and exposed breast spectra may suffice to evaluate the contribution of biochemical components from skin rather than an elaborate experiment with different anatomical sites. However, skin covers every anatomical site in the body, thus skin spectra will always have influence from subcutaneous anatomical

sites. Recording pure skin spectra *in vivo* is therefore very difficult. *Ex vivo* skin spectra may not provide an ideal comparison. Since the applicability of this model for breast-cancer studies hinges on recording correct transcutaneous breast spectra, irrefutable proof of breast spectra origin was established by analysis of different anatomical sites.

To validate the applicability of SB mice in breast cancer detection, frank breast tumors from ICRC mice were transplanted subcutaneously in inguinal breast of SB mice. Spectra were then recorded transcutaneously from these tumors. The mean tumor spectra and standard deviation is shown in Figure 2.4m. Comparison of mean normal breast and mean tumor spectra is illustrated in Figure 2.6a. Broad amide I and change in C=O ester band compared to normal breast are characteristic of tumor spectra. This can be attributed to increase in proteins and loss of lipid in tumors, corroborating earlier reports (9, 10). To further explore these differences, PC-LDA was carried out. The scatter plot of factors 1 and 2 shows discrete clusters of normal and tumor spectra (Figure 2.6b).

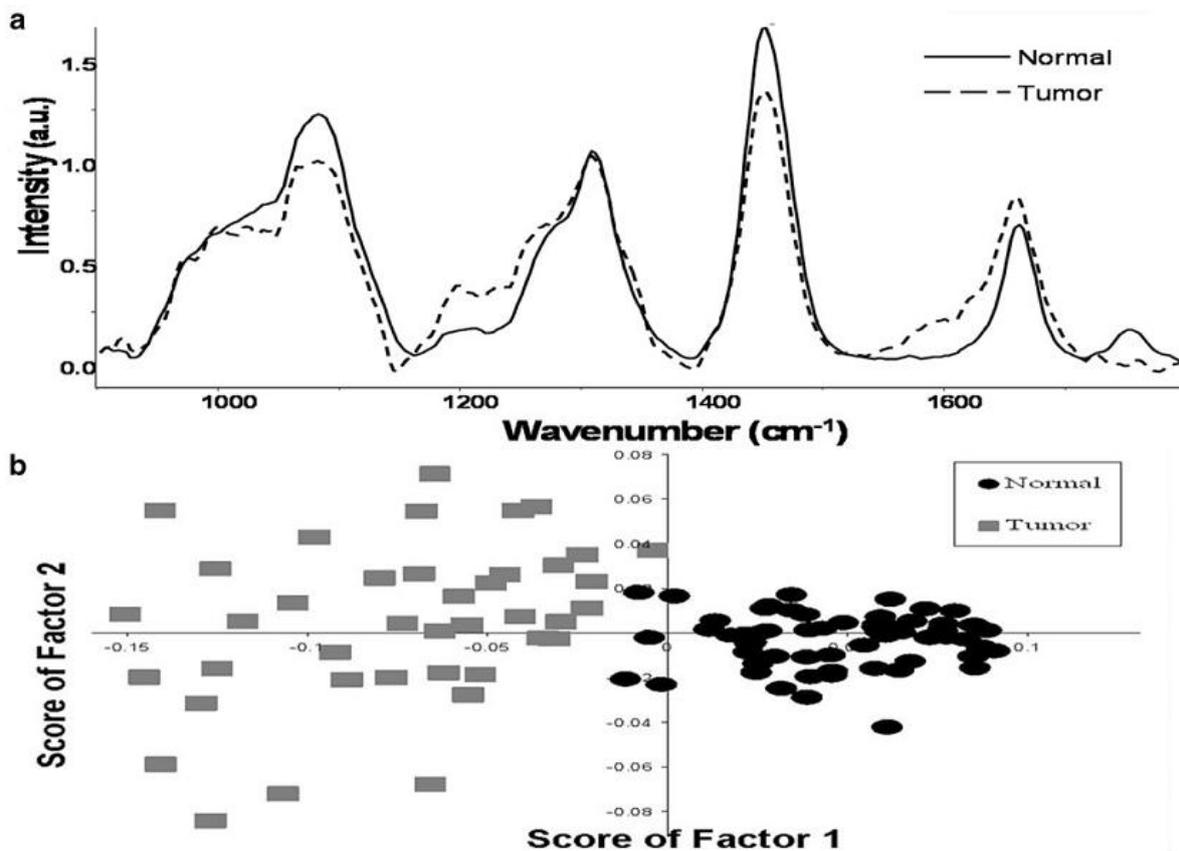


Figure 2.6 Classification of breast tumors using transcutaneous *in vivo* Raman spectroscopy. Figure shows a) mean corrected spectra recorded transcutaneously from normal breast and transplanted breast tumors of SB mice, b) PC-LDA scatter plot of factors 1 and 2, suggesting discrete clusters of transcutaneously recorded normal and tumor breast spectra.

The confusion matrix of PC-LDA model and LOOCV is shown in Table 2.2a and 2b respectively. 40 out of 40 spectra are correctly classified as tumor. 62/63 normal spectra are correctly classified while only 1/63 spectra misclassify with tumor. The classification efficiency for normal and tumor are 98 and 100% respectively. The discriminating efficiency between transcutaneous normal and tumor with SB mice model is better than previously reported model, where efficiency was 73% (107).

<b>a</b>			
	<b>Normal</b>	<b>Tumors</b>	<b>Classification Efficiency (%)</b>
<b>Normal</b>	<b>62/63</b>	<b>1</b>	<b>98</b>
<b>Tumors</b>	<b>0</b>	<b>40/40</b>	<b>100</b>
<b>b</b>			
<b>Normal</b>	<b>62/63</b>	<b>1</b>	<b>98</b>
<b>Tumors</b>	<b>0</b>	<b>40/40</b>	<b>100</b>

Table 2.2 Classification of breast tumors using transcutaneous *in vivo* Raman . Table shows confusion matrix for PC-LDA of breast normal and tumor spectra; a) model, and b) LOOCV.

### **III. Non-invasive follow up of physiological processes using Raman spectroscopy**

The sensitivity of Raman spectroscopy to biochemical changes has been established in previous section as well as in literature. Although suitable for detecting subtle malignancy associated changes in breast, the high sensitivity may also result in detection of normal physiological changes that affect breast. This may confound diagnosis and early detection of cancer. Thus, to establish the validity of this technique as a diagnostic/screening tool, a study of confounding variables is important. Pregnancy, lactation and aging, that induce massive changes in breast, are some of the confounding factors. In this subsection, the sensitivity of Raman spectroscopy to pregnancy, lactation and age-related changes was investigated. Effect of these confounding factors on diagnosis of breast cancer was also evaluated.

#### ***Pregnancy and lactation***

Several biochemical changes occur due to physiological processes like ageing, menstrual cycle, pregnancy and lactation. As age progresses, the mice undergoes different reproductive phases. During pregnancy and lactation, massive tissue remodeling occurs in breast (22). The mammary glands display many of the properties associated with tumor progression during pregnancy and lactation. For example, rapid proliferation of epithelial cells takes place during these phases. The lactating mammary gland also actively resists apoptotic signals. In addition, as the mammary gland undergoes these morphological changes, blood supply gets adjusted, and thus, like tumors, the mammary gland induces angiogenic remodeling. Thus, changes in breast during pregnancy and lactation have steps similar to carcinogenesis. Therefore, the potential of Raman spectroscopy to detect malignant changes in light of these confounding factors needs to be evaluated. This sub-section evaluates the sensitivity of

transcutaneous *in vivo* Raman spectroscopy to changes in breast of non pregnant, pregnant and lactating mice and its effect on breast tumor detection.

## **Materials and methods**

### **Animals**

Female SB mice were allowed to mate and successful mating was identified by appearance of vaginal plugs. Successful pregnancy was determined by observation of visible bulge in the abdominal region of mice (approximately 2 weeks post mating). Delivery of pups marked the beginning of lactation phase (approximately 3 weeks post mating). The different stages mentioned were established by a veterinarian. To minimize variability in data, same set of mice were used to record spectra immediately post mating, during pregnancy (2 weeks post mating) and lactation (1-2 days post delivery). 8 - 11 spectra per mouse were recorded transcutaneously from left and right inguinal breast of mice, resulting in 56 – 60 spectra per group. Each spectrum was recorded approximately 1 mm apart by using a precision stage. Only spectra from mice (n=6), which delivered live pups, were used for analysis. Spectra were also acquired from tumors of mice transplanted with tumors as described earlier.

### **Raman spectroscopy and Data Analysis**

The Raman instrumentation, spectra acquisition parameters and data analysis have been described in the previous sections (Section I and II Materials and methods).

## **Results and discussion**

### **Spectral analysis**

The spectral features of mean control breast spectrum (Figure 2.7 a)  $1743\text{ cm}^{-1}$  (C=O ester);  $1653\text{ cm}^{-1}$  (amide I);  $1440\text{ cm}^{-1}$  ( $\delta$  CH<sub>2</sub>);  $1301\text{ cm}^{-1}$  ( $\tau$ CH<sub>2</sub>); and  $1271\text{ cm}^{-1}$  (amide III) can be attributed to lipids. Broad amide I, change in features in  $1200\text{-}1400\text{ cm}^{-1}$  region of the mean tumor spectrum (Figure 2.7d) suggest dominance of proteins and DNA, as observed earlier. Mean spectra of pregnancy and lactating breast (Figure 2.7b and c respectively) exhibit subtle but significant variations in  $1340\text{ cm}^{-1}$  region.

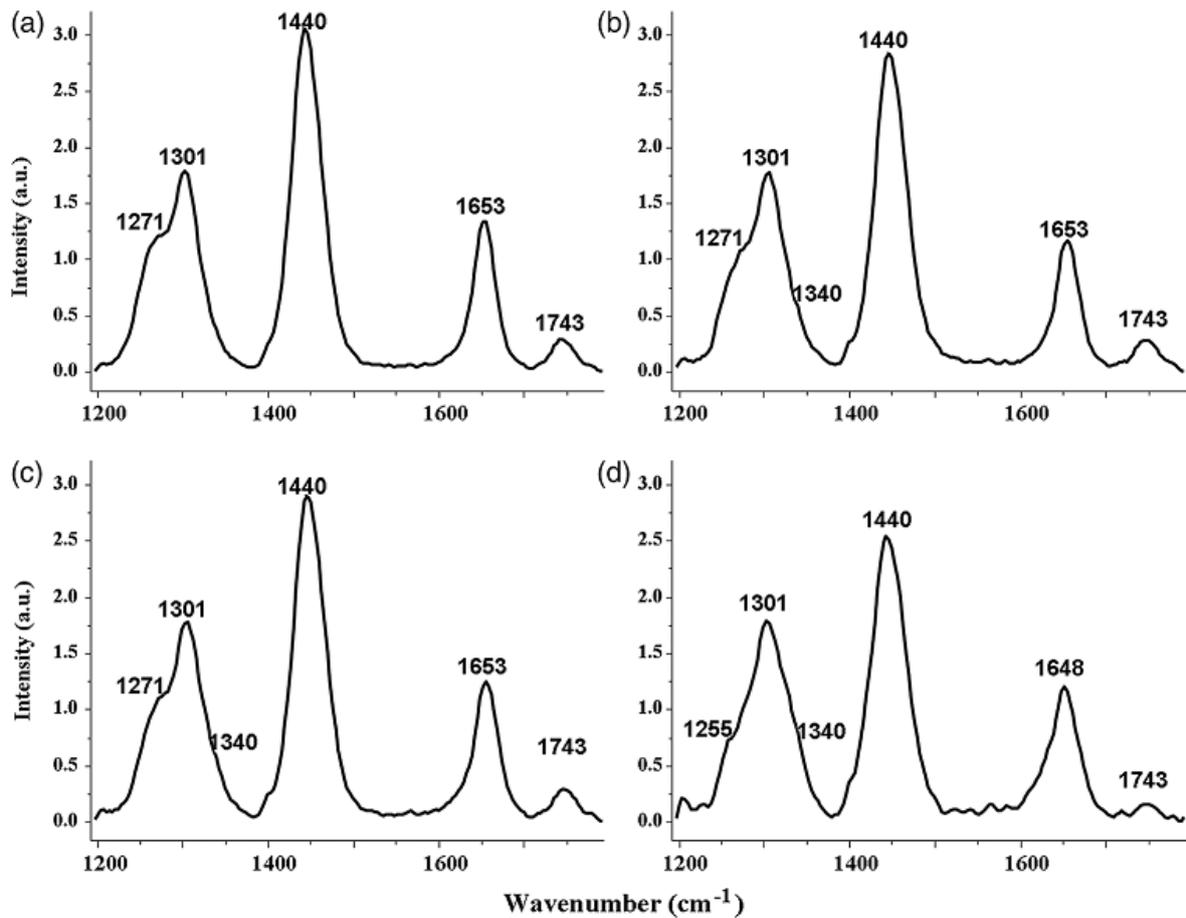


Figure 2.7 Pregnancy and lactation transcutaneous *in vivo* Raman spectroscopy study. Figure shows the mean *in vivo* Raman spectra of breast from, a) non pregnant, b) pregnant, c) lactating, d) tumor bearing mice interpolated in 1200-1800 cm<sup>-1</sup> range. Differences in lipid (1743), proteins (1648, 1271) and DNA (1340) are observed.

Difference spectra were computed subtracting mean control spectrum from mean pregnancy, lactation and tumor spectra, respectively (Figure 2.8a. 1 - 3). The negative peaks are due to control spectrum and positive peaks are due to pregnancy, lactation or tumor spectra. Difference pregnancy spectrum (Figure 2.8a. 1) exhibits loss of lipids ( $1268\text{ cm}^{-1}$ ,  $1743\text{ cm}^{-1}$ ), increase in DNA ( $1480\text{ cm}^{-1}$ ,  $1340\text{ cm}^{-1}$ ) and increase in proteins ( $1671\text{ cm}^{-1}$ ,  $1471\text{ cm}^{-1}$ ,  $1315\text{ cm}^{-1}$ ). Difference lactating spectrum (Figure 2.8a. 2) shows similar spectral features. Increase in proteins, DNA and decrease in lipids may be attributed to increase in number of cell nuclei (cell division) which is known to take place during pregnancy and lactation. Tumor difference spectra (Figure 2.8a. 3) suggests increase in proteins ( $1671\text{ cm}^{-1}$ ,  $1456\text{ cm}^{-1}$ ,  $1471\text{ cm}^{-1}$ ), increase in DNA ( $1480\text{ cm}^{-1}$ ,  $1340\text{ cm}^{-1}$ ) and decrease in lipids ( $1743\text{ cm}^{-1}$ ,  $1440\text{ cm}^{-1}$ ). Changes in lipids and DNA suggest cell division which is hallmark of tumorigenesis. Some positive bands ( $1630\text{ cm}^{-1}$  and  $1570\text{ cm}^{-1}$ ) may be ascribed to blood.

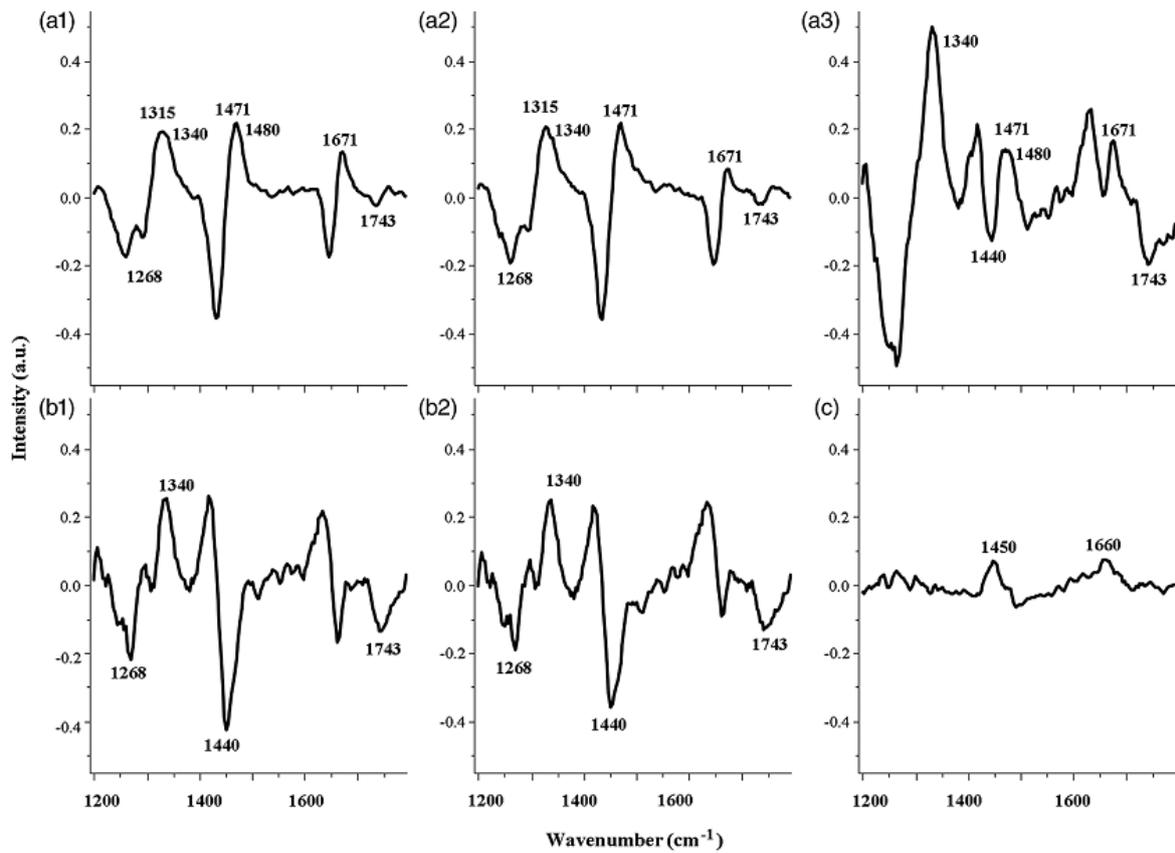


Figure 2.8 Pregnancy and lactation transcutaneous *in vivo* Raman spectroscopy study. Figure shows the difference spectrum; a 1), pregnancy – control a2), lactation – control a3); tumor – control b1), tumor – pregnancy b,2), tumor – lactation b3), lactation – pregnancy. The significant bands are labeled.

To understand difference between pathological and physiological conditions, mean pregnancy and mean lactating spectrum were subtracted (individually) from mean tumor spectrum. In this case, the positive bands are due to tumor and negative bands due to pregnancy/ lactation. Difference pregnancy (tumor – pregnancy) and difference lactation spectrum (tumor – lactation), shown in Figure 2.8b. 1 and Figure 2.8b. 2; suggest decrease in lipids ( $1743\text{ cm}^{-1}$ ,  $1440\text{ cm}^{-1}$ ) and increase in DNA ( $1340\text{ cm}^{-1}$ ) in tumor with respect to pregnancy or lactating condition. Positive bands  $1630\text{ cm}^{-1}$  and  $1570\text{ cm}^{-1}$  may be ascribed to blood. Difference physiological spectrum (Figure 2.8b.3) was also computed, by subtracting mean pregnancy spectrum from mean lactating spectrum, wherein positive peaks are due to lactation and negative peaks due to pregnancy. In this case, difference spectrum is very weak with respect to other difference spectra described above. Features  $1450\text{ cm}^{-1}$  and  $1660\text{ cm}^{-1}$  might indicate an increase in proteins in lactation with respect to pregnancy.

### **Classification of pregnancy/ lactation associated changes**

To explore the feasibility of classifying pregnant and lactating conditions from control, PC-LDA was used. Spectra interpolated in  $1200\text{-}1800\text{ cm}^{-1}$  range were used for analysis (several ranges were explored for the study- data not shown, best classification was obtained in the mentioned range). To avoid over fitting, 9 factors contributing 86% percent of correct classification; were used (Figure 2.9a). The 3-D plot of PC-LDA factors 1, 2 and 3 (Figure 2.9b) suggests classification of non pregnant (control) from pregnant and lactating mice breasts, while breast spectra of pregnant and lactating mice overlap.

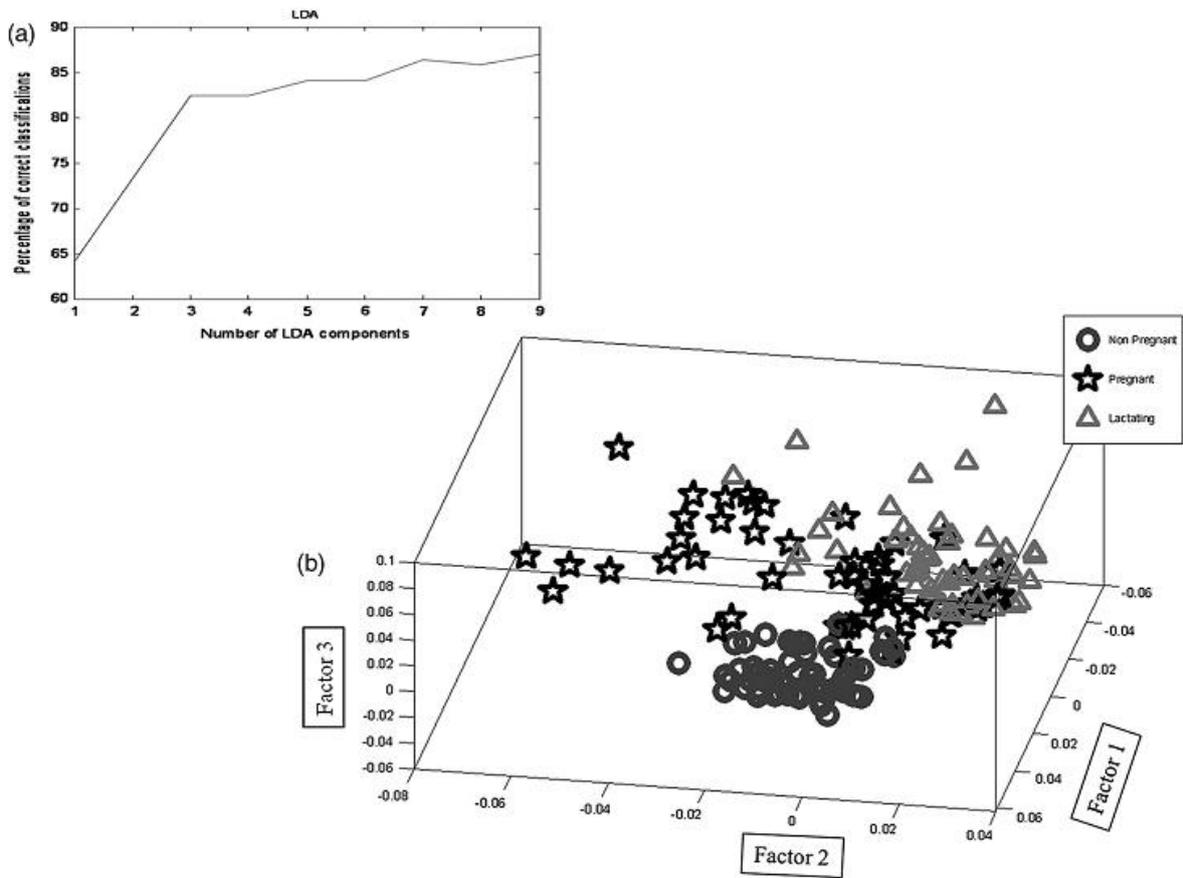


Figure 2.9 Pregnancy and lactation transcutaneous *in vivo* Raman spectroscopy study. Figure shows PC-LDA of non pregnant, pregnant and lactating mice breast spectra; a) screen plot, b) 3-D plot of PC-LDA Factors 1, 2 and 3, suggesting classification between different breast conditions.

The confusion matrix for PC-LDA model building is shown in Table 2.3a. In this analysis, 61 out of 61 spectra are correctly classified as control. 38/56 spectra are correctly classified as pregnant breast condition, whereas 4/56 misclassify as control and 14/56 misclassify as lactating condition. 54/59 spectra are correctly classified as lactating condition, while 5/59 misclassify as pregnancy condition. LOOCV was carried out to evaluate the results obtained by LDA. In analysis of LOOCV as shown in Table 2.3b, once again; 61/61 spectra correctly classify as control. 34/56 spectra are correctly classified as pregnant breast condition, whereas 4/56 misclassify as control and 18/56 misclassify as lactating condition. As mentioned earlier, pregnancy is a phase midway between normal and lactation. This probably explains few misclassifications with normal breast. Both pregnant and lactation phases represent changes in breast as a result of rapid cell proliferation. This may explain high misclassification observed between pregnant and lactating conditions. 52/59 spectra are correctly classified as lactating condition, while 7/59 misclassify as pregnancy condition. Classification efficiency of lactation is higher than pregnancy probably because lactation is characterized by cell differentiation and milk secretion in addition to cell proliferation. The classification efficiency of control, pregnant and lactating mice breast were 100%, 60% and 88% respectively.

<b>a</b>				
Condition (no. of spectra, no. of animals used)	Non Pregnant	Pregnant	Lactating	Classification Efficiency (%)
Non Pregnant (61, 6)	<b>61</b>	<b>0</b>	<b>0</b>	<b>100</b>
Pregnant (56, 6)	<b>4</b>	<b>38</b>	<b>14</b>	<b>68</b>
Lactating (59, 6)	<b>0</b>	<b>5</b>	<b>54</b>	<b>92</b>
<b>b</b>				
Non Pregnant (61, 6)	<b>61</b>	<b>0</b>	<b>0</b>	<b>100</b>
Pregnant (56, 6)	<b>4</b>	<b>34</b>	<b>18</b>	<b>61</b>
Lactating (59, 6)	<b>0</b>	<b>7</b>	<b>52</b>	<b>88</b>

Table 2.3 Pregnancy and lactation transcutaneous *in vivo* Raman spectroscopy study. Table shows the confusion matrix for PC-LDA of non pregnant, pregnant and lactating mouse

breast; a) model building, b) LOOCV, suggesting classification between these physiological conditions.

It is important to note that several changes take place in breast skin during pregnancy and lactation. In humans, skin pigmentation increases, striae appear on breast skin and skin gland secretion increase. Circulation to skin increase and veins in breast become more visible. Thus, there is a possibility that these changes may affect breast spectra. In the present study, since a hairless variant of Swiss albino mice which lack pigments were used, pigmentation is not a factor. No striae appearance was observed. With respect to blood flow and vascularization, in this study, no spectral bands attributable to blood were observed. Tumor development involves angiogenesis (increased blood vessels and blood flow), but spectral features of blood have not been reported in transcutaneous spectra of breast tumors. However, in the difference spectrum of current study; tumor – control, tumor – pregnancy and tumor – lactation, some bands may be ascribed to blood.

### **Classification of frank breast tumors from pregnancy/lactation**

The possibility of classifying frank tumors from pregnancy/lactation associated changes was explored using PC-LDA. For analysis, 3 factors contributing to 82% percent of correct classification; were used (Figure 2.10a). The 3-D plot of PC-LDA factors 1, 2 and 3 (Figure 2.10b) suggests classification of frank tumors from normal, pregnant and lactating mice.

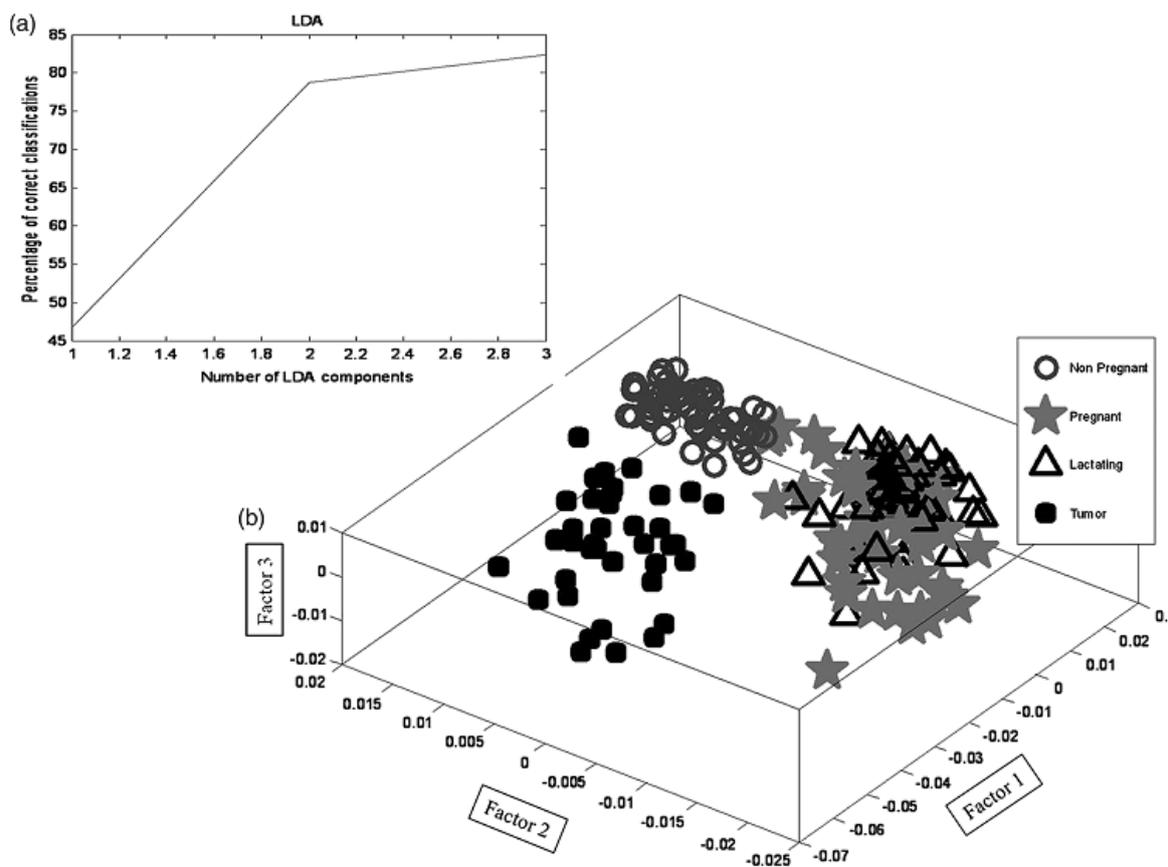


Figure 2.10 Pregnancy and lactation transcutaneous *in vivo* Raman spectroscopy study. Figure shows PC-LDA of tumor spectra and spectra of non pregnant, pregnant and lactating mice breasts; a) scree plot, b) 3-D plot of PC-LDA factors 1, 2 and 3, suggesting classification between different breast conditions and frank breast tumors.

Confusion matrix for model building and LOOCV is shown in Table 2.4a and 4b respectively. 61/61 spectra are correctly classified as control, whereas 34/56 and 41/59 respectively are correctly classified as pregnancy and lactation. 4/56 pregnancy spectra misclassify as control, while 18/59 misclassify as lactating. No misclassification with tumor is observed. 18/59 of lactation misclassify with pregnancy, whereas no misclassification with control or tumor are observed. These results mirror previous observations. Since pregnancy is a phase between control and lactation, misclassifications with both were observed. Pregnancy and lactation are both characterized by cell proliferation, hence the observed misclassification amongst them.

<b>a</b>					
Condition (no. of spectra, no. of animals used)	Non Pregnant	Pregnant	Lactating	Frank Breast Tumors	Classification Efficiency (%)
Non Pregnant (61, 6)	<b>61</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>100</b>
Pregnant (56, 6)	<b>4</b>	<b>35</b>	<b>17</b>	<b>0</b>	<b>63</b>
Lactating (59, 6)	<b>0</b>	<b>16</b>	<b>43</b>	<b>0</b>	<b>73</b>
Frank Breast Tumors (40, 2)	<b>1</b>	<b>0</b>	<b>0</b>	<b>39</b>	<b>98</b>
<b>b</b>					
Non Pregnant (61, 6)	<b>61</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>100</b>
Pregnant (56, 6)	<b>4</b>	<b>34</b>	<b>18</b>	<b>0</b>	<b>61</b>
Lactating (59, 6)	<b>0</b>	<b>18</b>	<b>41</b>	<b>0</b>	<b>69</b>
Frank Breast Tumors (40, 2)	<b>1</b>	<b>0</b>	<b>0</b>	<b>39</b>	<b>98</b>

Table 2.4 Pregnancy and lactation transcutaneous *in vivo* Raman spectroscopy study. Table shows the confusion matrix for PC-LDA of non pregnant, pregnant, lactating mouse breast

and frank breast tumors: a) model building, b) LOOCV, suggesting frank tumors can be classified from mentioned physiological conditions.

39 out of 40 tumor spectra classify correctly as tumor. Only 1/40 misclassify as control. It is known that tumors are a heterogeneous complex of necrotic centers, rapidly proliferating fronts and normal patches. This probably explains misclassification with normal breast. The classification efficiency of frank tumors from pregnancy/lactation condition is 97.5%. Results suggest minimal effect of pregnancy/lactation associated changes on detection of frank tumors using Raman spectroscopy.

### ***Aging***

Aging is one of the risk factors of breast cancer. Epidemiological evidence suggests that reproductive hormones play a role in breast cancer progression. These hormones stimulate cell division in breast tissue and result in accumulation of genetic damage. The accumulated genetic damage increases the risk of breast cancer with increasing age. This results in increased incidence of breast cancer in older women. In 1983, Pike *et. al.* proposed a model to explain the relationship between age, hormones and incidence of breast cancer, commonly referred to as Pike's model of 'breast tissue ageing' (20). According to this model, the process of 'breast ageing' is not uniform, but fluctuates with hormone level variations at a given age. Rate of 'breast ageing' is most rapid during menarche (approximately 13-15 years of human age), slows with each pregnancy (approximately 15-35 years), slows further during perimenopause (approximately 35-45 years), and is least after the menopause (approximately above 45 years). Rate of breast ageing also depends on several factors like a) age at menarche b) age at menopause c) number of pregnancies d) breast composition and e) mammographic density (108), which vary from individual to individual. A rapid, non-invasive technique to detect age-related changes in breast may help screen high risk population and assist medical counseling.

At birth, the mammary glands consist of stroma and mammary epithelium. During menarche, the epithelium forms Terminal End Buds (TEB) which invades into the fat pad resulting in branched ducts throughout the breasts. During pregnancy and lactation, further cell division and branching takes place. At end of lactation, mammary tissue involutes to regain pre-pregnancy condition. During perimenopause and menopause, atrophy (degeneration) of the mammary tissue occurs, resulting in considerable reduction in breast tissue post menopause. Sensitivity of Raman spectroscopy to these physiological changes may lead to misinterpretation of data and incorrect diagnosis. Therefore, the effect of physiological changes on diagnosis of cancer using Raman spectroscopy needs to be validated. Therefore, sensitivity of Raman spectroscopy to detect age-related changes in mouse breast and its effect on diagnosis of breast cancer has been evaluated in this sub-section.

## **Materials and methods**

### **Animals**

241 spectra were acquired transcutaneously from left and right inguinal breast of 19 SB mice. The age groups explored were 2 months (n=5), 4-6 months (n=4), 10-12 months (n=5) and 13-15 months (n=5). Approximately 10 spectra per mouse were acquired from 2, 10-12, 13-15 months old mice while 20 spectra per mouse were acquired from 4-6 months old mice were analyzed. Spectra were also acquired from frank tumors as discussed in the 'pregnancy and lactation' sub-section.

### **Raman spectroscopy and Data Analysis**

The Raman instrumentation, spectra acquisition parameters and data analysis have been described in the previous sections (Exploring rodent strains suitable for transcutaneous *in*

*vivo* spectroscopy and Classification of transcutaneous breast spectra from other anatomical sites and frank tumor).

## **Results and Discussion**

As mentioned earlier, this subsection aims to study age-related changes in mouse breast using Raman spectroscopy. Spectra were acquired transcutaneously from breast of mouse belonging different age groups. The age groups were chosen based on the different reproductive phases described in theory of 'breast tissue ageing'. The majority of females from most inbred strains first ovulate (menarche) naturally between 6 and 8 weeks after birth, while they attain menopause at 12-14 months of age. Thus, 2 months old mice and 13-15 months old mice were chosen to study changes in breast at menarche and menopause respectively. Other age groups, 4-6 months (mid reproductive phase) and 10-12 months (perimenopause) were chosen based on the menarche-menopause time points and mouse-human age comparison calculations.

### **Spectral analysis**

The spectral features of mean menarche breast spectrum (Figure 2.11a) -  $1743\text{ cm}^{-1}$  (C=O ester);  $1653\text{ cm}^{-1}$  (amide I);  $1440\text{ cm}^{-1}$  ( $\delta$  CH<sub>2</sub>);  $1301\text{ cm}^{-1}$  ( $\tau$ CH<sub>2</sub>); and  $1271\text{ cm}^{-1}$  (amide III) can be attributed to lipids. Mean tumor spectrum (Figure 2.11 e) show broad amide I and features in  $1200\text{-}1400\text{ cm}^{-1}$  region, suggesting dominance of proteins and DNA in tumor, as observed repeatedly earlier. Mean spectra of mid-reproductive, perimenopause and menopause breast (Figure 2.11b – d) exhibit subtle but significant changes in  $1340$ ,  $1440$ ,  $1653$  and  $1743\text{ cm}^{-1}$  region, probably indicating loss of lipids.

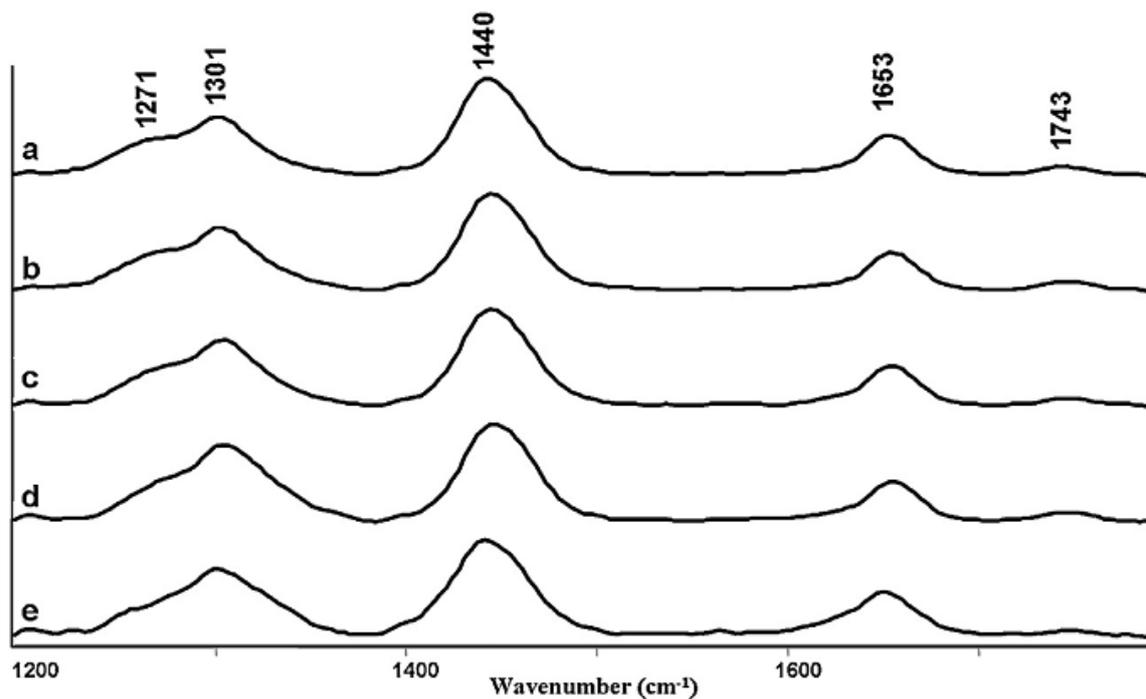


Figure 2.11 Aging transcutaneous *in vivo* Raman spectroscopy study. Figure shows the mean *in vivo* Raman spectra of mouse breast from, a) 2 months (menarche), b) 4-6 months (mid-reproductive), c) 10-12 months (perimenopause), d) 13-15 months (menopause) old mice and e) frank breast tumors interpolated in 1200-1800  $\text{cm}^{-1}$  range, suggesting subtle changes in protein and lipid content.

Difference spectrum was computed by subtracting mean menarche spectrum from mean mid-reproductive, perimenopause, menopause and tumor spectra, respectively (Figure 2.12a. 1 - 4). The negative peaks are due to menarche spectrum and positive peaks are due to mid-reproductive, perimenopause, menopause or tumor spectra. Mid-reproductive - menarche, perimenopause - menarche and menopause - menarche difference spectra suggest decrease in lipids (1268, 1301, 1650  $\text{cm}^{-1}$ ) and increase in DNA (1340, 1480  $\text{cm}^{-1}$ ) in mid-reproductive, perimenopause and menopause. Tumor - menarche spectra exhibit decrease in lipids (1268, 1301, 1440, 1743  $\text{cm}^{-1}$ ) and increase in DNA (1340  $\text{cm}^{-1}$ ), which corroborate previous reports.

To understand difference between pathological and physiological condition, mean mid-reproductive, perimenopause and menopause spectrum were subtracted from mean tumor spectrum. In this case, the positive bands are due to tumor and negative bands due to mid-reproductive, perimenopause or menopause. Tumor - mid-reproductive, tumor - perimenopause difference spectra (Figure 2.12b. 1-3) suggests decrease in lipids (1268, 1301, 1440, 1743  $\text{cm}^{-1}$ ) and increase in DNA (1340  $\text{cm}^{-1}$ ) whereas tumor - menopause difference spectrum shows decrease in lipids (1268, 1301, 1440, 1743  $\text{cm}^{-1}$ ) in tumor with respect to age groups mentioned.

Difference physiological spectrum (Figure 13 c 1-2) was also computed, by subtracting mean mid-reproductive spectrum from mean perimenopause and menopause spectrum, wherein negative peaks are due to mid-reproductive phase and positive peaks due to perimenopause and menopause. Decrease in lipids (1268, 1440 and 1743  $\text{cm}^{-1}$ ) during perimenopause and menopause with respect mid reproductive phase is observed in the above mentioned difference spectrum. Menopause - perimenopause difference spectrum (Figure 2.12d) also suggest decrease in lipids (1268, 1440, 1650  $\text{cm}^{-1}$ ).

Overall, analyses of difference spectra suggest loss of lipids and increase in DNA in tumor with respect to all age groups studied. Loss of lipids and increase in DNA during menopause, perimenopause and mid-reproductive with respect to menarche breast is also observed. Further, older age groups exhibit loss of lipid with respect to younger age groups.

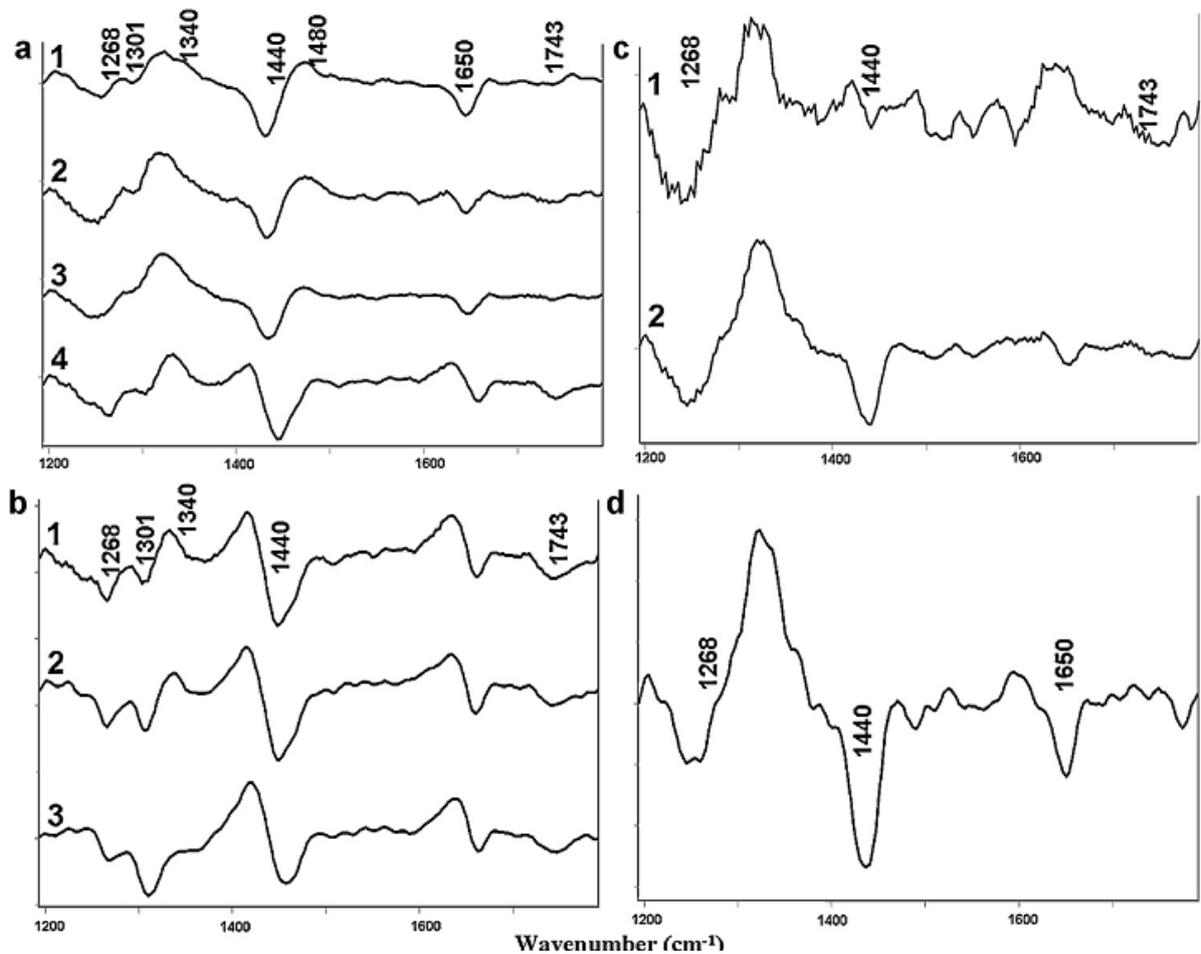


Figure 2.12 Aging transcutaneous *in vivo* Raman spectroscopy study. Figure shows the difference spectrum; a.1) Mid-reproductive – menarche, a.2) perimenopause – menarche a.3) menopause – menarche, a.4) tumor – menarche, b.1) tumor - mid-reproductive, b.2) tumor – perimenopause, b.3) tumor – menopause, c.1) menopause – mid-reproductive, c.2) perimenopause – mid-reproductive, d) menopause – perimenopause, highlighting changes in DNA, lipid and proteins between these age groups.

### **Classification of age-related changes in mouse breast**

To explore the feasibility of classifying different age groups, PC-LDA was used. Spectra interpolated in 1200-1800  $\text{cm}^{-1}$  range were analyzed. To avoid over fitting, 3 factors, contributing 80% correct classification were used (Figure 2.13a). The scatter plot of PC-LDA factors 1, 2 and 3 (Figure 2.13b) shows menarche cluster is distinct, while there is slight overlap between mid-reproductive phase, perimenopause and menopause clusters.

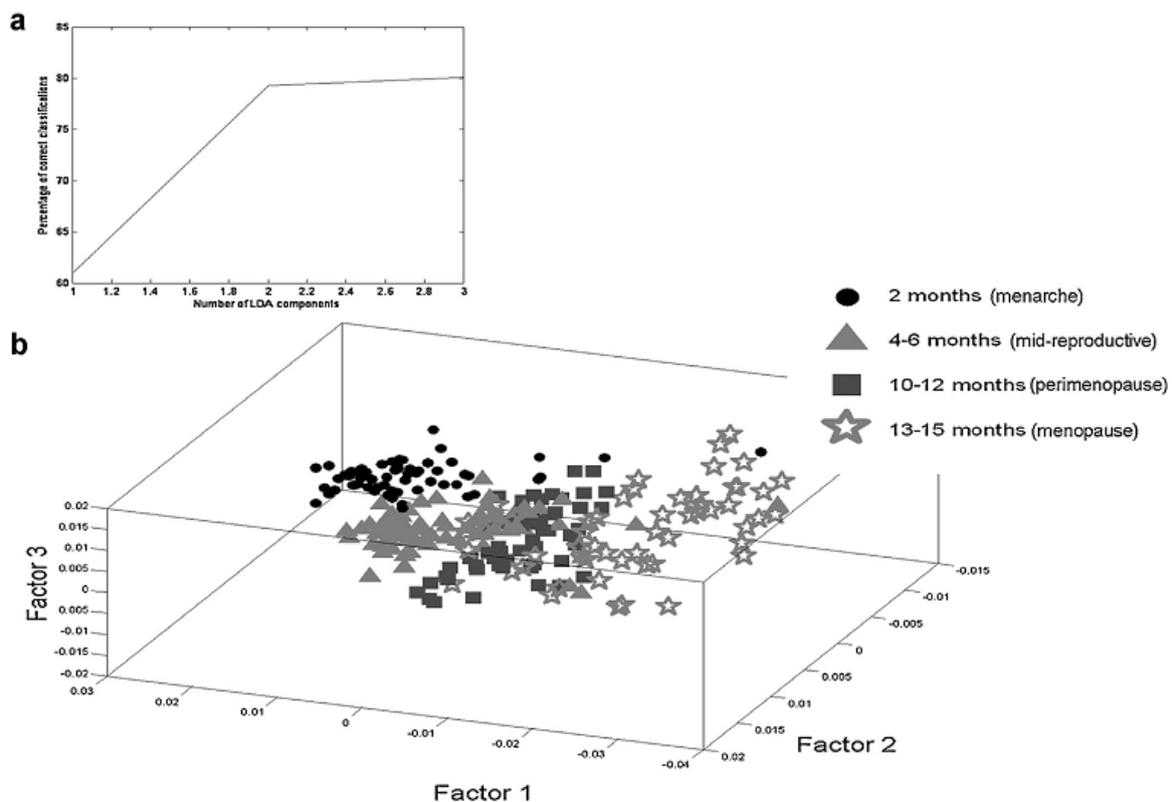


Figure 2.13 Aging transcutaneous *in vivo* Raman spectroscopy study. Figure shows the PC-LDA of 2 months (menarche), 4-6 months (mid-reproductive), 10-12 months (perimenopause) and 13-15 months (menopause) old mice breast spectra; a) scree plot showing number of factors used and variance contributed by the factors. b) 3-D plot of PC-LDA factors 1, 2 and 3, suggesting classification between different age groups.

a					
Age (No. of spectra, No. of animals)	2 months (menarche)	4-6 months (midreproductive)	10-12 months (perimenopause)	13-15 months (menopause)	Classification Efficiency (%)
2 months (62, 5)	60	1	1	0	97
4-6 months (71, 4)	1	57	12	1	80
10-12 months (54, 5)	2	12	37	3	69
13-15 months (54, 5)	0	6	9	39	72
b					
2 months (62, 5)	60	1	1	0	97
4-6 months (71, 4)	1	57	12	1	80

10-12 months (54, 5)	2	12	37	3	69
13-15 months (54, 5)	0	6	9	39	72

Table 2.5 Aging transcutaneous *in vivo* Raman spectroscopy study. Table shows the confusion matrix for PC-LDA of 2 months, 4-6 months, 10-12 months and months 13-15 old mice a) model building and b) LOOCV, again suggesting classification between age groups.

A similar trend is observed in the confusion matrix shown in Table 2.5. As observed in Table 2.5b, LOOCV yields average classification efficiency of 96.8 %, 80.3 %, 68.5 % and 72.2 % for 2 months (menarche), 4-6 months (mid reproductive phase), 10-12 months (perimenopause) and 13-15 months (menopause) old mouse breast, respectively. 12/71 (16%) spectra and 1/71 (1.4%) spectra from mid-reproductive phase misclassify with perimenopause and menopause respectively. 12/54 (22%) and 3/54 (5%) spectra from perimenopause misclassify with mid-reproductive phase and menopause, respectively. 6/54 (11%) and 9/54 (16%) spectra from menopause misclassify with mid-reproductive phase and perimenopause, respectively.

Misclassification between groups suggests homogeneity in tissues and similarity in biochemical properties of mid-reproductive phase, perimenopause and menopause breast. In contrast, only 1.6% spectra from menarche misclassify with mid-reproductive phase and perimenopause while no spectrum misclassify with menopause. This suggests that menarche is biochemically distinct. Thus, results indicate that ageing is most rapid during menarche, but slows down during other phases. As described earlier, Pike's model suggests that the rate of 'breast ageing' is rapid during menarche, but slows down during mid-reproductive phase, perimenopause and menopause. Thus, spectroscopic data correlates well with Pike's model of 'breast tissue ageing'. Overall, results suggest the possibility of detecting age-related changes in breast using transcutaneous *in vivo* Raman spectroscopy.

### **Classification of frank breast tumors from age-related changes**

As mentioned earlier, sensitivity of Raman spectroscopy to malignancy associated changes makes it ideal for diagnostic applications. However, biochemical changes also occur during physiological processes like ageing. Sensitivity of Raman spectroscopy to these physiological changes may lead to misinterpretation of data and incorrect diagnosis. Therefore, there is a

need to evaluate effect of age-related changes on diagnosis of cancer using Raman spectroscopy.

The possibility of classifying frank tumors from age-related changes was explored using PC-LDA. For analysis, 2 factors contributing to 83 % correct classification were used. The scatter plot of PC-LDA factors 1 and 2 (Figure 2.14b) suggests classification of tumors from age-related changes.

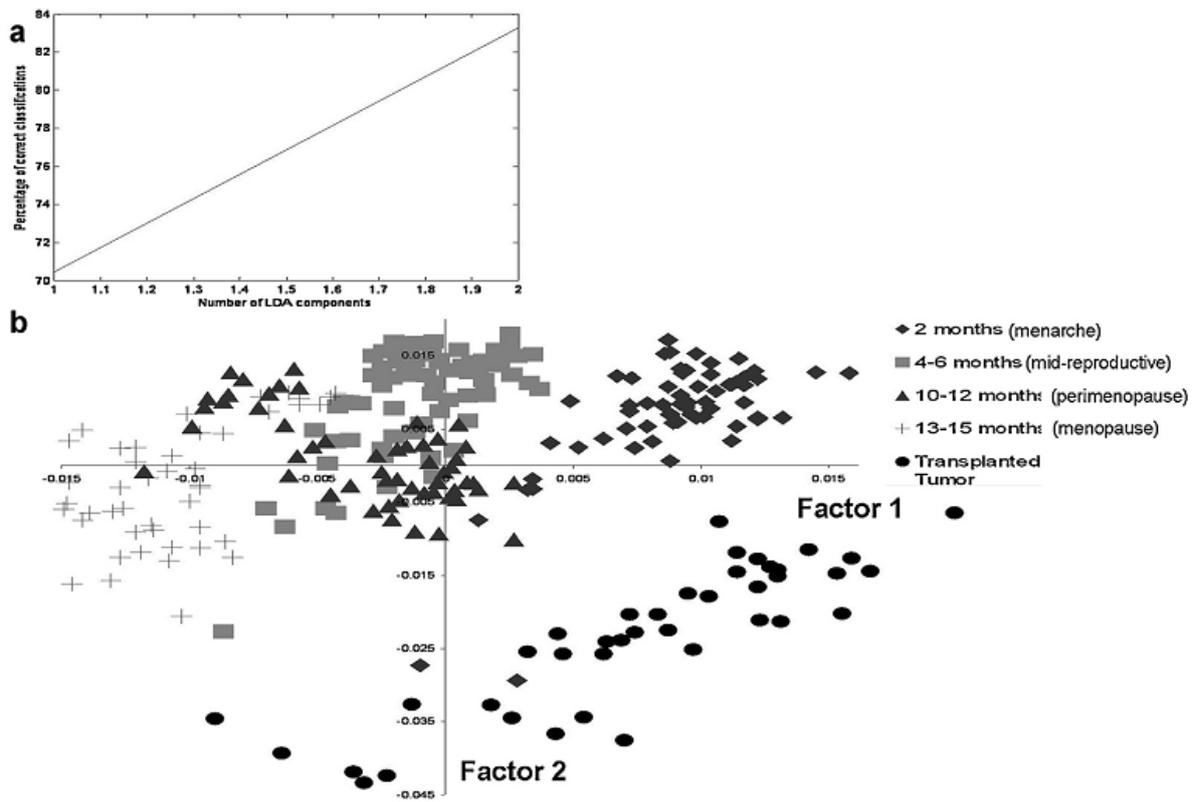


Figure 2.14 Aging transcutaneous *in vivo* Raman spectroscopy study. Figure shows the PC-LDA of frank breast tumors and spectra from breasts of mice belonging to different age groups; a) scree plot, b) 3-D plot of PC-LDA factors 1, 2 and 3, suggesting classification between different breast conditions and frank breast tumors.

a						
Age (No. of spectra, No. of animals)	2 months (menarche)	4-6 months (midreproductive)	10-12 months (perimenopause)	13-15 months (menopause)	Transplanted Tumor	Classification Efficiency (%)
2 months (62, 5)	56	0	4	0	2	90
4-6 months (71, 4)	0	54	14	3	0	76
10-12 months (54, 5)	0	10	40	4	0	74
13-15 months (54, 5)	0	7	1	46	0	85
Transplanted Tumor	2	0	0	0	38	95
b						
2 months (62, 5)	56	0	4	0	2	90
4-6 months	0	54	14	3	0	76

(71, 4)						
10-12 months (54, 5)	0	11	39	4	0	72
13-15 months (54, 5)	0	7	1	46	0	85
Transplanted Tumor (40, 2)	2	0	0	0	38	95

Table 2.6 Aging transcutaneous *in vivo* Raman spectroscopy study. Table shows the confusion matrix for PC-LDA of age groups compared with frank breast tumor; a) model building, and b) LOOCV, suggesting tumor can be classified from age related changes with 95% efficiency.

Confusion matrix for model building and LOOCV of models built using PC-LDA algorithm is shown in Table 2.6 a and b respectively. 38/40 spectra were correctly classified as tumor. Only 2/40 spectra from tumor misclassify with 2 months age group. Since tumors were transplanted in 2 months old mice, misclassifications between tumors and 2 months age-group were probably observed. Overall, results suggest that frank tumors can be classified from age-related variations with 95% efficiency.

#### **IV. Model suitable for breast carcinogenesis and *in vivo* spectroscopy**

This section explores strains and protocols suitable for induction of breast cancer from the point of view of *in vivo* spectroscopy.

##### **Materials and methods**

##### **Carcinogenesis**

Rats or mice were treated on 47<sup>th</sup> day after birth using several different protocols - intragastric administration (gavage) of DMBA (1mg/mouse) in SB mice, subcutaneous injection of N-Nitroso-N-methylurea (MNU, 0.1mg/mouse) in SB mice, intragastric administration of DMBA (65mg/kg) in SD rats, subcutaneous injection of MNU (1mg/rat) in SD rats, subcutaneous dusting of DMBA (1mg/rat) in SD rats, injection of DMBA (1mg/rat) into nipple of SD rats.

- a. Subcutaneous injection: A DMBA solution of 15mg/ml concentration dissolved in groundnut oil was prepared. An insulin syringe fitted with 16G needle was used to inject 80ul of DMBA solution. The needle was inserted in the 3<sup>rd</sup> left inguinal breast and the solution was dispersed subcutaneously between the 1<sup>st</sup> and 2<sup>nd</sup> left inguinal breast nipples

(Figure 16). The same protocol was used for controls except oil was injected instead of DMBA. 26 SD rats were treated using this protocol and 20 rats were kept as control.

b. Dusting: 1 mg DMBA was mixed with 2 mg cholesterol, cholesterol being the vehicle for delivery of DMBA. The rats were anaesthetized using ketamine and xylazine. A single incision was made using a surgical blade between 1<sup>st</sup> and 2<sup>nd</sup> left inguinal breast nipples. The skin was lifted and DMBA-cholesterol mixture was dusted over the mammary gland. The wound was sealed by Vetbond tissue adhesive. The control rats were dusted with cholesterol in the same manner. 4 SD rats were dusted with DMBA and 5 rats were kept as control.

c. Intragastric delivery (Gavage): 25 SD rats were administered 65mg/kg DMBA dissolved in groundnut oil, intragastrically by gavage. 9 rats were administered oil (Control). 18 SB mice were also administered DMBA (1mg/mouse) while 5 SB mice were administered oil using the same protocol.

d. Subcutaneous injection of MNU: 10 SB mice and SD rats were subcutaneously injected with MNU (0.1mg/mice and 1mg/rat). 5 mice and rats each were maintained as control.

## **Results and discussion**

Tumor incidence was low in SB mice (0 – 11%) using both protocols mentioned above. Of a total of 25 rats treated by gavage, 16 rats developed breast adenocarcinoma (~64%). The site of tumor development was random – tumors development was observed in all 12 mammary glands (Note that one tumor developed per rat. When taken all together and evaluated, there was no specific mammary gland that showed higher incidence compared to others). Since the site of tumor appearance cannot be predicted, to acquire spectra of breast from the site of

tumor before tumor appearance, all 12 glands need to be scanned spectroscopically. However, acquiring spectra from 12 glands is time consuming and hence impractical. Therefore, the model is unsuitable for sequential follow up studies.

Dusting protocol was implemented to achieve site-specific tumors. This way, since the site of tumor is known, spectra can be acquired before tumor appears. Dusting protocol successfully resulted in site specific tumors and high breast tumor (fibroadenoma) incidence (100%). However, the protocol involves surgical incision, which may result in injury. Resulting wound healing at the site of incision may influence Raman spectra. To avoid this, subcutaneous injection protocol was applied. Out of 28 rats treated by this protocol, 21 rats developed breast fibroadenoma at the site of injection (~75%). Although site-specific tumorigenesis was achieved, the tumors were benign fibroadenomas and not malignant breast adenocarcinomas.

In order to induce site-specific breast adenocarcinomas, nipple injection protocol was applied to SD rats. Using this protocol, 6 out of 10 rats developed breast adenocarcinoma at the site of injection (~60%). However, the time of tumor appearance ranged between 9 -24 weeks post treatment. A narrower range of tumor appearance times is required for meaningful analysis of progression data, since early detection time points may change with tumor appearance times. Thus, further standardization of this model is imperative for breast adenocarcinoma progression studies using transcutaneous *in vivo* Raman spectroscopy.

Species	Protocol	Carcinogen	Tumor incidence	latency	Type of tumor	Advantage	Disadvantage
SB mice	Gavage	DMBA	11%	20-24 weeks	adenocarcinoma	Mice best for <i>in vivo</i> spectroscopy	Tumor incidence low
SB mice	Subcutaneous injection	MNU	0%	NA	-	Mice best for <i>in vivo</i> spectroscopy	No tumors appeared
SD rats	Gavage	DMBA	64%	24-30 weeks	adenocarcinoma	Higher tumor incidence	Tumor appearance site not known, need to scan 12 breasts
SD rats	Subcutaneous injection	MNU	0%	NA	-	-	-
SD rats	Dusting	DMBA	100%	16-24 weeks	fibroadenoma	High tumor incidence, Tumor site specific	Surgical procedures required, tedious and surgical wound

							may affect Raman spectra
SD rats	Subcutaneous injection	DMBA	75%	16-24 weeks	fibroadenoma	Higher tumor incidence, site specific tumor, no surgical procedures required	Tumors are benign, not malignant
SD rats	Nipple injection	DMBA	50%	9-24 weeks	adenocarcinoma	Tumors are malignant, and site specific	Lower tumor incidence, range of tumor appearance times is high

Table 2.7 Summary of carcinogenesis protocols and outcome; subcutaneous injection is found best for benign fibroadenoma progression study while nipple injection protocol is best for malignant adenocarcinoma progression studies in breast

## Summary:

1. Several rodent species and strains were evaluated to test their suitability for *in vivo* spectroscopy. SB mice were found to be best for the purpose. Albino mice/SD rats were also found suitable, after shaving the hair in their breast region. Results indicated use of SB mice, shaved albino mice and shaved SD rats for further studies.
2. Breast could be uniquely classified from other anatomical sites using Transcutaneous *in vivo* Raman spectroscopy.
3. Transcutaneous *in vivo* Raman spectroscopy was used to follow up physiological processes - pregnancy, lactation and ageing in SB mice. Pregnancy and lactation could be distinguished from each other, controls and breast tumors. Mice belonging to different age groups could also be classified from each other and frank tumors. These observations confirmed feasibility of follow up studies using transcutaneous *in vivo* Raman spectroscopy. Results also suggested that the physiological factors do not adversely affect breast cancer diagnosis using transcutaneous *in vivo* Raman spectroscopy.
4. Several protocols were explored to find the one suitable for transcutaneous *in vivo* Raman spectroscopy based tumor progression study. Subcutaneous injection of carcinogen DMBA in SD rat breast for fibroadenoma and injection of DMBA into SD rat breast nipple for adenocarcinoma were found suitable for *in vivo* progression studies – owing to the ability of these protocols to induce site-specific tumors with high incidence.

Overall, the chapter laid groundwork for succeeding cancer progression studies described in the following chapter.

**CHAPTER 3 TO STUDY DEVELOPMENT OF  
BREAST NEOPLASMS INDUCED BY  
CHEMICAL CARCINOGEN USING RAMAN  
SPECTROSCOPY IN TARGET ORGANS AND  
BODY FLUIDS**

Findings of the previous chapter suggest two protocols suitable for sequential follow-up of breast neoplasm progression using transcutaneous *in vivo* Raman spectroscopy – subcutaneous injection of carcinogen DMBA in SD rat breast for fibroadenoma progression study and injection of DMBA in SD rat breast nipple for adenocarcinoma progression study. In this chapter, details of the progression studies are discussed. Section I describes transcutaneous *in vivo* Raman spectroscopic study of benign breast fibroadenoma progression. Section II details malignant breast adenocarcinoma progression study using transcutaneous *in vivo* Raman spectroscopy. Section III elaborates on urine-based Raman spectroscopic study on preadenocarcinoma condition. Section IV describes preadenocarcinoma condition study using serum-based Raman spectroscopy.

## **I. Study development of fibroadenoma using *in vivo* Raman spectroscopy**

Based on the premise that premalignant lesions may not be true indicators of cancer occurrence, and the only true indicator of tumorigenesis is the appearance of tumor itself, this subsection attempts to correlate spectral changes with success or failure to develop tumor in carcinogen treated rats. For this, breast tumor was induced by subcutaneous injection of carcinogen in breast. Spectra were acquired 0, 3, 8-10, 12-14 and 20 weeks post carcinogen treatment from treated rats and post oil treatment from controls. The spectra were analyzed using PCA and PC-LDA. PC-LDA models were also validated by independent test data. The test data consists of spectra acquired at varying rat ages (9 – 38 weeks old rats), from rats with varying times of tumor appearance (8 -22 weeks post carcinogen treatment) and different protocols for inducing carcinogenesis (intra-gastric administration and subcutaneous dusting of carcinogen); for rigorous evaluation of the model.

### **Materials and methods**

## **Animals**

A total of 81 SD rats were used in this study. 34 forty seven days old rats were administered oil (controls) while 47 forty seven days old rats were administered DMBA using protocols described earlier.

## **Carcinogenesis**

The carcinogenesis protocols have been described earlier. The number of animals used as test and controls for each protocol are listed below.

*Subcutaneous injection:* 26 SD rats were treated using this protocol and 20 rats were kept as control.

*Dusting:* 4 SD rats were dusted with DMBA and 5 rats were kept as control.

*Intragastric delivery (Gavage):* 17 SD rats were administered 65mg/kg DMBA dissolved in groundnut oil (Dhara, India) intragastrically by gavage. 9 rats were administered oil (Control).

## **Spectra acquisition**

### *Subcutaneous injection*

*Batch I:* Spectra were acquired from control and treated rats 0, 3, 8-10, 12-14 weeks post oil / carcinogen treatment. The rat breasts during these time points of spectra acquisition were visibly and palpably normal. Seven treated rats developed breast tumor (benign fibroadenoma) 18 weeks after carcinogen delivery. Since spectra were acquired from visibly and palpably normal breast prior tumor development (pretumor), the 0, 3, 8-10, 12-14 weeks spectra of these 7 rats were labeled pre-fibroadenoma (PF). Spectra acquired from rats that did not develop tumor even after 30 weeks post treatment were labeled no-fibroadenoma

(NF). Spectra were also acquired from frank fibroadenoma, labeled F, of 7 rats. The scheme of spectra acquisition is shown in Figure 3.1. The control and F spectra were used as training data set. The PF and NF spectra were used as test data set.

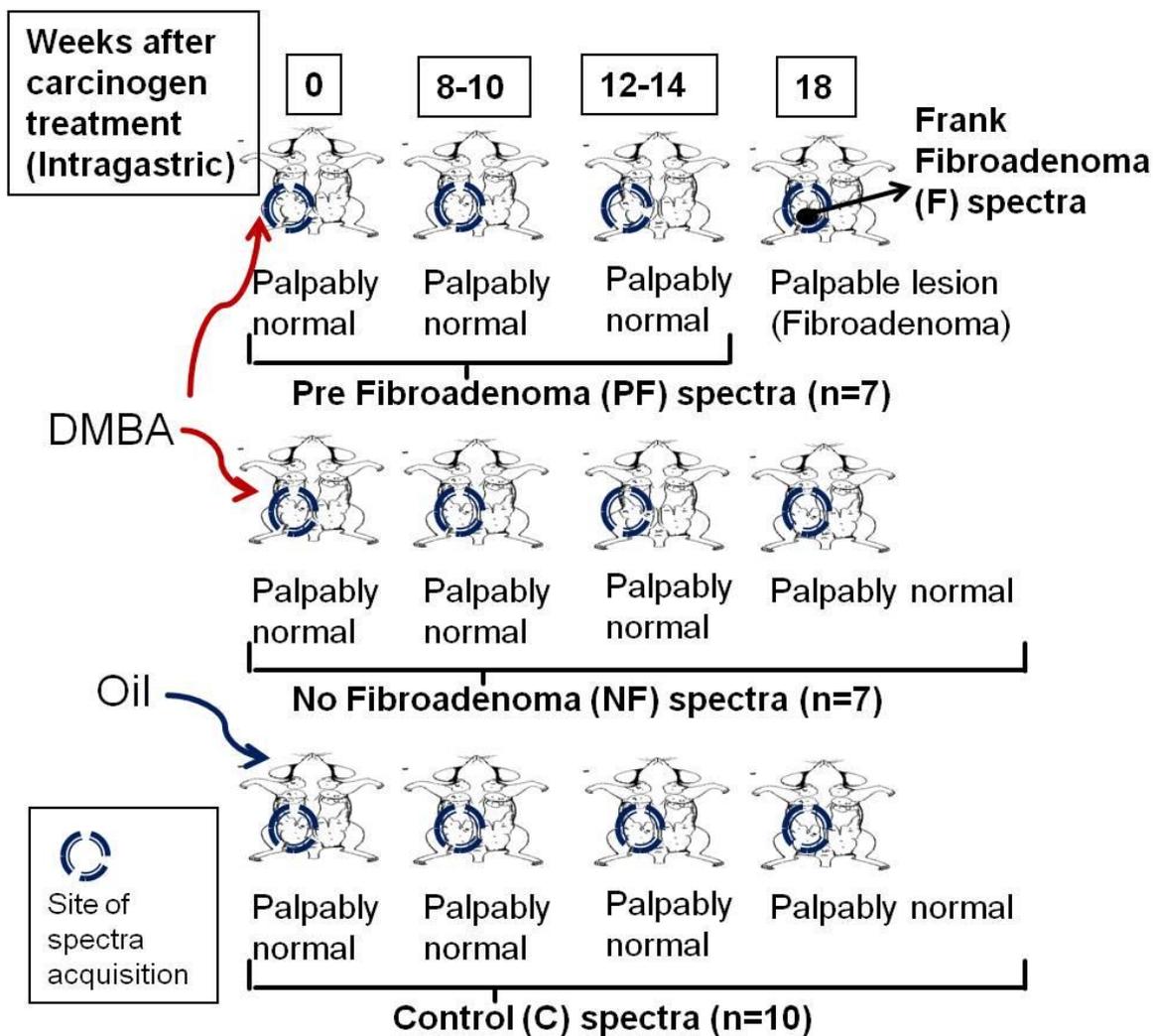


Figure 3.1 Scheme of spectra acquisition for transcutaneous *in vivo* Raman spectroscopic study of fibroadenoma progression. Figure shows time of spectra acquisition (weeks post carcinogen treatment), site of spectra acquisition and different groups of rats based on carcinogen treatment outcome – success or failure to develop tumor. Note that the first panel represents rats that successfully developed tumor, second panel shows those that did not develop tumor despite carcinogen treatment and third panel shows the control rats. Spectra have been labeled accordingly.

*Batch II:* Spectra were acquired from control and treated rats 3, 7 and 11 weeks post oil /carcinogen treatment. The treated rats developed fibroadenoma 22 weeks post treatment. The 3, 7 and 11 week spectra of rats that developed tumor were labeled PF. The PF and control spectra were used as test data set.

*Batch III:* Spectra were acquired from control and treated rats 3, 7 and 11 weeks post oil/DMBA treatment. All treated rats developed fibroadenoma 14 weeks post treatment. Since the treated rats eventually developed fibroadenoma, spectra from these rats were also grouped as PF. Both control and PF spectra were used as test data set.

*Dusting:*

Spectra were acquired from control and treated rats 3, 7 and 12 weeks post cholesterol (vehicle)/carcinogen treatment. During spectra acquisition, these rats were palpably and visibly normal. Although normal during spectra acquisition weeks, the treated rats developed fibroadenoma 14 weeks post carcinogen treatment and hence were labeled PF. These spectra as well as those from control rats were used as test data set.

*Gavage:*

Spectra were acquired from control (n=16) and treated (n=17) rats 0, 3, 8-10, 12-14 and 20 weeks post oil /carcinogen treatment. The treated rats did not develop fibroadenoma at the location scanned despite carcinogen treatment and hence were labeled no-fibroadenoma (NF). Both control and treated rat spectra were used as test data set.

The variation observed in the latency period of tumor development across different batches and protocols has been reported extensively in literature (109, 110).

The spectra were acquired at specific points with respect to the 1<sup>st</sup> inguinal nipple in every rat irrespective of carcinogenesis protocols (Figure 3.2) for all rats. This ensures that the same locations in breast are scanned throughout the study.

### **Raman spectroscopy and Data Analysis**

The Raman instrumentation, spectra acquisition parameters and data analysis have been described in the previous sections (Chapter 2 Section I and II Materials and methods).

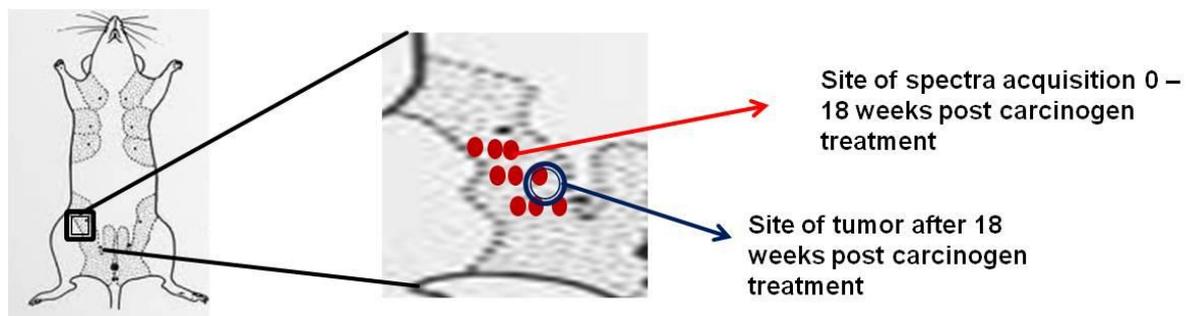


Figure 3.2 Sites of spectra acquisition for transcutaneous *in vivo* Raman spectroscopic study of fibroadenoma progression. Figure shows sites in breast with respect to nipple where spectra were acquired. Each red circle represents one spectrum acquisition site. These same sites were used for spectra acquisition over weeks from each rat. This ensures that the same location is scanned every time interval.

Since this is a follow up study wherein spectra were acquired from the same animal from 0 week (day of carcinogen delivery) till 14-22 week (time of tumor appearance), biopsy and subsequent histopathology of intermediate weeks could not be carried out.

## **Results and discussion**

### **Spectral analysis**

For spectral analysis, spectra acquired from subcutaneous injection batch I rats were used. Mean spectrum of control rats 0 week post carcinogen treatment ( $C_0$ ) was used as standard to compare mean spectra of control rats 3 week ( $C_3$ ), 8-10 week ( $C_{8-10}$ ), 12-14 week ( $C_{12-14}$ ) and 20 week ( $C_{20}$ ) post oil treatment, NF rats 3 week ( $NF_3$ ), 8-10 week ( $NF_{8-10}$ ), 12-14 week ( $NF_{12-14}$ ) and 20 week ( $NF_{20}$ ) post carcinogen treatment, PF rats 3 week ( $PF_3$ ), 8-10 week ( $PF_{8-10}$ ), 12-14 week ( $PF_{12-14}$ ) and 20 week ( $PF_{20}$ ) post carcinogen treatment and F (frank fibroadenoma). Difference spectra were calculated by subtracting  $C_0$  from  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ ,  $C_{20}$ ,  $NF_3$ ,  $NF_{8-10}$ ,  $NF_{12-14}$ ,  $NF_{20}$ ,  $PF_3$ ,  $PF_{8-10}$ ,  $PF_{12-14}$ ,  $PF_{20}$  and F spectra.

$C_0$  spectrum (Figure 3.3a.1) exhibit lipid features -  $1745\text{ cm}^{-1}$ ,  $1653\text{ cm}^{-1}$ ,  $1445\text{ cm}^{-1}$ ,  $1301\text{ cm}^{-1}$ ,  $1271\text{ cm}^{-1}$ ,  $1337\text{ cm}^{-1}$  DNA band and broad amide I.  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ ,  $C_{20}$  (Figure 3.3a.2 –a.5) show loss of  $1337\text{ cm}^{-1}$  band, sharper lipid ( $1653\text{ cm}^{-1}$ ,  $1445\text{ cm}^{-1}$  and  $1301\text{ cm}^{-1}$ ) and amide I bands. Several changes take place in breast as the rat ages as discussed earlier. Briefly, for clarity in spectral analysis, during puberty (from ~ 35 - 60 days after birth), mammary epithelial cells divide rapidly and invade into the fat pad resulting in branched ducts throughout the breasts. After initial intense proliferation, the cell proliferation becomes constant throughout the reproductive phase (~ 2 – 12 months after birth) at a rate lower than puberty. Cell division declines during perimenopause and menopause, breast tissue decreases (1.2-1.5 years after birth). Since  $C_0$  (day of carcinogen treatment) corresponds to 47<sup>th</sup> day

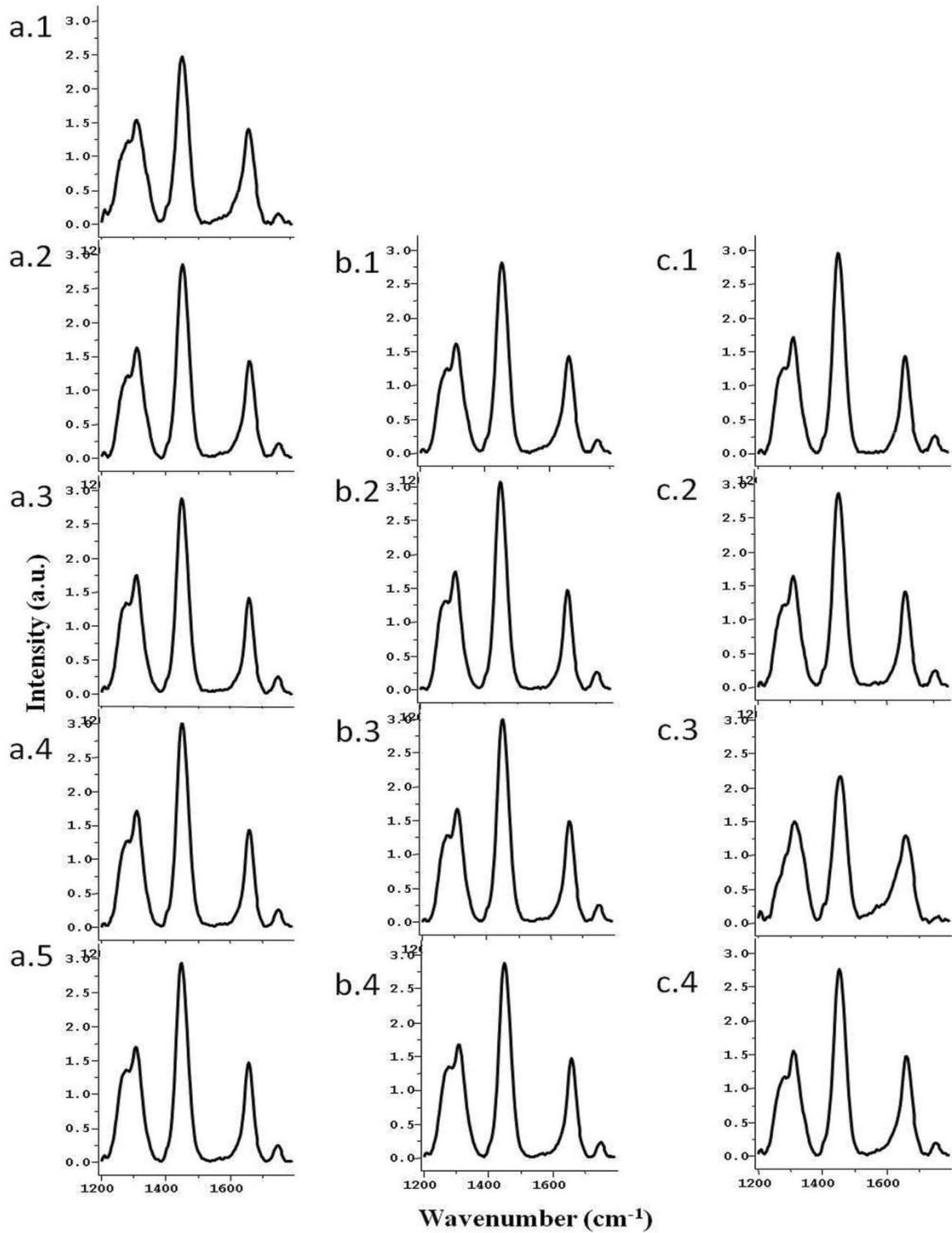
after birth, the rat is in puberty phase. Thus, rapid cell proliferation during puberty may explain DNA band and broad amide I in  $C_0$  mean spectrum, while decrease in the cell proliferation during reproductive phase compared to puberty probably explains loss of DNA band, sharper amide I and lipid bands  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ ,  $C_{20}$ . Difference spectra also support this finding. The  $C_3 - C_0$  difference spectrum (Figure 3.4a.1) exhibit positive protein bands at 1660 and 1450 and negative DNA band at  $1340\text{ cm}^{-1}$ . This suggests slight change in DNA and protein profile.  $C_{8-10} - C_0$  (Figure 3a.2) show positive lipid bands at 1740, 1440, 1301,  $1262\text{ cm}^{-1}$  negative DNA bands at 1337 and  $1480\text{ cm}^{-1}$ .  $C_{12-14} - C_0$  and  $C_{20} - C_0$  (Figure 3.4a.3 – a.4) difference spectra show profiles similar to  $C_{8-10} - C_0$ , suggesting increased lipid content and decreased DNA content during  $C_{8-10}$ ,  $C_{12-14}$ , and  $C_{20}$  compared to  $C_0$ . The profile of  $C_3$  is slightly different probably since it is a phase where puberty ends and reproductive phase begins.

The mean NF spectra,  $NF_3$ ,  $NF_{8-10}$ ,  $NF_{12-14}$ , and  $NF_{20}$  (Figure 3.3b.1-b.4) have characteristics similar to  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ , and  $C_{20}$  respectively. Loss of  $1337\text{ cm}^{-1}$  DNA band and sharper amide I in  $NF_3 - NF_{20}$  mean spectra compared to  $C_0$  indicates lower cell proliferation in these groups.  $NF_3 - C_0$ ,  $NF_{8-10} - C_0$ , and  $NF_{12-14} - C_0$  (Figure 3.4b.1 – b.4) also have profiles similar to  $C_3 - C_0$ ,  $C_{8-10} - C_0$  and  $C_{12-14} - C_0$  respectively. The profiles suggest increased lipid and decreased DNA compared to  $C_0$ . The similarity between controls and NF can be explained since NF rats do not develop fibroadenoma.

The mean PF spectra,  $PF_3$ ,  $PF_{8-10}$  and  $PF_{12-14}$  (Figure 3.3c.1 – c.4) show an additional shoulder band at  $1337\text{ cm}^{-1}$  (DNA) and broader amide I compared to their respective week control and NF rat mean spectra. This suggests higher cell proliferation in PF compared to control and NF rat mean spectra.  $PF_3 - C_0$ ,  $PF_{8-10} - C_0$ , and  $PF_{12-14} - C_0$  (Figure 3.4c.1-c.4) exhibit less intense DNA band compared to  $C_3 - C_0$ ,  $C_{8-10} - C_0$  and  $C_{12-14} - C_0$  and  $NF_3 - C_0$ ,  $NF_{8-10} - C_0$ ,

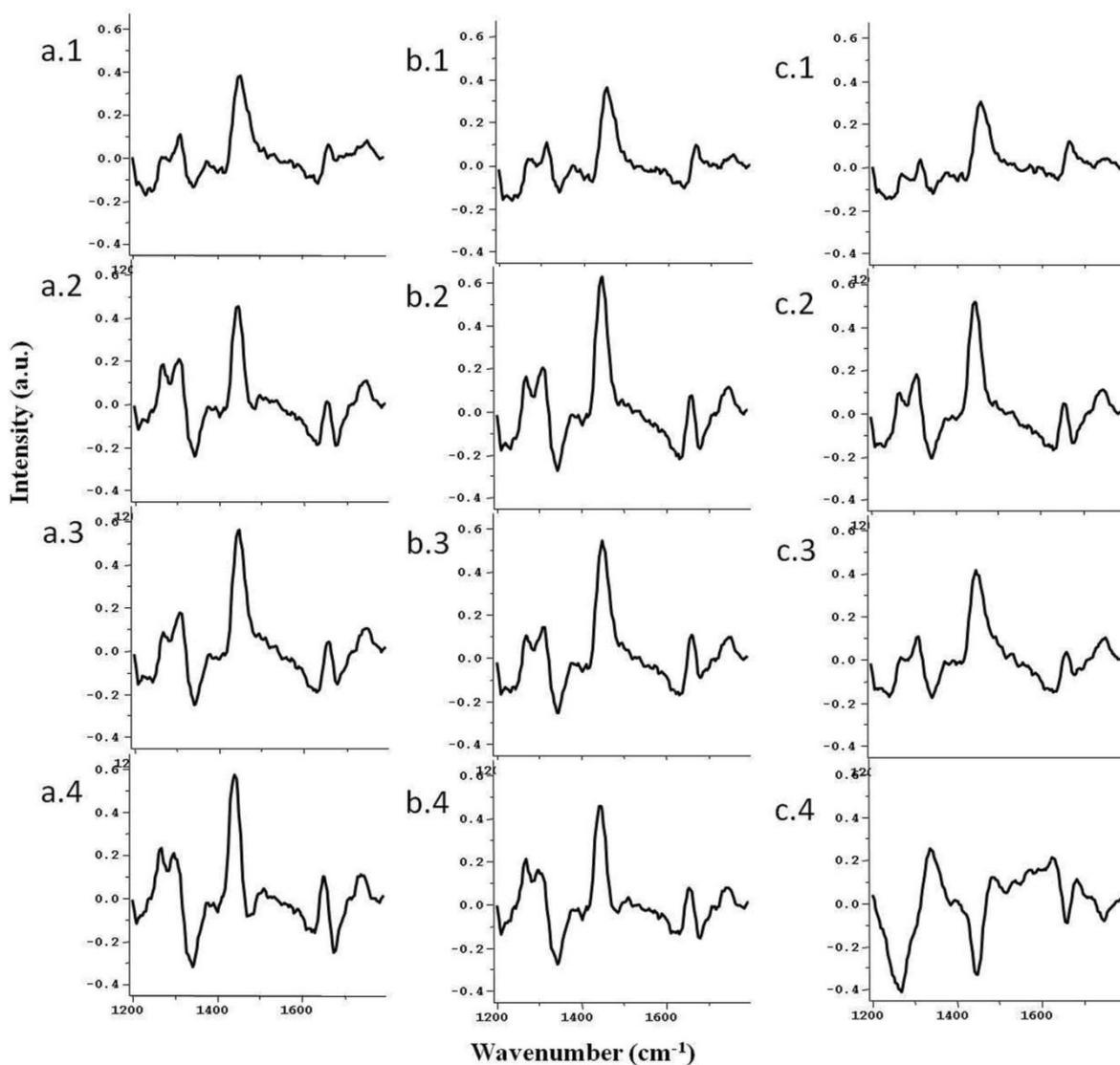
and  $NF_{12-14} - C_0$  respectively. This probably suggests higher cell proliferation in  $PF_3$ ,  $PF_{8-10}$ , and  $PF_{12-14}$  compared to  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ ,  $C_{20}$ , and  $NF_3$ ,  $NF_{8-10}$ ,  $NF_{12-14}$ ,  $NF_{20} - C_0$  respectively. The high cell proliferation may indicate changes preceding fibroadenoma development.

The mean F spectrum show broad amide I,  $\delta$   $CH_2$  band at  $1450\text{ cm}^{-1}$  and change in features in  $1200-1400\text{ cm}^{-1}$  region. F –  $C_0$  difference spectra show positive DNA bands at  $1480\text{ cm}^{-1}$  and  $1340\text{ cm}^{-1}$  and negative lipid bands at  $1260\text{ cm}^{-1}$ ,  $1440\text{ cm}^{-1}$  and  $1743\text{ cm}^{-1}$ . These features suggest low lipid content and high DNA content with respect to  $C_0$ . This suggests dominance of proteins and DNA and thus, higher proliferation compared to  $C_0$ . These findings corroborate well with earlier results.



**Figure 3.3** Transcutaneous *in vivo* Raman spectroscopic study fibroadenoma progression study. Figure shows the mean spectra of, a.1) C<sub>0</sub>, a.2) C<sub>3</sub>, a.3) C<sub>8-10</sub>, a.4) C<sub>12-14</sub>, 5) C<sub>20</sub>; b. 1) NF<sub>3</sub>, b.2) NF<sub>8-10</sub>, b.3) NF<sub>12-14</sub>, b.4) NF<sub>20</sub>, c.1) PF<sub>3</sub>, c.2) PF<sub>8-10</sub>, c.3) PF

12-14, c.4) F. (C: control, NF: No Fibroadenoma, PF: Pre Fibroadenoma, F: Fibroadenoma, weeks post carcinogen/oil treatment shown in subscripts). While C and NF spectra exhibit several similarities, PF spectra show subtle changes in lipid, protein and DNA content, and F spectra show loss of lipids and increase in DNA and proteins.



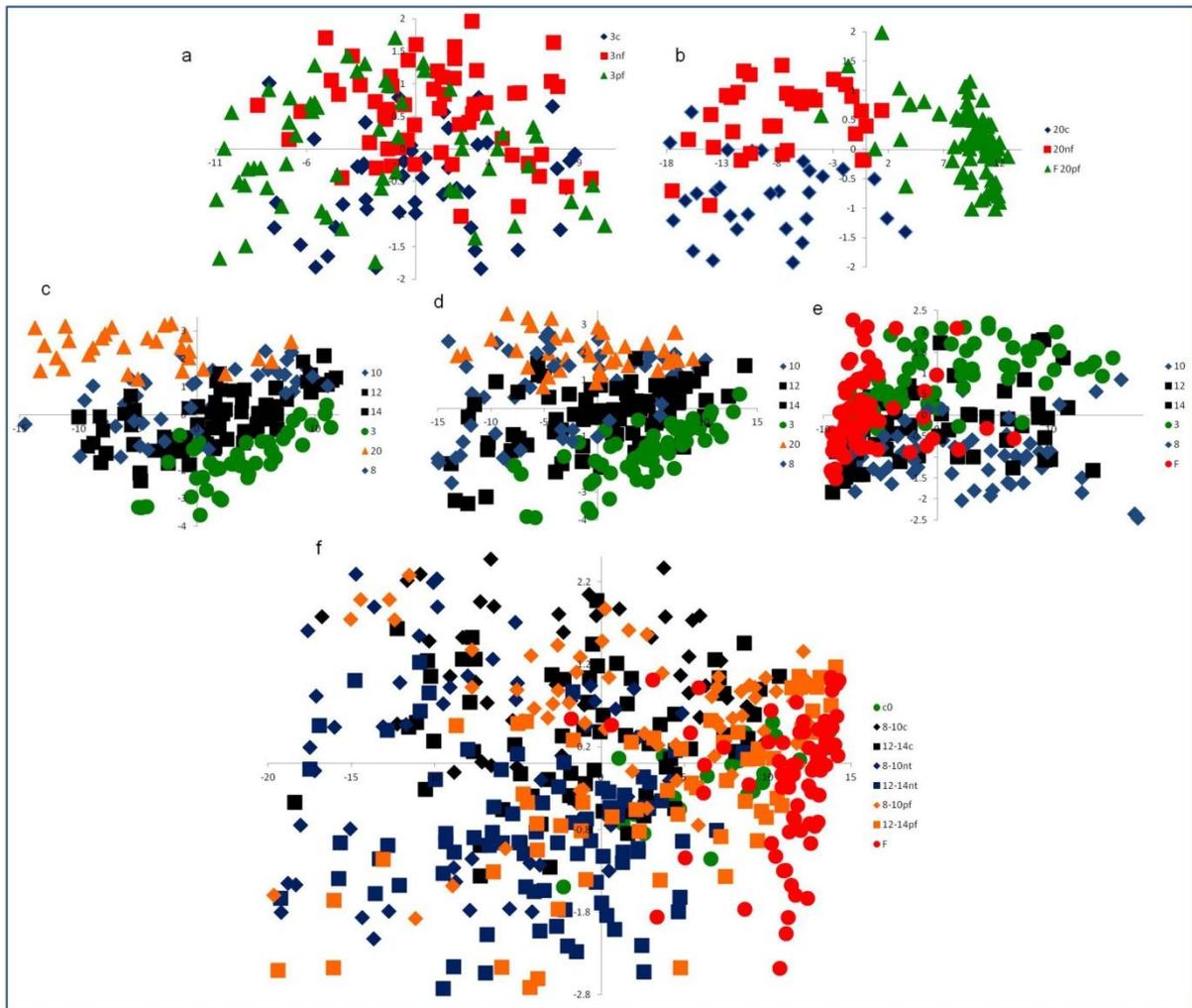
**Figure 3.4** Transcutaneous *in vivo* Raman spectroscopic study of fibroadenoma progression. The figure shows the difference spectra; a).1)  $C_3 - C_0$ , a.2)  $C_{8-10} - C_0$ , a.3)  $C_{12-14} - C_0$ , a.4)  $C_{20} - C_0$ , b.1)  $NF_3 - C_0$ , b.2)  $NF_{8-10} - C_0$ , b.3)  $NF_{12-14} - C_0$ , a.4)  $NF_{20} - C_0$ ; c.1)  $PF_3 - C_0$ , c.2)  $PF_{8-10} - C_0$ , c.3)  $PF_{12-14} - C_0$ , c.4)  $F - C_0$ . (C: control, NF: No Fibroadenoma, PF: Pre Fibroadenoma, F: Fibroadenoma, weeks post carcinogen/oil treatment shown in subscripts). The spectra suggest similarity between C and NF

**spectra. PF spectra are subtly different from C and NF spectra, while F spectra exhibit major differences in lipid, proteins and DNA content with respect to C.**

## Multivariate analysis

PCA: Preprocessed spectra interpolated in 1200-1800  $\text{cm}^{-1}$  range were subjected to PCA for delineating trends in the data set. PCA of initial weeks (Figure 3.5a) of all three groups – C<sub>3</sub>, NF<sub>3</sub> and PF<sub>3</sub> show no clustering. This suggests that C<sub>3</sub>, NF<sub>3</sub> and PF<sub>3</sub> spectra have similar features. PCA of final weeks (Figure 3.5b), C<sub>20</sub>, PF<sub>20</sub> and F show three clusters. F cluster is distinct, while C<sub>20</sub> and NF<sub>20</sub> overlap. This suggests that frank fibroadenoma spectra can be distinguished from C and NF rat spectra. PCA of C<sub>0</sub>, C<sub>3</sub>, C<sub>8-10</sub>, C<sub>12-14</sub>, C<sub>20</sub> (Figure 3.5c) show three overlapping clusters – C<sub>3</sub>, C<sub>8-14</sub> and C<sub>20</sub>. As explained earlier, C<sub>0</sub> is phase of rapid cell proliferation, while C<sub>3</sub> is the beginning of reproductive phase, wherein cell proliferation remains constant. However, in biological systems, phase boundaries – puberty to reproductive phase are not sharp. Hence some cell proliferation similar to C<sub>0</sub> is expected, which may explain the cluster of C<sub>3</sub>. C<sub>8-14</sub> belongs to the reproductive phase, hence the group form a single cluster. C<sub>20</sub>, although part of the reproductive phase, may have lower cell proliferation, since cell proliferation slowly declines as rat approaches perimenopause phase. This may explain C<sub>20</sub> cluster. PCA of NF<sub>0</sub>, NF<sub>3</sub>, NF<sub>8-10</sub>, NF<sub>12-14</sub>, NF<sub>20</sub> (Figure 3.5d) show clustering similar to controls, suggests spectral similarity between controls and NF. PCA of PF<sub>0</sub>, PF<sub>3</sub>, PF<sub>8-10</sub>, PF<sub>12-14</sub>, PF<sub>20</sub> and F (Figure 3.5e) show four overlapping clusters – PF<sub>3</sub>, PF<sub>8-10</sub>, PF<sub>12-14</sub> and F. This may be caused by change in cell proliferation patterns that lead to fibroadenoma development. It is also observed that many PF<sub>8-10</sub> and PF<sub>12-14</sub> spectra overlap with F cluster, probably indicating spectral features similar to F in PF<sub>8-10</sub> and PF<sub>12-14</sub> groups. It is important to note that PF<sub>3</sub>, PF<sub>8-10</sub> and PF<sub>12-14</sub> rats were visibly and palpably normal. Overlap of these groups with F group may thus be indicative of future fibroadenoma development. This possibility is supported by PCA of all groups (Figure 3.5f) – C<sub>0</sub>, C<sub>3</sub>, C<sub>8-10</sub>,

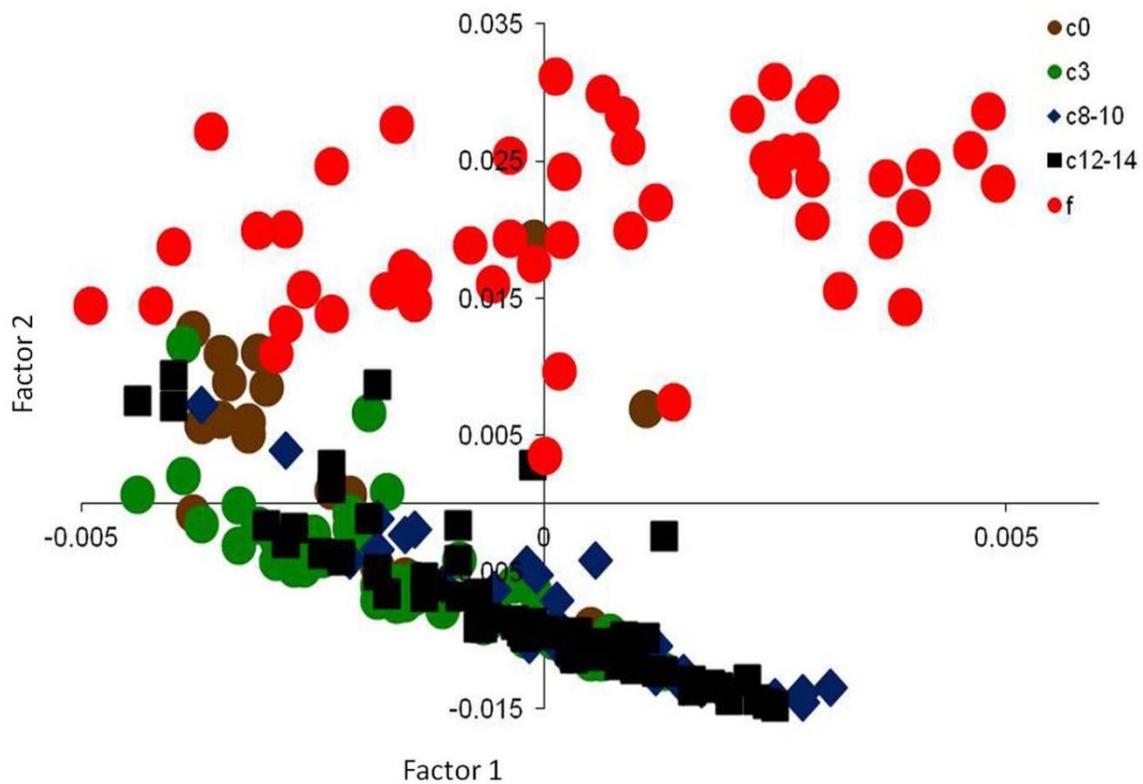
$C_{12-14}$ ,  $C_{20}$ ,  $NF_0$ ,  $NF_3$ ,  $NF_{8-10}$ ,  $NF_{12-14}$ ,  $NF_{20}$ ,  $PF_0$ ,  $PF_3$ ,  $PF_{8-10}$ ,  $PF_{12-14}$ ,  $PF_{20}$  and F. As seen in this figure, C and NF groups cluster away from F, while PF cluster overlap with F. Thus, results suggest possibility of distinguishing pretumor spectra from normal.



**Figure 3.5** Transcutaneous *in vivo* Raman spectroscopic study of fibroadenoma progression. Figure shows PCA of (a) initial weeks, C3 (blue), NF3(red), and PF3 (green); (b) final weeks, C20 (blue), NF20 (red), and F (green); (c) C3 (green), C8–10 (blue), C12–14 (black), and C20 (orange); (d) NF3 (green), NF8–10 (blue), NF12–14 (black), NF20 (orange); (e) PF3 (green), PF8–10 (blue), PF12–14 (black), and F (red); and (f) C0 (green), C3 (black), C8–10 (black), C12–14 (black), C20 (black), NF3 (blue), NF8–10 (blue), NF12–14 (blue), NF20 (blue), PF3 (orange), PF8–10 (orange), PF12–14 (orange), NF20 (orange), and F (red) (x-axis, score of factor 1; y-axis, score of factor 2).  
 - C: control, NF: No Fibroadenoma, PF: Pre Fibroadenoma, F: Fibroadenoma, weeks

post carcinogen/oil treatment shown in subscripts, colors in brackets correspond to legend colors. PCA suggests that C and NF cluster away from F. PF spectra majorly overlap with C and NF, but few PF spectra overlap with F. These spectra may be indicative of early fibroadenoma associated changes.

PC-LDA: To further explore the possibility of distinguishing pretumor spectra from control, PC-LDA was implemented. As mentioned earlier, several changes occur in breast during puberty and reproductive phase. Previous studies have shown that these age-related changes affect breast spectra. In view of this, PC-LDA model was trained using spectra from control rats of different age groups - C<sub>0</sub>, C<sub>3</sub>, C<sub>8-10</sub>, C<sub>12-14</sub> and F. Only subcutaneous injection Batch I spectra were used as training set, while the remaining – subcutaneous injection batch II and III, dusting and gavage, were used as test data set. The PC-LDA scatter plot of C<sub>0</sub>, C<sub>3</sub>, C<sub>8-10</sub>, C<sub>12-14</sub> and F (Figure 3.6) shows two distinct clusters - C and F. Some C<sub>0</sub> spectra overlap with F. As mentioned earlier, high rate of cell proliferation during C<sub>0</sub> and during fibroadenoma development may explain this.



**Figure 3.6** Transcutaneous *in vivo* Raman spectroscopic study of fibroadenoma progression. Figure shows PC-LDA model scatter plot of C<sub>3</sub>, C<sub>8-10</sub>, C<sub>12-14</sub>, C<sub>20</sub> and F (C: control, NF: No Fibroadenoma, PF: Pre Fibroadenoma, F: Fibroadenoma, weeks post carcinogen/oil treatment shown in subscripts). PC-LDA suggests clear classification between C and F

The confusion matrix for PC-LDA model building and LOOCV is shown in

Table **3.1**. After LOOCV, 12 out of 22 spectra classify correctly, while 6/22, 1/22 and 3/22 misclassify with C<sub>3</sub>, C<sub>8-10</sub> and F respectively. 22/49 C<sub>3</sub> spectra correctly classify, whereas 10/49, 7/49 and 10/49 misclassify with C<sub>0</sub>, C<sub>8-10</sub> and C<sub>12-14</sub>. 25/44 C<sub>8-10</sub> spectra classify correctly, while 2/44, 6/44 and 11/44 spectra misclassify with C<sub>0</sub>, C<sub>3</sub> and C<sub>10-12</sub> respectively. 16/69 C<sub>12-14</sub> spectra correctly classify, whereas 8/69, 15/69 and 20/69 misclassify with C<sub>0</sub>, C<sub>3</sub> and C<sub>8-10</sub> respectively. In case of F, 40/56 spectra correctly classify, whereas 16/56 misclassify with C<sub>0</sub>. Results suggest misclassification amongst control groups while F can be classified with 71% efficiency. Misclassification between F and C<sub>0</sub> may be explained based on rapid cell proliferation characteristic of these groups.

a) Model  (no. of animals used, no. of spectra)	$C_0$	$C_3$	$C_{8-10}$	$C_{12-14}$	F
$C_0$ (10, 22)	12	6	1	0	3
$C_3$ (10, 49)	10	22	7	10	0
$C_{8-10}$ (10, 44)	2	6	25	11	0
$C_{12-14}$ (10, 69)	8	15	30	16	0
F (7, 56)	16	0	0	0	40
b) LOOCV  (no. of animals used, no. of spectra)	$C_0$	$C_3$	$C_{8-10}$	$C_{12-14}$	F
$C_0$ (10, 22)	12	6	1	0	3

C <sub>3</sub> (10, 49)	10	22	7	10	0
C <sub>8-10</sub> (10, 44)	2	6	25	11	0
C <sub>12-14</sub> (10, 69)	8	15	30	16	0
F (7, 56)	16	0	0	0	40

**Table 3.1** Transcutaneous *in vivo* Raman spectroscopic study of fibroadenoma progression. Table shows confusion matrix for PC-LDA of C3, C8-10, C12-14, C20 and F; a) model, and b) LOOCV, suggesting clear classification between C and F (C: control, F: Fibroadenoma, weeks post carcinogen/oil treatment shown in subscripts).

PCA results suggest that control and NF groups cluster away from F while PF cluster overlap with F. To further support this observation, the PC-LDA model was evaluated with independent test data. As seen in Table 3.2a, none of the spectra from gavage C<sub>0</sub>, C<sub>3</sub>, C<sub>8</sub>, C<sub>12-14</sub>, C<sub>18</sub> are predicted as F. One spectrum out of 20 spectra from subcutaneous injection Batch II C<sub>7</sub> is predicted as F, while no spectra from C<sub>3</sub> and C<sub>11</sub> are predicted as F (Table 3.2b). None of the spectra from subcutaneous injection Batch III C<sub>3</sub>, C<sub>7</sub> and C<sub>11</sub> and dusting C<sub>3</sub> and C<sub>7</sub> are predicted as F (Table 3.2c). These results support the PCA findings that control spectra are different from F spectra.

a) Test Data (no. of animals used, no. of spectra)	$C_0$	$C_3$	$C_{8-10}$	$C_{12-14}$	$C_{20}$	F
$C_3(5, 20)$	4	9	3	4	0	0
$C_7(5, 20)$	3	7	5	4	0	1
$C_{11}(5, 24)$	5	12	3	4	0	0
b)	$C_0$	$C_3$	$C_{8-10}$	$C_{12-14}$	$C_{20}$	F
$C_3(5, 26)$	6	8	6	6	0	0
$C_7(4, 20)$	8	5	2	5	0	0
$C_{11}(2, 10)$	1	1	7	1	0	0
c)	$C_0$	$C_3$	$C_{8-10}$	$C_{12-14}$	$C_{20}$	F

C <sub>3</sub> (4, 20)	6	1	8	5	0	0
C <sub>7</sub> (5, 25)	3	10	5	7	0	0
d)	C <sub>0</sub>	C <sub>3</sub>	C <sub>8-10</sub>	C <sub>12-14</sub>	C <sub>20</sub>	F
C <sub>3</sub> (9, 41)	6	9	16	10	0	0
C <sub>8-10</sub> (4 20)	0	3	14	3	0	0
C <sub>12-14</sub> (6, 36)	5	2	22	7	0	0
C <sub>20</sub> (5, 20)	0	2	14	4	0	0

**Table 3.2 Transcutaneous *in vivo* Raman spectroscopic study of fibroadenoma progression. Tablee shows evaluation of PC-LDA model with C test data set; a) subcutaneous injection Batch II C, b) subcutaneous injection Batch III C, c) dusting C, and d) gavage C (C: control, weeks post carcinogen/oil treatment shown in subscripts). Results suggest no C spectra predicted as F.**

In case of NF (Table 3.3), 2/108 spectra of gavage NF<sub>3</sub> are predicted as F, while 1/275 spectra of gavage NF<sub>12-14</sub> is predicted as F. None of the spectra from gavage NF<sub>7</sub>, NF<sub>20</sub>, subcutaneous injection Batch II NF<sub>3</sub>, NF<sub>8-10</sub> and NF<sub>12-14</sub> are predicted as F. Thus, NF group spectra differ from F group spectra, as observed in PCA.

a) Test Data (no. of animals used, no. of spectra)	C <sub>0</sub>	C <sub>3</sub>	C <sub>8-10</sub>	C <sub>12-14</sub>	C <sub>20</sub>	F
NF <sub>3</sub> (7, 63)	12	29	10	12	0	0
NF <sub>8-10</sub> (6, 64)	1	7	44	12	0	0
NF <sub>12-14</sub> (7, 84)	5	26	39	14	0	0
b)	C <sub>0</sub>	C <sub>3</sub>	C <sub>8-10</sub>	C <sub>12-14</sub>	C <sub>20</sub>	F
NF <sub>3</sub> (7, 108)	44	29	14	19	0	2
NF <sub>8-10</sub> (5,	11	27	22	21	0	0

81)						
NF <sub>12-14</sub> (17, 275)	53	63	104	54	0	1
NF <sub>20</sub> (8, 126)	12	28	65	21	0	0

**Table 3.3 Transcutaneous *in vivo* Raman spectroscopic study of fibroadenoma progression. Table shows evaluation of PC-LDA model with NF test data set a) subcutaneous injection Batch I NF, and b) gavage NF (C: control, NF: No Fibroadenoma, F: Fibroadenoma, weeks post carcinogen/oil treatment shown in subscripts). Results suggest NF spectra rarely are predicted as F. Note that since all treated rats in dusting and subcutaneous injection Batch II and III developed tumors, they have no NF rats.**

In case of PF (

Table 3.4), 7/63, 5/81 and 8/60 spectra of subcutaneous injection batch I PF<sub>3</sub>, PF<sub>8-10</sub> and PF<sub>12-14</sub> are predicted as F while 6/64, 3/63 and 7/55 spectra from subcutaneous injection batch II are predicted as F. 3/49 and 2/34 spectra from subcutaneous injection batch III PF<sub>7</sub> and PF<sub>11</sub> respectively, 4/36, 3/36 and 7/35 spectra from dusting PF<sub>3</sub>, PF<sub>7</sub> and PF<sub>12</sub> respectively were predicted as F. Thus, PF spectra are predicted as F more frequently compared to control and NF. Thus, results corroborate the outcome of PCA. Overall, results suggest possibility of distinguishing pretumor spectra from controls.

a) Test Data (no. of animals used, no. of spectra)	C <sub>0</sub>	C <sub>3</sub>	C <sub>8-10</sub>	C <sub>12-14</sub>	C <sub>20</sub>	F
PF <sub>3</sub> (7, 63)	23	14	7	12	0	7
PF <sub>8-10</sub> (6, 61)	20	6	21	10	0	4
PF <sub>12-14</sub> (6, 60)	20	9	12	11	0	8
b)	C <sub>0</sub>	C <sub>3</sub>	C <sub>8-10</sub>	C <sub>12-14</sub>	C <sub>20</sub>	F
PF <sub>3</sub> (7, 64)	48	7	0	3	0	6
PF <sub>7</sub> (7, 63)	45	12	0	3	0	3
PF <sub>11</sub> (6, 55)	30	13	2	3	0	7
c)	C <sub>0</sub>	C <sub>3</sub>	C <sub>8-10</sub>	C <sub>12-14</sub>	C <sub>20</sub>	F
PF <sub>3</sub> (5, 46)	6	12	11	17	0	0

PF <sub>7</sub> (5, 49)	26	14	2	4	0	3
PF <sub>11</sub> (4, 34)	6	5	17	4	0	2
d)	C <sub>0</sub>	C <sub>3</sub>	C <sub>8-10</sub>	C <sub>12-14</sub>	C <sub>20</sub>	F
PF <sub>3</sub> (4, 36)	8	4	12	8	0	4
PF <sub>7</sub> (4, 36)	28	4	0	1	0	3
PF <sub>12</sub> (4, 35)	19	6	1	2	0	7

**Table 3.4 Transcutaneous *in vivo* Raman spectroscopic study of fibroadenoma progression. Table shows evaluation of PC-LDA model with PF test data set a) subcutaneous injection Batch I PF, b) subcutaneous injection Batch II PF, c) subcutaneous injection Batch III PF, and d) dusting PF (C: control, NF: No Fibroadenoma, F: Fibroadenoma, weeks post carcinogen/oil treatment shown in subscripts). Results suggest several PF spectra predicted as F. Thus, rats wherein spectra are predicted as F are positively correlated with future fibroadenoma appearance**

Although the results show that PF rat spectra are predicted as F, it should be noted that very few PF spectra are predicted as F while majority are predicted as C / NF. This is probably due to heterogeneity of the sample. Breast is a large organ, but tumor occurs in a small region. Thus, an afflicted breast has largely normal regions with a small abnormal region. Since, spectra were acquired from the several sites, they represent heterogeneity. As discussed by Malini et.al.(111) and Stone et.al.(64), a heterogeneous sample can be considered abnormal even if one spectrum is abnormal. Similar practice is used in histopathological assessment, wherein even if a single region of a single slide is abnormal, the whole sample is declared abnormal. Taking this into consideration, a rat wise analysis was done for the current data. All spectra from each rat irrespective of time of spectra acquisition were considered and even if one spectrum was found to be abnormal (predicted as F), the rat was declared abnormal i.e. will develop tumor. Using this criterion, 82% rats could be correctly predicted to develop tumor in future. Only 5% rats that did not develop tumor were predicted wrongly as abnormal. Thus, sensitivity and specificity of Raman spectroscopy to predict tumor occurrence in this study was 82% and 95% respectively. This further strengthens the evidence for feasibility of breast cancer screening using Raman spectroscopy.

Early detection of breast cancers results in improved prognosis, but the currently available screening tools have several disadvantages. A search for better screening techniques has instigated investigation in several diverse fields such as genomics, proteomics and optical spectroscopy(30, 35). Optical spectroscopic techniques have an edge over others as screening tools since these techniques are rapid, objective and amenable to *in vivo* applications. Several studies have shown the potential of different optical spectroscopic techniques in identification of premalignant lesions(33, 35, 37, 112). But, such premalignant lesions can only help risk

estimation and cannot predict actual tumor development(3). Therefore, the current study aims to analyze spectral data based on outcome – success or failure to develop tumor and establish a correlation between spectral changes and tumor appearance. In order to achieve this, rats were injected subcutaneously with carcinogen DMBA in their left inguinal mammary gland. The advantage of this methodology is that tumor appears approximately around the site of injection. Spectra were acquired from the whole left inguinal breast 0, 3, 8-10 and 12-14 weeks after carcinogen treatment from both treated rats and their corresponding controls. 7 treated rats developed tumor approximately 18 weeks post treatment (PF) while 7 rats did not develop any abnormality despite carcinogen treatment throughout the study (NF). The control rats also did not exhibit any abnormality throughout the study (C). Spectra were also acquired from frank tumor (F). PCA of all spectra acquired showed that C and NF formed clusters distinct from F and that, while majority of PF spectra overlapped with C/NF, several PF spectra overlapped with F.

To further evaluate this finding, a PC-LDA model was trained using C and F spectra. PC-LDA showed that F can be classified from C with 71% efficiency. The model was validated using independent test data. The test data consisted of spectra acquired at rat ages different from that used for training the model, varying time points from time of tumor appearance and different protocols for inducing carcinogenesis. Despite being subjected to a complex test data set, the model could correctly predict all controls as controls, showing high specificity of Raman spectroscopy. Several PF spectra were predicted as F corroborating with the results of PCA. Taking into consideration that tumor bearing breast is heterogeneous, which is largely normal with a small region harboring the tumor showing abnormality, a rat-wise analysis was performed, where in a rat was declared to ‘develop tumor in future’, even if one spectrum irrespective of time of spectra acquisition was predicted as F. Using this criterion, the

specificity and sensitivity of Raman spectroscopy in predicting tumor was found to be 95% and 82%, respectively. These results suggest the possibility of detecting tumor early.

Combined with developments in deep Raman spectroscopy(113-116), it may be possible to identify biochemical changes indicating tumor development at different depths. Moreover, technologies are now emerging that may allow quick imaging of the whole breast, replacing the tedious step by step spectra acquisition procedures. Recently, Schmäzlin et.al.(117) have reported Raman imaging with a fiber-coupled multichannel spectrograph, that allows capture of entire Raman image with one single exposure and chemical mapping without the need for scanning procedure. A Raman chemical map of the whole breast can be obtained using the fiber-coupled multichannel spectrograph system and abnormal spectra (that are predicted as F) can be identified. Images of breasts that are not likely to develop tumor will show normal map, whereas those that are in future going to develop tumor will show a map with largely normal areas harboring small areas of abnormality corresponding to the site of future tumor appearance. Thus, emerging technologies combined with sensitivity of Raman spectroscopy to pretumor changes may allow identification of precancerous changes in the whole breast volume. Further studies in this area may help develop this technique as an alternative/ adjunct breast cancer screening tool.

## **II. Study development of breast adenocarcinoma using *in vivo* Raman spectroscopy**

This subsection applies the methodology used for fibroadenoma progression study to investigate progression of malignant (adenocarcinoma) tumors in rats. Although fibroadenoma study lays the foundation for progression studies, malignant adenocarcinoma

progression is more relevant to the clinics. Therefore, this and the next two sections attempts to decipher spectral changes associated with adenocarcinoma development.

## **Materials and methods**

### **Animals**

A total of 15 SD rats were used in this study. 5 forty seven days old rats were administered oil (controls) while 10 forty seven days old rats were administered DMBA using the nipple injection protocol described earlier.

### **Spectra acquisition**

Spectra were acquired from control and treated rats 0, 2-4, 6-8, 10-12, 14-15 and 17-18 weeks post treatment. The rat breast during these time points of spectra acquisition was visibly and palpably normal. Six treated rats developed breast adenocarcinoma. The time of tumor appearance ranged between 9 – 24 weeks post carcinogen treatment. All spectra acquired before tumor appearance were labeled pretumor (PT). Spectra acquired from rats that did not develop tumor even after 32 weeks post treatment were labeled no-tumor (NT). Spectra were also acquired from frank adenocarcinoma, labeled Tumor/ T.

Raman instrument and Data analysis:

The same has been described in previous sections (Chapter 2 Sections I and II Materials and methods)

## **Results and discussion**

### **Spectral analysis**

Age-related spectral changes described previously were observed in control rats. Minor changes in lipids were observed in PT spectra whereas no changes were observed in NT spectra with respect to controls. Tumor spectra showed major decrease in lipids and increase in proteins and nucleic acids with respect to control.

### **Multivariate analysis**

PCA shows two major clusters – C and NT; PT and T (Figure 3.7). However, there is a considerable overlap between T and C. Moreover, there is no clear distinction between C, NT and PT. Thus, the experiment fails to discriminate pretumor spectra from control spectra. Probable reasons are wide range of tumor appearance times (9<sup>th</sup> -24<sup>th</sup> week post carcinogen treatment) and tumor incidence is ~ 50%. Further standardization of protocol such as a) stereotactic mechanism for injecting carcinogen at precise location and depth b) exploration of vehicles to deliver high concentration with low amount of fluid, are required for large scale studies.

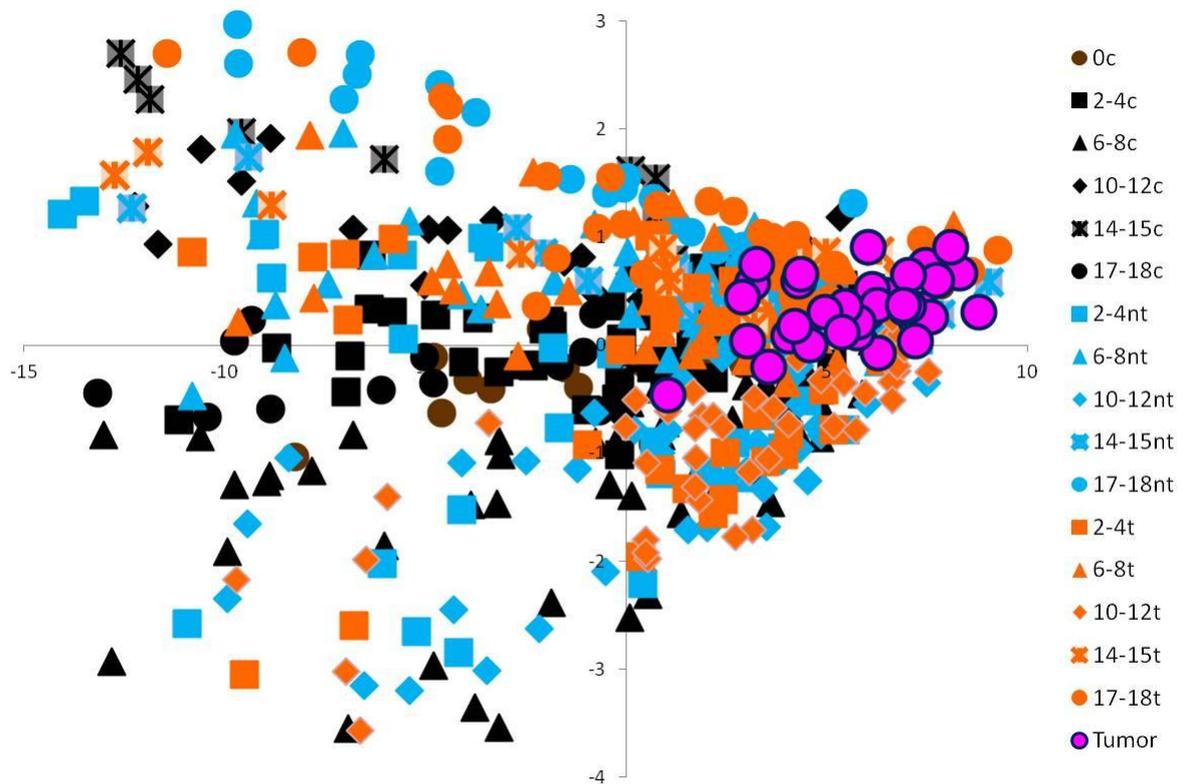


Figure 3.7 Transcutaneous in vivo Raman spectroscopic study of breast adenocarcinoma progression. Figure shows PCA of frank adenocarcinoma, pre adenocarcinoma and control spectra. Results suggest clusters of C/NT and T, while PT overlap with C/NT and T. However, the clusters are not clearly delineated. Further standardization of the model required for better experimental results

### **III. Study development of breast adenocarcinoma using urine based Raman spectroscopy**

One of the major disadvantages of transcutaneous *in vivo* studies is the requirement for pre-existing knowledge of the site of tumor development. This problem can be circumvented by use of body fluids like urine or serum. Urine and serum can provide information regarding biochemical changes throughout the body. Thus, there is no need to scan whole breast/ all breasts in anticipation of tumor. Moreover, body fluids based tests have several added advantages like accessibility, multiple sampling, easy handling, storage and transportation. The following subsection describes detection of pretumor using urine based Raman spectroscopy.

#### **Materials and methods**

##### **Animals**

A total of 42 SD rats were used in this study. 25 fifty days old SD rats were intragastrically administered 65 mg/kg DMBA dissolved in groundnut oil. 20 rats developed breast tumors (histopathologically confirmed adenocarcinoma of breast) approximately six months post carcinogen treatment. Seventeen 50 days old SD rat were administered oil (control). None of the control rats developed breast tumors.

##### **Urine collection**

The rats were restrained; airlifted and voided urine was collected in sterile petridishes. The urine was then transferred to sterile eppendorf tubes using a micropipette. Separate petridish and eppendorf were used for each rat urine sample. Approximately 150-200ul urine was

collected per rat. The urine samples were snap frozen immediately after collection and stored in  $-80^{\circ}\text{C}$ . Using this procedure, urine was collected from six groups:

- a) *Unprocessed control urine (n=9)*: urine samples of control rats were thawed and spectra were acquired from these samples.
- b) *Unprocessed tumor urine (n=9)*: urine samples of tumor bearing rats were thawed and spectra were acquired from these samples.
- c) *Concentrated control urine (n=8)*: urine samples of control were thawed, dehydrated in vacuum using Speed Vac<sup>TM</sup> and rehydrated with 40ul normal saline before spectra acquisition.
- d) *Concentrated tumor urine (n=7)*: urine samples of tumor bearing rats were thawed, dehydrated in vacuum using Speed Vac<sup>TM</sup> and rehydrated with 40ul normal saline before spectra acquisition.
- e) *Concentrated TT urine (n=4)*: urine samples were collected 5 months post carcinogen (DMBA) treatment from visibly and palpably normal rats. These rats were palpated every two weeks after urine collection. Approximately 1 month post urine collection (~ 6 months post carcinogen treatment), these rats developed breast tumors. Biopsy followed by histopathology confirmed the tumors to be adenocarcinoma. The urine samples collected from these rats were labeled 'Tumors Treated' and will hence forth be referred to as 'TT'. Before spectra acquisition, these samples were dehydrated and rehydrated as described above.
- f) *Concentrated NTT urine (n=5)*: urine samples were collected 5 months post carcinogen (DMBA) treatment from visibly and palpably normal rats. However, these rats failed to develop tumor even 8 months post carcinogen treatment. Urine samples

from these rats were labeled as 'No Tumors Treated' and henceforth are referred as 'NTT'. The urine was processed in the same way before spectra acquisition.

Urine samples from control, TT, NTT and tumor bearing rats were collected at the same time. Thus, all samples were collected from age matched rats. The protocol employed in the study has been depicted in Figure 3.8.

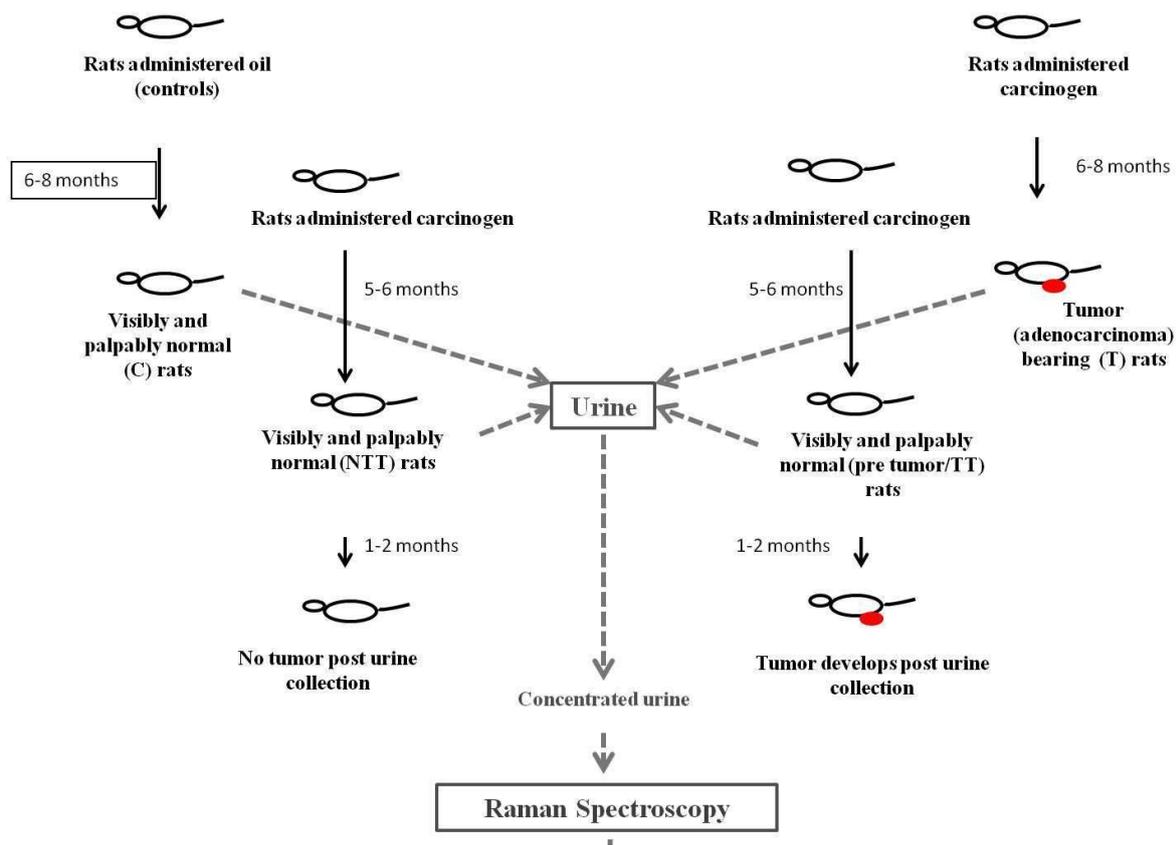


Figure 3.8 Figure shows methodology of urine based pre adenocarcinoma study using Raman spectroscopy

## Spectra acquisition

After passive thawing/rehydrating, samples were subjected to Raman spectroscopy by placing 40  $\mu\text{l}$  sample on calcium fluoride ( $\text{CaF}_2$ ) window and spectra were recorded using Fiber Optic Raman microprobe (Horiba-Jobin-Yvon, France). This Raman system consists of 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 (Figure Figure 3.9). Spectral acquisition details were: Excitation wavelength ( $\lambda_{\text{ex}}$ ) = 785 nm, laser power = 40 mW. Spectra were integrated for 10 seconds and averaged over 6 accumulations. On an average, 8 spectra were recorded from each sample to generate a total of 355 spectra under 6 groups, 81 spectra from unprocessed urine of control rats, 82 from unprocessed urine of tumor bearing rats, 64 spectra from concentrated urine of control rats, 56 from concentrated urine of tumor bearing rats, 40 spectra from concentrated urine of TT rats and 32 spectra from concentrated urine of NTT rats.



## Data Analysis

Data analysis has been described in the previous section (Chapter 2 Section II Materials and methods).

## Results and Discussion

### Spectral analysis

- a) *Unprocessed control and tumor urine*: Vector-normalized average spectrum (Figure 24a i) of control rat urine exhibit urea peaks at  $1004\text{cm}^{-1}$  (symmetrical C–N stretch) and  $1161\text{cm}^{-1}$  (attributed to NH<sub>2</sub> modes) and creatinine peaks at  $680\text{ cm}^{-1}$  (C–NH<sub>2</sub> and C=O stretching, ring vibrations) and  $850\text{ cm}^{-1}$  (C–NH<sub>2</sub> deformation and ring vibrations), as reported elsewhere (118). Mean tumor bearing rat urine spectra (Figure 3.10a i) showed differences in the intensities of several peaks, indicating difference in the concentration of the urine's biochemical components. Differences were seen in the intensity of specific peaks such as decreased intensity of the peaks of urea ( $1006\text{ cm}^{-1}$ ) and creatinine ( $680\text{ cm}^{-1}$ ) in the cancer group compared to control. To elucidate the spectral variations amongst groups, difference spectra were computed by subtracting mean control spectrum from mean tumor spectrum, respectively. The positive peaks of difference spectrum are from the mean tumor spectrum while negative peaks are from mean control spectrum. Tumor – control difference spectra (Figure 3.10a ii) also show a prominent positive urea peak at  $1006\text{ cm}^{-1}$  suggesting increased urea concentration in urine during cancer.

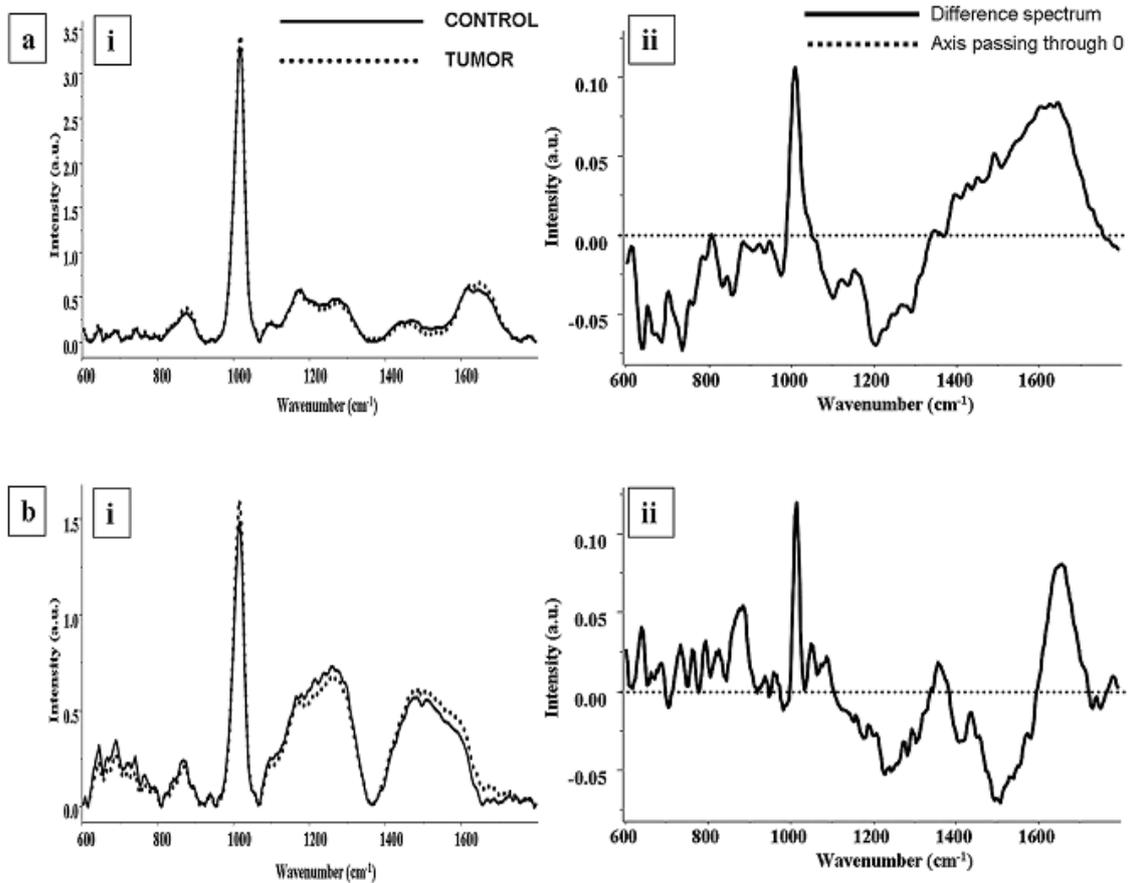


Figure 3.10 Urine based pre adenocarcinoma study. Figure shows a) i) mean spectra of unprocessed urine interpolated in 600-1800cm<sup>-1</sup> region from control and tumor bearing rats, ii) tumor – control difference spectrum, b) i) mean spectra of concentrated urine interpolated in 600-1800cm<sup>-1</sup> region from control and tumor bearing rats, and ii) tumor – control difference spectrum. Change in urea band intensity is clearly observed between control and tumor

b) *Concentrated control and tumor urine:* Mean concentrated control rat urine spectrum have features similar to unprocessed urine with additional peaks at 653, 756, 781, 885 and 925  $\text{cm}^{-1}$  (Figure 3.11b i). Mean tumor spectrum (Figure 3.11b i) show difference in the intensity of urea and creatinine peaks with respect to control. Tumor – control difference spectra show a prominent positive urea peak at 1006  $\text{cm}^{-1}$  suggesting increased urea concentration in urine during cancer (Figure 3.11e).

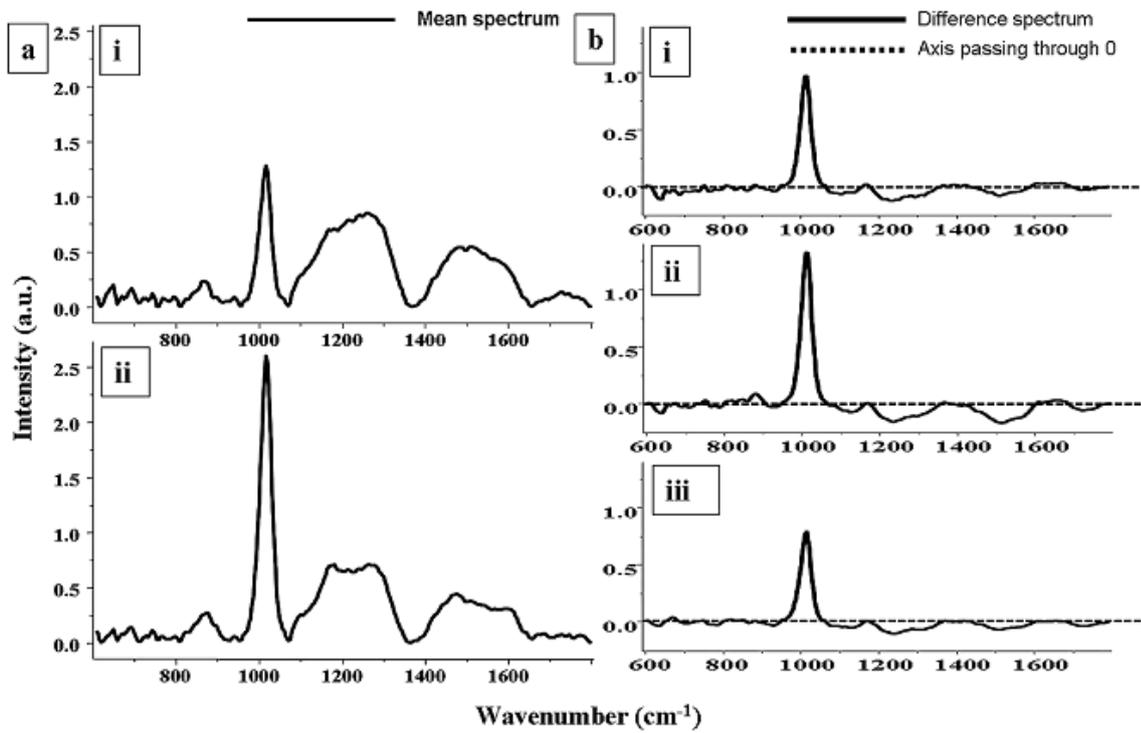


Figure 3.11 Urine based pre adenocarcinoma study. Figure shows a) Mean spectra of concentrated urine interpolated in 600-1800cm-1 region from i) NTT, and ii) TT rats, b) difference concentrated urine spectra i) NTT – control, ii) TT – control, and iii) TT - NTT. Change in urea intensity is again observed between the groups.

c) *Concentrated NTT and TT urine:* Mean concentrated NTT (Figure 3.11a i) and TT (Figure 3.11a ii) rat urine show difference in the intensity of urea peak. TT mean spectrum exhibit highest intensity compared to control and tumor while NTT mean spectrum show lowest concentration compared to all. TT – control (Figure 3.11b i) and Control – NTT (Figure 3.11b iii) difference spectra suggest higher urea concentration in TT compared to control and higher urea concentration in control compared to NTT. TT - NTT difference spectra (Figure 3.11b ii) also suggest increased urea concentration in TT compared to NTT.

## Multivariate analysis

- a) *Unprocessed control and tumor urine*: Preprocessed spectra interpolated in 600-1800  $\text{cm}^{-1}$  range were subjected to PCA for delineating trends in the data set. PCA variance plot and loadings are shown in Figure 3.12a and b. As can be seen in Figure 3.12a, cumulative variance covered by factor 2 and 3 are 81% and 84% respectively. Scatter plot of PCA factors (Figure 3.12c) shows a tendency towards classification of control and tumor bearing rat unprocessed urine.

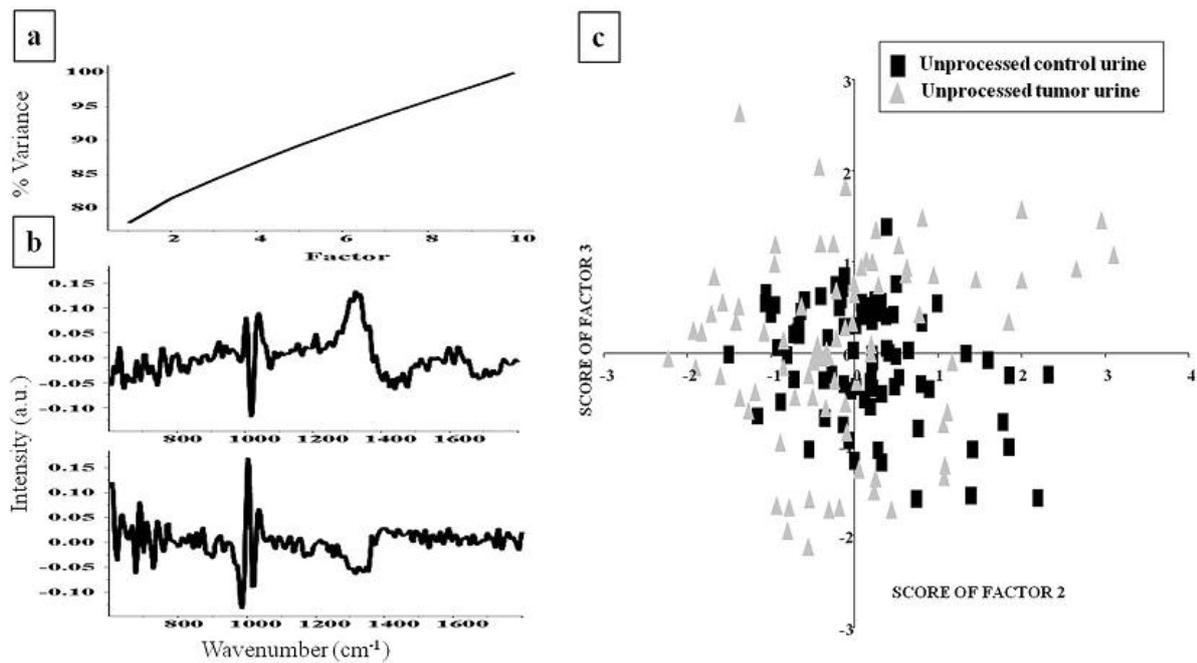


Figure 3.12 Urine based pre adenocarcinoma study. Figure shows PCA of unprocessed urine from control and tumor bearing rats; a) variance plot, b) loading factors 2 and 3, and c) scatter plot, suggesting classification between control and tumor.

To explore the feasibility of classifying the above groups, PC-LDA was used. To avoid over fitting, 9 factors contributing ~ 80 % percent of correct classification; were used (Figure 3.13a). The plot of PC-LDA factors 1, 2 and 3 (Figure 3.13b) show clusters of control and tumor unprocessed urine spectra.

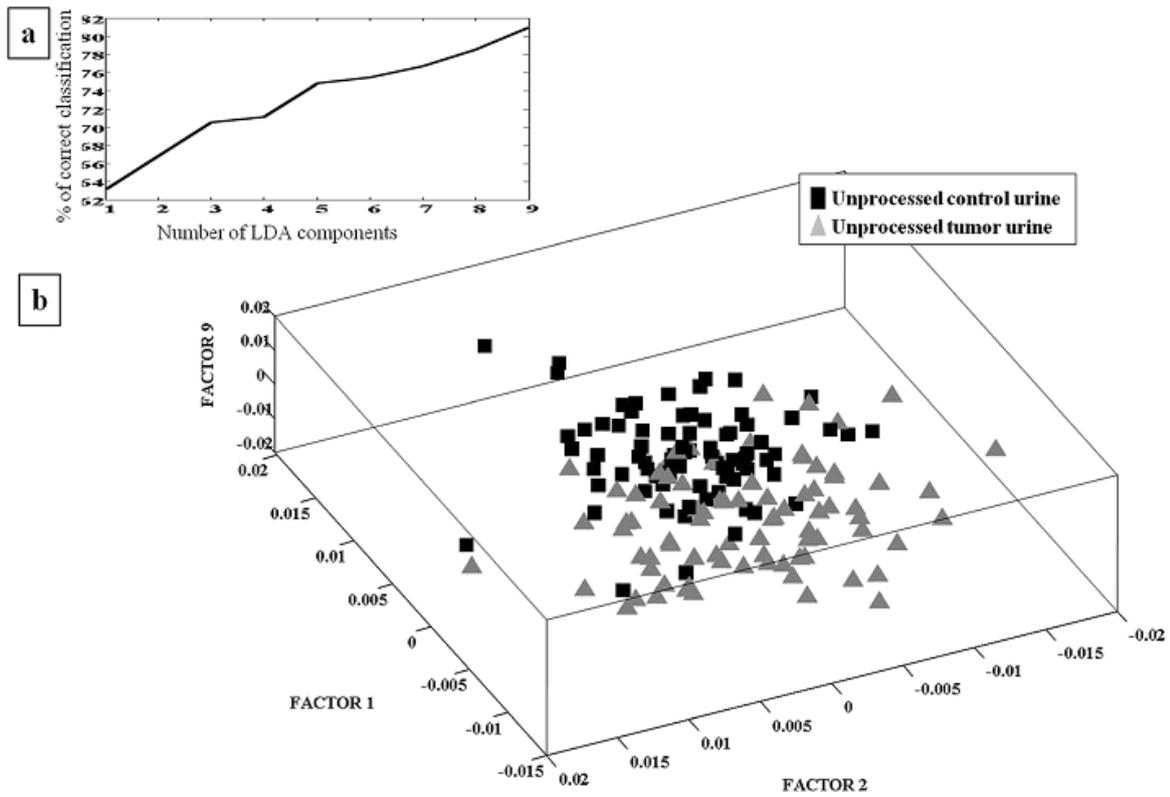


Figure 3.13 Urine based pre adenocarcinoma study. Figure shows PC-LDA of unprocessed urine from control and tumor bearing rats; a) scree plot, and b) scatter plot.

PC-LDA model was built and LOOCV was carried out to evaluate the results obtained by PC- LDA. In analysis of LOOCV as shown in Table 3.5a; 65/81 control spectra correctly classify as control while 16/81 misclassify as tumor; whereas 59/82 spectra are correctly classified as tumor while 23/82 spectra misclassify as control.

Urine is a complex colloidal solution consisting mainly of urea, creatinine, salts and colloids made of glycoprotein, proteins and mucopolysaccharides. Their concentration ranges from 9.3g/L (urea) to 0.67g/L (creatinine). The meager quantity present enhances the possibility of irregular distribution of the components mentioned. Further, as breast cancer progresses, minute concentration changes occur in limited number of urine components while the concentration of all other components of urine probably remain unchanged. These factors may contribute greatly to the misclassification observed amongst groups. Overall, the classification efficiency of control and tumor (using unprocessed urine samples) group was 80 % and 72 % respectively.

- b) *Concentrated control and tumor urine:* Spectra of control and tumor bearing rat concentrated urine interpolated in 600-1800  $\text{cm}^{-1}$  range were also subjected to PCA. PCA variance plot and loadings are shown in Figure 3.14a and b respectively. As can be seen in Figure 3.14a, cumulative variance covered by factor 2 and 3 are 82% and 84% respectively. Scatter plot of PCA factors (Figure 3.14c) shows clusters of concentrated control and tumor bearing rat urine.

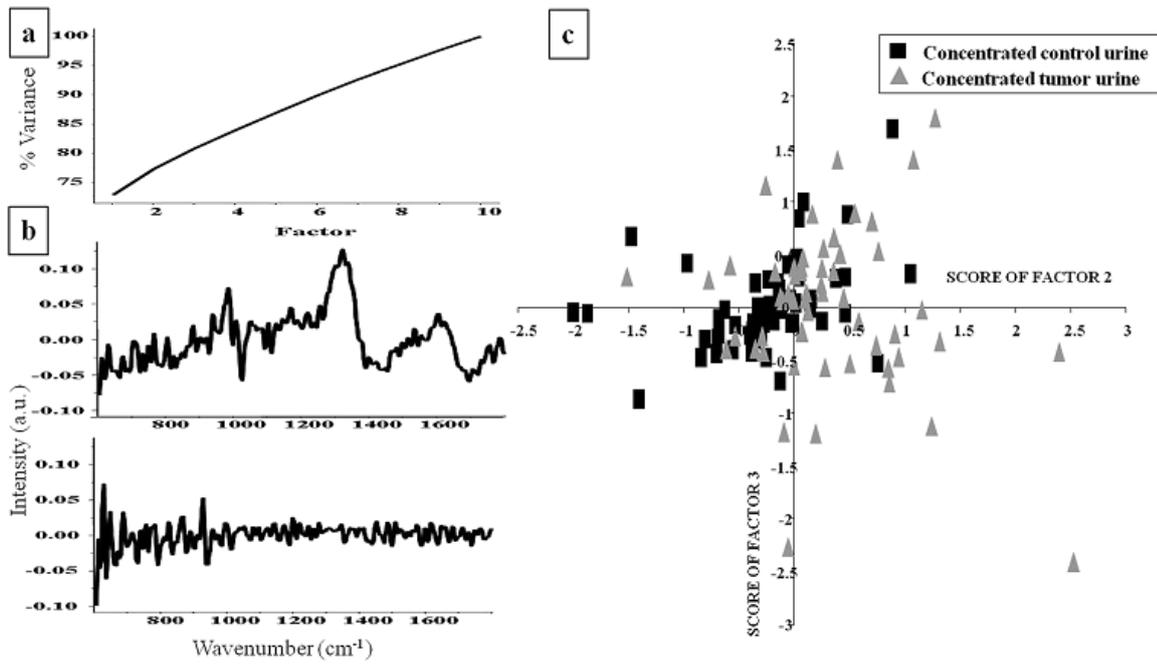


Figure 3.14 Urine based pre adenocarcinoma study. Figure shows PCA of concentrated urine from control and tumor bearing rats; a) variance plot, b) loading factors 2 and 3, and c) scatter plot, suggesting classification between the groups.

To explore the feasibility of classifying the above groups from control, PC-LDA was used. 4 factors contributing ~ 85 % percent of correct classification were applied (Figure 3.15a). The plot of PC-LDA factors 1, 2 and 3 (Figure 3.15b) shows well separated clusters of control and tumor spectra.

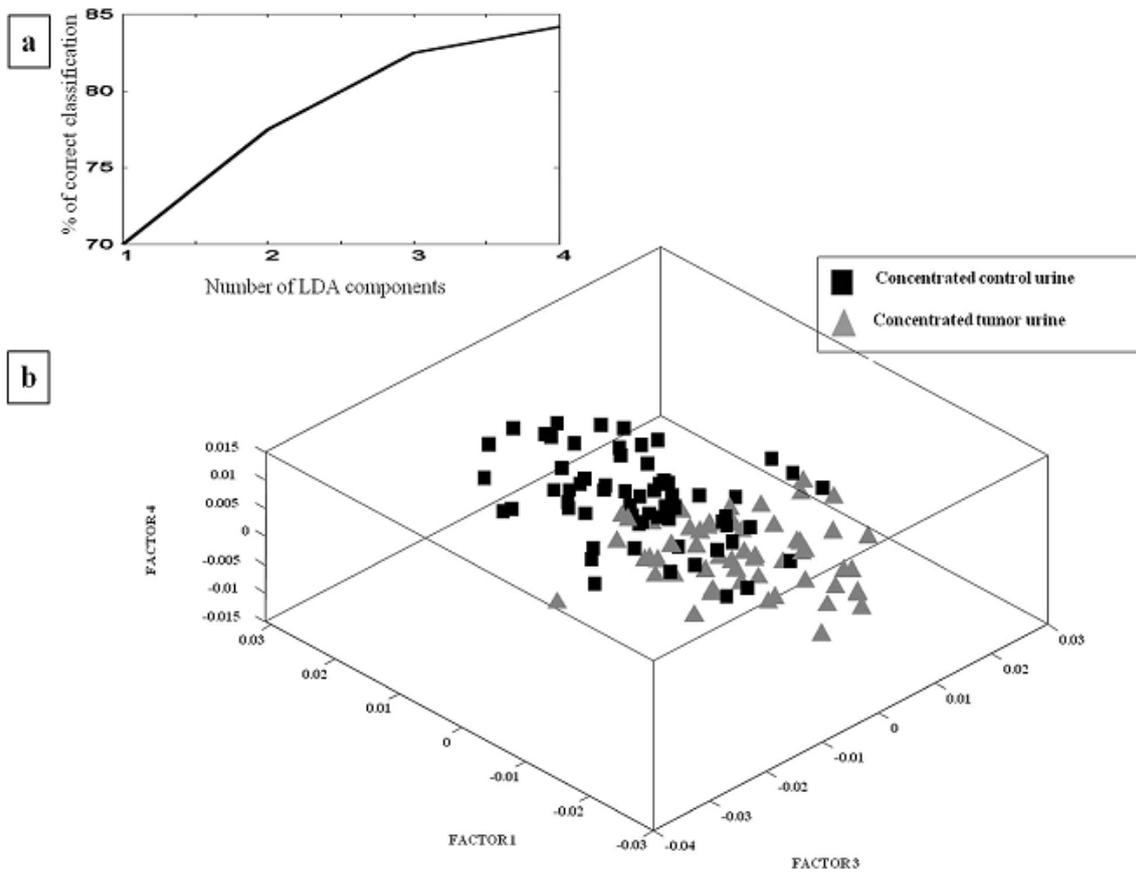


Figure 3.15 Urine based pre adenocarcinoma study. Figure shows PC-LDA of concentrated urine from control and tumor bearing rats; a) scree plot, b) scatter plot, showing clear classification.

LOOCV of results of PC-LDA model built (Table 3.5b); 50/64 control spectra correctly classify as control while 14/64 misclassify as tumor; whereas 51/56 spectra are correctly classified as tumor while 5/56 spectra misclassify as control. As discussed earlier, limiting concentration of urine components and their irregular distribution may explain the observed misclassification.

<b>a) LOOCV (No. of animals, No. of spectra)</b>	<b>Unprocessed urine control</b>	<b>Unprocessed urine tumor</b>
<b>Unprocessed urine control (9, 81)</b>	65 ( 80.24 % )	16
<b>Unprocessed urine tumor (9, 82)</b>	23	59 ( 71.95 % )
<b>b) LOOCV (No. of animals, No. of spectra)</b>		
<b>b) LOOCV (No. of animals, No. of spectra)</b>	<b>Concentrated urine control</b>	<b>Concentrated urine tumor</b>
<b>Concentrated urine control 8, 64)</b>	50 ( 78.12 % )	14
<b>Concentrated urine tumor (7, 56)</b>	5	51 ( 91.07% )

Table 3.5 Urine based pre adenocarcinoma study. Table shows PC-LDA LOOCV confusion matrix; a) unprocessed urine, b) concentrated urine; concentrated urine shows better classification (91%) compared to unprocessed urine (72%)



Although, in this case, the samples are concentrated, the total amount of components present in the sample analyzed is still very low. Since 150-200ul samples were concentrated and used for spectra acquisition, the total quantity of major component urea expected in one urine sample will be ~2ug. Thus, concentrating samples have higher quantity and probably more regular distribution of components compared to unprocessed urine, but the quantities being analyzed are meager and possibly results in the misclassification observed. Overall, the classification efficiency of control and tumor (unprocessed urine samples) group was 78 % and 91 % respectively. While the classification efficiency of control group in case of both unprocessed and concentrated control urine is ~80%, classification efficiency of tumor group in case of concentrated urine is higher (91%) compared to unprocessed urine (72%). Therefore, further studies were conducted using concentrated urine.

c) *Concentrated NTT and TT urine:* To further explore the sensitivity of urine based Raman spectroscopy in diagnosis of breast cancer and possibility of early detection, as mentioned earlier, urine samples were also collected prior tumor development. Spectra acquired from concentrated urine of control, tumor bearing, NTT and TT rats were preprocessed, interpolated in 600-1800  $\text{cm}^{-1}$  range and subjected to PCA and PC-LDA. The PCA variance plot and loading factors 1 and 3 are shown in Figure 3.16a and b respectively. The TT spectra in the PCA scatter plot (Figure 3.16c) shows a tendency of classification.

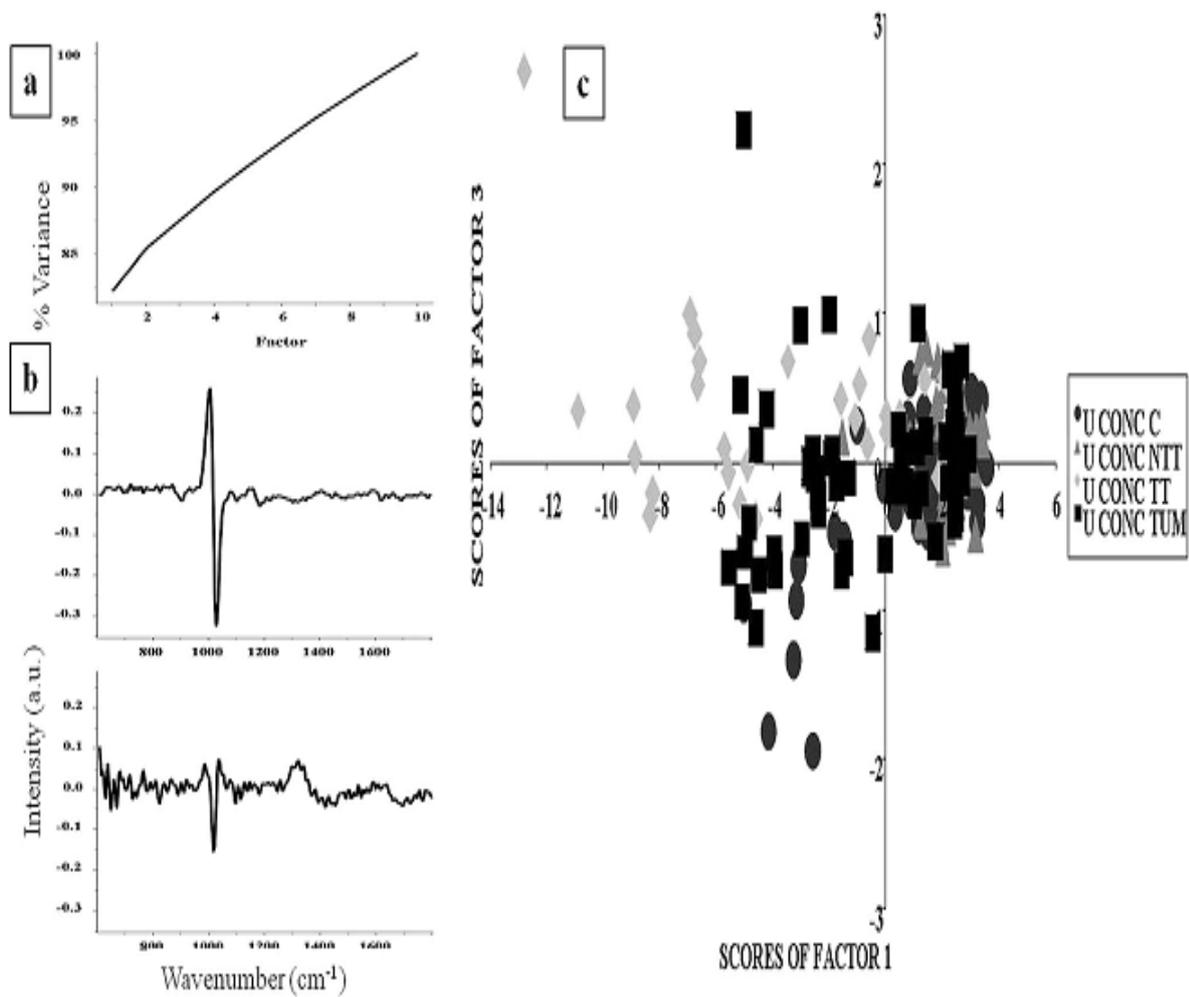


Figure 3.16 Urine based pre adenocarcinoma study. Figure shows PCA of concentrated urine from control, NTT, TT and tumor bearing rats; a) variance plot, b) loading factors 2 and 3, and c) scatter plot, suggesting two clusters C/NTT and TT/T.

The PC-LDA scatter plot (Figure 3.17b) of factors 1, 2 and 3 shows overlapping clusters of control, tumor, TT and NTT. It is however noteworthy, that control and NTT populate the left side of the plot whereas TT and tumor lie on the right side.

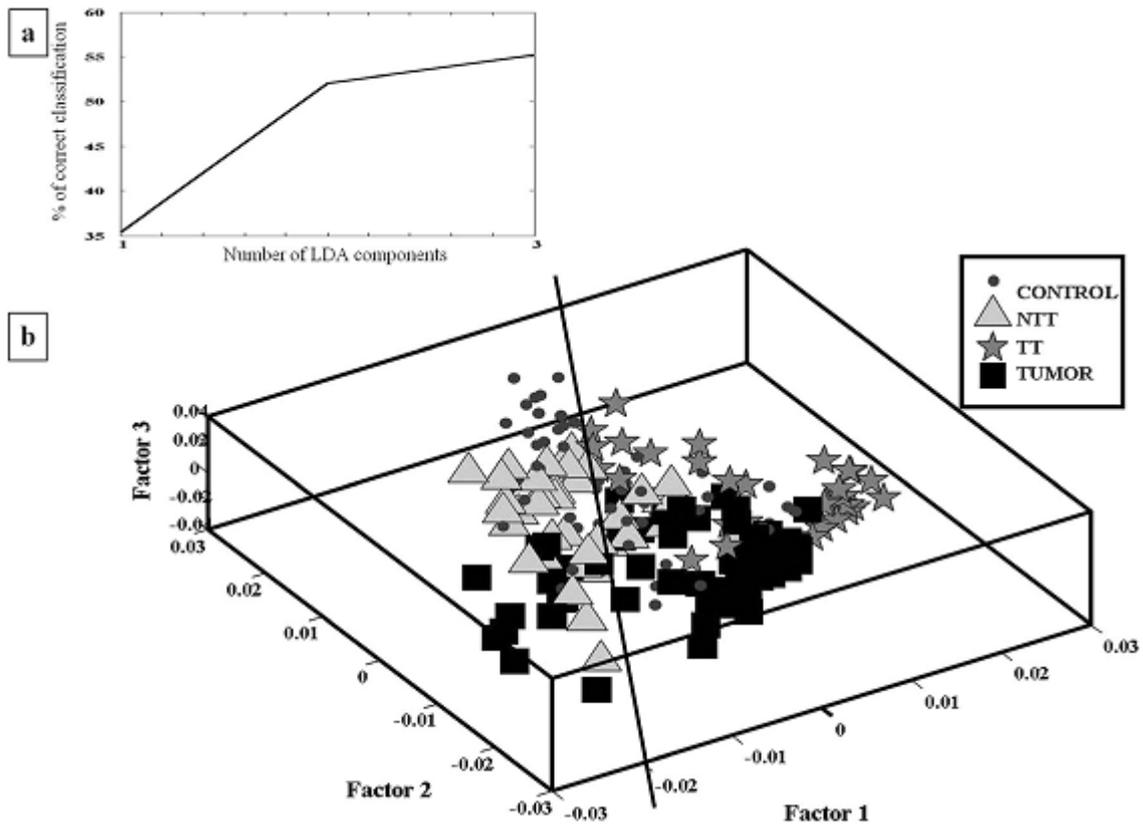


Figure 3.17 Urine based pre adenocarcinoma study. Figure shows PC-LDA of concentrated urine from control, NTT, TT and tumor bearing rats; a) scree plot, and b) scatter plot, again showing two clusters – C/NTT and TT/T

The results of PC-LDA in the form of confusion matrix are shown in Table 3.6a. 34/64 spectra are correctly classified as control, while 9/64 misclassified as NTT, 11/64 misclassified as TT and 10/64 misclassified as tumor. Seventeen of forty NTT spectra were correctly classified as NTT, while 15/40, 2/40 and 4/40 misclassified as control, TT and tumor respectively. In case of TT, 23/32 were correctly classified whereas 7/32 and 2/32 misclassified with control and tumor respectively. 30/56 tumor spectra classified correctly while 3/56, 11/56 and 12/56 misclassified with control, NTT and TT. The results of LOOCV are shown in Table 3.6b. As can be seen, 34/64 spectra are correctly classified as control, while 9/64 misclassified as NTT, 11/64 misclassified as TT and 10/64 misclassified as tumor. 17/40 NTT spectra were correctly classified as NTT, while 16/40, 2/40 and 5/40 misclassified as control, TT and tumor respectively. In case of TT, 22/32 were correctly classified whereas 8/32 and 2/32 misclassified with control and tumor respectively. 30/56 tumor spectra classified correctly while 3/56, 11/56 and 12/56 misclassified with control, NTT and TT.

<b>LOOCV (No. of animals, no. of spectra)</b>	<b>CONCENTRATED URINE CONTROL</b>	<b>CONCENTRATED URINE NTT</b>	<b>CONCENTRATED URINE TT</b>	<b>CONCENTRATED URINE TUMOR</b>
<b>CONCENTRATED URINE CONTROL (8, 64)</b>	<b>34 ( 53.12 % )</b>	9 (14.1%)	11 (17.2%)	10 (15.6%)
<b>CONCENTRATED URINE NTT (5, 40)</b>	16 (40%)	<b>17 ( 42.5 % )</b>	2 (5%)	5 (12.5%)
<b>CONCENTRATED URINE TT (4, 32)</b>	8 (25%)	0	<b>22 ( 62.5 % )</b>	2 (6.3%)
<b>CONCENTRATED URINE TUMOR (7, 56)</b>	3 (5.3%)	11 (19.6%)	12 (21.4%)	<b>30 (51.78 %)</b>

Table 3.6 Urine based pre adenocarcinoma study. Table shows confusion matrix for PC-LDA

LOOCV of control, NTT, TT and tumor, suggesting clear distinction between C/NTT, TT/T.

Despite misclassification amongst groups, 67.2% control spectra classify as either control or NTT, while 82.5% NTT spectra classify as either control or NTT. Control rats were not treated with carcinogen whereas NTT rats did not develop tumor in spite of carcinogen treatment. Therefore, control and NTT urine spectra represent ‘normal’ (non cancerous) condition. TT rats were rats that eventually developed tumor whereas tumor group rats had breast tumors at the time of urine collection. Thus, urine spectra of TT and tumor group rats represent ‘abnormal’ (cancerous) condition. As observed in Table 3.6b, 75% TT and 75% tumor spectra correctly classified as abnormal (TT/Tumor).

In a nutshell, results suggest that rats that did not develop tumor could be classified as ‘normal’ (with ~ 83% efficiency) even though these rats were treated with carcinogen and had high probability of developing tumor, while rats that did develop tumor after carcinogen treatment were classified as ‘abnormal’ (with 75% efficiency) using urine collected before any visible or palpable abnormality.

#### **IV. Study development of breast adenocarcinoma using serum based Raman spectroscopy**

##### *Progression study*

This subsection employs the same experimental design to for pre adenocarcinoma detection using serum based Raman spectroscopy.

##### **Materials and methods**

##### **Animals:**

Same as described in section III Materials and methods section of this chapter

### **Raman spectroscopy and Data Analysis**

The Raman instrumentation has been described in section III Materials and methods section of this chapter. Spectra acquisition parameters and data analysis have been described in the previous sections (Chapter 2 Section I and II Materials and methods).

### **Blood collection and serum separation**

Approximately 1.5 ml blood was collected from tail vein of each rat using scalp vein (Top winged infusion set, 22G). Samples were placed standing for 30 minutes to allow clot formation and then centrifuged at 3500 rpm for 10 minutes. 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 -80°C till use. Care was taken to avoid hemolysis.

## **Results and discussion**

### **Spectral analysis**

Spectra of C and NT rats were similar. Minor differences in protein bands with respect to control were observed in the PT rat spectra. Tumor spectra showed major difference from control serum (described in the next section).

### **Multivariate analysis**

PC-LDA scatter plot shows that cluster of C and NT is distinct from PT and T. Thus, PT spectra can be distinguished from C and NT (Figure 3.18). PC-LDA LOOCV results show 62, 67, 60 and 77% classification efficiency for C, NTT, TT and T, respectively. For a major

study, methodology for multiple blood collection to increase the efficiency of collection, and others factors as mentioned for urine studies need to be standardized.

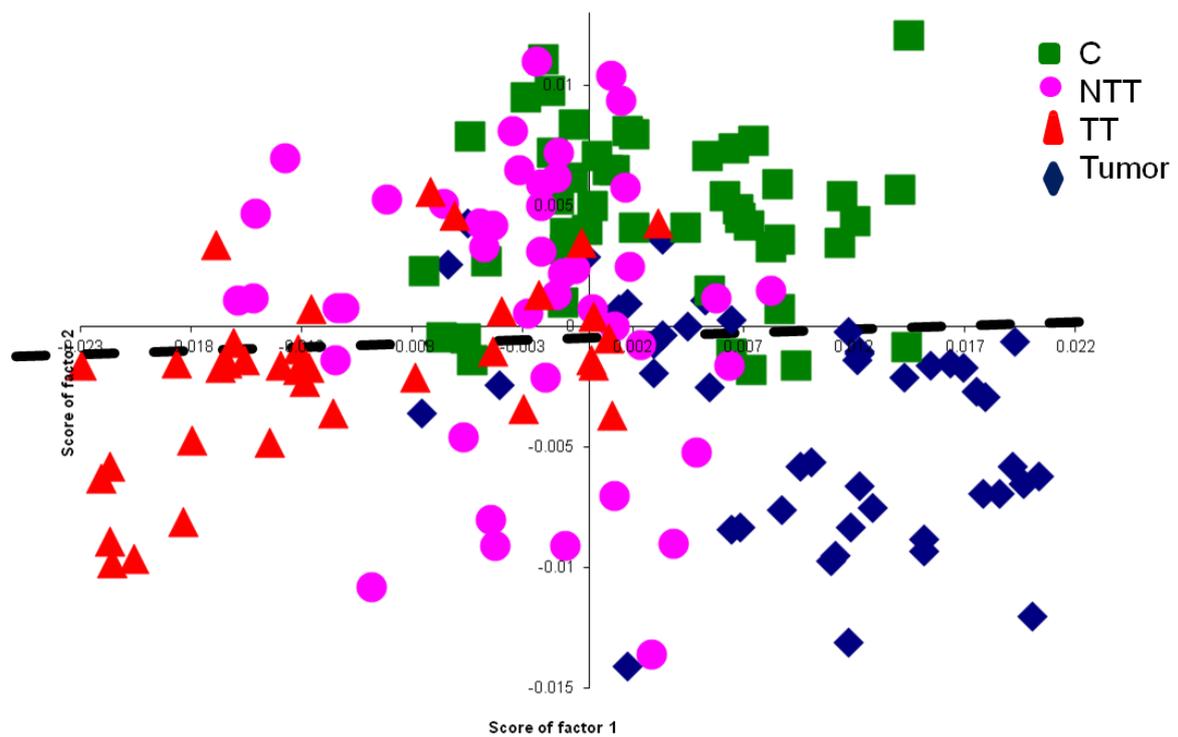


Figure 3.18 Serum based pre adenocarcinoma study. Figure shows PC-LDA scatter plot of C, NTT, TT and tumor serum samples, suggesting a discrete TT cluster. However, all groups show substantial overlap. Further standardization of serum collection protocol may help large-scale studies.

### ***Study of pre and post breast adenocarcinoma resection surgery serum***

This subsection looks at the feasibility of distinguishing serum obtained pre and post breast tumor resection surgery in rats. As mentioned earlier, despite advancements in treatment procedures, the long-term recurrence rate is still 21.4%. Several approaches are used to avoid recurrence. Surgical margin assessment is carried out to aid complete removal of tumor, while lymph nodes are evaluated and removed to contain spread of the disease. Further, breast imaging techniques like mammography, MRI, PET and ultrasound among others are used for surveillance of curative treatments and detection of asymptomatic local recurrence early. However, none of these approaches have achieved major success in preventing/predicting recurrence. Therefore, alternative prognostic/ treatment monitoring tools are the need of the hour.

Potential of Raman spectroscopy as an adjunct tool in breast cancer management has been widely reported. Haka et.al. have explored its application in treatment of breast cancer by demonstrating feasibility of *in vivo* margin assessment during breast cancer surgery. They showed ability of Raman spectroscopy to identify residual tumors by detecting a grossly invisible cancer that, upon pathologic review, required the patient to undergo a second surgical procedure (93). However, use of Raman spectroscopy on inaccessible organs like breast requires invasive operative procedures. Serum based Raman spectroscopy circumvents this problem. Apart from minimal invasiveness, serum/other body fluids based tests have several advantages like accessibility, multiple sampling, easy handling, storage and transportation. The ability of Raman spectroscopy to classify normal and cancer serum has been demonstrated (11, 119). However, no studies have reported spectral patterns of serum pre

and post surgical resection of tumors. Such studies may help explore spectral markers for breast cancer prognosis.

## **Materials and methods**

### **Animals**

A total of 16 SD rats were used in this study. Eleven 50 days old SD rats were intragastrically administered 65 mg/kg DMBA dissolved in groundnut oil. All 11 rats developed breast tumors (histopathologically confirmed adenocarcinoma of breast) six months post induction. Five 50 days old SD rat were administered oil (control). None of the control rats developed breast tumors. The protocol employed in the study has been depicted in Figure 3.19a.

### **Raman spectroscopy and Data Analysis**

The Raman instrumentation has been described in section III Materials and methods section of this chapter. Spectra acquisition parameters and data analysis have been described in the previous sections (Chapter 2 Section I and II Materials and methods).

### **Blood collection:**

The same has been described in Material and methods of section III of this chapter.

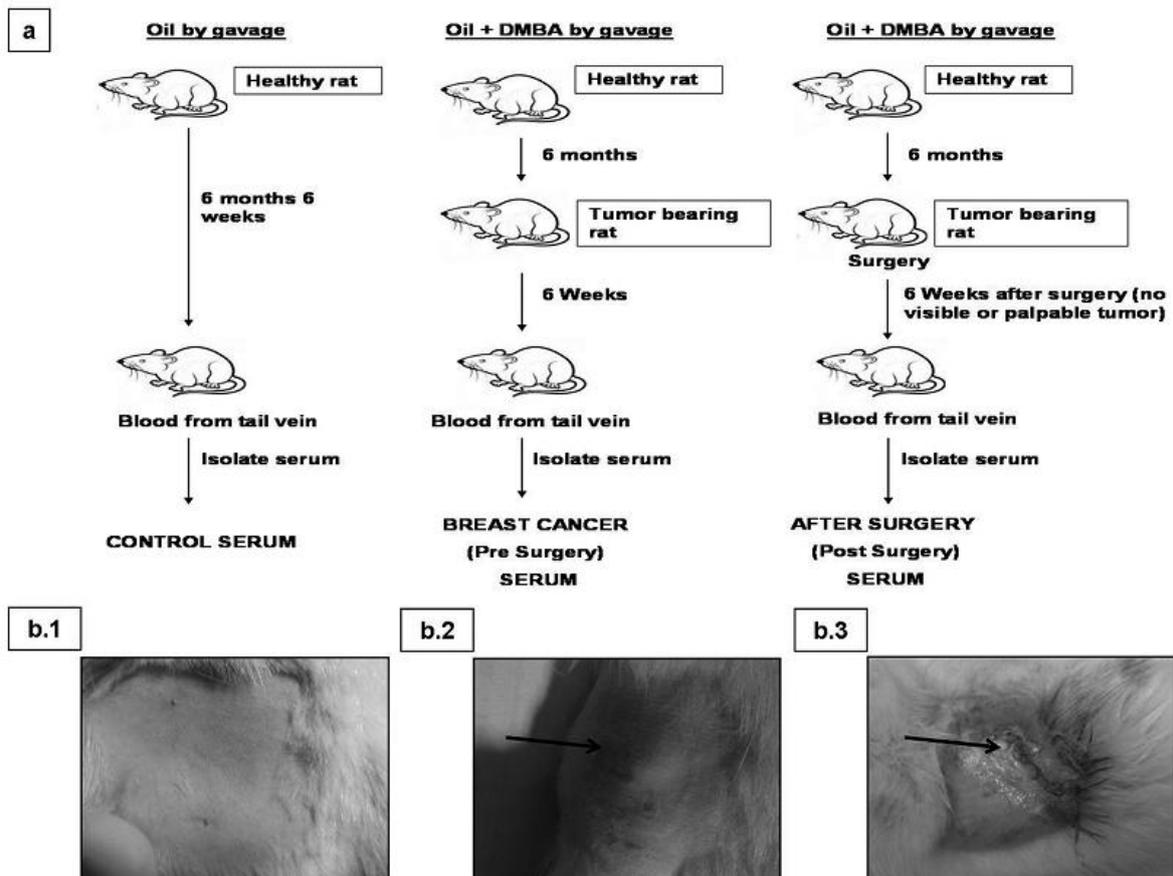


Figure 3.19 Pre and post surgery serum based Raman spectroscopy study. Figure shows a) protocol for induction of carcinogenesis, surgery and serum collection for control, ‘pre-surgical’ and ‘post-surgical’ groups, b) pictures of rat inguinal breast for control (b.1), pre-surgical (b.2) and post-surgical (b.3).

### **Tumor excision surgery**

Tumors of tumor bearing rats were surgically excised by an expert veterinarian (Figure 3.19b). The rats were then checked for recurrence regularly for six weeks. Six rats were visually and palpably normal six weeks post surgery and were further used for blood collection (post surgery samples). Collection of blood six weeks post surgery ensures minimal influence of inflammatory, surgical trauma and tissue loss associated changes in serum, since these changes do not persist for more than 5-7 days (120, 121). No additional drugs or treatment were administered to these rats, except topical application of 5% Povidone Iodine (Wockhardt Health Care) over surgical wound to prevent infection. Blood was also collected from age matched control and tumor bearing rats. One serum sample was collected for each rat and used for spectroscopy.

### **Results and discussion**

#### **Spectral analysis**

Vector-normalized average spectra of control (a.1), pre surgery (a.2) and post surgery (a.3) are presented in Figure 3.20. As is evident from the figure, contributions of proteins, DNA and amino acids like tyrosine, tryptophan, and phenylalanine 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.

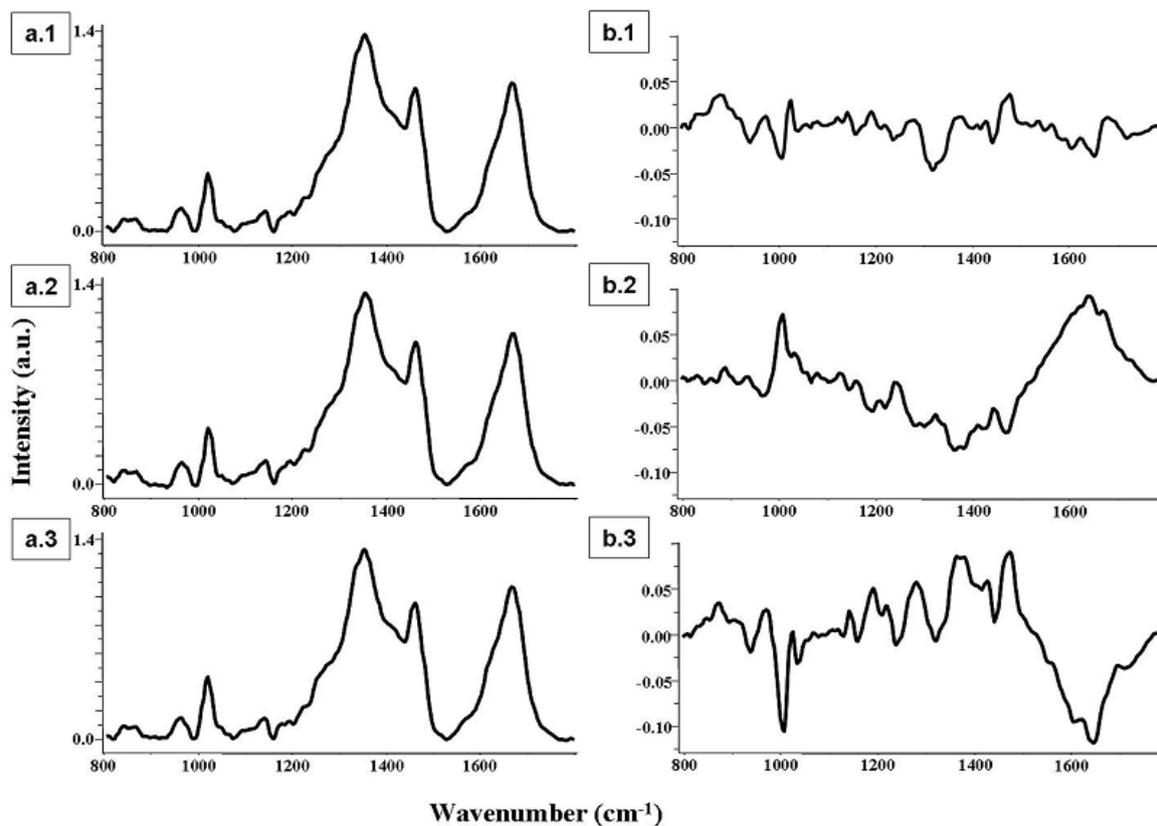


Figure 3.20 Pre and post surgery serum based Raman spectroscopy study. Figure shows mean and difference spectra; a) mean spectra of control (a.1), pre-surgical (a.2), and post-surgical (a.3), b) difference spectra: pre-surgical – control (b.1), post-surgical – control (b.2) and pre-surgical – post-surgical (b.3), suggesting changes in proteins, amino acids and DNA content of serum from different groups.

To elucidate the spectral variations amongst groups, difference spectra were computed by subtracting mean control spectrum from mean 'pre surgical' group and mean 'post surgical' group spectra, respectively (Figure 3.20b). The difference 'pre surgical' group spectrum (pre surgery-control) is presented in Figure 3.20b.1, where the positive peaks are from the mean 'pre surgical' group spectrum while negative peaks are from mean control spectrum. Positive peaks of proteins amide I ( $\sim 1675\text{ cm}^{-1}$ ) and amide III ( $\sim 1270\text{ cm}^{-1}$ ) and amino acids tyrosine and tryptophan (doublet at  $830\text{ cm}^{-1}$ ,  $850\text{ cm}^{-1}$ ) can be seen in 'pre surgical' group. Negative peak of phenylalanine ( $\sim 1002\text{ cm}^{-1}$ ) indicate decreased phenylalanine in 'pre surgical' group serum with respect to control. Decrease in phenylalanine concentration in breast cancer and oral cancer compared to control has been reported previously (122, 123). In contrast, difference 'post surgical' group spectrum shows strong positive peak ( $\sim 1002\text{ cm}^{-1}$ ) suggesting increased phenylalanine in serum post surgery (Figure 3.20b.2). Amide III at  $1240\text{ cm}^{-1}$  as opposed to  $1270\text{ cm}^{-1}$  in 'pre surgical' group difference spectrum probably indicates change in protein profile of serum pre and post surgery. To elucidate possible spectral differences between pre and post surgery sera, a difference spectrum was computed by subtracting post from pre surgery spectrum (Figure 3.20b.3). The strong negative phenylalanine peak ( $\sim 1002\text{ cm}^{-1}$ ) suggests increased phenylalanine post surgery.

### **Multivariate analysis**

Preprocessed interpolated in  $800\text{-}1800\text{ cm}^{-1}$  range spectra were subjected to PCA for delineating trends in the data set. PCA variance plot and loadings are shown in Figure 3.21a and b. As can be seen in Figure 35a, cumulative variance covered by factor 2 and 3 are 57.6% and 69.1% respectively. Scatter plot of PCA factors (Figure 3.21b) shows distinct 'pre' and 'post-surgical' clusters which group around control.



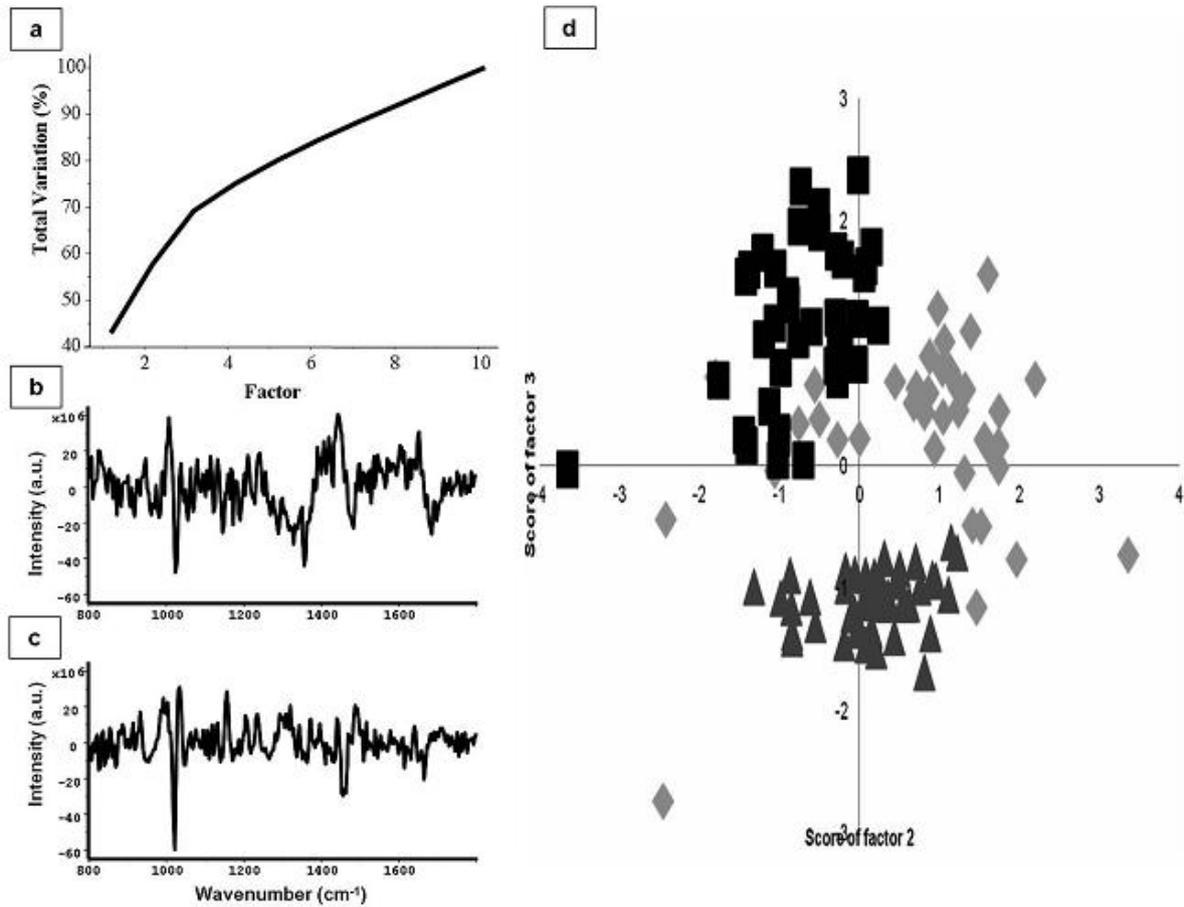


Figure 3.21 Pre and post surgery serum based Raman spectroscopy study. Figure shows PCA of control, pre-surgical and post-surgical groups; a) variance plot, b) loading factor 2, c) loading factor 3, d) scatter plot of factors 2 and 3, suggesting classification between the different groups mentioned.

To explore the feasibility of classifying ‘pre-surgical’ group and ‘post-surgical’ group from control, PC-LDA was used. Spectra were used for analysis. To avoid over fitting, 4 factors contributing ~ 89 % percent of correct classification; were used (Figure 3.22a). The plot of PC-LDA factors 1, 3 and 4 (Figure 3.22b) shows that ‘pre-surgical’ and ‘post-surgical’ groups cluster around control but are distinct from each other.

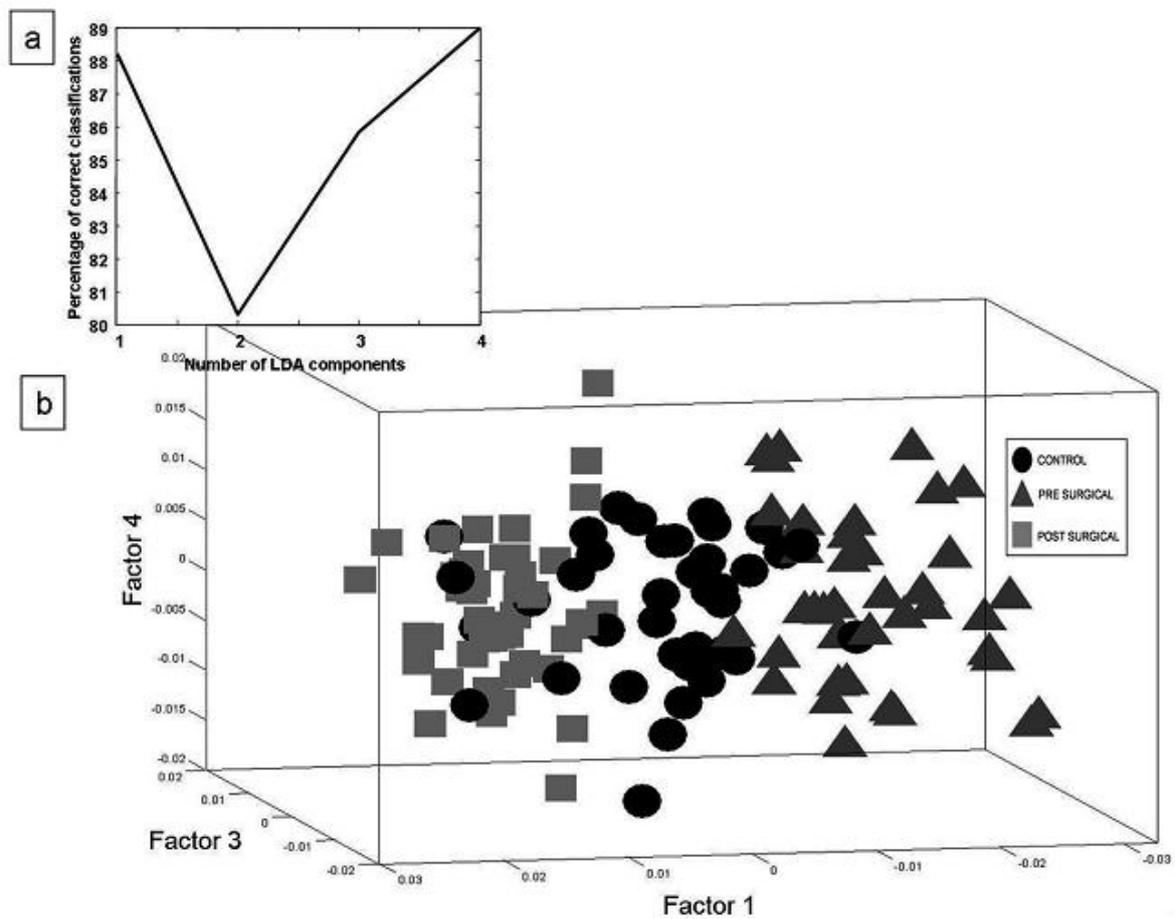


Figure 3.22 Pre and post surgery serum based Raman spectroscopy study. Figure shows PC-LDA of control, pre-surgical and post-surgical groups; a) scree plot, b) scatter plot of factors 1, 3 and 4, again suggesting classification between these groups.

The confusion matrix for PC-LDA model building is shown in Table 3.7a. In analysis of LOOCV as shown in Table 3.7b; 33/40 spectra correctly classify as control while 34/39 'pre surgical' group spectra are correctly classified. Several studies have shown major protein profile changes in breast cancer serum with respect to control (124). These changes may explain the observed difference between control and 'pre surgical' group serum. 4/40 control spectra misclassify as 'pre surgical' group whereas 5/39 'pre surgical' group spectra misclassify as control. Although several changes occur during carcinogenesis, not all proteins change. This may explain misclassification between the groups. Further, serum is a colloidal dispersion (125), wherein the microvolume (50 $\mu$ l) being probed is heterogenous at microscopic level. Such a small volume corresponds to limiting concentrations of serum constituents like proteins, lipids, electrolytes and thus a possibility of irregular distribution of these solutes even after several rounds of mixing. This heterogeneity may also explain the misclassification between these groups.

<b>LOOCV (No. of animals, no. of spectra)</b>	<b>Control</b>	<b>Before Surgery</b>	<b>After Surgery</b>
<b>Control (5, 40)</b>	<b>33 (82.5%)</b>	4	3
<b>Before Surgery (5, 39)</b>	5	<b>34 (87.2%)</b>	0
<b>After Surgery (6, 48)</b>	4	0	<b>44 (91.6%)</b>

Table 3.7 : Pre and post surgery serum based Raman spectroscopy study. Table shows confusion matrix for PC-LDA LOOCV of control, pre-surgical and post-surgical groups, suggesting classification efficiency of ~86% between control, pre and post surgery serum spectra.

44/48 'post surgical' group spectra are correctly classified. None of 'post surgical' group spectra misclassify with 'pre surgical' group. Literature suggests increase in several proteins; CSF1, THSB2, IL6, IL7, IL16, FasL and VEGF-B post surgery (126). These changes may be responsible for classification from control and 'pre surgical' group. 4/48 'post surgical' group spectra misclassify with control. This may be due to loss of tumor associated factors post surgery or serum heterogeneity or both.

The classification efficiency of control, 'pre surgical' group and 'post surgical' group was 82 %, 87 % and 91 % respectively. As discussed earlier, the only difference between 'pre surgical' group and 'post surgical' group is presence or absence of tumor. Thus, the high classification efficiency of 'pre surgical' group and 'post surgical' group and absence of misclassification amongst these groups suggest a possible role of tumor-associated factors in classification. Spectral identification of tumor-associated factors may help determine risk of recurrence.

## **Summary**

1. The transcutaneous *in vivo* fibroadenoma progression study suggested possibility of a) distinguishing pretumor spectra from control spectra b) prediction of tumor appearance in test rats
2. Although a similar methodology was applied to study adenocarcinoma, pre adenocarcinoma could not be distinguished from control using *in vivo* spectroscopy. This was possibly due to deficiency in the model used. Further standardization of the model may help achieve better results

3. Pre adenocarcinoma condition could however be identified and distinguished from control using urine based Raman spectroscopy. Serum-based Raman spectroscopy also showed encouraging results in discriminating pre adenocarcinoma condition from rats. However, further standardizations in serum collection and storage may yield better results. Serum-based Raman spectroscopy could distinguish ‘pre’ and ‘post’ breast adenocarcinoma surgery conditions.

Thus, the chapter suggests possibility of identifying spectra indicative of tumor development before tumor appearance.

**CHAPTER 4    STUDY RAMAN SPECTRAL  
SIGNATURES OF EXPERIMENTAL LUNG  
METASTASIS FROM BREAST CANCER CELL  
LINE**

This chapter describes fiber-optic based Raman spectroscopy of primary lung tumor, primary breast tumor, normal lung, normal breast and breast metastatic lesions in lungs of mouse models and classification between primary and metastatic lesions using multivariate statistical analysis.

### **Rationale:**

As elaborated earlier, the efficacy of breast cancer treatment is low. This can majorly be attributed to metastatic spread of breast cancer. Metastatic relapse remains incurable with average survival less than 2 years (3). It is estimated that ~6% of patients diagnosed with breast cancer have metastatic disease at the time of diagnosis and 20% to 50% patients first diagnosed with primary breast cancer will eventually develop metastatic disease, the most common sites of distant metastases being bone (41.1%), lung (22.4%), liver (7.3%), and brain (7.3%). Differential diagnosis between primary lung lesion and breast metastatic lesion in lung has been reported to be especially difficult (99, 127-137). These lesions are histopathologically, morphologically and radiographically similar. Further, both stain positively for cytokeratin (CK) 7 and negative for CK20.1. In addition, many lung carcinomas do not stain for thyroid transcription factor (TTF)-1, a known strategy to identify lung malignancy and on the other hand, some breast cancers do not stain for estrogen receptor (ER), an indicator for presence of breast cancer cells. This further compounds the problem of distinguishing primary lung lesions from breast metastatic lesions. It is pertinent to note that differential diagnosis is vital for effective therapeutic intervention and favorable prognosis.

Lung cancer causes the highest cancer related mortality worldwide. The number of deaths due to lung cancer is as high as the combined deaths caused by next four most fatal cancers – breast, prostate, colon and pancreas. Approximately 1.8 million lung cancer cases and 1.59

million lung cancer related deaths were estimated in 2012 (1). Most of the lung cancer cases are detected at advanced stages of the disease, resulting in 5-year survival rate as low as 16%. Studies have shown better prognosis with early detection of lung cancer (6). Lung cancer screening using sputum cytology and chest radiography did not result in reduction of advanced lung cancer cases or deaths (138, 139). The current USPSTF guidelines recommend annual screening for lung cancer with low-dose computed tomography (LDCT) in adults (140). Randomized studies have indicated significantly fewer lung cancer deaths in cohort screened using LDCT compared to control group (100). One of the disadvantages of LDCT is difficulty in confident diagnosis of pulmonary metastasis (141), especially since lung is the most common site of metastasis with approximately 50% extrathoracic cancer patients exhibiting lung metastasis (142). Conventional radiography also fails to distinguish primary from metastatic lesions in 2/3<sup>rd</sup> of cases (143). Detection of metastases is however vital for effective therapeutic intervention and favorable prognosis. Hence, there is a need for sensitive, rapid, objective, cost effective alternate tools for diagnosis of metastasis.

Raman spectroscopy has also been explored to study metastasis. Oliviera *et. al.* have shown feasibility of differentiating primary and metastatic cutaneous melanoma while Terentis *et.al.* have demonstrated discrimination of live human metastatic melanoma cells from skin fibroblasts using Raman microspectroscopy (144, 145). Distinction between metastatic and non metastatic cell lines has been shown using microspectroscopy (97, 146). Fullwood *et.al.* have used Raman and immersion Raman spectroscopy to study metastatic brain tumors and have explored the possibility of identifying primary sites of origin (98). Stone *et.al.* have studied lymph node metastasis in breast cancer using Raman spectroscopy (14, 64, 77).

Recently, Short *et.al.* have reported development of a probe to collect real time *in vivo* lung spectra and have successfully acquired spectra of lungs (147). This may prove to be an

invaluable non-surgical adjunct to LDCT, wherein lesions detected by LDCT may be categorized into primary lung cancer and pulmonary metastases using Raman spectroscopy. Literature suggests that breast is the most common cancer that causes lung metastasis (142). Several Raman spectroscopic studies on breast cancer have been reported in literature. The ability of this technique to classify normal breast tissues from benign and malignant tissues has been shown *ex-vivo*. Feasibility of distinguishing normal breast from breast tumors in rats has been demonstrated both *ex vivo* and *in vivo* (10, 38, 63, 76, 81, 82, 85, 148). *In vivo* spectroscopy for surgical margin assessment during partial mastectomy surgery, deep spectroscopy of breast tissues and detection of microcalcification to detect breast cancer early has also been demonstrated (12, 88, 93, 113, 115).

In light of real time *in vivo* probe ability and extensive literature on Raman spectroscopy based detection of breast cancer, lung cancer and metastatic condition, the current study aims to evaluate ability of fiberoptic based Raman system to distinguish breast metastatic lesions in lung from primary lung tumors. In this study, primary lung tumors and breast metastatic lesions were induced in lungs of mouse models. Spectra were then acquired from normal lung, primary lung tumor and breast metastatic lesions induced in lungs. To ensure robust analysis, primary breast tumors were also induced in mice and spectra from normal breast and primary breast tumor were also incorporated in the study. Spectra from all five groups were analyzed using PCA and PC-LDA.

## **Materials and Methods**

### **Animals**

Tumors from Mouse Mammary Tumor Virus (MMTV) - induced spontaneous tumorigenesis model, C3H Jax mouse were harvested after sacrificing the mouse by cervical dislocation and

used to acquire spectra of primary breast tumor (n=4). Lung adenoma was induced by intraperitoneal injection of benzo[a]pyrene (B[a]P) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) once a week for 8 weeks in AJ mice (149). Mice were sacrificed after 28 weeks by cervical dislocation and excised lungs were used to acquire spectra of primary lung tumor (n=4). Spectra were also acquired from normal breast (n=5) and normal lung tissues (n=6). Breast metastasis in lung was induced by intravenous injection of C3H Jax tumor single cell suspension ( $4 \times 10^6$  cells) into new 8 weeks old C3H Jax mice. After 3 weeks, the mice were sacrificed by cervical dislocation, lungs harvested (n=8) and used for spectroscopy.

### **Sectioning and H&E staining:**

The tissues were cut into small pieces. One spectrum was acquired from each piece. Immediately after spectra acquisition, the spot where laser hit the tissue was marked with India ink and fixed with 2% glacial acetic acid. Paraffin embedded blocks was prepared using established protocols. Sections were obtained from the marked spot and H&E staining was carried out for these sections. These were then evaluated by a pathologist. Since sections were obtained from the region where laser interacted with the tissue, the pathology and spectra can be directly correlated.

### **Raman spectroscopy and Data Analysis**

The Raman instrumentation, spectra acquisition parameters and data analysis have been described in the previous sections (Chapter 2 Section I and II Materials and methods).

## **Results and Discussion**

### **Spectral analysis**

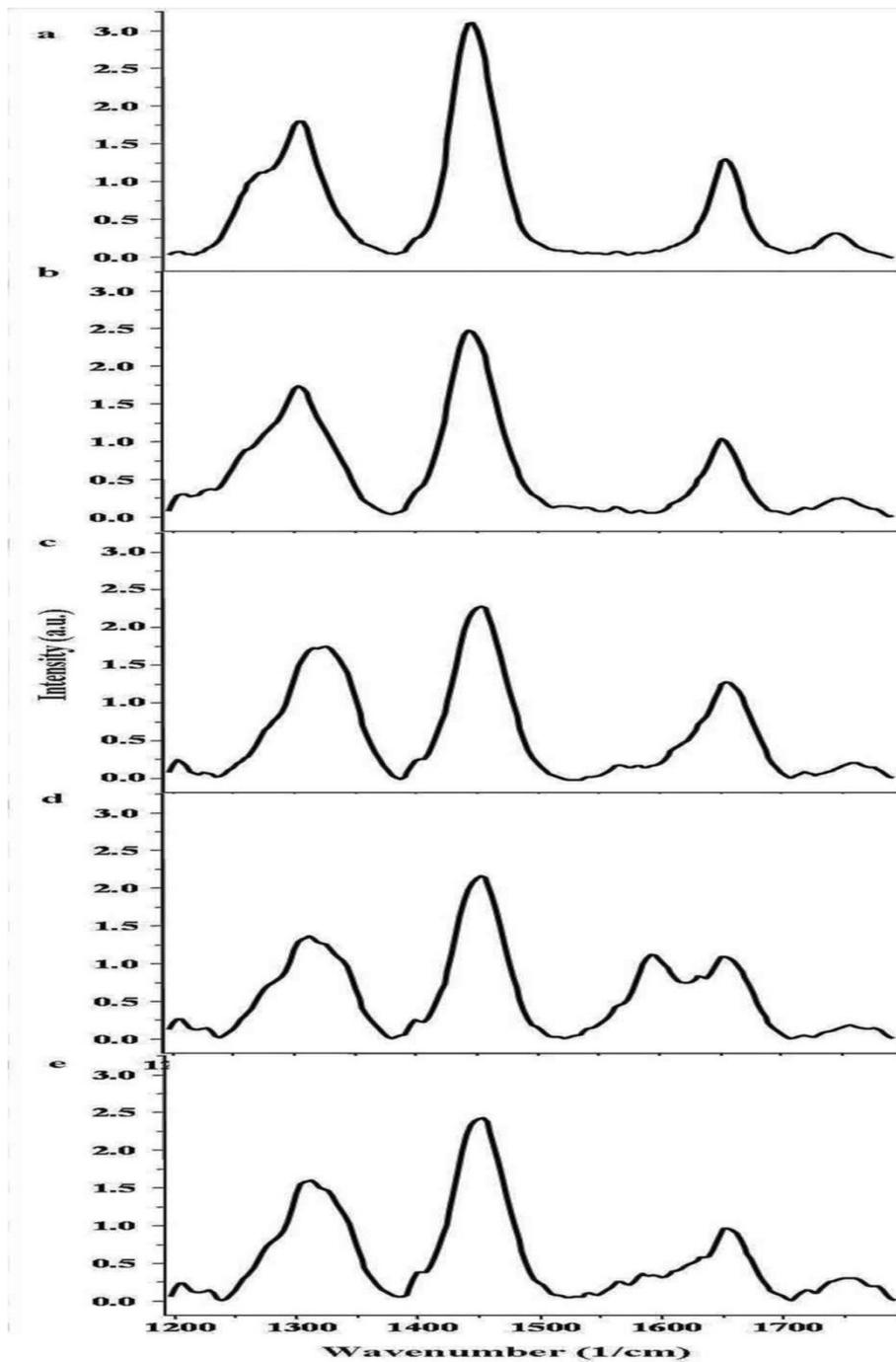
*Mean spectra:* The spectral features of mean control breast spectrum (

**Figure 4.1 a)** - 1743  $\text{cm}^{-1}$  (C=O ester); 1653  $\text{cm}^{-1}$  (amide I) ; 1440  $\text{cm}^{-1}$  ( $\delta$  CH<sub>2</sub>); 1301  $\text{cm}^{-1}$  ( $\tau$ CH<sub>2</sub>); and 1271  $\text{cm}^{-1}$  can be attributed to lipids. Mean breast tumor spectrum (Figure 37 b) show broad amide I and features in 1200-1400  $\text{cm}^{-1}$  region, suggesting dominance of proteins and DNA in tumor. These findings corroborate well with earlier studies (9, 63). The mean control lung spectrum (

**Figure 4.1d)** exhibits features at 1650  $\text{cm}^{-1}$  (amide I), 1311  $\text{cm}^{-1}$ , 1335  $\text{cm}^{-1}$ , 1450  $\text{cm}^{-1}$  ( $\delta$  CH<sub>2</sub>), 1301  $\text{cm}^{-1}$  ( $\tau$ CH<sub>2</sub>) and 1590  $\text{cm}^{-1}$  as reported earlier (106). The 1590  $\text{cm}^{-1}$  band has been suspected to be carbon particles, since the mice were sacrificed using CO<sub>2</sub> asphyxiation in the reported study. However, since in this study, cervical dislocation was used to sacrifice mice, the 1590  $\text{cm}^{-1}$  along with 1311  $\text{cm}^{-1}$  may be attributable to cytochrome. Cytochromes are present abundantly in lungs since the organ is involved with the oxygen transfer process. Mean lung adenoma spectrum (

**Figure 4.1e)** show loss of 1590  $\text{cm}^{-1}$  band with respect to control. Broad amide I and amide III with respect to control breast, breast tumor, control lung and lung tumor mean spectrum (

**Figure 4.1c)** is observed in mean breast metastasis spectra.



**Figure 4.1 Discrimination of primary lung from metastatic breast lesions. Figure shows mean spectra interpolated in 1200-1800cm<sup>-1</sup> region of a) breast control, b) primary breast tumor, c) breast cancer metastasis in lung, d) lung control, and e) primary lung tumor, suggesting difference in protein, lipid, DNA and 1590 cm<sup>-1</sup> band.**

*Difference spectra:* To elucidate the spectral variations amongst groups, difference spectra were computed. Breast control – breast tumor (Figure 4.2ai) suggests lower amount of proteins (negative peak at 1671, 1456, 1471  $\text{cm}^{-1}$ ) and DNA (negative peak at 1480, 1340  $\text{cm}^{-1}$ ) and higher lipid content (positive peaks at 1743, 1440  $\text{cm}^{-1}$ ) in breast control compared to tumors. Control breast – breast metastasis difference spectrum (Figure 4.2aii) suggests higher lipid content (positive peaks at 1740, 1440  $\text{cm}^{-1}$ ) and lower DNA content (negative peaks at 1340, 1480  $\text{cm}^{-1}$ ) in control with respect to metastasis. Breast tumor – breast metastasis difference spectrum (Figure 4.2aiii) suggests decreased DNA (negative peaks at 1340, 1470  $\text{cm}^{-1}$ ) in breast tumor compared to metastasis.

Lung control – lung tumor difference spectrum (Figure 4.2bi) suggest lower lipid content (negative peaks at 1740, 1440 and 1301  $\text{cm}^{-1}$ ) in lung control compared to lung tumor. The same has been demonstrated in other studies (106). The loss of 1311  $\text{cm}^{-1}$  band is another characteristic of lung tumor spectra. Lung control – breast metastasis (Figure 4.2bii) and lung tumor – breast metastasis difference spectrum (Figure 4.2biii) suggest lower DNA content (negative peaks at 1340, 1470  $\text{cm}^{-1}$ ) in lung control and lung tumor compared to metastasis.

Breast control – lung control difference spectrum (Figure 4.2ci) suggest lipid dominance in breast control while 1590  $\text{cm}^{-1}$  characterizes lung control. Breast control – lung tumor difference spectra (Figure 4.2cii) suggests lower DNA content in breast control compared to lung tumor (negative peaks at 1480, 1340  $\text{cm}^{-1}$ ). Breast control – lung control difference spectra (Figure 4.2ciii) also highlight the characteristic 1590  $\text{cm}^{-1}$  band of lungs. Breast tumor – lung tumor spectra (Figure 4.2civ) suggest increased DNA content in lung tumor compared to breast tumor.

Overall, protein and DNA content is least in control tissues, comparatively higher in primary tumors and highest in metastatic lesions. Lipid content is highest in control breast, comparatively lesser in lung tumor and is least in control lung, primary breast cancer and metastatic lesions. The 1590  $\text{cm}^{-1}$  band is characteristic of control lung that disappears in lung tumor.

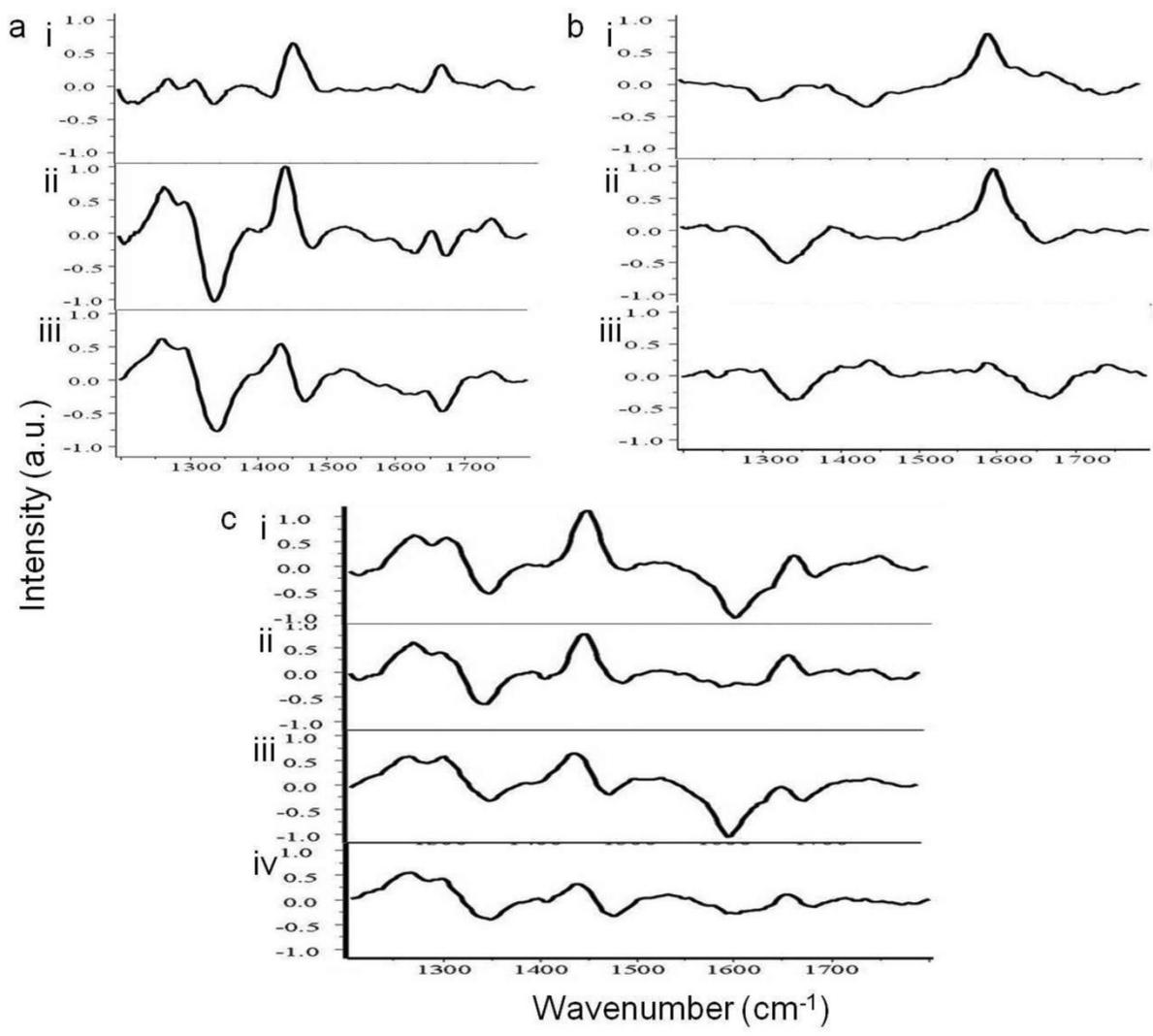


Figure 4.2 Discrimination of primary lung from metastatic breast lesions. Figure shows difference spectra; a) i)breast control – breast tumor, ii) breast control – breast metastasis, iii) breast tumor – breast metastasis, b) i) lung control – lung tumor, ii) lung control – breast metastasis, iii) lung tumor – breast metastasis, c) i) breast control – lung control, ii) breast control – lung tumor, iii) breast tumor – lung control, and iv) breast tumor - lung tumor. These suggest difference between protein, DNA and lipid content in different groups.

## Multivariate analysis

Preprocessed interpolated in 1200-1800  $\text{cm}^{-1}$  range spectra were subjected to PCA for delineating trends in the data set. PCA variance plot and loadings are shown in Figure 3a and b. As can be seen in Figure 4.3a, cumulative variance covered by factor 1, 3 and 4 are 82.3%, 92.7 and 94.6% respectively. Scatter plot of PCA factors (Figure 4.3c) shows distinct clusters of control breast, breast tumor, control lung and breast metastasis.

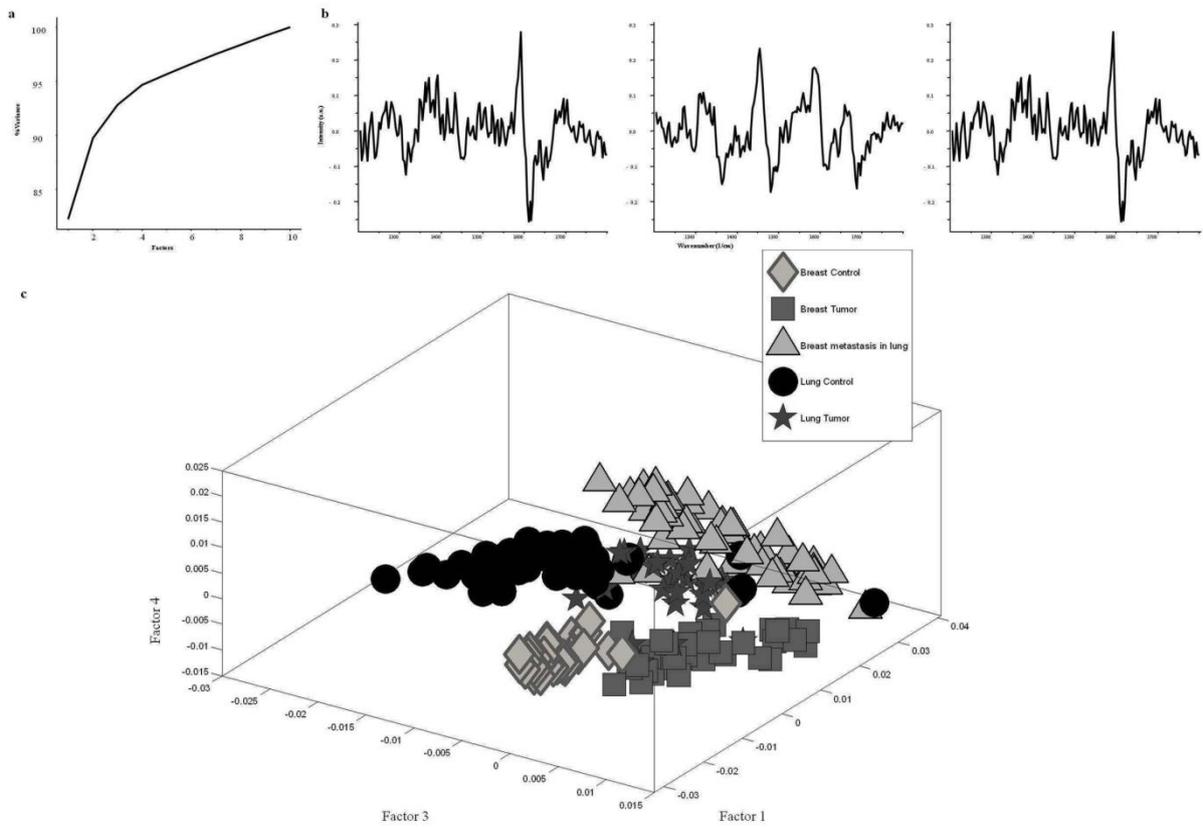


Figure 4.3 Discrimination of primary lung from metastatic breast lesions. Figure shows PCA a) variance plot, b) Loading factors 1, 3 and 4, and c) Scatter plot, suggesting classification between different groups.

The lung tumor cluster lies in the center and is close to control lung cluster and breast metastasis cluster. Ability of Raman spectroscopy to distinguish normal breast and breast cancer has been reported earlier (9, 63). Earlier studies in mouse model have shown that ex-vivo spectra of lung and breast differ considerably (106). Distinct spectral identity of breast from several anatomical sites *in vivo* has also been demonstrated (148). Thus, results of this study corroborate with earlier studies. Results also suggest that breast metastasis can be distinguished from normal breast and breast cancer using Raman spectroscopy. Thus, PCA suggests possibility of distinguishing breast cancer metastasis in lung from both breast and lung primary tumors.

To further explore the feasibility of classifying these different groups, PC-LDA was used. To avoid over fitting, 3 factors (72) contributing ~ 85 % percent of correct classification; were used (Figure 4.4a). The plot of PC-LDA factors 1, 2 and 3 (Figure 4.4b) shows clustering pattern similar to PCA.

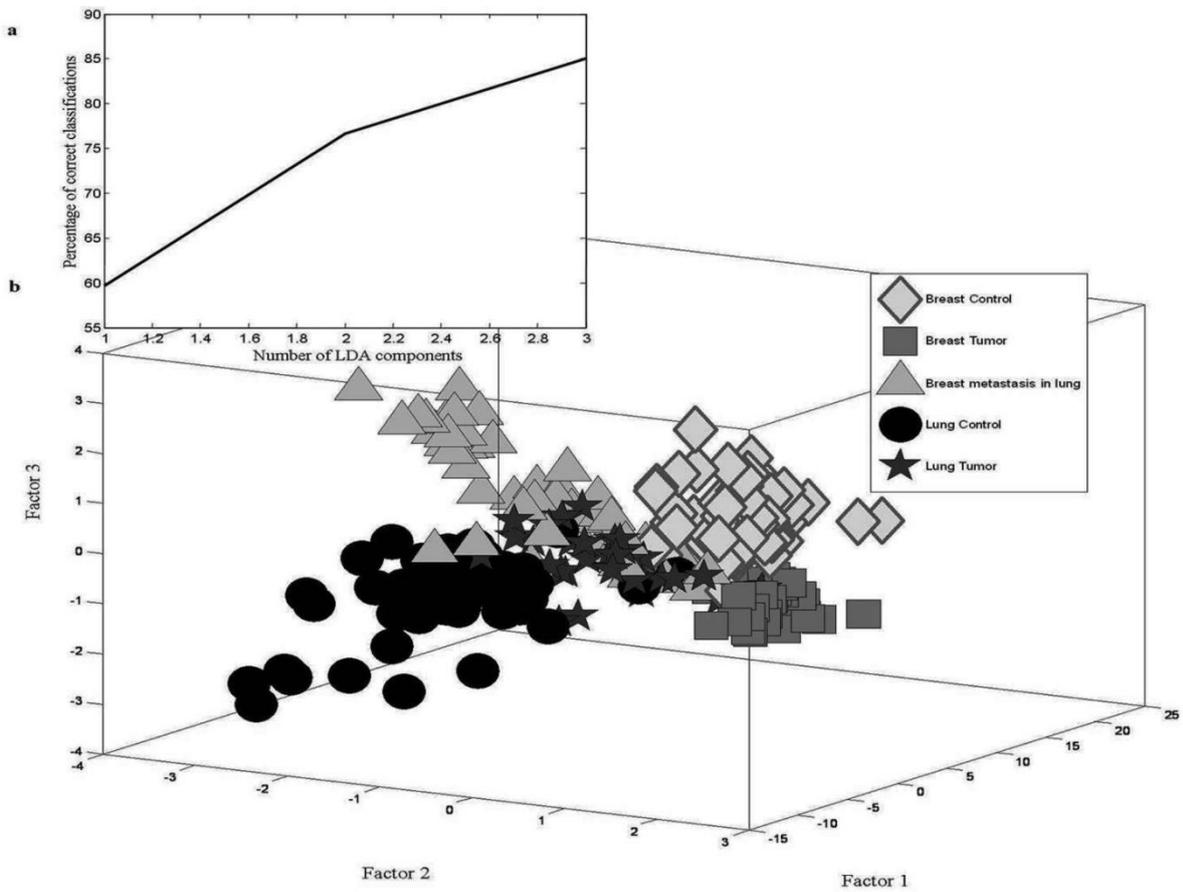


Figure 4.4 Discrimination of primary lung from metastatic breast lesions. Figure shows PC-LDA a) scree plot, b) scatter plot, again suggesting classification between different groups.

The confusion matrix for PC-LDA model building is shown in Table 4.1a. In this analysis, 59 out of 62 spectra were correctly classified as breast control, while 1/62 and 2/62 spectra misclassified with breast tumor and lung tumor respectively. 38/40 breast tumor spectra are correctly classified, while 2 misclassify as breast control. In case of breast metastasis in lungs, 45/63 spectra correctly classify, 3/63 misclassify with breast tumor and 15/63 spectra misclassify with lung tumor. 44/49 lung control spectra are correctly classified as lung control whereas 4/49 misclassified with lung tumor and 1/49 misclassified with breast tumor. 25/34 lung tumor spectra were correctly classified, while 2/34 misclassified with lung control, 3/34 misclassified with breast cancer metastasis and 4/34 misclassified with breast tumor.

a) MODEL (No. of animals, No. of spectra)	Breast Control	Breast Tumor	Breast cancer Metastasis in lungs	Lung Control	Lung Tumor
Breast Control (5,62)	59	1	0	0	2
Breast Tumor (4,40)	2	38	0	0	0
Breast Metastasis in lungs (4,63)	0	3	45	0	15
Lung Control (6,49)	0	1	0	44	4
Lung Tumor (4,34)	0	4	3	2	25

b) LOOCV (No. of animals, No. of spectra)	Breast Control	Breast Tumor	Breast cancer Metastasis in lungs	Lung Control	Lung Tumor
Breast Control (5,62)	59 (95%)	1	0	0	2
Breast Tumor (4,40)	2	38 (95%)	0	0	0
Breast Metastasis in lungs (4,63)	0	4	45 (71%)	0	14
Lung Control (6,49)	0	1	0	44 (90%)	4
Lung Tumor (4,34)	0	4	4	2	24

					(71%)
--	--	--	--	--	-------

c) Groups	Sensitivity (%)	Specificity (%)
Breast Control	95.2	97.9
Breast Tumor	95	90.8
Breast Metastasis in lungs	71.4	95.8
Lung Control	89.8	98
Lung Tumor	70.6	82.9

d) Test Prediction (No. of animals, No. of spectra)	Breast Control	Breast Tumor	Breast Metastasis in lungs	Lung Control	Lung Tumor
Breast Metastasis in lungs (4,88)	0 (0%)	11 (12.5%)	56 (63.6%)	0 (0%)	21 (23.8%)

**Table 4.1 Discrimination of primary lung from metastatic breast lesions. Table shows PC-LDA confusion matrix for a) model building, b) LOOCV, and c) LOOCV - sensitivity and specificity, and d) independent test prediction, suggesting feasibility of identifying metastatic lesion with ~ 64% efficiency.**

LOOCV was carried out to evaluate the results obtained by PC- LDA. In analysis of LOOCV as shown in Table 4.1b; 59 out of 62 spectra were correctly classified as breast control, while 1/62 and 2/62 spectra misclassified with breast tumor and lung tumor respectively. Correct classification of breast spectra suggests distinctness of breast spectra. Misclassification of breast control and breast tumor may be due to heterogeneity of breast tumor. As mentioned earlier and observed in difference spectra, there is an increase in lipids in lung tumor compared to control lung. Since, control breast predominantly consists of lipids, a misclassification between control breast and lung tumor is possible. 38/40 breast tumor spectra are correctly classified, while 2 misclassify as breast control. Heterogeneity of tumors may explain misclassification with normal breast. In case of breast metastasis in lungs, 45/63 spectra correctly classify, 4/63 misclassify with breast tumor and 14/63 spectra misclassify with lung tumor. Misclassification of breast metastasis in lungs with breast tumor may be due to signals from breast tumor cells lodged in lungs. High misclassification with lung tumor may be due to architectural similarity of primary and metastatic tumor. 44/49 lung control spectra are correctly classified as lung control whereas 4/49 misclassified with lung tumor and 1/49 misclassified with breast tumor. Misclassification between lung control and lung tumor may be attributed to heterogeneity of lung tumors. As explained earlier, misclassification between lung control and breast tumor may be due to their low lipid content. 24/34 lung tumor spectra were correctly classified, while 2/34 misclassified with lung control, 4/34 misclassified with breast cancer metastasis and 4/34 misclassified with breast tumor. Heterogeneity of lung tumor may explain misclassification with lung control. Architectural similarity amongst tumors may explain misclassification between lung tumor,

breast cancer metastasis and breast tumor. After LOOCV, breast metastasis in lung could be identified with ~71% sensitivity and ~96% specificity (Table 4.1c).

In order to ascertain the robustness of the model, test prediction using spectra from breast metastasis tissues from four independent animals was carried out. The results of test prediction are shown in Table 4.1c. 56 out of 88 spectra are correctly predicted as breast metastasis in lungs, while 21/88 and 11/88 was wrongly predicted as lung tumor and breast tumor respectively. As mentioned earlier, prediction as lung tumor may be due to architectural similarity between lung tumor and breast metastasis in lungs. Prediction as breast tumor may be due to presence of breast tumor cells that have metastasized into lungs.

### **Summary**

1. Metastatic breast lesions could be induced by injecting with syngenic tumor cells into blood stream of C3H /J mice. Spectra were acquired using fiber-optic probe based Raman spectroscopy from these metastatic lesions and studied
2. Metastatic breast lesion spectra could be distinguished from primary breast and lung tumor as well as breast and lung control spectra

Overall, the chapter showed feasibility of acquiring spectra from metastatic lesions using fiber-Raman and the possibility of distinguishing them from primary lung and breast tumors as well as controls

## **CHAPTER 5 THESIS SUMMARY**

Breast cancer is the most fatal cancer among women worldwide. Few of the most important problems that plague breast cancer management are lack of sensitive techniques for early detection of this cancer, which can substantially reduce cancer mortality and identification of metastatic breast cancer, that render most of the treatments ineffective. This thesis attempts to address these issues using a rapid, sensitive technique that is well suited for *in vivo* applications – Raman spectroscopy. The work is divided into three objectives. Point-wise summary of each objective is listed below.

1. To develop a rodent model of breast neoplasms suitable for characterization by Raman spectroscopy.

- i. The best model for acquiring transcutaneous *in vivo* breast spectra is SB mice. Albino rats and mice also give comparable results after shaving hair on their mammary glands. Transcutaneous *in vivo* Raman spectroscopy thus removes the necessity for animal sacrifice for acquiring spectra and allows follow up studies.
- ii. In SB mice, breast could be classified from other anatomical sites and frank breast tumor using transcutaneous *in vivo* Raman spectroscopy.
- iii. Using transcutaneous *in vivo* Raman spectroscopy in SB mice, physiological processes such as pregnancy, lactation and ageing can be investigated. Frank tumors can also be classified from normal breast. This further shows the feasibility of follow up studies.
- iv. SB mice is not, however, a model suitable for study of breast cancer progression. Rats treated with carcinogen using subcutaneous injection protocol are best suited for using *in vivo* breast fibroadenoma progression study. After further standardization of protocol, it may be possible to study breast adenocarcinoma progression using rats

treated by injecting carcinogen into nipples of their mammary glands. These protocols are vital for study of cancer progression.

2. To study development of breast neoplasms induced by chemical carcinogen using Raman spectroscopy in target organs and body fluids.
  - i. It is possible to distinguish pretumor spectra from normal breast spectra using transcutaneous *in vivo* Raman spectroscopy. Study of breast fibroadenoma progression showed that, on executing rat-wise analysis, tumors can be predicted with 82% sensitivity and 95% specificity. This suggests possibility of developing Raman spectroscopy as a non invasive screening tool.
  - ii. Urine and serum – based Raman spectroscopy may also be used to distinguish pretumor condition from normal. ‘Pre’ and ‘post’ breast adenocarcinoma resection surgery serum spectra can be also be distinguished. These studies highlight the potential applications of bio-fluid based Raman spectroscopy in breast cancer management. Serum-based Raman spectroscopy may also be used to distinguish pre and post breast adenocarcinoma resection surgery condition.
3. To study Raman spectral signatures of experimental lung metastasis from breast cancer cell line.
  - i. It is possible to distinguish breast metastatic lesions in lungs from normal breast and lung as well as primary lung and breast tumors. Combined with LDCT and bronchoscope aided spectroscopy, lung lesions may be classified into primary and metastatic tumors and help guide therapy.

Summarizing, the major part of the study involved exploring suitable models and carcinogenesis protocols for study of cancer progression. In a major investigation using 81 SD rats, possibility of distinguishing pre fibroadenoma spectra from control and predicting fibroadenoma appearance in test rats was shown. Preliminary studies showing possibility of distinguishing rats which would later develop tumor from controls was also shown using biofluid-based Raman spectroscopy. Feasibility of distinguishing metastatic lesions from primary tumors using fiber-optic Raman spectroscopy was demonstrated for the first time. Thus, the thesis provides several leads towards Raman spectroscopy applications in breast cancer management.

Several technological leaps in Raman spectroscopy have been achieved over the last few years that may prove vital to its applications in breast cancer management. Matousek et.al and Stone et.al. (58, 113-116) have demonstrated deep Raman spectroscopy that can help acquire spectra from any region of the human female breast, opening up avenues for thorough breast cancer screening. Schmäzlin et.al.(117) have reported Raman imaging with a fiber-coupled multichannel spectrograph, that allows capture of entire Raman image with one single exposure and chemical mapping without the need for scanning procedure. Such techniques would greatly reduce spectra acquisition times, and hence render Raman spectroscopy a more efficacious mass screening tool. In future, efforts need to be directed towards amalgamation of deep Raman spectroscopy and single exposure imaging with the experimental design and data analysis protocols discussed in this thesis. With respect to metastasis, studies ought to be initiated in human subjects for testing the veracity of using fiber-optic based Raman spectroscopy to distinguish primary and metastatic lesions.

## **REFERENCES**

1. "Globocan 2012," [http://globocan.iarc.fr/Pages/fact\\_sheets\\_cancer.aspx](http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx) (February, 2015).
2. P. P. Rosen et al., *Rosen's breast pathology*, Fourth edition. ed.
3. C. B. Matsen, and L. A. Neumayer, "Breast cancer: a review for the general surgeon," *JAMA surgery* **148**(10), 971-979 (2013).
4. A. B. Nover et al., "Modern breast cancer detection: a technological review," *International journal of biomedical imaging* **2009**(902326 (2009).
5. S. M. Ismail et al., "Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia," *BMJ* **298**(6675), 707-710 (1989).
6. N. A. Howlader N, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds), "SEER Cancer Statistics Review, 1975-2011," *National Cancer Institute. Bethesda, MD* (2013).
7. J. A. Shaw et al., "Genomic analysis of circulating cell-free DNA infers breast cancer dormancy," *Genome research* **22**(2), 220-231 (2012).
8. D. Evers et al., "Optical spectroscopy: current advances and future applications in cancer diagnostics and therapy," *Future oncology* **8**(3), 307-320 (2012).
9. M. Chowdary et al., "Biochemical correlation of Raman spectra of normal, benign and malignant breast tissues: a spectral deconvolution study," *Biopolymers* **91**(7), 539-546 (2009).
10. A. S. Haka et al., "Diagnosing breast cancer by using Raman spectroscopy," *Proceedings of the National Academy of Sciences of the United States of America* **102**(35), 12371-12376 (2005).

11. J. Pichardo-Molina et al., "Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients," *Lasers in medical science* **22**(4), 229-236 (2007).
12. I. Barman et al., "Application of Raman spectroscopy to identify microcalcifications and underlying breast lesions at stereotactic core needle biopsy," *Cancer research* **73**(11), 3206-3215 (2013).
13. X. Bi et al., "Evaluating HER2 amplification status and acquired drug resistance in breast cancer cells using Raman spectroscopy," *Journal of biomedical optics* **19**(2), 025001 (2014).
14. J. Horsnell et al., "Raman spectroscopy--a new method for the intra-operative assessment of axillary lymph nodes," *The Analyst* **135**(12), 3042-3047 (2010).
15. B. H. Jun et al., "Multifunctional silver-embedded magnetic nanoparticles as SERS nanoprobe and their applications," *Small* **6**(1), 119-125 (2010).
16. M. D. Keller et al., "Development of a spatially offset Raman spectroscopy probe for breast tumor surgical margin evaluation," *Journal of biomedical optics* **16**(7), 077006-077006-077008 (2011).
17. A. Saha et al., "Raman spectroscopy: a real-time tool for identifying microcalcifications during stereotactic breast core needle biopsies," *Biomedical optics express* **2**(10), 2792-2803 (2011).
18. L. Xiao et al., "Imaging of epidermal growth factor receptor on single breast cancer cells using surface-enhanced Raman spectroscopy," *Analytica chimica acta* **843**(73-82) (2014).
19. C. Zheng et al., "The use of Au@SiO<sub>2</sub> shell-isolated nanoparticle-enhanced Raman spectroscopy for human breast cancer detection," *Analytical and bioanalytical chemistry* **406**(22), 5425-5432 (2014).

20. M. C. Pike et al., "'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer," (1983).
21. D. Ramsay et al., "Anatomy of the lactating human breast redefined with ultrasound imaging," *Journal of Anatomy* **206**(6), 525-534 (2005).
22. M. M. Richert et al., "An atlas of mouse mammary gland development," *Journal of mammary gland biology and neoplasia* **5**(2), 227-241 (2000).
23. M. Murray, "Nonneoplastic Alterations of the Mammary Epithelium Can Mimic Atypia," *Archives of Pathology & Laboratory Medicine* **133**(5), 722-728 (2009).
24. A. Bombonati, and D. C. Sgroi, "The molecular pathology of breast cancer progression," *The Journal of pathology* **223**(2), 307-317 (2011).
25. L. S. DeBruin, and P. D. Josephy, "Perspectives on the chemical etiology of breast cancer," *Environmental health perspectives* **110 Suppl 1**(119-128 (2002).
26. K. McPherson, C. Steel, and J. Dixon, "Breast cancer—epidemiology, risk factors, and genetics," *BMJ* **321**(7261), 624-628 (2000).
27. A. Qaseem et al., "Screening mammography for women 40 to 49 years of age: a clinical practice guideline from the American College of Physicians," *Annals of internal medicine* **146**(7), 511-515 (2007).
28. U. P. S. T. Force, "Screening for breast cancer: US Preventive Services Task Force recommendation statement," *Annals of internal medicine* **151**(10), 716 (2009).
29. L. J. Graham et al., "Current approaches and challenges in monitoring treatment responses in breast cancer," *Journal of Cancer* **5**(1), 58-68 (2014).
30. P. R. Srinivas et al., "Proteomics in early detection of cancer," *Clinical Chemistry* **47**(10), 1901-1911 (2001).

31. S. Duraipandian et al., "Real-time Raman spectroscopy for in vivo, online gastric cancer diagnosis during clinical endoscopic examination," *Journal of biomedical optics* **17**(8), 081418-081418 (2012).
32. S. K. Teh et al., "Near-infrared Raman spectroscopy for gastric precancer diagnosis," *Journal of Raman Spectroscopy* **40**(8), 908-914 (2009).
33. T. Denkçeken et al., "Elastic light single-scattering spectroscopy for the detection of cervical precancerous ex vivo," *Biomedical Engineering, IEEE Transactions on* **60**(1), 123-127 (2013).
34. S. Duraipandian et al., "Near-infrared-excited confocal Raman spectroscopy advances in vivo diagnosis of cervical precancer," *Journal of biomedical optics* **18**(6), 067007-067007 (2013).
35. M. M. Stephen 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**(1), 278 (2013).
36. G. A. Wagnieres, W. M. Star, and B. C. Wilson, "In vivo fluorescence spectroscopy and imaging for oncological applications," *Photochemistry and photobiology* **68**(5), 603-632 (1998).
37. I. J. Bigio, "Elastic light scattering spectroscopy for the detection of early cancer and pre-cancer," *LEOS Annual Meeting Conference Proceedings, 2009. LEOS'09. IEEE* 268-269 (2009).
38. H. Abramczyk, and B. Brozek-Pluska, "Raman imaging in biochemical and biomedical applications. Diagnosis and treatment of breast cancer," *Chemical reviews* **113**(8), 5766-5781 (2013).

39. S. B. Cartaxo et al., "FT-Raman spectroscopy for the differentiation between cutaneous melanoma and pigmented nevus," *Acta Cirurgica Brasileira* **25**(4), 351-356 (2010).
40. L. P. Choo-Smith et al., "Medical applications of Raman spectroscopy: from proof of principle to clinical implementation," *Biopolymers* **67**(1), 1-9 (2002).
41. H. Ding et al., "Development of Raman spectral markers to assess metastatic bone in breast cancer," *Journal of biomedical optics* **19**(11), 111606 (2014).
42. D. I. Ellis et al., "Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool," *The Analyst* **138**(14), 3871-3884 (2013).
43. P. Larkin, *Infrared and raman spectroscopy : principles and spectral interpretation*, Elsevier, Amsterdam ; Boston (2011).
44. F. S. Parker, *Applications of infrared, raman, and resonance raman spectroscopy in biochemistry*, Plenum Press, New York (1983).
45. M. D. Morris, and P. Matousek, *Emerging Raman applications and techniques in biomedical and pharmaceutical fields*, Springer, Heidelberg ; New York (2010).
46. N. B. Colthup, L. H. Daly, and S. E. Wiberley, *Introduction to infrared and Raman spectroscopy*, 3rd ed., Academic Press, Boston (1990).
47. J. Cazes, Ed., *Analytical instrumentation handbook*, 3 ed., CRC press, Boca Raton, FL (2004).
48. I. R. Lewis, and H. G. M. Edwards, *Handbook of Raman spectroscopy : from the research laboratory to the process line*, Marcel Dekker, New York (2001).
49. E. V. Efremov, F. Ariese, and C. Gooijer, "Achievements in resonance Raman spectroscopy: Review of a technique with a distinct analytical chemistry potential," *Analytica chimica acta* **606**(2), 119-134 (2008).

50. B. B. Johnson, and W. L. Peticolas, "The resonant Raman effect," *Annual Review of Physical Chemistry* **27**(1), 465-521 (1976).
51. H. Zhou et al., "SERS detection of bacteria in water by in situ coating with Ag nanoparticles," *Analytical chemistry* **86**(3), 1525-1533 (2014).
52. L. F. Sallum et al., "Determination of acetylsalicylic acid in commercial tablets by SERS using silver nanoparticle-coated filter paper," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **133**(107-111) (2014).
53. S. Schlücker, "Surface-Enhanced Raman spectroscopy: Concepts and chemical applications," *Angewandte Chemie International Edition* **53**(19), 4756-4795 (2014).
54. P. L. Stiles et al., "Surface-enhanced Raman spectroscopy," *Annu. Rev. Anal. Chem.* **1**(601-626) (2008).
55. C. L. Evans, and X. S. Xie, "Coherent anti-Stokes Raman scattering microscopy: chemical imaging for biology and medicine," *Annu. Rev. Anal. Chem.* **1**(883-909) (2008).
56. L. G. Rodriguez, S. J. Lockett, and G. R. Holtom, "Coherent anti-stokes Raman scattering microscopy: A biological review," *Cytometry Part A* **69**(8), 779-791 (2006).
57. W. M. Tolles et al., "A review of the theory and application of coherent anti-Stokes Raman spectroscopy (CARS)," *Applied Spectroscopy* **31**(4), 253-271 (1977).
58. K. Buckley, and P. Matousek, "Non-invasive analysis of turbid samples using deep Raman spectroscopy," *The Analyst* **136**(15), 3039-3050 (2011).
59. U. Utzinger et al., "Near-infrared Raman spectroscopy for in vivo detection of cervical precancers," *Applied spectroscopy* **55**(8), 955-959 (2001).

60. I. Latka et al., "Fiber optic probes for linear and nonlinear Raman applications—Current trends and future development," *Laser & Photonics Reviews* **7**(5), 698-731 (2013).
61. S. Stewart et al., "Raman imaging," *Annual Review of Analytical Chemistry* **5**(337-360) (2012).
62. H. Abramczyk, and B. Brozek-Pluska, "Raman imaging in biochemical and biomedical applications. Diagnosis and treatment of breast cancer," *Chemical reviews* **113**(8), 5766-5781 (2013).
63. M. V. Chowdary et al., "Discrimination of normal, benign, and malignant breast tissues by Raman spectroscopy," *Biopolymers* **83**(5), 556-569 (2006).
64. M. Sattlecker et al., "Investigation of support vector machines and Raman spectroscopy for lymph node diagnostics," *The Analyst* **135**(5), 895-901 (2010).
65. H. Abdi, and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics* **2**(4), 433-459 (2010).
66. A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR)* **31**(3), 264-323 (1999).
67. C. Chatfield, and A. J. Collins, *Introduction to multivariate analysis*, Springer (2013).
68. D. Pena, "Multivariate data analysis," *McGraw Hill-Interamericana de Espana SAU* (2002).
69. M. J. Schervish, "A review of multivariate analysis," *Statistical Science* 396-413 (1987).
70. H. Seal, "Multivariate statistical analysis for biologists," *Methuen* (1964).
71. B. G. Tabachnick, and L. S. Fidell, "Using multivariate statistics," (2001).

72. J. G. Kelly et al., "Biospectroscopy to metabolically profile biomolecular structure: a multistage approach linking computational analysis with biomarkers," *Journal of proteome research* **10**(4), 1437-1448 (2011).
73. S. P. Singh et al., "In vivo Raman spectroscopy of oral buccal mucosa: a study on malignancy associated changes (MAC)/cancer field effects (CFE)," *The Analyst* **138**(14), 4175-4182 (2013).
74. E. M. Kanter et al., "Application of Raman spectroscopy for cervical dysplasia diagnosis," *Journal of biophotonics* **2**(1-2), 81-90 (2009).
75. S. K. Teh et al., "Diagnostic potential of near-infrared Raman spectroscopy in the stomach: differentiating dysplasia from normal tissue," *British journal of cancer* **98**(2), 457-465 (2008).
76. R. Alfano et al., "Human breast tissues studied by IR Fourier transform Raman spectroscopy," *Lasers in the life sciences* **4**(1), 23-28 (1991).
77. J. Smith et al., "Raman spectral mapping in the assessment of axillary lymph nodes in breast cancer," *Technology in cancer research & treatment* **2**(4), 327-332 (2003).
78. J. Kneipp et al., "Characterization of breast duct epithelia: a Raman spectroscopic study," *Vibrational Spectroscopy* **32**(1), 67-74 (2003).
79. R. E. Kast et al., "Raman spectroscopy can differentiate malignant tumors from normal breast tissue and detect early neoplastic changes in a mouse model," *Biopolymers* **89**(3), 235-241 (2008).
80. Q. Li, Q. Gao, and G. Zhang, "Classification for breast cancer diagnosis with Raman spectroscopy," *Biomedical optics express* **5**(7), 2435-2445 (2014).
81. C. H. Liu et al., "Resonance Raman and Raman spectroscopy for breast cancer detection," *Technology in cancer research & treatment* **12**(4), 371-382 (2013).

82. H. Abramczyk et al., "Raman 'optical biopsy' of human breast cancer," *Progress in biophysics and molecular biology* **108**(1-2), 74-81 (2012).
83. Y. Yang et al., "Label-free imaging of human breast tissues using coherent anti-Stokes Raman scattering microscopy," *SPIE BiOS 79032G-79032G-79036* (2011).
84. Y. Yang et al., "Differential diagnosis of breast cancer using quantitative, label-free and molecular vibrational imaging," *Biomedical optics express* **2**(8), 2160-2174 (2011).
85. H. Abramczyk et al., "The hallmarks of breast cancer by Raman spectroscopy," *Journal of Molecular Structure* **924-926**(175-182 (2009).
86. Z. A. Nima et al., "Single-walled carbon nanotubes as specific targeting and Raman spectroscopic agents for detection and discrimination of single human breast cancer cells," *Journal of biomedical optics* **18**(5), 55003 (2013).
87. N. P. Damayanti et al., "Differentiation of cancer cells in two-dimensional and three-dimensional breast cancer models by Raman spectroscopy," *Journal of biomedical optics* **18**(11), 117008 (2013).
88. N. C. Dingari et al., "Development and comparative assessment of Raman spectroscopic classification algorithms for lesion discrimination in stereotactic breast biopsies with microcalcifications," *Journal of biophotonics* **6**(4), 371-381 (2013).
89. L. Liang et al., "Exploring type II microcalcifications in benign and premalignant breast lesions by shell-isolated nanoparticle-enhanced Raman spectroscopy (SHINERS)," *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy* **132**(397-402 (2014).
90. A. Saha et al., "Precision of Raman spectroscopy measurements in detection of microcalcifications in breast needle biopsies," *Analytical chemistry* **84**(15), 6715-6722 (2012).

91. S. Lee et al., "Rapid and sensitive phenotypic marker detection on breast cancer cells using surface-enhanced Raman scattering (SERS) imaging," *Biosensors & bioelectronics* **51**(238-243 (2014).
92. C. M. Krishna et al., "Combined Fourier transform infrared and Raman spectroscopic approach for identification of multidrug resistance phenotype in cancer cell lines," *Biopolymers* **82**(5), 462-470 (2006).
93. A. S. Haka et al., "In vivo margin assessment during partial mastectomy breast surgery using raman spectroscopy," *Cancer research* **66**(6), 3317-3322 (2006).
94. P. N. Goel et al., "Investigating the effects of Pentoxifylline on human breast cancer cells using Raman spectroscopy," *Journal of Innovative Optical Health Sciences* (2014).
95. H. Abramczyk et al., "Oncologic photodynamic diagnosis and therapy: confocal Raman/fluorescence imaging of metal phthalocyanines in human breast cancer tissue in vitro," *The Analyst* **139**(21), 5547-5559 (2014).
96. P. D. Chowdary et al., "Molecular histopathology by spectrally reconstructed nonlinear interferometric vibrational imaging," *Cancer research* **70**(23), 9562-9569 (2010).
97. M. Marro et al., "Molecular monitoring of epithelial-to-mesenchymal transition in breast cancer cells by means of Raman spectroscopy," *Biochimica et biophysica acta* **1843**(9), 1785-1795 (2014).
98. L. M. Fullwood et al., "Investigating the use of Raman and immersion Raman spectroscopy for spectral histopathology of metastatic brain cancer and primary sites of origin," *Analytical Methods* **6**(12), 3948-3961 (2014).
99. W. G. Cahan, and E. Castro, "Significance of a solitary lung shadow in patients with breast cancer," *Annals of surgery* **181**(2), 137 (1975).

100. P. B. Bach et al., "Benefits and harms of CT screening for lung cancer: a systematic review," *Jama* **307**(22), 2418-2429 (2012).
101. I. H. Russo, and J. Russo, "Mammary gland neoplasia in long-term rodent studies," *Environmental health perspectives* **104**(9), 938 (1996).
102. D. Medina, "Of mice and women: A short history of mouse mammary cancer research with an emphasis on the paradigms inspired by the transplantation method," *Cold Spring Harbor perspectives in biology* **2**(10), a004523 (2010).
103. J. Kulkarm, V. Lalitha, and S. Bhide, "Weak carcinogenic effect of masheri, a pyrolysed tobacco product, in mouse skin tumor models," *Journal of cancer research and clinical oncology* **115**(2), 166-169 (1989).
104. K. Karande et al., "Intracytoplasmic type A particles from mammary tumours and leukaemias of strain ICRC mice," *British journal of cancer* **39**(2), 132 (1979).
105. C. A. Squier, P. Cox, and P. W. Wertz, "Lipid content and water permeability of skin and oral mucosa," *Journal of investigative dermatology* **96**(1), 123-126 (1991).
106. N. Huang et al., "Full range characterization of the Raman spectra of organs in a murine model," *Optics express* **19**(23), 22892-22909 (2011).
107. L. Raniero et al., "In and ex vivo breast disease study by Raman spectroscopy," *Theoretical Chemistry Accounts* **130**(4-6), 1239-1247 (2011).
108. N. F. Boyd et al., "Mammographic densities and breast cancer risk," *Cancer Epidemiology Biomarkers & Prevention* **7**(12), 1133-1144 (1998).
109. I. Benakanakere et al., "Natural and synthetic progestins accelerate 7, 12-dimethylbenz [a] anthracene-initiated mammary tumors and increase angiogenesis in Sprague-Dawley rats," *Clinical cancer research* **12**(13), 4062-4071 (2006).
110. J. Russo, and I. H. Russo, "Experimentally induced mammary tumors in rats," *Breast cancer research and treatment* **39**(1), 7-20 (1996).

111. R. Malini et al., "Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: a Raman spectroscopy study," *Biopolymers* **81**(3), 179-193 (2006).
112. L. Yang et al., "Separation of normal and premalignant cervical epithelial cells using confocal light absorption and scattering spectroscopic microscopy ex vivo," *BioMed Research International* **2011**((2011).
113. R. Baker et al., "Depth profiling of calcifications in breast tissue using picosecond Kerr-gated Raman spectroscopy," *The Analyst* **132**(1), 48-53 (2007).
114. P. Matousek, and N. Stone, "Prospects for the diagnosis of breast cancer by noninvasive probing of calcifications using transmission Raman spectroscopy," *Journal of biomedical optics* **12**(2), 024008 (2007).
115. N. Stone et al., "Subsurface probing of calcifications with spatially offset Raman spectroscopy (SORS): future possibilities for the diagnosis of breast cancer," *The Analyst* **132**(9), 899-905 (2007).
116. N. Stone, and P. Matousek, "Advanced transmission Raman spectroscopy: a promising tool for breast disease diagnosis," *Cancer research* **68**(11), 4424-4430 (2008).
117. E. Schmäzlin et al., "Raman Imaging with a Fiber-Coupled Multichannel Spectrograph," *Sensors* **14**(11), 21968-21980 (2014).
118. J. A. M. Bispo et al., "Correlating the amount of urea, creatinine, and glucose in urine from patients with diabetes mellitus and hypertension with the risk of developing renal lesions by means of Raman spectroscopy and principal component analysis," *Journal of biomedical optics* **18**(8), 087004-087004 (2013).
119. A. Sahu et al., "Serum based diagnosis of asthma using Raman spectroscopy: an early phase pilot study," *Plos one* **8**(11), e78921 (2013).

120. J. Desborough, "The stress response to trauma and surgery," *British journal of anaesthesia* **85**(1), 109-117 (2000).
121. B. Chernow et al., "Hormonal responses to graded surgical stress," *Archives of internal medicine* **147**(7), 1273-1278 (1987).
122. H.-S. Lai et al., "Plasma free amino acid profile in cancer patients," *Seminars in cancer biology* 267-276 (2005).
123. Y. Miyagi et al., "Plasma free amino acid profiling of five types of cancer patients and its application for early detection," *PLoS One* **6**(9), e24143 (2011).
124. B. K. Kim et al., "The multiplex bead array approach to identifying serum biomarkers associated with breast cancer," *Breast Cancer Research* **11**(2), R22 (2009).
125. J. Boyle, "Lehninger principles of biochemistry: Nelson, D., and Cox, M," Wiley Online Library (2005).
126. L. G. Perez-Rivas et al., "Serum protein levels following surgery in breast cancer patients: a protein microarray approach," *International journal of oncology* **41**(6), 2200-2206 (2012).
127. S. Raab et al., "Adenocarcinoma in the lung in patients with breast cancer. A prospective analysis of the discriminatory value of immunohistology," *American journal of clinical pathology* **100**(1), 27-35 (1993).
128. S. Raab, C. Sturgis, and M. Wick, "Metastatic tumors in the lung: a practical approach to diagnosis," *Practical Pulmonary Pathology. Churchill Livingstone, Philadelphia* (2005).
129. J. J. Casey et al., "The solitary pulmonary nodule in the patient with breast cancer," *Surgery* **96**(4), 801-805 (1984).
130. D. J. Dabbs et al., "Detection of estrogen receptor by immunohistochemistry in pulmonary adenocarcinoma," *The Annals of thoracic surgery* **73**(2), 403-406 (2002).

131. E. Y. Chang et al., "The evaluation and treatment implications of isolated pulmonary nodules in patients with a recent history of breast cancer," *The American journal of surgery* **191**(5), 641-645 (2006).
132. P. L. Wagner et al., "Immunohistochemical expression of estrogen and progesterone receptors in primary pulmonary neuroendocrine tumors," *Archives of pathology & laboratory medicine* **132**(12), 1889 (2008).
133. E. Sasaki et al., "Breast-specific expression of MGB1/mammaglobin: an examination of 480 tumors from various organs and clinicopathological analysis of MGB1-positive breast cancers," *Modern pathology* **20**(2), 208-214 (2007).
134. Y. Takeda et al., "Analysis of expression patterns of breast cancer-specific markers (mammaglobin and gross cystic disease fluid protein 15) in lung and pleural tumors," *Archives of pathology & laboratory medicine* **132**(2), 239 (2008).
135. T. Okasaka et al., "Stepwise examination for differential diagnosis of primary lung cancer and breast cancer relapse presenting as a solitary pulmonary nodule in patients after mastectomy," *Journal of surgical oncology* **98**(7), 510-514 (2008).
136. J. Herbst et al., "Evidence-based criteria to help distinguish metastatic breast cancer from primary lung adenocarcinoma on thoracic frozen section," *American journal of clinical pathology* **131**(1), 122-128 (2009).
137. R. T. Vollmer, "Primary Lung Cancer vs Metastatic Breast Cancer A Probabilistic Approach," *American journal of clinical pathology* **132**(3), 391-395 (2009).
138. P. W. Payne et al., "Sputum Screening by Quantitative Microscopy A Reexamination of a Portion of the National Cancer Institute Cooperative Early Lung Cancer Study," *Mayo Clinic Proceedings* 697-704 (1997).
139. M. R. Melamed, "Lung cancer screening results in the National Cancer Institute New York study," *Cancer* **89**(S11), 2356-2362 (2000).

140. A. M. Davis, and A. S. Cifu, "Lung cancer screening," *JAMA* **312**(12), 1248-1249 (2014).
141. J. B. Seo et al., "Atypical Pulmonary Metastases: Spectrum of Radiologic Findings 1," *Radiographics* **21**(2), 403-417 (2001).
142. P. K. woodward, F. Dehdashti, and C. E. Putman, "Radiologic diagnosis of extrathoracic metastases to the lung," *Oncology-Huntington* **12**(3), 431-440 (1998).
143. N. Pavlidis et al., "Diagnostic and therapeutic management of cancer of an unknown primary," *European Journal of Cancer* **39**(14), 1990-2005 (2003).
144. A. F. d. Oliveira et al., "Differential diagnosis in primary and metastatic cutaneous melanoma by FT-Raman spectroscopy," *Acta Cirurgica Brasileira* **25**(5), 434-439 (2010).
145. A. C. Terentis et al., "Confocal Raman microspectroscopy discriminates live human metastatic melanoma and skin fibroblast cells," *Journal of Raman Spectroscopy* **44**(9), 1205-1216 (2013).
146. C. Nieva et al., "The lipid phenotype of breast cancer cells characterized by Raman microspectroscopy: towards a stratification of malignancy," *PLoS One* **7**(10), e46456 (2012).
147. M. A. Short et al., "Development and preliminary results of an endoscopic Raman probe for potential in vivo diagnosis of lung cancers," *Optics letters* **33**(7), 711-713 (2008).
148. T. Bhattacharjee et al., "Swiss bare mice: a suitable model for transcutaneous in vivo Raman spectroscopic studies of breast cancer," *Lasers Med Sci* **29**(1), 325-333 (2014).

149. S. S. Hecht et al., "Inhibition of lung tumorigenesis in A/J mice by N-acetyl-S-(N-2-phenethylthiocarbamoyl)-L-cysteine and myo-inositol, individually and in combination," *Carcinogenesis* **23**(9), 1455-1461 (2002).

## **Published paper reprints**

# Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

## **Transcutaneous *in vivo* Raman spectroscopic studies in a mouse model: evaluation of changes in the breast associated with pregnancy and lactation**

Tanmoy Bhattacharjee  
Girish Maru  
Arvind Ingle  
C. Murali Krishna



SPIE

# Transcutaneous *in vivo* Raman spectroscopic studies in a mouse model: evaluation of changes in the breast associated with pregnancy and lactation

Tanmoy Bhattacharjee,<sup>a</sup> Girish Maru,<sup>b</sup> Arvind Ingle,<sup>c</sup> and C. Murali Krishna<sup>a</sup>

<sup>a</sup>Advanced Center for Treatment Research and Education in Cancer, Chilkapati Laboratory, Kharghar, Navi Mumbai 410210, India

<sup>b</sup>Advanced Center for Treatment Research and Education in Cancer, Maru Laboratory, Kharghar, Navi Mumbai 410210, India

<sup>c</sup>Advanced Center for Treatment Research and Education in Cancer, Laboratory Animal Facility, Kharghar, Navi Mumbai 410210, India

**Abstract.** Raman spectroscopy (RS) has been extensively explored as an alternative diagnostic tool for breast cancer. This can be attributed to its sensitivity to malignancy-associated biochemical changes. However, biochemical changes due to nonmalignant conditions like benign lesions, inflammatory diseases, aging, menstrual cycle, pregnancy, and lactation may act as confounding factors in diagnosis of breast cancer. Therefore, in this study, the efficacy of RS to classify pregnancy and lactation-associated changes as well as its effect on breast tumor diagnosis was evaluated. Since such studies are difficult in human subjects, a mouse model was used. Spectra were recorded transcutaneously from the breast region of six Swiss bare mice postmating, during pregnancy, and during lactation. Data were analyzed using multivariate statistical tool Principal Component–Linear Discriminant Analysis. Results suggest that RS can differentiate breasts of pregnant/lactating mice from those of normal mice, the classification efficiencies being 100%, 60%, and 88% for normal, pregnant, and lactating mice, respectively. Frank breast tumors could be classified with 97.5% efficiency, suggesting that these physiological changes do not affect the ability of RS to detect breast tumors. © 2013 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.18.4.047004]

Keywords: pregnancy; lactation; transcutaneous; *in vivo*; Raman spectroscopy.

Paper 12712R received Nov. 3, 2012; revised manuscript received Feb. 5, 2013; accepted for publication Mar. 19, 2013; published online Apr. 15, 2013.

## 1 Introduction

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death in females worldwide, accounting for 23% (1.38 million) of the total new cancer cases and 14% (458,400) of the total cancer deaths in 2008.<sup>1</sup> Literature suggests improved prognosis with early detection of breast cancer.<sup>2</sup> However, currently available screening/diagnostic tools suffer from several disadvantages like tedious sample preparation, long output times, and interobserver variance.<sup>3,4</sup> Rapid, objective, and preferably noninvasive alternate screening/diagnostic techniques are hence being extensively explored. Raman spectroscopy (RS) is one such tool, which has shown promising results in the diagnosis of several cancers,<sup>5</sup> including breast cancers.<sup>6,7</sup> RS is an inelastic scattering process where the energy of photon scattered by the sample is different from the incident photon due to transfer of energy to or from vibrational modes of molecules in the sample.<sup>8</sup> Since bands of Raman spectrum are characteristic of specific molecular vibrations unique to a molecule, RS can provide the chemical fingerprint of a sample. However, biochemical changes also occur due to physiological processes like aging, menstrual cycle, pregnancy, and lactation.

Continuous changes take place in the breast as age progresses and the mice undergo different reproductive phases. During pregnancy and lactation, massive tissue remodeling occurs in the breast,<sup>9</sup> emphasizing the need to study these

processes spectroscopically. At birth, the mammary glands consist of stroma-connective tissue, fibroblasts, the mammary fat pad, and epithelial cords which is a small, branched ductal network of mammary epithelium that invades from nipple into fat pad.<sup>10</sup> During puberty, the epithelium forms terminal end buds, which invade the fat pad resulting in branched ducts throughout the breasts. The final developmental stages of the mammary glands occur during pregnancy and lactation. Massive proliferation of ductal cells and alveolar buds takes place. The epithelial to adipocyte ratio increases and capillaries are found within the connective tissue surrounding each individual alveolus. During the second half of pregnancy, the alveolar buds progressively cleave and differentiate into individual alveoli that will ultimately transform into milk-secreting lobules during lactation. The mammary glands display many of the properties associated with tumor progression during pregnancy and lactation. For example, rapid proliferation of epithelial cells takes place during these phases. The lactating mammary gland also actively resists apoptotic signals.<sup>11</sup> In addition, as the mammary gland undergoes these morphological changes, its blood supply gets adjusted, and thus, like tumors, the mammary gland induces angiogenic remodeling.<sup>12</sup> Thus, changes in the breast during pregnancy and lactation have steps similar to carcinogenesis. Therefore, the potential of RS to detect malignant changes in light of these confounding factors needs to be evaluated. The ability of RS to differentiate benign lesions from normal and malignant conditions has been demonstrated.<sup>6,7</sup> The effect of aging on breast cancer diagnosis has also been explored.<sup>6</sup>

Address all correspondence to: C. Murali Krishna, Advanced Center for Treatment Research and Education in Cancer, Tata Memorial Center, Chilkapati Laboratory, Sector 22, Kharghar, Navi Mumbai 410210, India. Tel: +91-22-27405039; Fax: 91-22-2740 5085; E-mail: pittu1043@gmail.com or mchilakapati@actrec.gov.in

However, sensitivity of RS to other physiological changes (menstrual cycle, pregnancy, lactation) is yet to be ascertained.

The current study aims to evaluate the sensitivity of transcutaneous *in vivo* RS to changes in breasts of nonpregnant, pregnant, and lactating mice and its effect on breast tumor detection. In the study, spectra were recorded transcutaneously from breast of nonpregnant (control), pregnant, and lactating mice. To test the ability of this technique to differentiate between malignancy and pregnancy- or lactation-associated changes, transcutaneous spectra were also recorded from subcutaneously transplanted frank breast tumors. Data were analyzed using multivariate statistical tools. The findings of the study are discussed in the paper.

## 2 Materials and Methods

### 2.1 Animals

Female Swiss bare mice,<sup>13</sup> a hairless version of Swiss albino mice, were allowed to mate, and successful mating was identified by appearance of vaginal plugs. Successful pregnancy was determined by observation of visible bulge in the abdominal region of mice (approximately 2 weeks postmating). Delivery of pups marked the beginning of the lactation phase (approximately 3 weeks postmating). These different stages were established by a veterinarian. To minimize variability in data, the same set of mice were used to record spectra immediately postmating, during pregnancy (2 weeks postmating), and lactation (1 to 2 days postdelivery). Between 8 and 11 spectra per mouse were recorded transcutaneously from left and right inguinal breast of mice, resulting in 56 to 60 spectra per group. Each spectrum was recorded approximately 1 mm apart by using a precision stage. Only spectra from mice ( $n = 6$ ) who delivered live pups were used for analysis. The study was approved by Institutional Animal Ethics Committee. All animals were housed under standard laboratory conditions, fed a diet of

in-house-prepared pellets, and provided with water *ad libitum*. The protocol employed in the study has been depicted in Fig. 1.

### 2.2 Tumor Transplantation

To record tumor spectra, tumors from the well-known mouse mammary tumor virus-induced spontaneous tumorigenesis model Indian Cancer Research Center mice<sup>14</sup> were extracted, cut into small pieces using a scalpel, washed in normal saline, and grafted (two tumors per mice) subcutaneously in the inguinal breast mammary fat pad region of the Swiss bare mice. Subcutaneous transplantation mimics the *in situ* location of breast tumors. The incisions made for transplantation were well away from the site of transplantation. This ensures that incisions in the skin do not influence the spectroscopic readings. Ten spectra were then recorded transcutaneously from four tumors transplanted in mice ( $n = 2$ ) a day after surgery, resulting in a total of 40 spectra in the tumor group.

### 2.3 Raman Spectroscopy

All spectra were recorded using HE-785 (Jobin-Yvon-Horiba, France) Raman spectrometer described elsewhere.<sup>15</sup> Briefly, this system consists of a diode laser (Process Instruments) of 785 nm wavelength as excitation source, a high efficiency spectrograph with fixed 950 gr/mm grating coupled with a charge-coupled device (CCD) (Synapse). The spectrograph has no movable parts and the resolution is  $\sim 4 \text{ cm}^{-1}$ . A commercially available (InPhotonics Inc., Downy St., USA) probe consisting of a 105  $\mu\text{m}$  excitation fiber and a 200  $\mu\text{m}$  collection fiber (numerical aperture = 0.40) was used to couple excitation source and detection system. The estimated spot size and depth of field as per the manufacturer's specifications is 105  $\mu\text{m}$  and 1 mm, respectively. Spectral acquisition parameters were  $\lambda_{\text{ex}}$  785 nm, laser power 80 mW, spectra were integrated for 15 s and averaged over three accumulations.

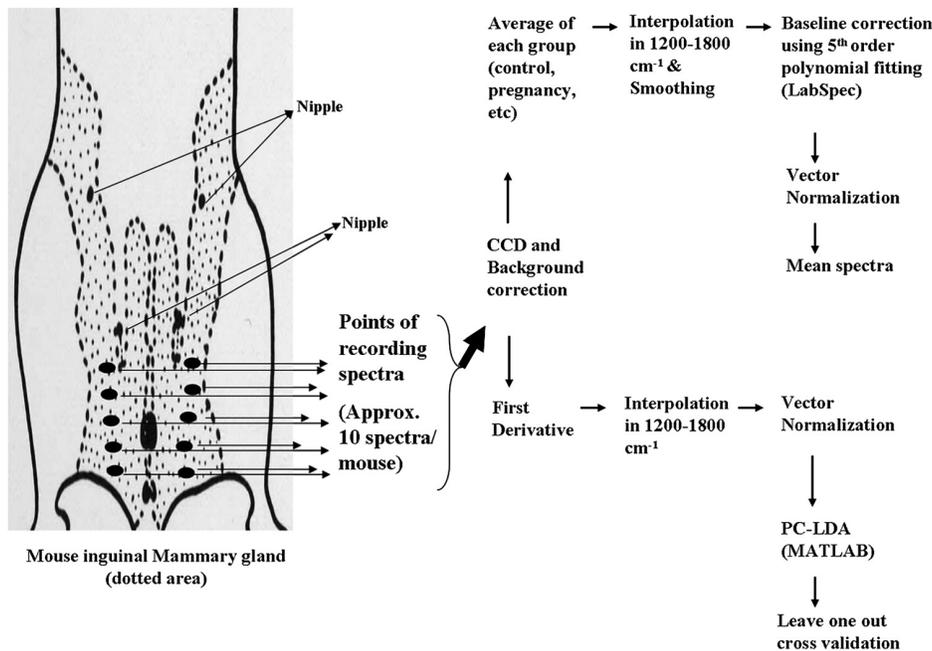


Fig. 1 Schematic representation of spectra acquisition, data preprocessing, and multivariate classification protocols.

## 2.4 Data Analysis

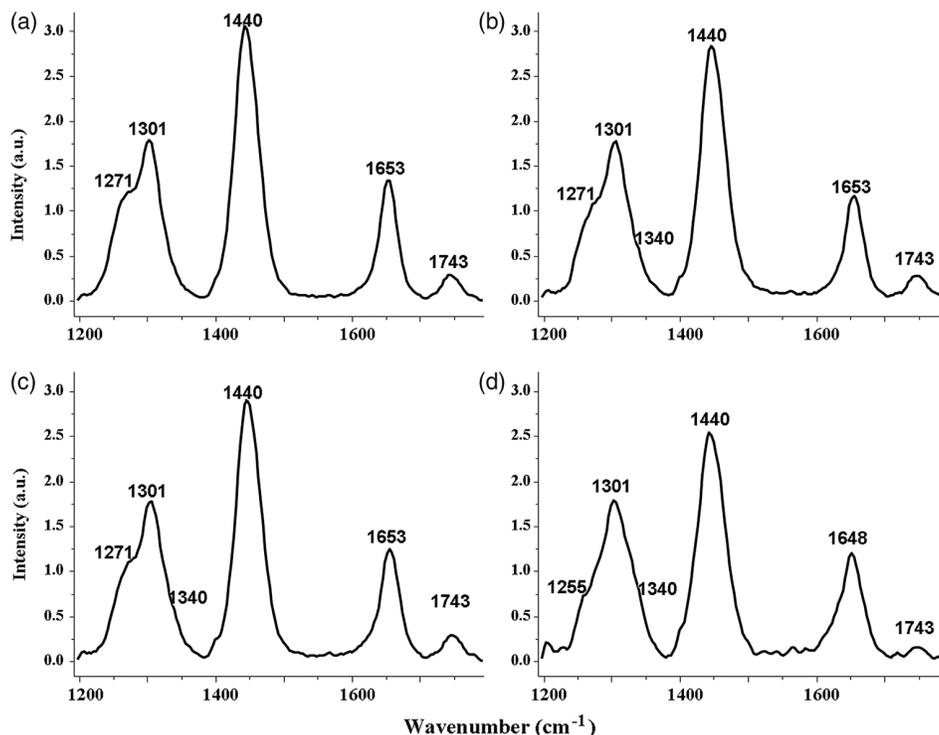
The protocol for data analysis<sup>16–18</sup> is as follows. Spectra recorded from mouse breasts of different groups were pre-processed by correcting for CCD response with a National Institute of Standards and Technology certified SRM 2241 material and subtraction of background signals from optical elements. To remove interference of the slow-moving background, first derivative of the preprocessed Raman spectra was calculated (Savitzky-Golay, window size 3), interpolated in the 1200 to 1800  $\text{cm}^{-1}$  range (Raman fingerprint region), and vector normalized. Classification of different groups was achieved using the multivariate analysis tool Principal Component-Linear Discriminant Analysis (PC-LDA)<sup>19,20</sup> implemented in MATLAB (Mathwork Inc.) based in-house software.<sup>21</sup>

First derivatives of preprocessed spectra were subjected to supervised PC-LDA. PCA is the routinely used method for data compression and visualization. It describes data variance by identifying a new set of orthogonal features, which are called principal components (PCs) that are linear combinations of original data variables. These PCs are calculated by identifying eigenvectors for the covariance matrix of mean-centered data. Because of their orthogonal characteristics, first few PCs are enough to represent maximum data variance. And for visual discrimination, we project each of the spectra in the newly formed coordinate space of these selected PCs. While PCA aims to identify features that represent variance among complete data, LDA provides data classification based on an optimized criterion which is aimed for more class separability. LDA is a method of choice when input data have higher within-class variance that could lead to development of PCs that are inappropriate for visual discrimination. The classification criterion is identified using the scatter measure of within-class and

between-class variances. LDA transformations are further identified as eigenvector matrix of this classification criterion. With the help of this LDA transform matrix, any test spectra can be classified to a class by iteratively calculating Euclidean/Root Mean Square or Mahalanobis distance of transformed test spectra and the mean of transformed input data set. In this study, we have employed Mahalanobis distance for class prediction, since it handles nonlinearity well.<sup>22</sup> LDA can be used in companion with PCA (PC-LDA) to further increase the performance 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).

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 the separation between the groups obtained by taking into account contribution of all factors selected for analysis. These results can also be depicted in the form of scatter plots generated by plotting combinations of scores of factors. Plotting different combinations of factor scores give a visual understanding of classification patterns in the data.

Average spectra were computed from the background subtracted spectra (without derivatization) for each class and baseline corrected by fitting a fifth-order polynomial function. The spectra were smoothed postaveraging using LabSpec 4.18 (average method, window size 3), for representing the mean spectra (Fig. 2). These baseline-corrected, vector-normalized spectra were used for spectral comparisons and for computing



**Fig. 2** Mean *in vivo* Raman spectra of breast from nonpregnant (a), pregnant (b), lactating (c), and tumor-bearing (d) mice interpolated in 1200 to 1800  $\text{cm}^{-1}$  range.

difference spectra. Standard deviation was also calculated to illustrate intragroup variability (Fig. 3).

### 3 Results and Discussion

RS is sensitive to biochemical changes which, apart from detecting malignancy-associated changes in a breast, may also detect normal physiological changes that affect the breast. To establish the validity of this technique as a diagnostic tool, a study of confounding variables is important. Pregnancy/lactation, a process that induces massive changes in the breast, is one of the confounding factors. In this study, the sensitivity of RS to differentiate breasts of normal, pregnant, and lactating mice and its effect on the diagnosis of breast cancer was evaluated. The role of pregnancy and lactation in breast cancers has been extensively explored in mouse models.<sup>23,24</sup> Thus, a mouse model was used in the study. Moreover, such studies are difficult in human subjects; hence animal models are preferred.

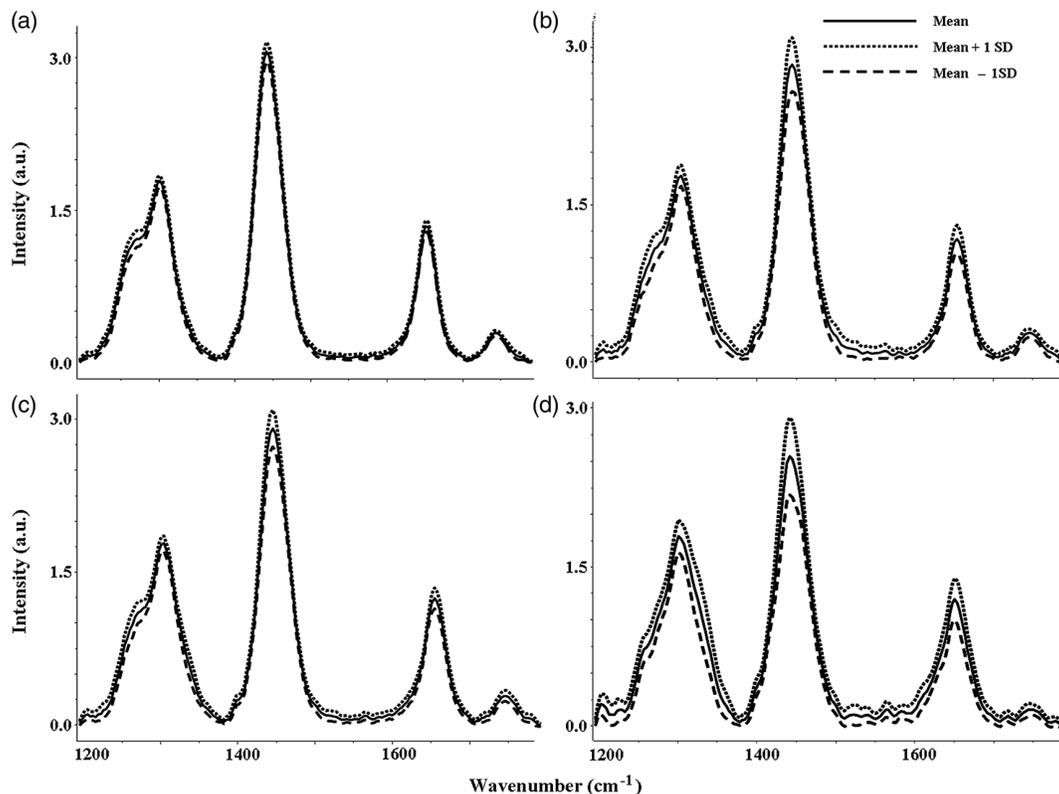
#### 3.1 Spectral Analysis

The spectral features of the mean control breast spectrum [Fig. 2(a)]  $1743\text{ cm}^{-1}$  ( $C=O$  ester);  $1653\text{ cm}^{-1}$  (amide I);  $1440\text{ cm}^{-1}$  ( $\delta$  CH<sub>2</sub>);  $1301\text{ cm}^{-1}$  ( $\tau$  CH<sub>2</sub>); and  $1271\text{ cm}^{-1}$  (amide III) can be attributed to lipids. Broad amide I, a change in features in the  $1200$  to  $1400\text{ cm}^{-1}$  region of the mean tumor spectrum [Fig. 2(d)], suggests the dominance of proteins and DNA. A normal breast consists of mammary epithelium supported by a mammary fat pad rich in lipids, whereas a tumor is characterized by changes in protein profiles, an increase in cell proliferation, and changes in breast architecture. This

explains lipid dominance in the control and variation in protein, increase in DNA, and loss of lipids in the tumors. These findings corroborate well with earlier studies.<sup>6,7,25</sup> The mean spectra of pregnancy and lactating breasts [Fig. 2(b) and 2(c), respectively] exhibit subtle but significant variations in the  $1340\text{ cm}^{-1}$  region.

There are several methods to study the spectral variations among groups; independent component analysis,<sup>6</sup> curve deconvolution,<sup>7</sup> and difference spectrum<sup>26</sup> are amongst the widely used. In our study, difference spectrum was computed by subtracting mean control spectrum from mean pregnancy, lactation, and tumor spectra, respectively [Fig. 4(a1) through 4(a3)]. The negative peaks are due to the control spectrum and the positive peaks are due to pregnancy, lactation, or tumor spectra. The difference pregnancy spectrum [Fig. 4(a2)] shows the following changes: a loss of lipids ( $1268\text{ cm}^{-1}$ ,  $1743\text{ cm}^{-1}$ ), an increase in DNA ( $1480\text{ cm}^{-1}$ ,  $1340\text{ cm}^{-1}$ ), and an increase in proteins ( $1671\text{ cm}^{-1}$ ,  $1471\text{ cm}^{-1}$ ,  $1315\text{ cm}^{-1}$ ). The difference lactating spectrum [Fig. 4(a2)] shows similar spectral features. Increase in proteins and DNA with a decrease in lipids may be attributed to an increase in the number of cell nuclei (cell division), which is known to take place during pregnancy and lactation. Tumor difference spectra [Fig. 4(a3)] suggests an increase in proteins ( $1671\text{ cm}^{-1}$ ,  $1456\text{ cm}^{-1}$ ,  $1471\text{ cm}^{-1}$ ), an increase in DNA ( $1480\text{ cm}^{-1}$ ,  $1340\text{ cm}^{-1}$ ), and a decrease in lipids ( $1743\text{ cm}^{-1}$ ,  $1440\text{ cm}^{-1}$ ), which corroborates previous reports.<sup>6,7</sup> Changes in lipids and DNA suggest cell division, which is hallmark of tumorigenesis.<sup>11</sup> Some positive bands ( $1630\text{ cm}^{-1}$  and  $1570\text{ cm}^{-1}$ ) may be ascribed to blood.<sup>27</sup>

To understand the difference between pathological and physiological conditions, mean pregnancy and mean lactating



**Fig. 3** Mean and standard deviation of transcutaneously recorded breast spectra from nonpregnant (a), pregnant (b), lactating (c), and tumor-bearing (d) mice interpolated in  $1200$  to  $1800\text{ cm}^{-1}$  range.

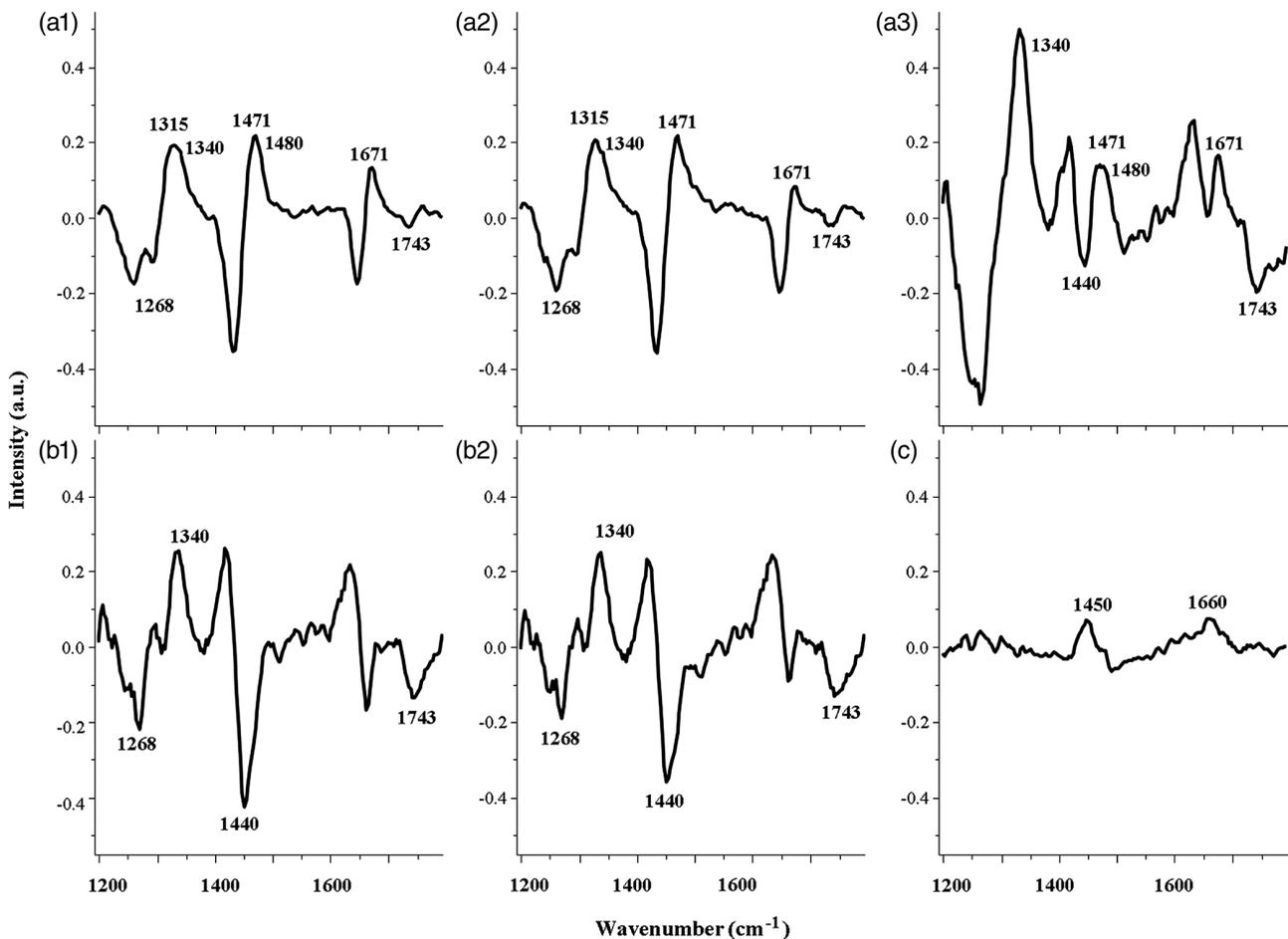
spectrum were subtracted (individually) from the mean tumor spectrum. In this case, the positive bands are due to tumors and the negative bands due to pregnancy or lactation. The difference pregnancy (tumor – pregnancy) and the difference lactation (tumor – lactation) spectrum, shown in Fig. 4(b1) and 4(b2), suggests decrease in lipids ( $1743\text{ cm}^{-1}$ ,  $1440\text{ cm}^{-1}$ , and  $1268\text{ cm}^{-1}$ ) and an increase in DNA ( $1340\text{ cm}^{-1}$ ) in tumors with respect to pregnancy or lactating conditions. Positive bands  $1630\text{ cm}^{-1}$  and  $1570\text{ cm}^{-1}$  may be ascribed to blood. Difference in the physiological spectrum [Fig. 4(c)] was also computed by subtracting the mean pregnancy spectrum from the mean lactating spectrum, wherein positive peaks are due to lactation and negative peaks due to pregnancy. In this case, the difference spectrum is very weak with respect to other difference spectra described above. Features  $1450\text{ cm}^{-1}$  and  $1660\text{ cm}^{-1}$  might indicate an increase in proteins in lactation with respect to pregnancy. The spectral assignments are based on available literature.<sup>27</sup>

### 3.2 Classification of Pregnancy- and Lactation-Associated Changes

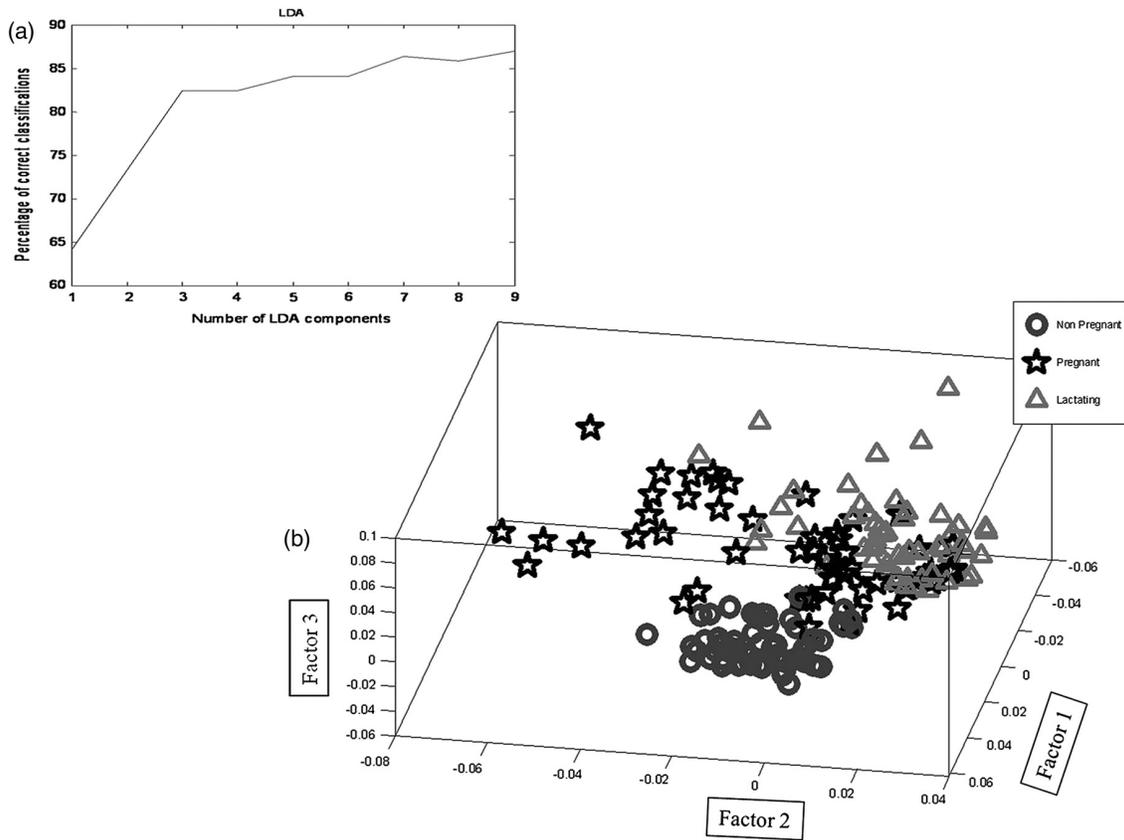
To explore the feasibility of differentiating pregnant and lactating conditions from control, PC-LDA was used. Spectra interpolated in the  $1200$  to  $1800\text{ cm}^{-1}$  range were used for analysis

(several ranges were explored for the study, best classification was obtained in the mentioned range). To avoid overfitting, nine factors<sup>28</sup> contributing 86% of correct classifications were used [Fig. 5(a)]. The three-dimensional (3-D) plot of PC-LDA factors 1, 2, and 3 [Fig. 5(b)] suggests classification of nonpregnant (control) mice breasts from pregnant and lactating mice breasts, while the breast spectra of pregnant mice and lactating mice overlap.

The confusion matrix for the PC-LDA model building is shown in Table 1(a). In this analysis, 61 out of 61 spectra are correctly classified as control. Thirty-eight out of 56 spectra are correctly classified as pregnant breast condition, whereas 4 out of 56 are misclassified as control and 14 out of 56 are misclassified as lactating condition. Fifty-four of 59 spectra are correctly classified as lactating condition, whereas 5 of 59 are misclassified as pregnancy condition. LOOCV was carried out to evaluate the results obtained by LDA. LOOCV builds a model based on all observations but one and tests the left out observation against the model built; this is repeated until all observations are left out once. The performance is estimated in terms of classification efficiency, which is the percentage of spectra from each group that are correctly classified. In analysis of LOOCV as shown in Table 2(b), once again, 61 out of 61 spectra were correctly classified as control. Thirty-four of 56 spectra are correctly classified as pregnant breast condition, whereas 4 of 56



**Fig. 4** Difference spectrum; pregnancy – control (a1); lactation – control (a2); tumor – control (a3); tumor – pregnancy (b1); tumor – lactation (b2); lactation – pregnancy (c).



**Fig. 5** PC-LDA to explore differences in mouse breast of nonpregnant, pregnant, and lactating mice: Scree plot (a) and 3-D plot of PC-LDA factors 1, 2, and 3 (b) suggesting classification between different breast conditions.

were misclassified as control and 18 of 56 were misclassified as lactating condition. As mentioned earlier, pregnancy is a phase midway between normal and lactation. This probably explains few misclassifications with normal breast. Both pregnant and lactation phases represent changes in breast as a result of rapid cell proliferation. This may explain high misclassification observed between pregnant and lactating conditions. Fifty-two of 59 spectra are correctly classified as lactating condition, while

7 of 59 were misclassified as pregnancy condition. The classification efficiency of lactation is higher than for pregnancy probably because lactation is characterized by cell differentiation and milk secretion in addition to cell proliferation. The classification efficiency of control, pregnant, and lactating mice breasts was 100%, 61%, and 88%, respectively.

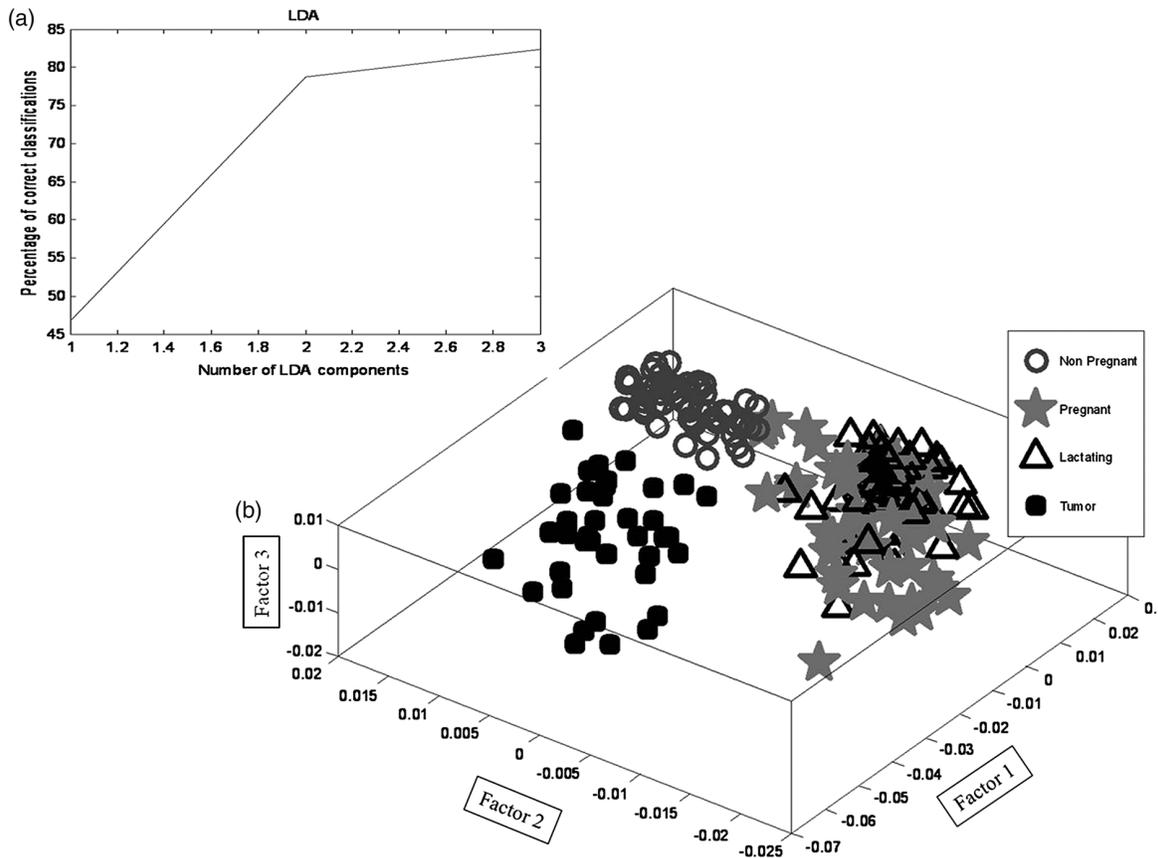
It is important to note that several changes take place in breast skin during pregnancy and lactation. In humans, skin

**Table 1** Confusion matrix for leave-one-out cross-validation of nonpregnant, pregnant, and lactating mouse breast: model building (a) and LOOCV (b) (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions). Sample size is shown in brackets.

Condition (no. of spectra, no. of animals used)	Nonpregnant	Pregnant	Lactating	Classification efficiency (%)
<b>(a)</b>				
Nonpregnant (61, 6)	61	0	0	100
Pregnant (56, 6)	4	38	14	68
Lactating (59, 6)	0	5	54	92
<b>(b)</b>				
Nonpregnant (61, 6)	61	0	0	100
Pregnant (56, 6)	4	34	18	61
Lactating (59, 6)	0	7	52	88

**Table 2** Confusion matrix for leave-one-out cross-validation of nonpregnant, pregnant, lactating mouse breast, and frank breast tumors: model building (a) and LOOCV (b) (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions). Sample size is shown in brackets.

Condition (no. of spectra, no. of animals used)	Nonpregnant	Pregnant	Lactating	Frank breast tumors	Classification efficiency (%)
(a)					
Nonpregnant (61, 6)	61	0	0	0	100
Pregnant (56, 6)	4	35	17	0	63
Lactating (59, 6)	0	16	43	0	73
Frank breast tumors (40, 2)	1	0	0	39	98
(b)					
Nonpregnant (61, 6)	61	0	0	0	100
Pregnant (56, 6)	4	34	18	0	61
Lactating (59, 6)	0	18	41	0	69
Frank breast tumors (40, 2)	1	0	0	39	98



**Fig. 6** PC-LDA to explore differences between breast spectra of tumor-bearing, nonpregnant, pregnant, and lactating mice: Scree plot (a) and 3-D plot of PC-LDA factors 1, 2, and 3 (b) suggesting classification between different breast conditions and frank breast tumors.

pigmentation increases, striae appear on breast skin, and skin gland secretions increase. Circulation to the skin increases and veins in the breast become more visible.<sup>29</sup> Thus, there is a possibility that these changes may affect breast spectra. In the present study, since a hairless variant of Swiss albino mice which lack pigments were used, pigmentation is not a factor. No striae appearance was observed. With respect to blood flow and vascularization, in this study, no spectral bands attributable to blood were observed. Tumor development involves angiogenesis (increased blood vessels and blood flow), but spectral features of blood have not been reported in transcutaneous spectra of breast tumors.<sup>30</sup> However, in the difference spectrum of current study—tumor – control, tumor – pregnancy, and tumor – lactation—some bands may be ascribed to blood.

### 3.3 Classification of Frank Breast Tumors from Pregnancy/Lactation

The possibility of classifying frank tumors from pregnancy- or lactation-associated changes was explored using PC-LDA. For analysis, three factors contributing to 82% of correct classifications were used [Fig. 6(a)]. The 3-D plot of PC-LDA factors 1, 2, and 3 [Fig. 6(b)] suggests classification of frank tumors from normal, pregnant, and lactating mice. The confusion matrices for model building and LOOCV are shown in Tables 2(a) and 2(b), respectively. Sixty-one out of 61 spectra are correctly classified as control, whereas 34 out of 56 and 41 out of 59, respectively, are correctly classified as pregnancy and lactation. Four out of 56 pregnancy spectra were misclassified as control, while 18 out of 59 were misclassified as lactating. No misclassification with respect to tumors was observed. Eighteen out of 59 of lactation misclassified with pregnancy, whereas no misclassifications with control or tumor are observed. These results mirror previous observations (Sec. 3.2). Since pregnancy is a phase between control and lactation, misclassifications with both were observed. Pregnancy and lactation are both characterized by cell proliferation, hence the observed misclassifications amongst them.

Thirty-nine out of 40 tumor spectra classify correctly as tumor. Only 1 of 40 misclassified as control. It is known that tumors are a heterogeneous complex of necrotic centers, rapidly proliferating fronts, and normal patches. This probably explains misclassification with normal breast. The classification efficiency of frank tumors from pregnancy/lactation conditions is 97.5%. Results suggest minimal effect of pregnancy- and lactation-associated changes on detection of frank tumors using RS.

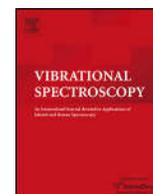
## 4 Conclusion

RS is a rapid, objective, and potentially noninvasive technique that can provide a “molecular fingerprint” of a sample. Hence, it has been extensively explored as a potential diagnostic tool for breast cancer. However, normal physiological changes like pregnancy and/or lactation affect the biochemical composition of the breast. To establish the validity of any diagnostic tool, a study of such confounding variables is important. In this study, the ability of this technique to identify pregnancy- and/or lactation-associated changes and its effect on tumor detection was evaluated. Results suggest that this technique can identify the above changes. Further, these physiological changes do not affect ability of RS to detect tumors. Further studies with precancerous and benign conditions are warranted.

## References

1. A. Jemal et al., “Global cancer statistics,” *CA Cancer J. Clin.* **61**(2), 69–90 (2011).
2. N. Howlader et al., Ed., *SEER Cancer Statistics Review, 1975–2008 (Vintage 2009 Populations)*, National Cancer Institute, Bethesda, Maryland (2011), [http://seer.cancer.gov/csr/1975\\_2009\\_pops09/](http://seer.cancer.gov/csr/1975_2009_pops09/).
3. S. M. Ismail et al., “Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia,” *Br. Med. J.* **298**(6675), 707–710 (1989).
4. J. B. Sorensen et al., “Interobserver variability in histopathologic subtyping and grading of pulmonary adenocarcinoma,” *Cancer* **71**(10), 2971–2976 (1993).
5. D. J. Evers et al., “Optical spectroscopy: current advances and future applications in cancer diagnostics and therapy,” *Future Oncol.* **8**(3), 307–320 (2012).
6. S. Haka et al., “Diagnosing breast cancer by using Raman spectroscopy,” *Proc. Natl. Acad. Sci. U.S.A.* **102**(35), 12371–12376 (2005).
7. M. V. P. Chowdary et al., “Biochemical correlation of Raman spectra of normal, benign and malignant breast tissues: a spectral deconvolution study,” *Biopolymers* **91**(7), 539–546 (2009).
8. C. V. Raman and K. S. Krishnan, “A new type of secondary radiation,” *Nature* **121**(3048), 501–502 (1928).
9. S. R. Oakes, H. N. Hilton, and C. J. Ormandy, “Key stages in mammary gland development—The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium,” *Breast Cancer Res.* **8**(207), 1–10 (2006).
10. M. M. Richert et al., “An atlas of mouse mammary gland development,” *J. Mammary Gland Biol.* **5**(2), 227–241 (2000).
11. B. S. Wiseman and Z. Werb, “Stromal effects on mammary gland development and breast cancer,” *Science* **296**(5570), 1046–1049 (2002).
12. V. Djonov, A. C. Andres, and A. Ziemiecki, “Vascular remodelling during the normal and malignant life cycle of the mammary gland,” *Microsc. Res. Tech.* **52**(2), 182–189 (2001).
13. R. A. Bhisey, P. L. Veturkar, and A. Borges, “Comparison of sensitivity of Swiss albino and hairless Swiss bare mice to two stage skin tumorigenesis,” *Indian J. Exp. Biol.* **25**(2), 90–96 (1987).
14. S. R. Pai and P. G. Save, “Influence of high dose of estradiol on ICRC mouse bearing transplanted mammary tumours,” *Indian J. Med. Res.* **64**(12), 1788–1789 (1976).
15. S. P. Singh et al., “Raman spectroscopy in head and neck cancers: towards oncological applications,” *J. Cancer Res. Ther.* **8**(6), S126–S132 (2012).
16. A. Sahu et al., “Raman spectroscopy of oral buccal mucosa: a study on age-related physiological changes and tobacco-related pathological changes,” *Technol. Cancer Res. Treat.* **11**(6), 528–623 (2012).
17. A. Nijssen et al., “Discriminating basal cell carcinoma from perilesional skin using high wave-number Raman spectroscopy,” *J. Biomed. Opt.* **12**(3), 034004 (2007).
18. S. Koljenovic et al., “Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy,” *Lab. Invest.* **82**(10), 1265–1277 (2002).
19. C. Kendall et al., “Evaluation of Raman probe for oesophageal cancer diagnostics,” *Analyst* **135**(12), 3038–3041 (2010).
20. T. J. Harvey et al., “Factors influencing the discrimination and classification of prostate cancer cell lines by FTIR microspectroscopy,” *Analyst* **134**(6), 1083–1091 (2009).
21. A. Ghanate et al., “Comparative evaluation of spectroscopic models using different multivariate statistical tools in a multicancer scenario,” *J. Biomed. Opt.* **16**(2), 025003 (2011).
22. S. Balakrishnama and A. Ganapathiraju, “Linear discriminant analysis—A brief tutorial,” (1998), [http://lcv.stat.fsu.edu/research/geometrical\\_representations\\_of\\_faces/PAPERS/lda\\_theory.pdf](http://lcv.stat.fsu.edu/research/geometrical_representations_of_faces/PAPERS/lda_theory.pdf).
23. D. Medina, “Mammary developmental fate and breast cancer risk,” *Endocr. Relat. Cancer* **12**(3), 483–495 (2005).
24. J. Russo et al., “The protective role of pregnancy in breast cancer,” *Breast Cancer Res.* **7**(3), 131–142 (2005).
25. R. E. Kast et al., “Raman spectroscopy can differentiate malignant tumors from normal breast tissue and detect early neoplastic changes in a mouse model,” *Biopolymers* **89**(3), 235–241 (2008).
26. M. V. Chowdary et al., “Discrimination of normal, benign, and malignant breast tissues by Raman spectroscopy,” *Biopolymers* **83**(5), 556–569 (2006).

27. F. S. Parker, *Application of Infrared, Raman, and Resonance Raman spectroscopy in Biochemistry*, Plenum Press, New York (1983).
28. J. Kelly et al., "Biospectroscopy to metabolically profile biomolecular structure: a multistage approach linking computational analysis with biomarkers," *J. Proteome Res.* **10**(4), 1437–1448 (2011).
29. C. Schaefer and P. Peters, "Dermatological medicine and local therapeutics," Chapter 2.10 in *Drugs During Pregnancy and Lactation*, C. Schaefer, Ed., pp. 105–106, Elsevier Science, Amsterdam, The Netherlands (2001).
30. L. Raniero et al., "In and *ex vivo* breast disease study by Raman spectroscopy," *Theor. Chem. Acc.* **130**(4–6), 1239–1247 (2011).



## Transcutaneous *in vivo* Raman spectroscopy: Detection of age-related changes in mouse breast

T. Bhattacharjee<sup>a</sup>, G. Maru<sup>b</sup>, A. Ingle<sup>c</sup>, C. Murali Krishna<sup>a,\*</sup>

<sup>a</sup> Chilakapati Lab, ACTREC, Kharghar, Navi-Mumbai 410210, India

<sup>b</sup> Maru Lab, ACTREC, Kharghar, Navi-Mumbai 410210, India

<sup>c</sup> Laboratory Animal Facility, ACTREC, Kharghar, Navi-Mumbai 410210, India

### ARTICLE INFO

#### Article history:

Received 7 November 2012

Received in revised form 12 April 2013

Accepted 15 April 2013

Available online 20 April 2013

#### Keywords:

Aging

Raman spectroscopy

Transcutaneous

*in vivo*

Mouse model

### ABSTRACT

The risk of female breast cancer increases as age progresses. This can be explained by Pike's model, which suggests that the process of aging in breast is not uniform. 'Breast ageing' is most rapid during menarche, slows with each pregnancy, slows further during perimenopause, and is least after the menopause. In this study, the feasibility of using transcutaneous *in vivo* Raman spectroscopy to detect age-related changes in mouse breast and its effect on tumor detection were explored. Spectra acquired transcutaneously from breast of 2 (menarche), 4–6 (mid reproductive phase), 10–12 (perimenopause), 13–15 month (menopause) old mice and frank breast tumors were analyzed using principal component-linear discriminant analysis (PC-LDA). A classification efficiency of ~80% was achieved for different age groups. Further, it was observed that the number of misclassifications among age groups increase as age progresses. For example, 3% spectra from menarche misclassify with other age groups, while 19–28% from mid-reproductive, perimenopause and menopause misclassify with each other. Misclassifications between groups indicate homogeneity in tissues. Thus, results suggest that menarche breast is biochemically distinct while breast during mid-reproductive, perimenopause and menopause is relatively homogenous. This probably indicates that the rate of aging is rapid during menarche, but slows down during other phases. Thus, spectroscopic data correlate with Pike's model. Although sensitive to age-related changes, RS could classify tumors with 95% efficiency. Overall, results suggest possibility of distinguishing age-related changes using RS without affecting ability to classify tumors.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death in females worldwide, accounting for 23% (1.38 million) of the total new cancer cases and 14% (458,400) of the total cancer deaths in 2008 [1]. In 2012, 226,870 new invasive breast cancer cases and 39,510 breast cancer deaths were estimated to occur in US [2]. In spite of extensive research, the etiology of human breast cancer remains largely unknown. Risk factors of breast cancer include hereditary, hormonal, and environmental factors [3]. Epidemiological evidence [4] suggests that reproductive hormones play a role in breast cancer progression [5]. These hormones stimulate cell division in breast tissue and result in accumulation of genetic damage. The accumulated

genetic damage increases the risk of breast cancer with increasing age. This results in increased incidence of breast cancer in older women. In 1983, Pike *et al.* proposed a model to explain the relationship between age, hormones and incidence of breast cancer, commonly referred to as Pike's model of 'breast tissue ageing' [6]. According to this model, the process of 'breast ageing' is not uniform, but fluctuates with hormone level variations at a given age. Rate of 'breast ageing' is most rapid during menarche (approximately 13–15 years of human age), slows with each pregnancy (approximately 15–35 years), slows further during perimenopause (approximately 35–45 years), and is least after the menopause (approximately above 45 years). Rate of breast aging also depends on several factors like (a) age at menarche, (b) age at menopause, (c) number of pregnancies, (d) breast composition, and (e) mammographic density [7], which vary from individual to individual. A rapid, non-invasive technique to detect age-related changes in breast may help screen high risk population and assist medical counseling.

Optical spectroscopic techniques like transillumination spectroscopy have been explored for breast cancer risk assessment [8]. Evaluation of risk factors like mammographic density and breast changes in young (18–21 years) and old (31–40 years)

\* Corresponding author at: Scientific Officer 'F' and Principal Investigator, Chilakapati Laboratory, Advanced Center for Treatment Research and Education in Cancer (ACTREC), Tata Memorial Center (TMC), Kharghar, Sector '22', Navi Mumbai 410210, India. Tel.: +91 22 2740 5039; fax: +91 22 2740 5085.

E-mail addresses: [mchilakapati@actrec.gov.in](mailto:mchilakapati@actrec.gov.in), [pittu1043@gmail.com](mailto:pittu1043@gmail.com) (C.M. Krishna).

women [9,10] have been reported using this technique. Raman spectroscopy (RS), a rapid, objective and non-invasive optical spectroscopic technique; has been explored extensively as an alternate/adjunct tool for diagnosis of cervix, lung, gastrointestinal, brain, oral [11–15] and several other cancers [16]. Several RS based breast cancer studies have also been reported in literature [17–24]. In 1991, Alfano et al. first reported differences between Raman spectra of normal and malignant breast [17] followed by several other groups [18–20]. Since then, surgical margin evaluation [21] and deep Raman spectroscopy have been demonstrated [22,23]. *In vivo* studies such as margin assessment during partial mastectomy breast surgery have also been reported [24].

RS is an inelastic scattering process where energy of photon scattered by the sample is different from the incident photon due to transfer of energy to or from vibrational modes of molecules in the sample. Since bands of Raman spectrum are characteristic of specific molecular vibrations unique to a molecule, RS can provide chemical fingerprint of a sample. Sensitivity of RS to malignancy associated biochemical changes makes it ideal for diagnostic applications. However, biochemical changes also occur during physiological processes like aging, pregnancy, lactation and menstrual cycle. At birth, the mammary glands consist of stroma and mammary epithelium. During menarche, the epithelium forms Terminal End Buds (TEB) which invades into the fat pad resulting in branched ducts throughout the breasts. During pregnancy and lactation, further cell division and branching takes place. At end of lactation, mammary tissue involutes to regain pre-pregnancy condition [25]. During perimenopause and menopause, atrophy (degeneration) of the mammary tissue occurs, resulting in considerable reduction in breast tissue post menopause [26,27]. Sensitivity of RS to these physiological changes may lead to misinterpretation of data and incorrect diagnosis. Therefore, the effect of physiological changes on diagnosis of cancer using RS needs to be validated. Previously, we have shown that pregnancy and lactation do not affect diagnosis of breast cancer using RS [28].

The current study aims to evaluate the ability of RS to detect age-related changes in mouse breast and its effect on diagnosis of breast cancer. Spectra were acquired transcutaneously from breast of mice belonging to different age groups and tumor bearing mice. Data were analyzed using principal component-linear discriminant analysis (PC-LDA). Findings of the study are reported in this paper.

## 2. Materials and methods

### 2.1. Animals

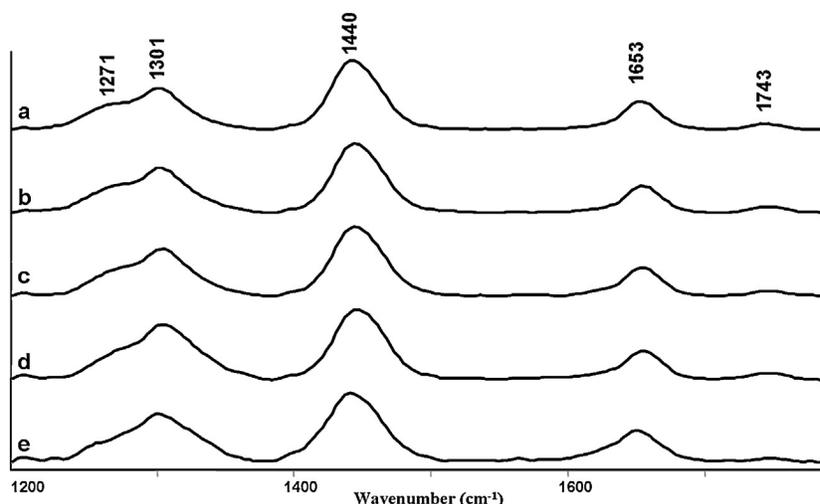
Breast-aging studies are difficult in human subjects; hence animal models have been explored. Mouse models have been extensively used to elucidate human aging process [29] and study influence of hormones on breast cancer [30]. Therefore, a mouse model was chosen for this study. 241 spectra were acquired transcutaneously from left and right inguinal breast of 19 Swiss Bare mice, a hairless mutant of Swiss albino mice [31]. Feasibility of transcutaneous studies with minimal skin interference has been previously reported [28,32,33]. The age groups explored were 2 months ( $n=5$ ), 4–6 months ( $n=4$ ), 10–12 months ( $n=5$ ) and 13–15 months ( $n=5$ ). Approximately 10 spectra per mouse were acquired from 2, 10–12, 13–15 months old mice while 20 spectra per mouse were acquired from 4 to 6 months old mice were analyzed. The study was approved by Institutional Animal Ethics Committee. All animals were housed under standard laboratory conditions, fed a diet of in-house-prepared pellets and provided with water *ad libitum*.

### 2.2. Tumor transplantation

Tumor transplantation was carried out using established protocols [34] by an expert veterinarian. Briefly, tumors from well known mouse mammary tumor virus (MMTV) – induced spontaneous tumorigenesis model, the Indian Cancer Research Center (ICRC) mouse [35] were extracted, cut into small pieces using a scalpel, and washed in normal saline. Histopathological analysis of one such piece confirmed tumorigenesis. The rest were grafted subcutaneously, 2 tumors per mice, in inguinal breast of 2 months old Swiss Bare mice. Subcutaneous transplantation mimics *in situ* location of breast tumors. Further, the incisions made for transplantation were well away from site of transplantation to ensure that they do not influence the spectroscopic readings. Spectra were acquired a day after transplantation from 2 tumor bearing mice. 10 spectra per tumor were collected, resulting in total 40 spectra in tumor group.

### 2.3. Raman spectroscopy

All spectra were recorded using HE-785 Raman spectrometer (Jobin-Yvon-Horiba, France) described elsewhere [36]. Briefly,



**Fig. 1.** Mean *in vivo* Raman spectra of mouse breast from (a) 2 months (menarche), (b) 4–6 months (mid-reproductive), (c) 10–12 months (perimenopause) (d) 13–15 months (menopause) old mice and (e) frank breast tumors interpolated in 1200–1800  $\text{cm}^{-1}$  range.

this system consists of a diode laser (Process Instruments) of 785 nm wavelength as excitation source, a high efficiency (HE) spectrograph with fixed 950 gr/mm grating coupled with a charged coupled device (CCD) (Synapse). The spectrograph has no movable parts and resolution is  $\sim 4\text{ cm}^{-1}$ . Commercially available fiber-optic probe (InPhotonics Inc., Downy St., USA), consisting of 105  $\mu\text{m}$  excitation fiber and a 200  $\mu\text{m}$  collection fiber (numerical aperture = 0.40) was used to couple excitation source and detection system. The estimated spot size and depth of field as per manufacturer's specifications are 105  $\mu\text{m}$  and 1 mm, respectively. Spectral acquisition parameters used for the study were:  $\lambda_{\text{ex}}=785\text{ nm}$ , laser power = 80 mW. Spectra were integrated for 15 s and averaged over 3 accumulations.

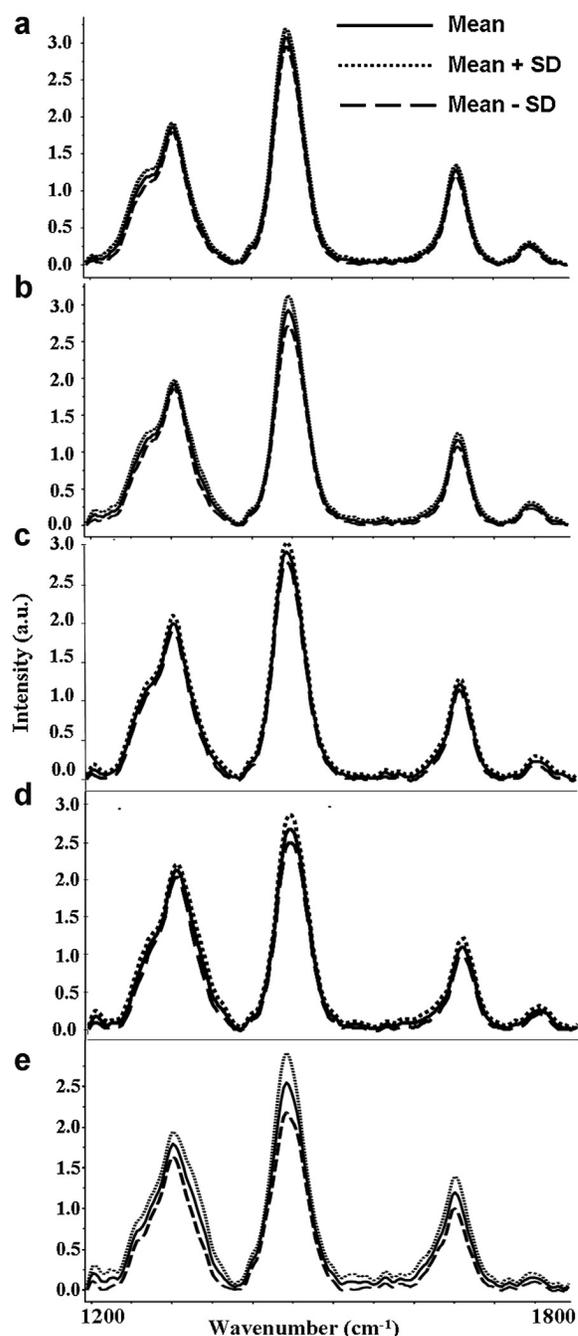
#### 2.4. Data analysis

Spectra of different groups were preprocessed by correcting for CCD response with a NIST certified SRM 2241 material and subtraction of background signals from optical elements [37]. To remove interference of the slow moving background, first derivative of the preprocessed Raman spectra were calculated (Savitzky–Golay, window size 3) [38,39], interpolated in 1200–1800  $\text{cm}^{-1}$  range (Raman fingerprint region) and vector normalized. Analysis of the preprocessed spectra was carried out using multivariate analysis tool PC-LDA implemented in MATLAB (Mathwork Inc.) based in-house software [40]. Principal component analysis (PCA) is routinely used method for data compression and visualization while LDA provides data classification based on an optimized criterion for more class separability. LDA can be used in conjunction with PCA (PC-LDA) to further increase performance efficiency of classification. PC-LDA models were cross validated using leave one spectrum out cross validation (LOOCV). LOOCV builds a model based on all spectra but one, and tests the left out 'spectrum' against the model built. This is repeated until all spectra have been left out once. In this mode of cross validation, each spectrum is treated as an individual sample. In view of the inherent heterogeneity of breast [25–27] and prospective applications of RS in surgical demarcation and optically guided biopsies, where it is necessary to delineate normal from abnormal points; such an approach is desirable [41].

Average spectra were computed from the background subtracted spectra (without derivatization) for each class and baseline corrected by fitting a fifth order polynomial function (Fig. 1). Standard Deviation was also calculated to illustrate intra-group variability (Fig. 2). These baseline corrected, vector normalized spectra were used for spectral comparisons and for computing difference spectra.

### 3. Results and discussion

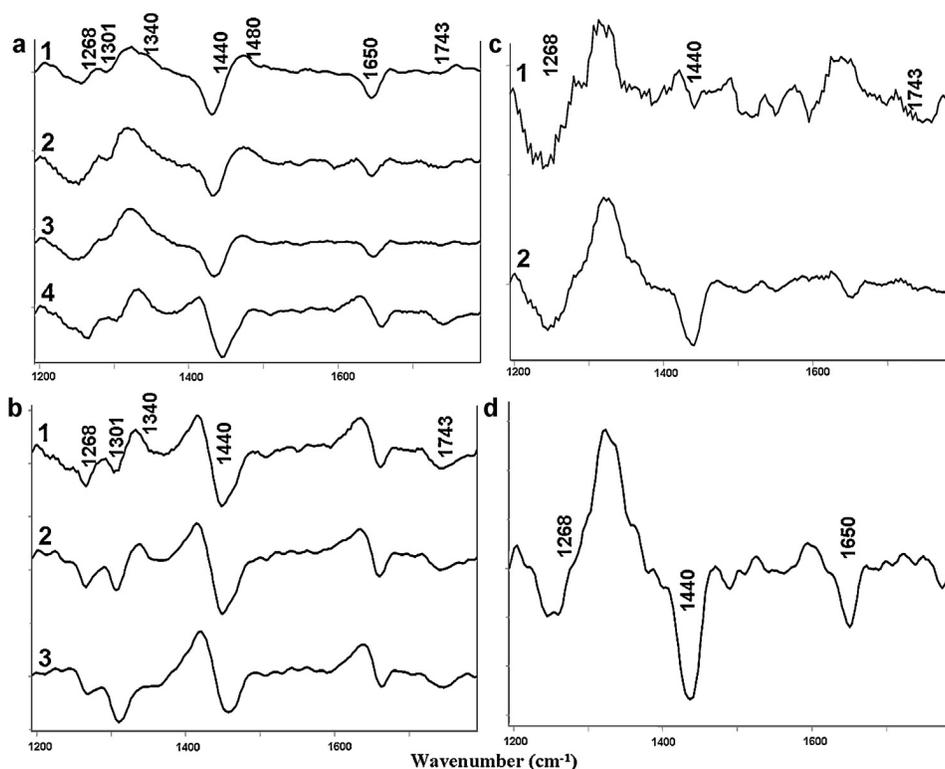
RS has been used for studying age-related changes in bone, collagen and ocular tissues [42–44]. The current study aims to study age-related changes in mouse breast using RS. Spectra were acquired transcutaneously from breast of mouse belonging to different age groups. The age groups were chosen based on the different reproductive phases described in theory of 'breast tissue ageing'. The majority of females from most inbred strains first ovulate (menarche) naturally between 6 and 8 weeks after birth [45], while they attain menopause at 12–14 months of age. Thus, 2 months old mice and 13–15 months old mice were chosen to study changes in breast at menarche and menopause respectively. Other age groups, 4–6 months (mid reproductive phase) and 10–12 months (perimenopause) were chosen based on the menarche–menopause time points and mouse–human age comparison calculations [46].



**Fig. 2.** Mean and standard deviation of *in vivo* mouse breast Raman spectra from (a) menarche, (b) mid-reproductive, (c) perimenopause (d) menopause mice and (e) frank breast tumors interpolated in 1200–1800  $\text{cm}^{-1}$  range.

#### 3.1. Spectral analysis

The spectral features of mean menarche breast spectrum (Fig. 1a) – 1743  $\text{cm}^{-1}$  (C=O); 1653  $\text{cm}^{-1}$  (amide I); 1440  $\text{cm}^{-1}$  ( $\delta\text{CH}_2$ ); 1301  $\text{cm}^{-1}$  ( $\tau\text{CH}_2$ ); and 1271  $\text{cm}^{-1}$  (amide III) can be attributed to lipids. Mean tumor spectrum (Fig. 1e) show broad amide I and features in 1200–1400  $\text{cm}^{-1}$  region, suggesting dominance of proteins and DNA in tumor. Normal breast consists of mammary epithelium and lipid rich mammary fat pad. This explains lipid dominance in normal breast. Tumor is characterized by changes in protein profiles, breast architecture and increase in cell proliferation. This may explain variation in protein, increase in DNA and loss of lipids in tumor. These findings corroborate well



**Fig. 3.** Difference spectrum; (a.1) mid-reproductive–menarche (a.2) perimenopause–menarche (a.3) menopause–menarche (a.4) tumor–menarche (b.1) tumor–mid-reproductive (b.2) tumor–perimenopause (b.3) tumor–menopause (c.1) menopause–mid-reproductive (c.2) perimenopause–mid-reproductive (d) menopause–perimenopause.

with earlier studies [18–20,47]. Mean spectra of mid-reproductive, perimenopause and menopause breast (Fig. 1b–d) exhibit subtle but significant changes in 1340, 1440, 1653 and 1743  $\text{cm}^{-1}$  region, probably indicating loss of lipids.

There are several methods for detailed study of spectral variations among groups. Independent component analysis [18], curve deconvolution [19] and difference spectrum [20] are amongst the widely used. In this study, difference spectrum was computed by subtracting mean menarche spectrum from mean mid-reproductive, perimenopause, menopause and tumor spectra, respectively (Fig. 3a 1–4). The negative peaks are due to menarche spectrum and positive peaks are due to mid-reproductive, perimenopause, menopause or tumor spectra. Mid-reproductive–menarche, perimenopause–menarche and menopause–menarche difference spectra suggest decrease in lipids (1268, 1301, 1650  $\text{cm}^{-1}$ ) and increase in DNA (1340, 1480  $\text{cm}^{-1}$ ). Tumor–menarche spectra exhibit decrease in lipids (1268, 1301, 1440, 1743  $\text{cm}^{-1}$ ) and increase in DNA (1340  $\text{cm}^{-1}$ ), which corroborate previous reports [18–20]. Changes in lipids and DNA in tumor suggest cell division – a hallmark of tumorigenesis [48].

To understand difference between pathological and physiological condition, mean mid-reproductive, perimenopause and menopause spectrum were subtracted from mean tumor spectrum. In this case, the positive bands are due to tumor and negative bands due to mid-reproductive, perimenopause or menopause. Tumor–mid-reproductive, tumor–perimenopause difference spectra (Fig. 3b 1–3) suggests decrease in lipids (1268, 1301, 1440, 1743  $\text{cm}^{-1}$ ) and increase in DNA (1340  $\text{cm}^{-1}$ ) whereas tumor–menopause difference spectrum shows decrease in lipids (1268, 1301, 1440, 1743  $\text{cm}^{-1}$ ) in tumor with respect to age groups mentioned.

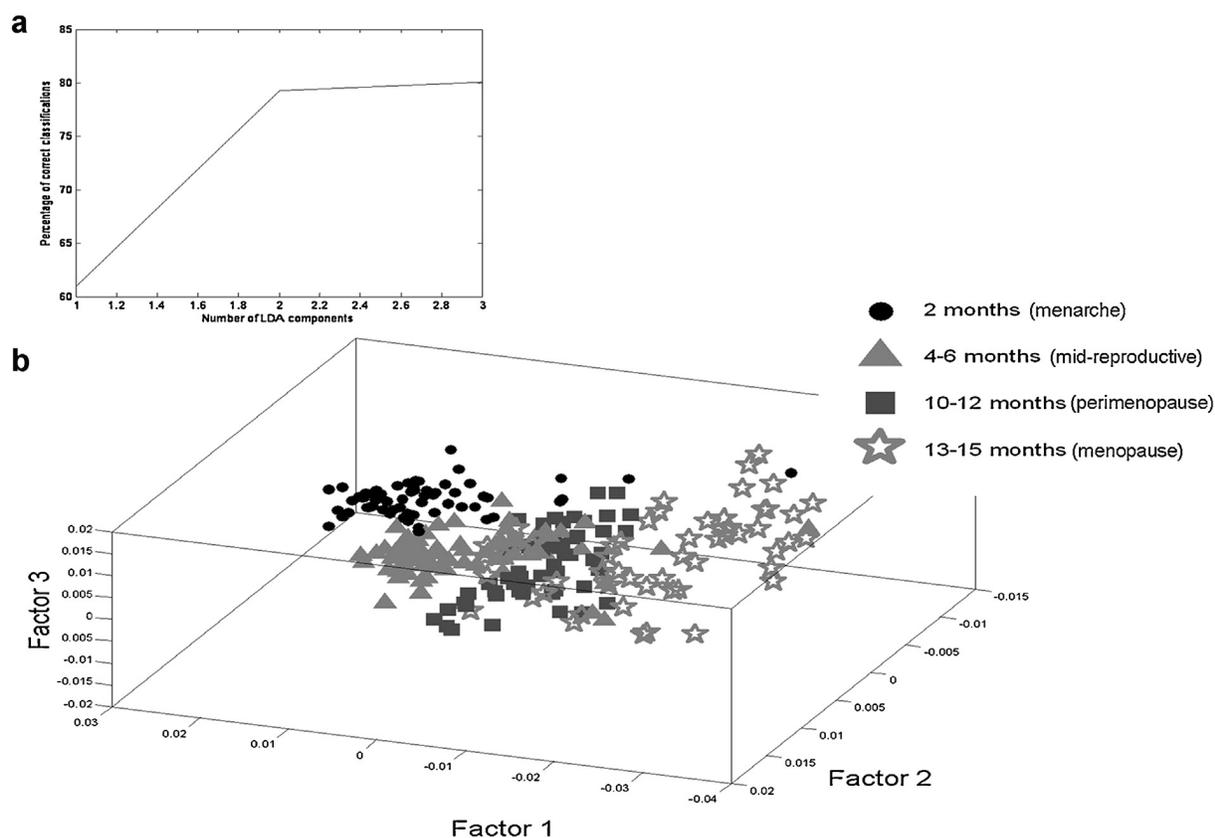
Difference physiological spectrum (Fig. 3c 1–2) was also computed, by subtracting mean mid-reproductive spectrum from mean

perimenopause and menopause spectrum, wherein negative peaks are due to mid-reproductive phase and positive peaks due to perimenopause and menopause. Decrease in lipids (1268, 1440 and 1743  $\text{cm}^{-1}$ ) during perimenopause and menopause with respect mid reproductive phase is observed in the above mentioned difference spectrum. Menopause–perimenopause difference spectrum (Fig. 3d) also suggest decrease in lipids (1268, 1440, 1650  $\text{cm}^{-1}$ ).

Overall, analyses of difference spectra suggest loss of lipids and increase in DNA in tumor with respect to all age groups studied. Loss of lipids and increase in DNA during menopause, perimenopause and mid-reproductive with respect to menarche breast is also observed. Further, older age groups exhibit loss of lipid with respect to younger age groups. The spectral assignments are based on available literature [49].

### 3.2. Classification of age-related changes in mouse breast

To explore the feasibility of classifying different age groups, PC-LDA was used. Spectra interpolated in 1200–1800  $\text{cm}^{-1}$  range were analyzed. Several ranges were explored for the study. Best classification was obtained in the mentioned range. To avoid over fitting, 3 factors [50] contributing 80% correct classification were used (Fig. 4a). The scatter plot of PC-LDA factors 1, 2 and 3 (Fig. 4b) shows menarche cluster is distinct, while there is slight overlap between mid-reproductive phase, perimenopause and menopause clusters. A similar trend is observed in the confusion matrix shown in Table 1. As observed in Table 1b, LOOCV yields average classification efficiency of 96.8%, 80.3%, 68.5% and 72.2% for 2 months (menarche), 4–6 months (mid reproductive phase), 10–12 months (perimenopause) and 13–15 months (menopause) old mouse breast, respectively. 12/71 (16%) spectra and 1/71 (1.4%) spectra from mid-reproductive phase misclassify with perimenopause



**Fig. 4.** PC-LDA for age-wise differences in breast of 2 months (menarche), 4–6 months (mid-reproductive), 10–12 months (perimenopause) and 13–15 months (menopause) old mice (a) Scree plot showing number of factors used and variance contributed by the factors (b) 3D plot of PC-LDA factors 1, 2 and 3 suggesting classification between different age groups.

and menopause respectively. 12/54 (22%) and 3/54 (5%) spectra from perimenopause misclassify with mid-reproductive phase and menopause, respectively. 6/54 (11%) and 9/54 (16%) spectra from menopause misclassify with mid-reproductive phase and perimenopause, respectively. Misclassification between groups suggests homogeneity in tissues and similarity in biochemical properties [51,52] of mid-reproductive phase, perimenopause and menopause breast. In contrast, only 1.6% spectra from menarche misclassify with mid-reproductive phase and perimenopause while no spectrum misclassify with menopause. This suggests that menarche is biochemically distinct. Thus, results indicate that aging is most rapid during menarche, but slows down during other phases. As described earlier, Pike's model suggests that the rate of 'breast ageing' is rapid during menarche, but slows down

during mid-reproductive phase, perimenopause and menopause. Thus, spectroscopic data correlates well with Pike's model of 'breast tissue ageing'. Overall, results suggest the possibility of detecting age-related changes in breast using transcutaneous *in vivo* Raman spectroscopy.

### 3.3. Classification of frank breast tumors from age-related changes

As mentioned earlier, sensitivity of RS to malignancy associated changes makes it ideal for diagnostic applications. However, biochemical changes also occur during physiological processes like aging. Sensitivity of RS to these physiological changes may lead to misinterpretation of data and incorrect diagnosis. Therefore, there

**Table 1**

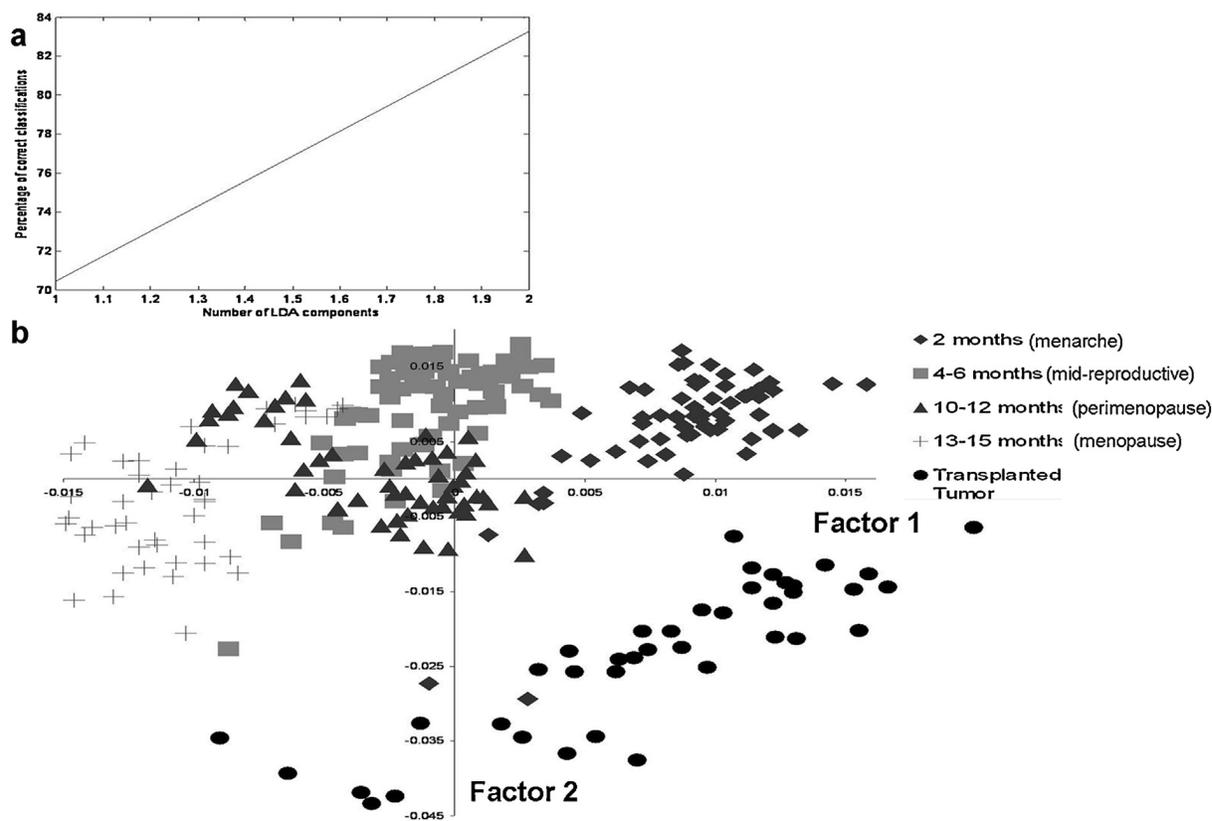
Confusion matrix for (a) model building and (b) leave-one-out cross validation of 2 months, 4–6 months, 10–12 months and months 13–15 old mice (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions).

Age (no. of spectra, no. of animals)	2 months (menarche)	4–6 months (mid-reproductive)	10–12 months (perimenopause)	13–15 months (menopause)	Classification efficiency (%)
<b>a</b>					
2 months (62, 5)	60	1	1	0	97
4–6 months (71, 4)	1	57	12	1	80
10–12 months (54, 5)	2	12	37	3	69
13–15 months (54, 5)	0	6	9	39	72
<b>b</b>					
2 months (62, 5)	60	1	1	0	97
4–6 months (71, 4)	1	57	12	1	80
10–12 months (54, 5)	2	12	37	3	69
13–15 months (54, 5)	0	6	9	39	72

**Table 2**

Confusion matrix for (a) model building and (b) leave-one-out cross validation of age groups compared with frank breast tumor.

Age (no. of spectra, no. of animals)	2 months (menarche)	4–6 months (mid-reproductive)	10–12 months (perimenopause)	13–15 months (menopause)	Transplanted tumor	Classification efficiency (%)
<b>a</b>						
2 months (62, 5)	56	0	4	0	2	90
4–6 months (71, 4)	0	54	14	3	0	76
10–12 months (54, 5)	0	10	40	4	0	74
13–15 months (54, 5)	0	7	1	46	0	85
Transplanted tumor	2	0	0	0	38	95
<b>b</b>						
2 months (62, 5)	56	0	4	0	2	90
4–6 months (71, 4)	0	54	14	3	0	76
10–12 months (54, 5)	0	11	39	4	0	72
13–15 months (54, 5)	0	7	1	46	0	85
Transplanted tumor (40, 2)	2	0	0	0	38	95

**Fig. 5.** PC-LDA to explore differences between frank breast tumors and age-related changes in breast (a) Scree plot (b) Scatter plot of PC-LDA factors 1 and 2 suggesting classification between different breast conditions and frank breast tumors.

is a need to evaluate effect of age-related changes on diagnosis of cancer using RS.

The possibility of classifying frank tumors from age-related changes was explored using PC-LDA. For analysis, 2 factors contributing to 83% correct classification was used (Fig. 5a). The scatter plot of PC-LDA factors 1 and 2 (Fig. 5b) suggests classification of tumors from age-related changes. Confusion matrix for model building and LOOCV of models built using PC-LDA algorithm is shown in Table 2 a and b respectively. 38/40 spectra were correctly classified as tumor. Only 2/40 spectra from tumor misclassified with 2 months age group. Since tumors were transplanted in 2 months old mice, misclassifications between tumors and 2 months age-group were probably observed. Overall, results suggest that frank tumors can be classified with 95% efficiency (Table 2b).

#### 4. Conclusions

Age-related variations in mouse breast were probed using transcutaneous *in vivo* Raman spectroscopy and its effect on tumor detection was evaluated. Results suggest possibility of identifying age-related changes in breast using this technique. However, age-related changes do not affect tumor detection, which is in agreement with previous reports [18].

#### Acknowledgements

Authors would like to acknowledge Mr. Pramod N. Tawde for his help in animal-handling and spectra acquisition and Ms. Aditi Sahu for help in editorial makeover. The Raman spectrometer employed in the study was procured from project

entitled “Development of *in vivo* laser Raman spectroscopy methods for diagnosis of oral precancerous and cancerous conditions” (BT/PRI11282/MED/32/83/2008), Department of Biotechnology, Government of India.

## References

- [1] A. Jemal, F. Bray, M.M. Center, J. Ferlay, E. Ward, D. Forman, *CA Cancer J. Clin.* 1 (2) (2011) 69–90.
- [2] R. Siegel, D. Naishadham, A. Jemal, *CA Cancer J. Clin.* 62 (2012) 10–29.
- [3] L.S. DeBruin, P.D. Josephy, *Environ. Health. Perspect.* 110 (1) (2002) 119–128.
- [4] B. Zumoff, J. Fishman, H.L. Bradlow, Leon Heilman, *Cancer Res.* 35 (1975) 3365–3373.
- [5] S.G. Korenman, *Lancet* 15 (8170) (1980) 700–701.
- [6] M.C. Pike, M.D. Krailo, B.E. Henderson, J.T. Casagrande, D.G. Hoel, *Nature* 303 (1983) 767–770.
- [7] O.M. Ginsburg, L.J. Martin, N.F. Boyd, *Br. J. Cancer* 99 (2008) 1369–1374.
- [8] K.M. Blackmore, J.A. Knight, L. Lilge, *Cancer Epidemiol. Biomarkers Prev.* 7 (5) (2008) 1043–1050.
- [9] K.M. Blackmore, J.A. Knight, J. Wong, S. Tharmalingam, L. Lilge, *Med. Phys.* 37 (2) (2010) 419–426.
- [10] K.M. Blackmore, J.A. Knight, R. Jong, L. Lilge, *Br. J. Radiol.* 80 (2007) 545–556.
- [11] V. Chang, P. Cartwright, S. Bean, G. Palmer, R. Bentley, N. Ramanujam, *Neoplasia* 1 (4) (2009) 325–332.
- [12] H. Yamazaki, S. Kaminaka, E. Kohda, M. Mukai, H. Hamaguchi, *Radiat. Med.* 21 (1) (2003) 1–6.
- [13] S. Teh, W. Zheng, K. Ho, M. Teh, K. Yeoh, Z. Huang, *Br. J. Cancer* 98 (2) (2008) 457–465.
- [14] Y. Zhou, C.H. Liu, Y. Sun, Y. Pu, S. Boydston-White, Y. Liu, R.R. Alfano, *J. Biomed. Opt.* 17 (11) (2012) 116021.
- [15] S.P. Singh, A. Deshmukh, P. Chaturvedi, C.M. Krishna, *J. Biomed. Opt.* 17 (10) (2012) 105002.
- [16] D.J. Evers, B.H.W. Hendriks, G.W. Lucassen, T.J.M. Ruers, *Future Oncol.* 8 (3) (2012) 307–320.
- [17] R.R. Alfano, C.H. Liu, W.L. Sha, H.R. Zhu, D.L. Akins, J. Cleary, R. Prudente, E. Celmer, *Laser Life Sci.* 4 (1) (1991) 23–28.
- [18] S. Haka, K.E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R.R. Dasari, M.S. Feld, *Proc. Natl. Acad. Sci. U.S.A.* 2 (35) (2005) 12371–12376.
- [19] M.V.P. Chowdary, K.K. Kumar, S. Mathew, L. Rao, C. Murali Krishna, J. Kurien, *Biopolymers* 1 (7) (2009) 539–546.
- [20] M.V. Chowdary, K.K. Kumar, J. Kurien, P. Mathew, C.M. Krishna, *Biopolymers* 3 (5) (2006) 556–569.
- [21] M.D. Keller, E. Vargis, N. Granja, R.H. Wilson, M.A. Mycek, M.C. Kelley, A. Mahadevan-Jansen, *J. Biomed. Opt.* 16 (7) (2011) 077006.
- [22] N. Stone, R. Baker, K. Rogers, A.W. Parker, P. Matousek, *Analyst* 132 (2007) 899–905.
- [23] N. Stone, P. Matousek, *Cancer Res.* 8 (11) (2008) 4424–4430.
- [24] A. Haka, Z. Volynskaya, J.A. Gardecki, J. Nazemi, J. Lyons, D. Hicks, M. Fitzmaurice, R.R. Dasari, J.P. Crowe, M.S. Feld, *Cancer Res.* 6 (6) (2006) 3317–3322.
- [25] M.M. Richert, K.L. Schwertfeger, J.W. Ryder, S.M. Anderson, J. Mammary Gland Biol. Neoplasia 5 (2) (2000) 227–241.
- [26] M. Murray, *Arch. Pathol. Lab. Med.* 133 (2009) 722–728.
- [27] J. Russo, I.H. Russo, *Maturitas* 49 (2004) 2–15.
- [28] T. Bhattacharjee, G. Maru, A. Ingle, C.M. Krishna, *J. Biomed. Opt.* 18 (4) (2013) 047004, <http://dx.doi.org/10.1117/1.JBO.18.4.047004>.
- [29] L. Demetrius, *Ann. N.Y. Acad. Sci.* 1067 (2006) 66–82.
- [30] D. Medina, *Cold Spring Harb. Perspect. Biol.* 2 (2010) a004523.
- [31] R.A. Bhisey, P.L. Veturkar, A. Borges, *Indian J. Exp. Biol.* 5 (2) (1987) 90–96.
- [32] T. Bhattacharjee, P. Kumar, G. Maru, A. Ingle, C.M. Krishna, *J. Cancer Res. Ther.* 8 (S1) (2012) S34.
- [33] L. Raniero, R.A. Canevari, L.N.Z. Ramalho, F.S. Ramalho, E.A.P. dos Santos, R.A. Bitar, K.J. Jalkanen, H.S. Martinho, A.A. Martin, *Theor. Chem. Acc.* 130 (4) (2011) 1239–1247.
- [34] M. Sluysers, R.V. Nie, *Cancer Res.* 34 (1974) 3253–3257.
- [35] S.R. Pai, P.G. Save, *Indian J. Med. Res.* 64 (12) (1976) 1788–1792.
- [36] S.P. Singh, A. Deshmukh, P. Chaturvedi, C.M. Krishna, *J. Cancer Res. Ther.* 8 (2012) S126–S132.
- [37] C. Gobinet, V. Vrabie, M. Manfait, O. Piot, *IEEE Trans. Biomed. Eng.* 6 (5) (2009) 1342–1371.
- [38] A. Sahu, A. Deshmukh, A.D. Ghanate, S.P. Singh, P. Chaturvedi, C.M. Krishna, *Technol. Cancer Res. Treat.* 1 (6) (2012) 528–623.
- [39] A. Nijssen, K. Maquelin, P.J. Caspers, T.C.B. Schut, M.H.A. Neumann, G.J. Puppels, *J. Biomed. Opt.* 12 (2007) 034004-1–034004-7.
- [40] A.D. Ghanate, S. Kothiwale, S.P. Singh, D. Bertrand, C.M. Krishna, *J. Biomed. Opt.* 6 (2) (2011) 1–9.
- [41] R. Malini, K.V. Krishna, J. Kurien, K.M. Pai, L. Rao, V.B. Kartha, C.M. Krishna, *Biopolymers* 81 (2006) 179–193.
- [42] J.J. Freeman, B. Wopenka, M.J. Silva, J.D. Pasteris, *Calcif. Tissue Int.* 68 (3) (2001) 156–162.
- [43] K.A. Dehring, A.R. Smukler, B.J. Roessler, Michael D. Morris, *Appl. Spectrosc.* 0 (4) (2006) 366–372.
- [44] Y.S. Ozaki, A.T. Mizunol, K. Itoh, K.S. Iriyama, *J. Biol. Chem.* 62 (32) (1987) 15445–15551.
- [45] D.G. Whittingham, M.J. Wood, in: H.L. Foster, J.D. Small, J.G. Fox (Eds.), *Mouse in Biomedical Research*, Academic Press, New York, 1983, pp. 156–157.
- [46] K. Flurkey, J.M. Currer, D.E. Harrison, in: J.G. Fox, S.W. Barthold, M.T. Davisson (Eds.), *The Mouse in Biomedical Research*, Elsevier, Burlington, MA, 2007, pp. 637–672.
- [47] R.E. Kast, G.K. Serhatkulu, A. Cao, A.K. Pandya, H. Dai, J.S. Thakur, V.M. Naik, R. Naik, M.D. Klein, G.W. Auner, R. Rabah, *Biopolymers* 89 (3) (2008) 235–241.
- [48] B.S. Wiseman, Z. Werb, *Science* 296 (5570) (2002) 1046–1049.
- [49] F.S. Parker, *Application of Infrared, Raman and Resonance Raman spectroscopy in Biochemistry*, Plenum Press, New York, NY, 1983.
- [50] J. Kelly, J. Trevisan, A. Scott, P. Crmichael, H. Pollock, P. Martin-Hirsch, F. Martin, *J. Proteome Res.* 10 (2011) 1437–1448.
- [51] N. Bergner, *Chemom. Intell. Lab. Syst.* 117 (2012) 224–232.
- [52] M. Francisco-Fernandez, *Chemom. Intell. Lab. Syst.* 118 (2012) 159–172.

# Swiss bare mice: a suitable model for transcutaneous in vivo Raman spectroscopic studies of breast cancer

T. Bhattacharjee · Piyush Kumar · G. Maru · A. Ingle · C. Murali Krishna

Received: 17 December 2012 / Accepted: 8 May 2013  
© Springer-Verlag London 2013

**Abstract** Breast cancer is the most common cancer affecting females worldwide. As early detection results in better prognosis, screening tools for breast cancer are being explored. Raman spectroscopy, a rapid, objective, and noninvasive tool, has shown promising results in the diagnosis of several cancers including breast cancer. For development as a screening tool, a study of spectral signatures associated with breast cancer progression is imperative. However, such studies are not possible in human subjects. Hence, there is a need for a suitable animal model, which is conducive to transcutaneous in vivo Raman spectroscopic measurements of breast with minimal interference from skin and hair and has contribution from functional mammary epithelium of breast. In this study, rodent models like C57, Swiss albino, Swiss bare, agouti mice, and Sprague–Dawley rats were evaluated. Among these models, transcutaneous breast spectra of hairless Swiss bare mice have the best signal-to-noise ratio and were closest to reported ex vivo as well as intraoperative in vivo human breast spectra. Principal

component–linear discriminant analysis of several anatomical sites confirms minimal skin interference and suggests contribution from functional mammary epithelium of breast. Moreover, transcutaneous spectra from normal breast and breast tumors of Swiss bare mice could be classified with 99 % efficiency, which is better than the previous reports. Thus, Swiss bare mice model may be better suited for transcutaneous in vivo Raman spectroscopic studies of breast physiology and pathology, especially breast cancer. Prospectively, in addition to cancer progression, breast-to-bone metastasis can also be studied, since these anatomical sites can be uniquely classified.

**Keywords** Transcutaneous · In vivo · Raman spectroscopy · Animal model · Breast cancer

## Introduction

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death in females worldwide, accounting for 23 % (1.38 million) of new cancer cases and 14 % (458,400) of the total cancer deaths in 2008 [1]. Literature suggests improved prognosis with early detection [2]. However, due to disadvantages like tedious sample preparation, long output times, and interobserver variances [3, 4] associated with the currently available screening/diagnostic tools, rapid, objective, and preferably noninvasive alternate screening/diagnostic techniques are being extensively explored. One such tool, Raman spectroscopy (RS), has shown promising results in the diagnosis of several cancers [5] including breast cancer [6, 7]. The ability of this technique to classify normal breast tissues from benign and malignant tissues has been established ex vivo [8]. In vivo margin assessment during breast cancer surgery has also been demonstrated [9]. Spectroscopic distinction between metastatic and nonmetastatic cell lines has been shown [10]. Studies to evaluate possibility of distinguishing invasive carcinoma and

---

T. Bhattacharjee and Piyush Kumar have contributed equally to this work.

T. Bhattacharjee · P. Kumar · C. M. Krishna  
Chilakapati Lab, ACTREC, Kharghar,  
Navi Mumbai 410210, India

G. Maru  
Maru Lab, ACTREC, Kharghar,  
Navi Mumbai 410210, India

A. Ingle  
Laboratory Animal Facility, ACTREC, Kharghar,  
Navi Mumbai 410210, India

C. M. Krishna (✉)  
Chilakapati Laboratory, Advanced Center for Treatment Research  
and Education in Cancer (ACTREC), Tata Memorial Center  
(TMC), Kharghar, Sector '22',  
Navi Mumbai 410210, India  
e-mail: mchilakapati@actrec.gov.in

C. M. Krishna  
e-mail: pittu1043@gmail.com

ductal in situ carcinoma have been attempted using cryopreserved sections [11]. Due to its sensitivity to subtle biochemical changes and ability to identify malignancy, development of RS as an “early cancer detection” tool is the focus of current biomedical optics-based research.

“Early detection” research requires spectra of cancer progression stages and clinically normal conditions which will eventually develop into cancer. However, most patients hospitalized are in advanced stages of breast cancer. Thus, these studies are very difficult in human subjects. Conventionally, such research has relied on animal model systems. Animal models are of paramount importance in different aspects of biomedical research. Studies on initiation, progression, and prevention for several diseases have been successfully carried out on animals. These models form the basis of testing efficacy and side effects of new drugs. They also provide a platform to study basic physiological processes and functions of the body [12, 13]. A model system conducive to in vivo Raman spectroscopic measurements may help study breast cancer progression and identify spectral markers for “early detection” of cancer.

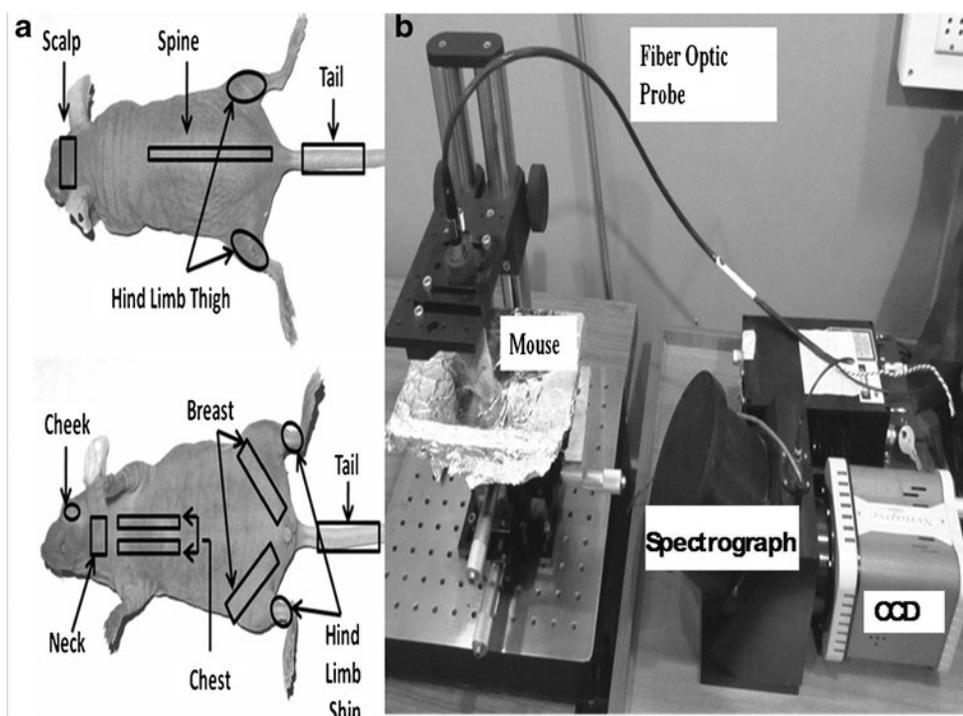
The aim of the current study was to identify a suitable animal model for transcutaneous in vivo Raman spectroscopic studies. A plethora of animal models, especially rodents, have been used in basic and clinical research, due to several advantages over large animals. Rodents like mice and rats have short life cycle, can be easily bred and obtained in large numbers, can be conveniently handled, can be manipulated surgically, and can be transplanted with human cancer cell lines (in case of immunocompromised

animals) and tumors. Further, rats and mice have been extensively used for pathological, genetic, and proteomic studies of breast cancer [12, 13]. Thus, rodent models were considered for the study. From spectroscopic point of view, an ideal model would be one that (1) enables convenient and efficient in vivo spectral measurements from every region of breast, (2) does not require prior incision or any other surgical procedure, (3) suffers minimally from interferences like hair, (4) has no/negligible contribution from biomolecules of anatomical layers like skin, (5) spectra are of definite breast origin, and (6) functional mammary epithelium, which is the site of cancer, contributes to breast spectra in addition to mammary fat pad. In this study, several rodent species and strains were screened for compliance with the above criteria. Spectra were recorded transcutaneously from inguinal breast of dark-haired mice, agouti mice, white-haired mice, white-haired rats, shaved rats, and hairless mice. After choosing a probable model, uniqueness of breast spectra compared to other anatomical sites was studied using principal component–linear discriminant analysis (PC-LDA). The efficiency of classification between transcutaneous normal and tumor spectra was also evaluated. Findings of the study are discussed in the paper.

## Materials and methods

**Animals** Approximately 10 spectra were recorded transcutaneously from inguinal breast of C57 mice ( $n=2$ ), agouti mice ( $n=2$ ), Swiss albino mice ( $n=2$ ), Swiss bare

**Fig. 1** Illustration of **a** anatomical sites used in the study and **b** Raman spectrometer

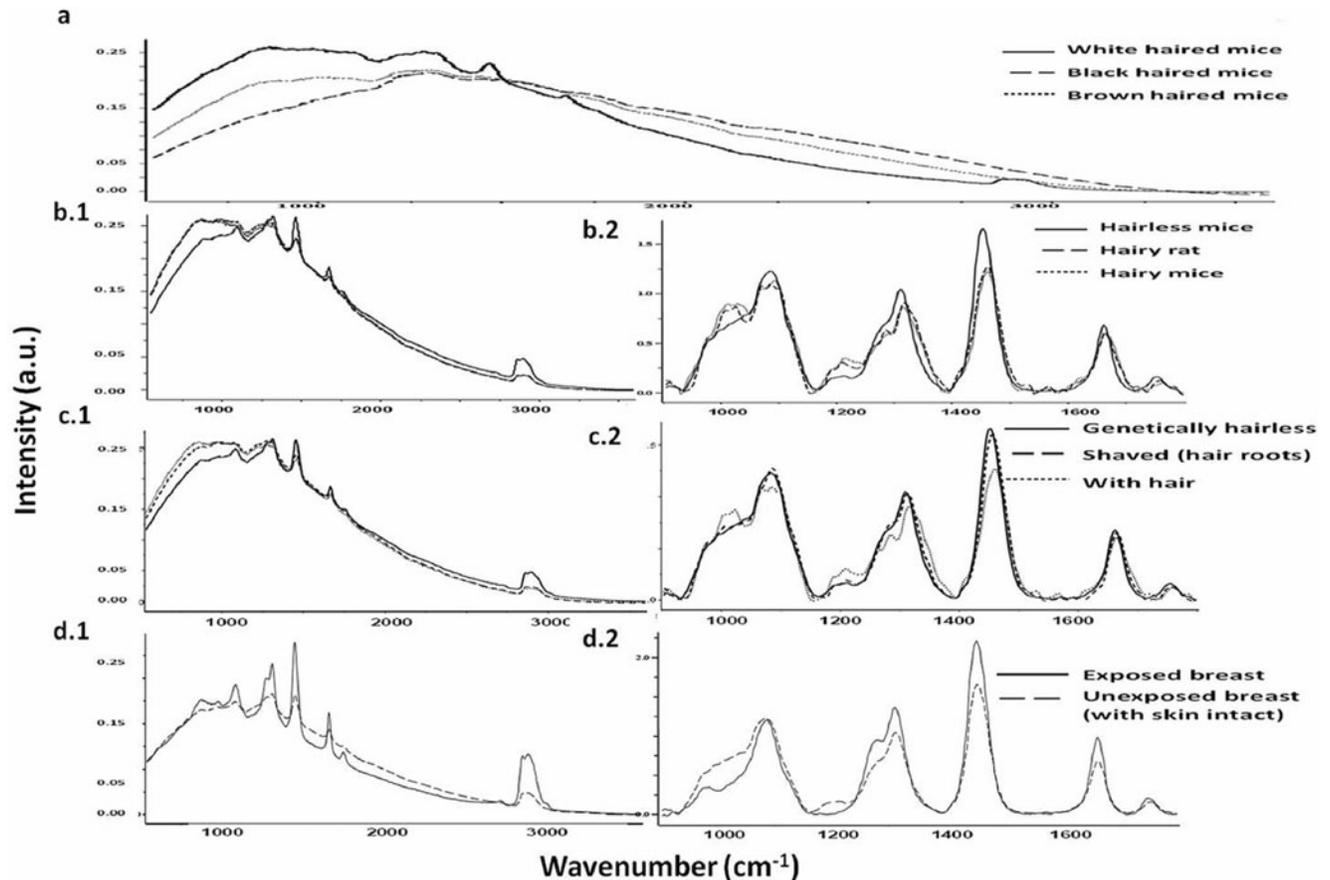


(SB) mice ( $n=2$ ), a hairless mutant of Swiss albino mice [14] and Sprague–Dawley (SD) rats ( $n=2$ ) for choosing the best species/strain. Skin in the inguinal breast region of SB mice was incised to expose breast and has been referred to as “exposed breast” in the rest of the write-up. Ten spectra were recorded from “exposed breast” of two SB mice. Spectra were also recorded from inguinal breast of rats after shaving. Further, approximately six spectra were recorded from breast, scalp, cheek, neck, chest, thigh, shin, spine, and tail of SB mice ( $n=10$ ) to evaluate the uniqueness of breast spectra. The anatomical sites mentioned are illustrated in Fig. 1a. The study was approved by the Institutional Animal Ethics Committee. All animals were housed under standard laboratory conditions, fed with a diet of in-house-prepared pellets, and provided with water ad libitum.

**Tumor transplantation** Indian Cancer Research Center (ICRC) mouse, a well-known mouse mammary tumor virus-

induced spontaneous breast tumorigenesis model [15], was used to extract tumors. Tumor transplantation was carried out using established protocols [16] by an expert veterinarian. Briefly, ICRC mouse breast tumor was extracted, cut into small pieces using a scalpel, and washed in normal saline. Histopathological analysis of one such piece confirmed tumorigenesis. The rest were grafted (two tumors per mice) subcutaneously in inguinal breast of SB mice. Spectra were recorded transcutaneously from tumors transplanted in mice ( $n=2$ ).

**Raman spectroscopy** All spectra were recorded using a Raman spectrometer (Fig. 1b) described elsewhere [17]. Briefly, this system consists of a diode laser (PI-ECL-785-300-FC, Process Instruments) of 785 nm wavelength as excitation source, a high-efficiency spectrograph (HE-785, HORIBA Jobin Yvon, France) with a fixed 950 gr/mm grating coupled with a charge-coupled



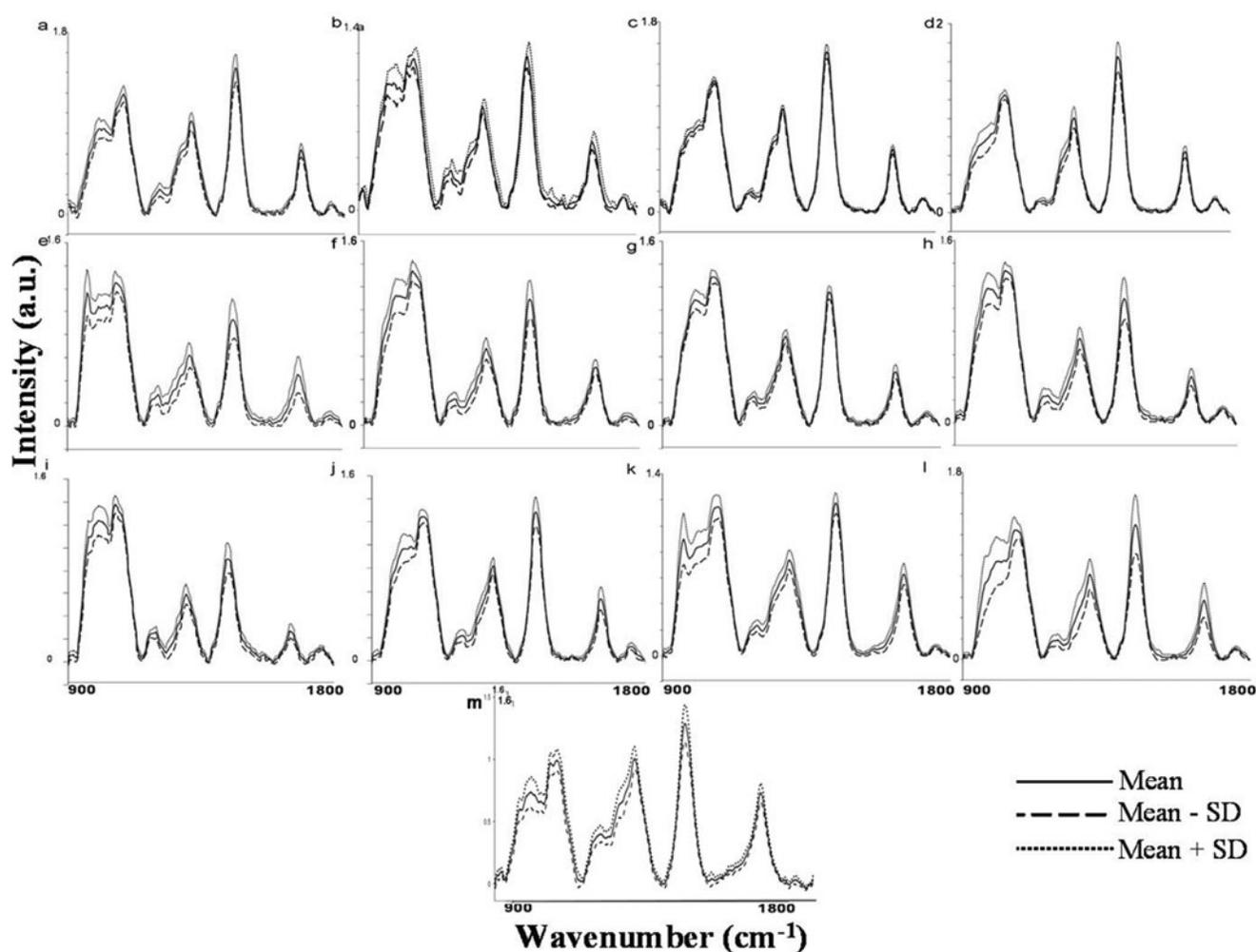
**Fig. 2** (a) Typical transcutaneous raw spectra from inguinal breast of white-, brown-, and black-haired mice. White-haired mice give the best signal-to-noise ratio. (b.1) Typical transcutaneous raw spectra and (b.2) mean corrected spectra of hairless mice, white-haired mice, and white-haired rats ( $900\text{--}1800\text{ cm}^{-1}$ ); hairless mice spectra have the best signal-to-noise ratio. (c.1) Typical transcutaneous raw spectrum and (c.2) mean corrected spectrum ( $900\text{--}1800\text{ cm}^{-1}$ ) from breast of

unshaved rats (with hair), shaved rats (with roots of hair), and genetically hairless mice (no hair/hair roots); hairless mice have the best signal-to-noise ratio. (d.1) Typical raw spectrum and (d.2) mean corrected spectrum ( $900\text{--}1800\text{ cm}^{-1}$ ) of exposed breast and transcutaneous (with skin) breast of hairless SB mice; transcutaneous breast spectrum closely resemble exposed breast spectrum

device -CCD (CCD-1024X256-BIDD-SYN, Synapse). The spectrograph has no movable parts, and spectral resolution is  $\sim 4 \text{ cm}^{-1}$ . Commercial RamanProbe (RPS 785/12-5, InPhotonics Inc., Downey St., USA) [18], consisting of an excitation and a collection fiber (NA-0.40) of diameters 105 and 200  $\mu\text{m}$ , respectively, was used to couple excitation source and detection system. This probe utilizes a backscattering ( $\theta=180^\circ$ ) sampling geometry. The estimated spot size and depth of penetration as per the manufacturer's specifications are 105  $\mu\text{m}$  and 1 mm, respectively. Spectral acquisition parameters were as follows:  $\lambda_{\text{ex}}=785 \text{ nm}$ , laser power 80 mW, and spectra were integrated for 15 s and averaged over three accumulations.

**Data analysis** The protocol for data analysis is as follows: Spectra of different groups were preprocessed by

correcting for CCD response with a NIST-certified SRM 2241 material and subtraction of background signals from optical elements [19]. To remove interference of the slow-moving background, the first derivative of the preprocessed Raman spectra was calculated (Savitzky–Golay, window size 3) [20–22], interpolated in the range of 900–1800  $\text{cm}^{-1}$  (Raman fingerprint region), and vector-normalized. Analysis of the preprocessed spectra was carried out using multivariate analysis tool PC-LDA implemented in MATLAB (MathWorks Inc.)-based in-house software [23]. Raw spectra, corrected for CCD and background, were baseline-corrected using fifth-order polynomial fit in LabSpec 4.18. Mean spectra, computed from these baseline-corrected spectra, were used for spectral comparison. Standard deviations were computed to illustrate the errors in measurement across all groups.



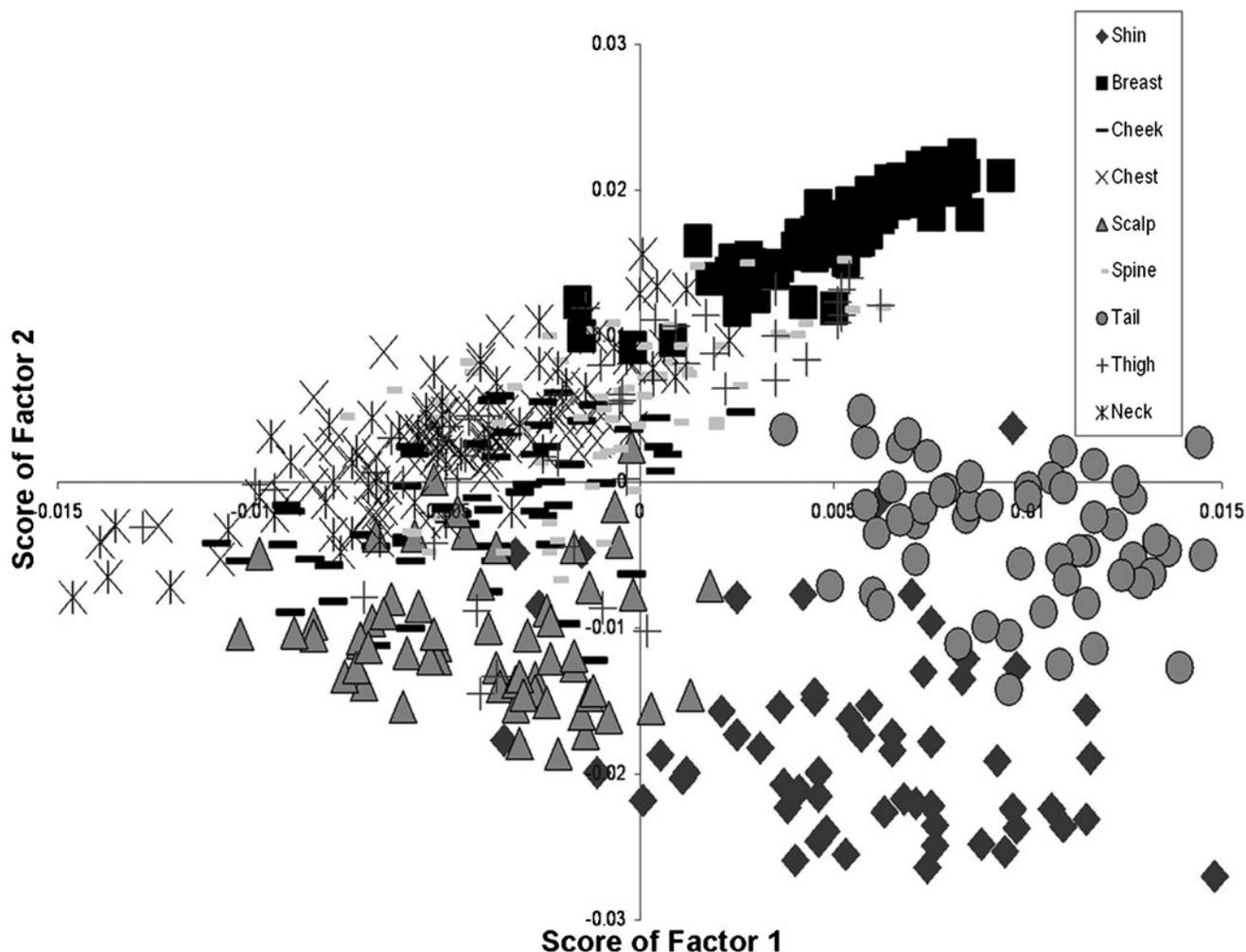
**Fig. 3** Mean and standard deviation of transcutaneously acquired spectra (900–1800  $\text{cm}^{-1}$ ) from **a** hairy mouse breast, **b** hairy rat breast, **c** shaved rat breast, **d** hairless mouse breast, **e** hairless mouse shin, **f** hairless mouse cheek, **g** hairless mouse chest, **h** hairless mouse scalp, **i**

hairless mouse spine, **j** hairless mouse tail, **k** hairless mouse thigh, **l** hairless mouse neck, and **m** tumor transplanted subcutaneously in hairless mouse breast region

## Results and discussion

*Evaluation of hairless Swiss bare mice as a suitable transcutaneous model for in vivo Raman spectroscopy of breast cancer* In the first step, spectra of animals with different hair colors were acquired. Typical raw spectra acquired transcutaneously from breast of black-haired C57 mice, brown-haired agouti mice, and white-haired Swiss albino mice are shown in Fig. 2(a). As seen in the figure, spectra of colored mice exhibit very high background, whereas white-haired mice gave the best signal-to-noise ratio with low background. Thus, further studies were carried out using white-haired animals. Next, typical raw spectrum (Fig. 2(b.1)) and corrected mean spectrum (Fig. 2(b.2)) of white-haired SD rats, white-haired Swiss albino mice, and hairless SB mice were evaluated. Among these rats and mice, transcutaneous breast spectra of hairless SB mice have the best signal-to-noise ratio and are close to the breast spectra reported in literature [6–8, 24], the major Raman bands being C=O band of esters,  $\delta\text{CH}_2$

bend, two features in amide III, and sharp amide I. This suggests that hair may interfere with transcutaneous breast spectra acquisition and, hence, hairless mice would be more suitable for such studies. Further, to confirm the effect of hair on spectra, hair was shaved off from the breast region of SD rats and spectra were recorded transcutaneously. Better signal-to-noise ratio was observed in typical raw spectrum of hairless mice compared to their shaved and unshaved rat counterparts (Fig. 2(c.1)). The difference between shaved rats and genetically hairless SB mice may be due to roots of hair present in shaved rats, which are absent in genetically hairless mice. Despite high background, corrected mean spectrum of breast post-shaving shows spectral similarity to that of genetically hairless mice spectrum, whereas hairy SD rats show several additional spectral features which could be due to contribution of hair (Fig. 2(c.2)). This further supports the possibility that hair interferes with spectra acquisition. Further, as shown in subpanels d.1 and d.2 of Fig. 2, transcutaneous breast spectrum of genetically hairless SB mice resembles exposed breast



**Fig. 4** PC-LDA scatterplot of factors 1 and 2 for analysis of anatomical sites' spectra from hairless SB mice. Discrete clusters are observed for breast, shin tail, and scalp, while neck, spine, thigh, cheek, and chest show misclassification

spectrum. The standard deviations along with mean of spectra from hairy mice, hairy rats, shaved rats, and SB mice are shown in Fig. 3a–d.

Overall, spectra of hairless SB mice gave the best signal-to-noise ratio among their black-haired, brown-haired, white-haired, and shaved counterparts. This suggests that SB mice are better suited for transcutaneous *in vivo* Raman spectroscopy of breast.

#### *Evaluation of distinctness of transcutaneous breast spectra from other anatomical sites*

In the previous section, feasibility of acquiring transcutaneous breast spectra was demonstrated. The spectra show predominance of lipids. However, skin also contains lipids [25]. Therefore, the transcutaneously recorded breast spectra may have (a) maximum contribution from skin with little contribution from breast, (b) approximately equal contribution from skin and breast/in some ratio, and (c) maximum contribution from breast with minimum influence of skin. It is important to establish the origin of spectra for breast-related studies, since if (a) is true, transcutaneous breast studies cannot be carried out, while if (b) is true, sufficient information may not be acquired.

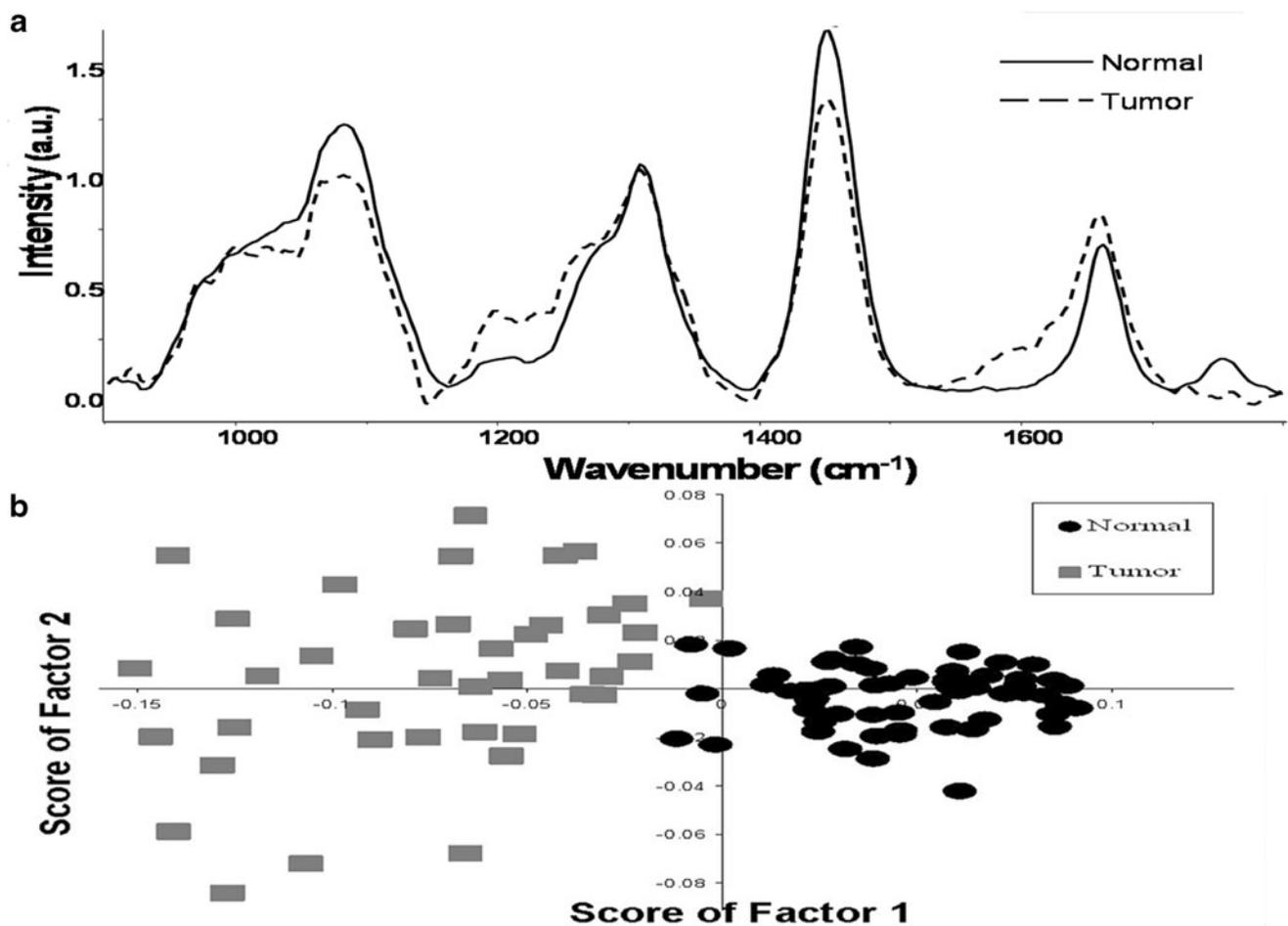
To address this, spectra were acquired transcutaneously from breast, scalp, cheek, neck, chest, thigh, shin, spine, and tail of SB mice. The mean spectra and standard deviation of all sites interpolated in the range 900–1800  $\text{cm}^{-1}$  are shown in Fig. 3e–l. A sharp phosphate band indicative of bone

is observed in shin and tail, while lower intensities of the same are observed in cheek and scalp. Predominant lipid features indicated by strong band corresponding to  $\text{CH}_2$  deformation,  $\text{CH}_2$  twisting, and stretching of  $\text{C}=\text{C}$  and  $\text{C}=\text{O}$  are seen in breast and to a lesser extent in neck and spine. Muscle signatures such as amide III,  $\text{CH}_2$  twisting, deformation of  $\text{CH}_2$ ,  $\text{CH}_3$ , and amide I are seen in cheek, spine, neck, chest, and thigh spectra. Tail spectrum consists of collagen-specific features like broad amide I, amide III band indicative of protein backbone, and protein-specific bands of  $\text{CH}_2$  and  $\text{CH}_3$  deformation.

To explore the differences between these spectra, PC-LDA was used. The scatter plot of PC-LDA factors 1 and 2, shown in Fig. 4, suggests classification of breast, shin, scalp, and tail. Spectra of other sites overlap. Confusion matrix for PC-LDA model and leave-one-out cross-validation (LOOCV) is shown in Table 1. Out of 63 breast spectra, 59 could be correctly identified, while 4/63 spectra misclassified with spine. No breast spectra misclassified with other anatomical sites, although all anatomical sites are covered with skin. This suggests that the influence of lipids on breast spectra is not contributed by skin lipids but by lipids in the mammary fat pad. Thus, the possibility (c) is confirmed, suggesting that useful information regarding breast and breast-associated physiological or pathological changes can be gleaned using transcutaneous *in vivo* RS.

**Table 1** PC-LDA of anatomical sites—confusion matrix for (a) model and (b) leave-one-out cross-validation (diagonal elements represent true positive predictions)

Sites (no. of spectra)	Shin	Breast	Cheek	Chest	Scalp	Spine	Tail	Thigh	Neck	Classification efficiency (%)
<b>a</b>										
Shin (58)	49	0	2	0	3	0	4	0	0	84
Breast (63)	0	59	0	0	0	4	0	0	0	94
Cheek (60)	0	0	29	5	8	5	0	10	3	48
Chest (60)	0	0	7	15	0	6	0	8	24	25
Scalp (60)	1	0	8	1	48	0	0	2	0	80
Spine (61)	0	9	9	3	1	25	0	7	7	41
Tail (60)	1	0	0	0	0	2	57	0	0	95
Thigh (66)	0	12	6	8	6	20	0	5	9	8
Neck (62)	0	3	11	8	2	15	0	4	19	31
<b>b</b>										
Shin (58)	48	0	2	0	3	0	5	0	0	83
Breast (63)	0	59	0	0	0	4	0	0	0	94
Cheek (60)	0	0	27	7	8	5	0	10	3	45
Chest (60)	0	0	7	15	0	6	0	8	24	25
Scalp (60)	1	0	8	1	48	0	0	2	0	80
Spine (61)	0	9	9	3	1	25	0	7	7	41
Tail (60)	1	0	0	0	0	2	57	0	0	95
Thigh (66)	0	12	6	8	6	20	0	5	9	8
Neck (62)	0	3	11	11	2	15	0	4	16	26



**Fig. 5** **a** Mean corrected spectra recorded transcutaneously from normal breast and transplanted breast tumors of SB mice. **b** PC-LDA scatterplot of factors 1 and 2 showing discrete clusters of transcutaneously recorded normal and tumor breast spectra

Secondly, it is important to ascertain that spectra of breast are influenced by functional mammary tissue in addition to lipids in mammary fat pad. Breast consists of the mammary fat pad, connective tissues, fibroblasts, and branched ductal network of mammary epithelium [26]. The mammary epithelium is the functional part of the breast which forms milk glands during lactation and is the site of breast cancer. The mammary fat pad and connective tissue provide support to the functional breast tissue. Thus, the breast spectra should have contributions from mammary epithelium in addition to lipids of mammary fat pad. To address this, breast spectra were compared with spectra of neck, spine, and cheek. These sites have lipids in the form of adipose deposits. If lipids are the only contributing factor, breast should misclassify with these lipid rich sites. However, as seen in Table 1, breast spectra do not match neck and cheek spectra, while only 4/63 misclassify with spine. Uniqueness of breast spectra from neck, spine, and cheek possibly indicates contribution of mammary tissue-specific signals along with mammary fat pad to the breast spectra.

The study of anatomical sites provides further insights into the sensitivity of RS to biochemical components. The sites that can be uniquely identified by transcutaneous in vivo RS are breast, shin, tail, and scalp, their classification efficiencies being 94, 83, 95, and 80 %, respectively. These sites have distinct biochemical composition: breast rich in

**Table 2** PC-LDA of breast normal and tumor spectra—confusion matrix for (a) model and (b) leave-one-out cross-validation (diagonal elements represent true positive predictions)

	Normal	Tumors	Classification efficiency (%)
<b>a</b>			
Normal	62/63	1	98
Tumors	0	40/40	100
<b>b</b>			
Normal	62/63	1	98
Tumors	0	40/40	100

lipids and proteins, tail rich in collagen, shin rich in phosphates, and scalp rich in phosphates and proteins. These results further augment the evidence for chemical sensitivity of transcutaneous *in vivo* Raman spectroscopy.

It may be argued that comparison of skin, transcutaneous breast, and exposed breast spectra may suffice to evaluate the contribution of biochemical components from skin rather than an elaborate experiment with different anatomical sites. However, skin covers every anatomical site in the body; thus, skin spectra will always have influence from subcutaneous anatomical sites. Recording pure skin spectra *in vivo* is therefore very difficult. *Ex vivo* skin spectra may not provide an ideal comparison. Since the applicability of this model for breast cancer studies hinges on recording correct transcutaneous breast spectra, irrefutable proof of breast spectra origin was established by analysis of different anatomical sites.

*Classification of spectra from normal breast and breast tumors* To validate the applicability of SB mice in breast cancer detection, frank breast tumors from ICRC mice were transplanted subcutaneously in inguinal breast of SB mice. Spectra were then recorded transcutaneously from these tumors. The mean tumor spectra and standard deviation are shown in Fig. 3m. Comparison of mean normal breast and mean tumor spectra is illustrated in Fig. 5a. Broad amide I and change in C=O ester band compared to normal breast are characteristics of tumor spectra. This can be attributed to increase in proteins and loss of lipid in tumors, corroborating earlier reports [6–9]. To further explore these differences, PC-LDA was carried out. The scatterplot of factors 1 and 2 shows discrete clusters of normal and tumor spectra (Fig. 5b). The confusion matrix of PC-LDA model and LOOCV is shown in Table 2. Out of 40 spectra, 40 are correctly classified as tumor; 62/63 normal spectra are correctly classified, while only 1/63 spectra misclassified with tumor. The classification efficiency for normal and tumor is 98 and 100 %, respectively. The discriminating efficiency between transcutaneous normal and tumor with SB mice model is better than a previously reported model, where efficiency was 73 % [27].

## Conclusion

The aim of the study was to determine a suitable model for convenient transcutaneous breast spectra measurements enabling breast cancer studies. Results suggest that breast of white-haired animals give good spectra, but better spectra can be acquired from hairless mice. A study of different anatomical sites was carried out to evaluate the contribution of skin and functional mammary epithelium in breast

spectra. The origin of spectra was confirmed to be from breast, by showing that transcutaneous breast spectra were classified with 94 % efficiency from other anatomical sites, although all of them are covered with skin. To prove the applicability of this model for cancer studies, we show that transcutaneous normal breast spectra can be classified from transcutaneous breast tumor spectra with 99 % efficiency, which is higher than previous reports. Thus, SB mice model may be the best suited model for transcutaneous *in vivo* Raman spectroscopic studies of breast physiology and pathology, especially cancer progression. Prospectively, in addition to cancer progression studies, breast to bone metastasis can also be studied *in vivo*, since these anatomical sites can be uniquely classified.

**Acknowledgments** The authors would like to acknowledge Mr. Pramod N. Tawde for his help in animal handling and spectra acquisition. The authors PK and TB would like to acknowledge the Council of Scientific and Industrial Research and ACTREC, Tata Memorial Center, respectively, for the fellowship. The Raman spectrometer employed in the study was procured from DBT project BT/PRI11282/MED/32/83/2008, entitled “Development of *in vivo* laser Raman spectroscopy methods for diagnosis of oral precancerous and cancerous conditions,” Department of Biotechnology, Government of India.

## References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. *CA Cancer J Clin* 61(2):69–90
2. Howlader N (ed) (2001) SEER cancer statistics review. National Cancer Institute, Bethesda, MD, pp 1975–2009
3. Ismail SM, Colclough AB, Dinnen JS, Eakins D, Evans DM, Gradwell E, O'Sullivan JP, Summerell JM, Newcombe RG (1989) Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia. *BMJ* 298:707–710
4. Sørensen JB, Hirsch FR, Gazdar A, Olsen JE (1993) Interobserver variability in histopathologic subtyping and grading of pulmonary adenocarcinoma. *Cancer* 71:2971–2976
5. Evers DJ, Hendriks B, Lucassen G, Ruers T (2012) Optical spectroscopy: current advances and future applications in cancer diagnostics and therapy. *Future Oncol* 8(3):307–320
6. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS (2005) Diagnosing breast cancer by using Raman spectroscopy. *Proc Natl Acad Sci USA* 102(35):12371–12376
7. Chowdary MV, Kumar KK, Mathew S, Rao L, Krishna CM, Kurien J (2009) Biochemical correlation of Raman spectra of normal, benign and malignant breast tissues: a spectral deconvolution study. *Biopolymers* 91:539–546
8. Chowdary MV, Kumar KK, Kurien J, Mathew SP, Krishna CM (2006) Discrimination of normal, benign and malignant breast tissues by Raman spectroscopy. *Biopolymers* 83(5):556–569
9. Haka AS, Volynskaya Z, Gardecki JA, Nazemi J, Lyons J, Hicks D, Fitzmaurice M, Dasari RR, Crowe JP, Feld MS (2006) *In vivo* margin assessment during partial mastectomy breast surgery using Raman spectroscopy. *Cancer Res* 15(6):3317–3322
10. Nieva C, Marro M, Santana-Codina N, Rao S, Petrov D (2012) The lipid phenotype of breast cancer cells characterized by Raman

- microspectroscopy: towards a stratification of malignancy. *PLoS One* 7(10):e46456
11. Clare SE, Neubauer H, Schuetz C, Kurek R, Solomayer E, Dr W, Dye DF, Zaleski JM, Goulet RJ, Fehm T (2006) Raman spectroscopy to distinguish progression stages in breast cancer. *J Clin Oncol* 24(18S):10619
  12. Matulka LA, Wagner K (2005) Models of breast cancer. *Drug Discov Today: Dis Models* 2(1):1–6
  13. Wagner K (2004) Models of breast cancer: quo vadis, animal modeling? *Breast Cancer Res* 6:31–38
  14. Bhisey RA, Veturkar PL, Borges A (1987) Comparison of sensitivity of Swiss albino and hairless Swiss bare mice to two stage skin tumorigenesis. *Indian J Exp Biol* 25(2):90–96
  15. Pai SR, Save PG (1976) Influence of high dose of estradiol on ICRC mouse bearing transplanted mammary tumors. *Indian J Med Res* 64(12):1788–1792
  16. Sluysers M, Nie RV (1974) Estrogen receptor content and hormone-responsive growth of mouse mammary tumors. *Cancer Res* 34:3253–3257
  17. Singh SP, Deshmukh A, Chaturvedi P, Krishna CM (2012) Raman spectroscopy in head and neck cancers: towards oncological applications. *J Cancer Res Ther* 8:S126–132
  18. InPhotonics Inc.. RamanProbe. <http://www.inphotonics.com/probes.htm>. Accessed 16 March 2013
  19. Gobinet C, Vrabie V, Manfait M, Piot O (2009) Preprocessing methods of Raman spectra for source extraction on biomedical samples: application on paraffin-embedded skin biopsies. *IEEE Trans Biomed Eng* 56(5):1371–1342
  20. Sahu A, Deshmukh A, Ghanate AD, Singh SP, Chaturvedi P, Krishna CM (2012) Raman spectroscopy of oral buccal mucosa: a study on age-related physiological changes and tobacco-related pathological changes. *Technol Cancer Res Treat* 11(6):528–623
  21. Nijssen A, Maquelin K, Caspers PJ, Schut TCB, Neumann MHA, Puppels GJ (2007) Discriminating basal cell carcinoma from perilesional skin using high wave-number Raman spectroscopy. *J Biomed Opt* 12:034004–1–034004–7
  22. Koljenović S, Choo-Smith LP, Schut TCB, Kros JM, van den Berge HJ, Puppels GJ (2002) Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy. *Lab Invest* 82:1265–1277
  23. Ghanate AD, Kothiwale S, Singh SP, Bertrand D, Krishna CM (2011) Comparative evaluation of spectroscopic models using different multivariate statistical tools in a multicancer scenario. *J Biomed Opt* 16(2):1–9
  24. Kast RE, Serhatkulu GK, Cao A, Pandya AK, Dai H, Thakur JS, Naik VM, Naik R, Klein MD, Auner GW, Rabah R (2008) Raman spectroscopy can differentiate malignant tumors from normal breast tissue and detect early neoplastic changes in a mouse model. *Biopolymers* 89(3):235–41
  25. Wilkinson DI, Karasek MA (1966) Skin lipids of a normal and a mutant (asebic) mouse strain. *J of Invest Dermatol* 47:449–455
  26. Richert MM, Schwertfeger KL, Ryder JW, Anderson SM (2000) An atlas of mouse mammary gland development. *J Mammary Gland Biol Neoplasia* 5(2):227–241
  27. Raniero L, Canevari RA, Ramalho LNZ, Ramalho FS, dos Santos EAP, Bitar RA, Jalkanen KJ, Martinho HS, Martin AA (2011) In and ex vivo breast disease study by Raman spectroscopy. *Theor Chem Acc* 130(4):1239–1247

## FULL ARTICLE

**Raman spectroscopy of serum: A study on ‘pre’ and ‘post’ breast adenocarcinoma resection in rat models**

Tanmoy Bhattacharjee<sup>1</sup>, Aarif Khan<sup>1</sup>, Piyush Kumar<sup>1</sup>, Arvind Ingle<sup>2</sup>, Girish Maru<sup>3</sup>,  
and C. Murali Krishna<sup>\*, 1</sup>

<sup>1</sup> Chilakapati Lab, ACTREC, Tata Memorial Centre, Navi Mumbai, India

<sup>2</sup> Laboratory Animal Facility, ACTREC, Tata Memorial Centre, Navi Mumbai, India

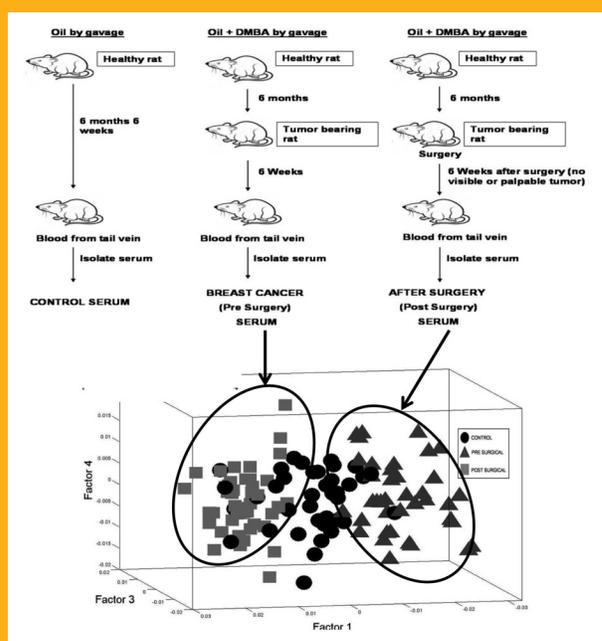
<sup>3</sup> Maru Lab, ACTREC, Tata Memorial Centre, Navi Mumbai, India

Received 7 April 2014, revised 15 May 2014, accepted 16 June 2014

Published online 15 July 2014

**Key words:** Raman spectroscopy, animal model, serum, tumor resection surgery, pre and post surgical resection, breast adenocarcinomas, PC-LDA

Risk of recurrence is a major problem in breast cancer management. Currently available prognostic markers have several disadvantages including low sensitivity and specificity, highlighting the need for new prognostic techniques. One of the candidate techniques is serum-based Raman spectroscopy (RS). In this study, feasibility of using RS to distinguish ‘pre’ from ‘post’ breast tumor resection serum in rats was explored. Spectral analysis suggests change in proteins and amino acid profiles in ‘post’ compared to ‘pre-surgical’ group. Principal-Component-Linear-Discriminant-Analysis shows 87% and 91% classification efficiency for ‘pre’ and ‘post-surgical’ groups respectively. Thus, the study further supports efficacy of RS for theranostic applications.

**1. Introduction**

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death in females worldwide, accounting for 23% (1.38 million) of the

total new cancer cases and 14% (458,400) of the total cancer deaths in 2008 [1]. Despite advances in breast cancer therapy and management, the risk of

\* Corresponding author: e-mail: mchilakapati@actrec.gov.in, pittu1043@gmail.com, Telephone: +91-22-2740 5039

recurrence due to residual tumor remains 20–30% [2, 3]. Currently used prognostic factors are axillary lymph nodal status, tumor size, tumor histologic grade, lymphatic and vascular invasion, proliferative markers, Estrogen Receptor (ER)/progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) status [4, 5]. However, these clinicopathological parameters remain imperfect prognostic classifiers partly owing to tumor heterogeneity. Molecular gene expression and protein markers have been studied rigorously as prognostic determinators; but have been shown to have low sensitivity and specificity [6, 7]. Holistic approaches such as genomics and proteomics that analyze multiple genes/proteins simultaneously have shown encouraging results and may help better prognosis prediction compared to single markers [8, 9]. However, tedious sample preparation and long output times are some disadvantages of the above. Rapid techniques requiring least/no sample preparation that can help multi-marker studies are hence being extensively explored.

RS, a rapid, objective technique that requires little sample preparation, is based on an inelastic scattering process where the energy of photon scattered by the sample is different from the incident photon due to transfer of energy to or from vibrational modes of molecules in the sample [10]. Since bands of Raman spectrum are characteristic of specific molecular vibrations unique to a molecule, RS can provide chemical fingerprint/biochemical profile of a sample. RS has shown promising results in diagnosis of cervix, lung, gastrointestinal, brain, oral, skin, colon [11–17] and several other cancers [18] including breast cancers [19–24]. The ability of this technique to classify normal breast tissues from benign and malignant tissues has been established *ex-vivo* [25]. Feasibility of distinguishing normal breast from breast tumors in rats has been demonstrated both *ex vivo* and *in vivo* [26, 27]. Spectroscopic distinction between metastatic and non metastatic cell lines has been shown [28]. Studies to evaluate possibility of distinguishing invasive carcinoma and ductal *in situ* carcinoma have been attempted using cryopreserved sections [29]. These studies demonstrate sensitivity of this technique to subtle changes in biochemical profile. Since subtle biochemical changes may shed light on possible treatment outcome, RS may help in prognostic applications.

RS has been used to study effect of drugs [30], radiation in cancer cell lines [31] as well as predict treatment response in cervical [32] and lung cancers [33]. Haka et al. have demonstrated *in vivo* margin assessment during breast cancer surgery. Significantly, they detected a grossly invisible cancer that, upon pathologic review, required the patient to undergo a second surgical procedure [34], suggesting ability of RS to identify residual tumors. However, use of RS on inaccessible organs like breast requires

invasive operative procedures. Serum based RS circumvents this problem. Apart from minimal invasiveness, serum/other body fluids based tests have several advantages like accessibility, multiple sampling, easy handling, storage and transportation. The ability of RS to classify normal and cancer serum has been demonstrated [35–39]. Possibility of serum based asthma treatment response monitoring using RS has been reported [40]. However, no studies have reported spectral patterns of serum pre and post surgical resection of tumors. Such studies may help explore spectral markers for breast cancer prognosis.

The current study aims to explore ability of RS to distinguish pre and post-surgical resection serum. In this study, spectra were acquired from serum of tumor bearing rats (pre surgical group), serum of rats after surgical resection of tumor (post surgical group) and control rats and analyzed using PC-LDA. Results of the study are presented and discussed in this manuscript.

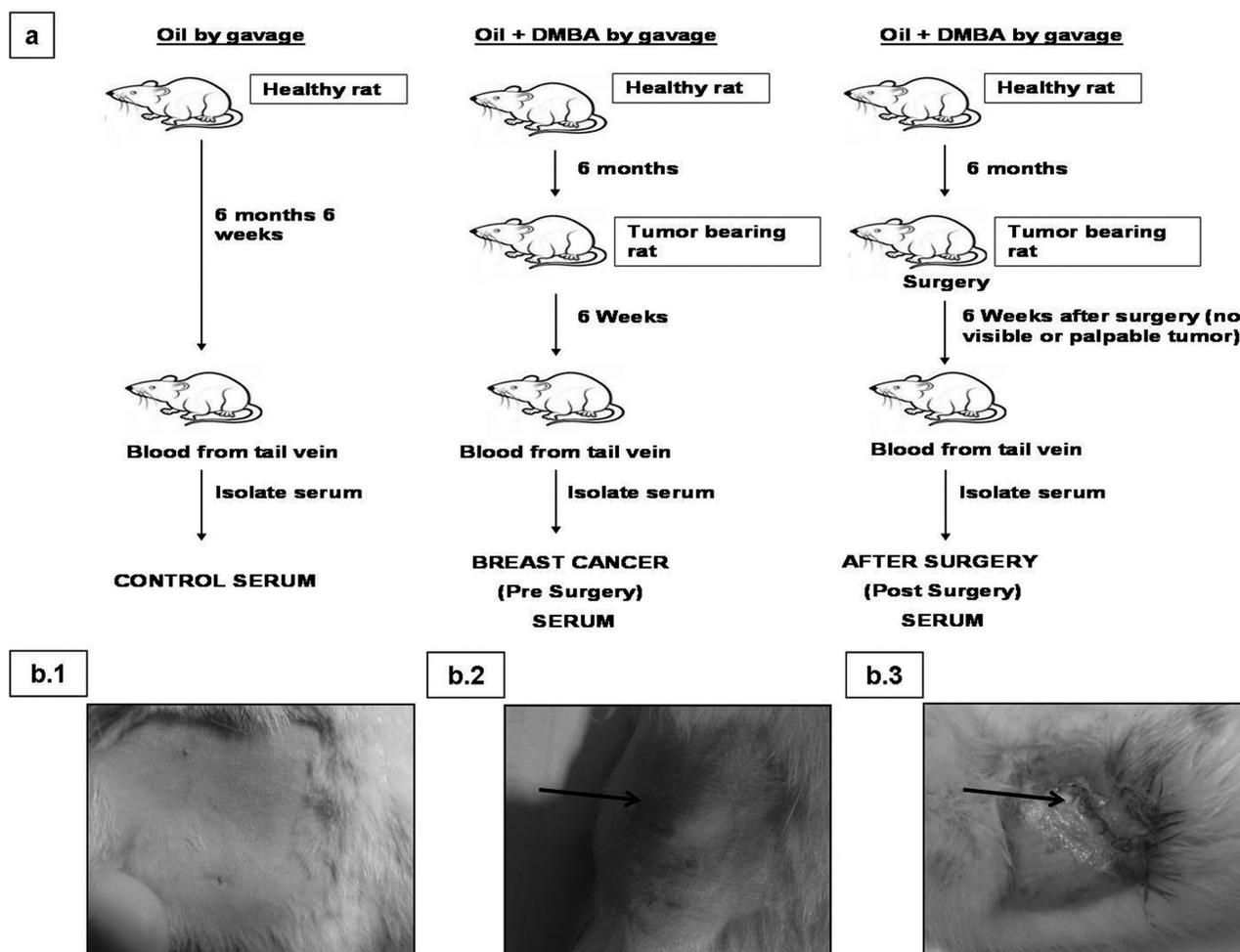
## 2. Materials and methods

### 2.1 Animals

A total of 16 Sprague-Dawley (SD) rats were used in this study. Eleven 50 days old SD rats were administered 65 mg/kg 7, 12 dimethyl benzanthracene (DMBA, Sigma-Aldrich, purity 95%) dissolved in groundnut oil (Dhara, India) intragastrically by gavage. All 11 rats developed breast tumors (histopathologically confirmed adenocarcinoma of breast) six months post induction. Five 50 day old SD rats were administered oil (control). None of the control rats developed breast tumors. The study was approved by Institutional Animal Ethics Committee, ACTREC endorsed by the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA), Government of India. All animals were housed under standard laboratory conditions, fed a diet of in-house-prepared pellets and provided with water *ad libitum*. The protocol employed in the study has been depicted in Figure 1a.

### 2.2 Tumor excision surgery

Tumors of tumor bearing rats were surgically excised by an expert veterinarian (Figure 1b). The rats were then checked for recurrence regularly for six weeks. Six rats were visually and palpably normal six weeks post surgery and were further used for blood collection (post surgery samples). Collection of blood six weeks post surgery ensures minimal in-



**Figure 1** Study protocol (a) induction of carcinogenesis, surgery and serum collection for control, 'pre-surgical' and 'post-surgical' groups (b) pictures of rat inguinal breast for control (b.1), pre-surgical (b.2) and post-surgical (b.3).

fluence of inflammatory, surgical trauma and tissue loss associated changes in serum, since these changes do not persist for more than 5–7 days [41–44]. No additional drugs or treatment were administered to these rats, except topical application of 5% Povidone Iodine (Wockhardt Health Care) over surgical wound to prevent infection. Studies have suggested minimal systemic absorption of Povidone Iodine on topical application [45]. Blood was also collected from age matched control and tumor bearing rats. One serum sample was collected for each rat and used for spectroscopy.

### 2.3 Blood collection and serum separation

Approximately 1.5 ml blood was collected from tail vein of each rat using scalp vein (Top winged infusion set, 22G). Samples were placed standing for 30

minutes to allow clot formation and then centrifuged at 3500 rpm for 10 minutes. 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  $-80^{\circ}\text{C}$  till use. Care was taken to avoid hemolysis. Using this procedure, serum was acquired from 5 control, 5 tumor bearing (pre surgery) and 6 surgically operated (post surgery) rats 6 weeks after successful surgery with no palpable and visible recurrence. Thus, samples were collected from age matched rats. Serum from same rats before and after surgery could not be acquired for the following reasons – hemolysis of sample collected before surgery; inflammation, recurrence or detection of palpable tumor in breast not under study after surgery; or hemolysis of sample collected after surgery. To avoid adding unknown variables, the study was strictly restricted to serum of rats which a) showed no hemolysis b) had only one tumor c) no new tumor appeared during the duration of the

study, either visibly or palpably d) did not show inflammation post surgery.

## 2.4 Spectra acquisition

After passive thawing, samples were subjected to Raman spectroscopy by placing 50  $\mu\text{l}$  serum on calcium fluoride ( $\text{CaF}_2$ ) window and spectra were recorded using Fiber Optic Raman microprobe (Horiba-Jobin-Yvon, France). This Raman system consists of 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 ( $\lambda_{\text{ex}}$ ) = 785 nm, laser power = 40 mW. Spectra were integrated for 10 seconds and averaged over 6 accumulations. On an average, 8 spectra were recorded from each sample to generate a total of 128 spectra under 3 groups, 40 spectra from control rats, 39 from tumor bearing rats (pre surgery) and 48 spectra from surgically operated rats (post surgery).

## 2.5 Spectral pre-processing

Raman spectra from all serum samples were corrected for Charge coupled device (CCD) response with a National Institute of Science and Technology (NIST) certified Standard Reference Material 2241 (SRM 2241) followed by the subtraction of background signals from optical elements and substrate. To remove interference of the slow moving background, first derivatives of spectra (Savitzky-Golay method and window size 3) were computed [46, 48]. Spectra were interpolated in 800–1800  $\text{cm}^{-1}$  region, vector-normalized and used as input for multivariate analysis.

## 2.6 Multivariate analysis

First derivative, vector normalized spectra were subjected to multivariate unsupervised Principal Component Analysis (PCA) and supervised Principal

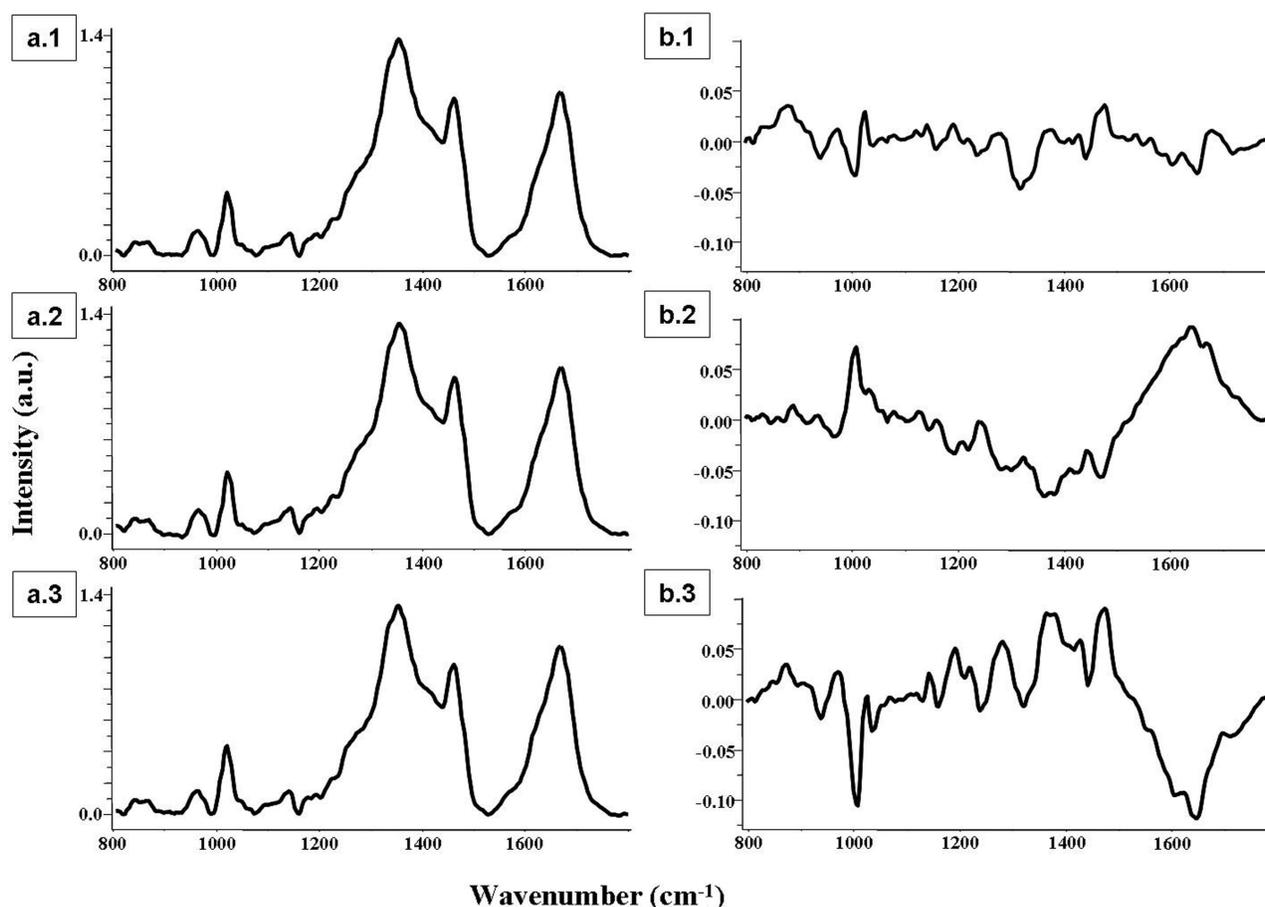
Component-Linear Discriminant Analysis (PC-LDA). PCA is a routinely used method for data compression and visualization while LDA provides data classification based on an optimized criterion which is aimed for better class separation. LDA can be used in companion with PCA to increase efficiency of classification. For this, PCA scores obtained using a set of significant 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. LDA models were validated by Leave-one-out cross-validation (LOOCV). LOOCV is a type of rotation estimation used mainly for smaller datasets i.e. a technique useful for assessing performance of a predictive model with a hypothetical validation set when an explicit validation set is not available. Algorithms for these analyses were implemented in MATLAB (Mathworks Inc., USA) based software using in-house codes [49].

Mean spectra were computed from the background subtracted spectra prior to derivatization for each class, by averaging  $Y$ -axis variations keeping  $X$ -axis constant for each class, and baseline corrected by fitting a fifth order polynomial function. These baselines corrected spectra were vector normalized and then used for computing mean spectra. Difference spectra were also calculated by subtracting mean spectra of control group from 'pre' and 'post surgical' group.

## 3. Results and discussion

### 3.1 Spectral analysis

Vector-normalized average spectra of control (a.1), pre surgery (a.2) and post surgery (a.3) are presented in Figure 2. As is evident from the figure, contributions of proteins, DNA and amino acids like tyrosine, tryptophan, and phenylalanine 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. To elucidate the spectral variations amongst groups, difference spectra were computed by subtracting mean control spectrum from mean 'pre surgical' group and mean 'post surgical' group spectra, respectively (Figure 1b). Subtraction of mean spectra is one of the conventional ways of looking at spectral differences, it provides differences over a selected spectral range and thus understanding of the moieties that may have been modified is facilitated. The difference 'pre surgical' group spectrum (pre surgery-control) is presented in Figure 2b.1, where the positive peaks are

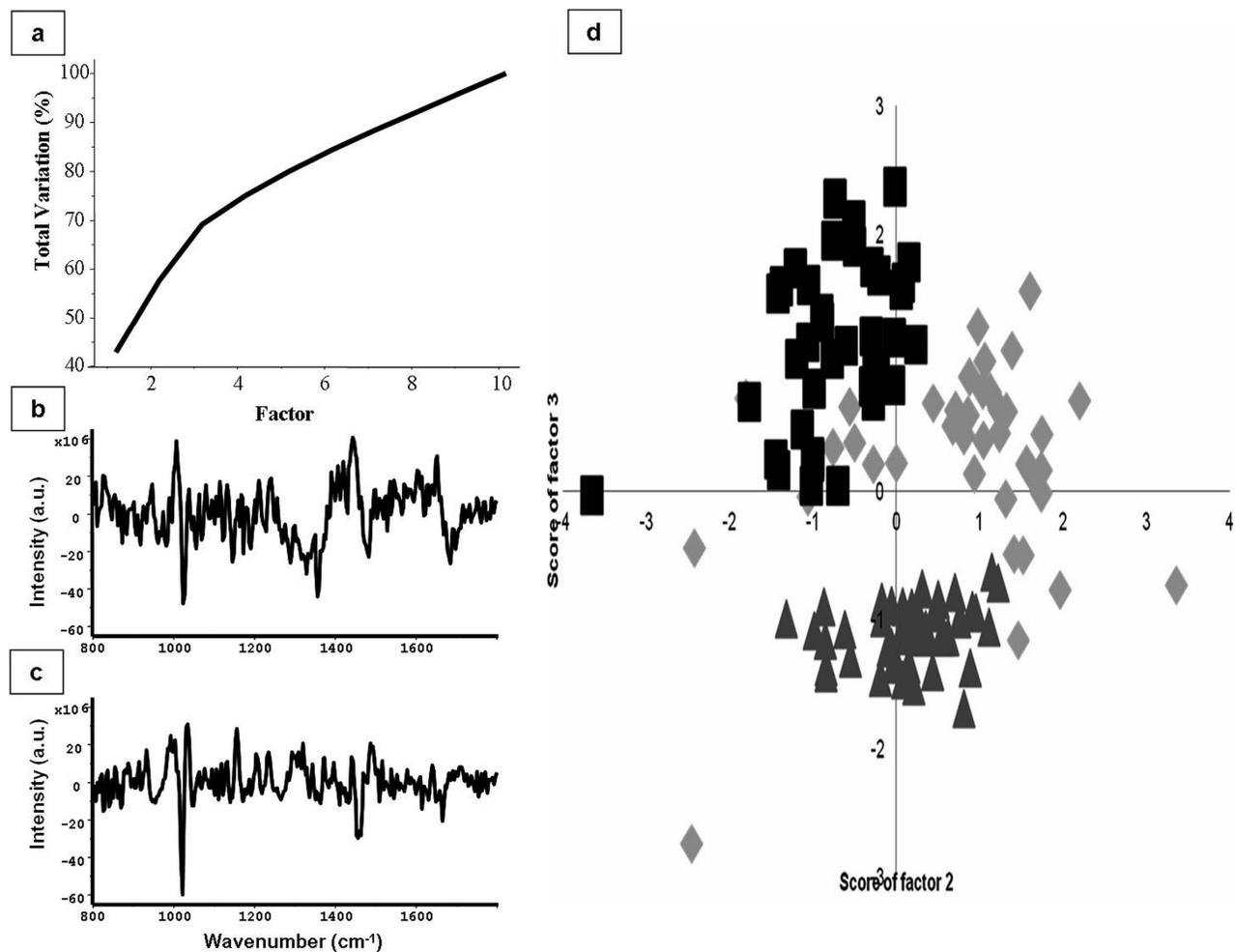


**Figure 2** Mean and Difference spectra (a) mean spectra of control (a.1), pre-surgical (a.2) and post-surgical (a.3) (b) difference spectra; pre-surgical – control (b.1), post-surgical – control (b.2) and pre-surgical – post-surgical (b.3).

from the mean 'pre surgical' group spectrum while negative peaks are from mean control spectrum. Positive peaks of proteins amide I ( $\sim 1675\text{ cm}^{-1}$ ) and amide III ( $\sim 1270\text{ cm}^{-1}$ ) and amino acids Tyr and Trp (doublet at  $830\text{ cm}^{-1}$ ,  $850\text{ cm}^{-1}$ ) can be seen in 'pre surgical' group. Negative peak of phenylalanine ( $\sim 1002\text{ cm}^{-1}$ ) indicate decreased phenylalanine in 'pre surgical' group serum with respect to control. Decrease in phenylalanine concentration in breast cancer and oral cancer compared to control has been reported previously [50, 51]. In contrast, difference 'post surgical' group spectrum shows strong positive peak ( $\sim 1002\text{ cm}^{-1}$ ) suggesting increased phenylalanine in serum post surgery. Amide III at  $1240\text{ cm}^{-1}$  as opposed to  $1270\text{ cm}^{-1}$  in 'pre surgical' group difference spectrum probably indicates change in protein profile of serum pre and post surgery. To elucidate possible spectral differences between pre and post surgery sera, a difference spectrum was computed by subtracting post from pre surgery spectrum (Figure 2b.3). The strong negative phenylalanine peak ( $\sim 1002\text{ cm}^{-1}$ ) suggests increased phenylalanine post surgery.

### 3.2 Multivariate analysis

Preprocessed interpolated in  $800\text{--}1800\text{ cm}^{-1}$  range spectra were subjected to PCA for delineating trends in the data set. PCA variance plot and loadings are shown in Figure 3a and b. As can be seen in Figure 3a, cumulative variance covered by factor 2 and 3 are 57.6% and 69.1% respectively. Scatter plot of PCA factors (Figure 3b) shows distinct 'pre' and 'post-surgical' clusters which group around control. To explore the feasibility of classifying 'pre-surgical' group and 'post-surgical' group from control, PC-LDA was used. To avoid over fitting, 4 factors [52] contributing  $\sim 89\%$  percent of correct classification; were used (Figure 4a). The plot of PC-LDA factors 1, 3 and 4 (Figure 4b) shows that 'pre-surgical' and 'post-surgical' groups cluster around control but are distinct from each other. The confusion matrix for PC-LDA model building is shown in Table 1(a). Leave-one-out-cross validation (LOOCV) was carried out to evaluate the results obtained by PC-LDA. LOOCV builds a model based on all observations



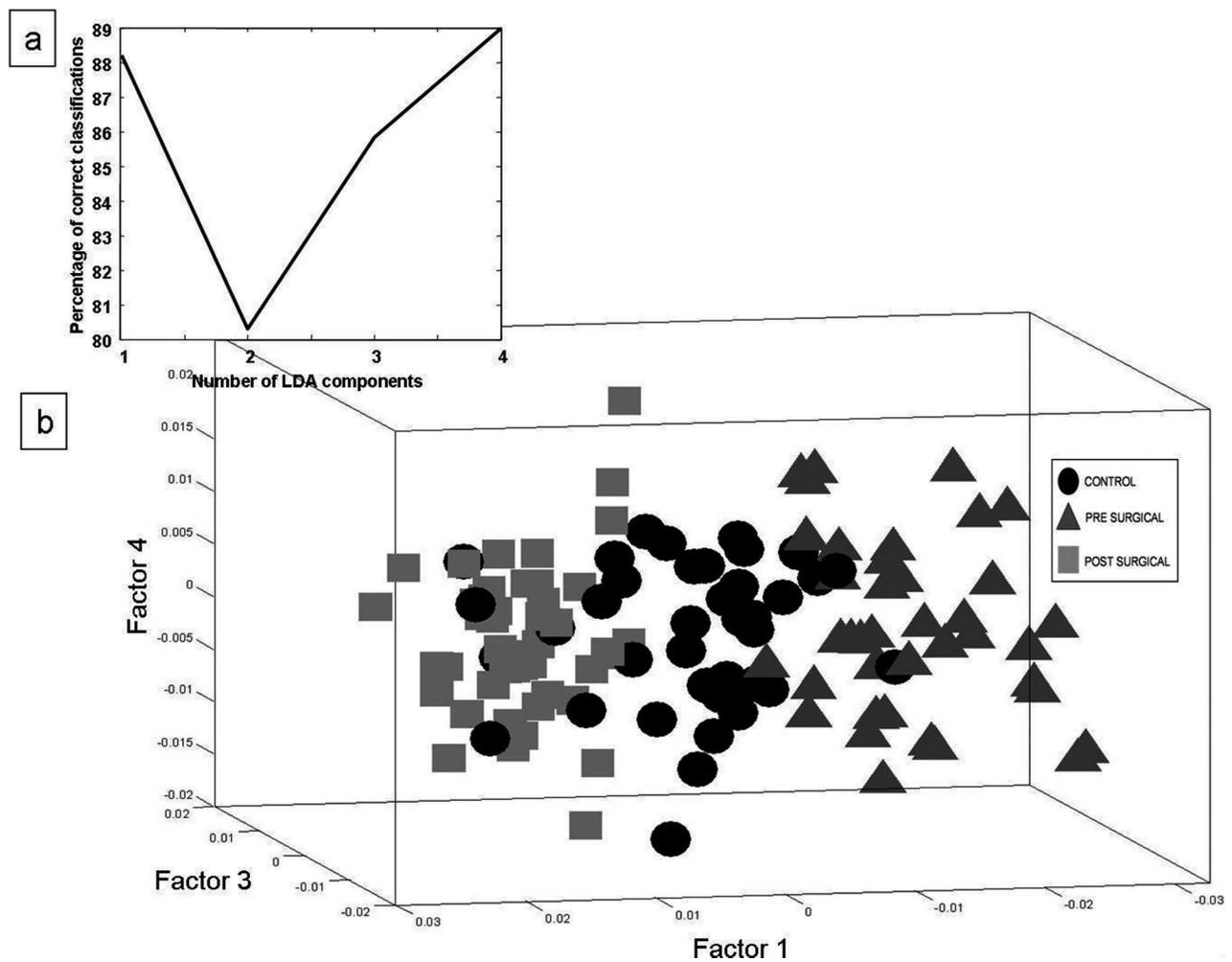
**Figure 3** PCA of control, pre-surgical and post-surgical groups (a) variance plot (b) loading factor 2 (c) loading factor 3 (d) scatter plot of factors 2 and 3.

but one, and tests the left out observation against the model built; this is repeated until all observations are left out once. The performance is estimated in terms of classification efficiency, which is percentage of spectra from each group that are correctly classified. In analysis of LOOCV as shown in Table 1(b); 33/40 spectra correctly classify as control while 34/39 'pre surgical' group spectra are correctly classified. Several studies have shown major protein profile changes in breast cancer serum with respect to control [53, 54]. These changes may explain the observed difference between control and 'pre surgical' group serum. 4/40 control spectra misclassify as 'pre surgical' group whereas 5/39 'pre surgical' group spectra misclassify as control. Although several changes occur during carcinogenesis, not all proteins change. This may explain misclassification between the groups. Further, serum is a colloidal dispersion [55, 56], wherein the microvolume (50  $\mu$ l) being probed is heterogenous at microscopic level. Such a small volume corresponds to limiting concentrations of

serum constituents like proteins, lipids, electrolytes and thus a possibility of irregular distribution of these solutes even after several rounds of mixing. This heterogeneity may also explain the misclassification between these groups.

44/48 'post surgical' group spectra are correctly classified. None of 'post surgical' group spectra misclassify with 'pre surgical' group. Literature suggests increase in several proteins; CSF1, THSB2, IL6, IL7, IL16, FasL and VEGF-B post surgery [57]. These changes may be responsible for classification from control and 'pre surgical' group. 4/48 'post surgical' group spectra misclassify with control. This may be due to loss of tumor associated factors post surgery or serum heterogeneity or both.

The classification efficiency of control, 'pre surgical' group and 'post surgical' group was 82%, 87% and 91% respectively. As discussed earlier, the only difference between 'pre surgical' group and 'post surgical' group is presence or absence of tumor. Thus, the high classification efficiency of 'pre surgi-



**Figure 4** PC-LDA of control, pre-surgical and post-surgical groups (a) scree plot (b) scatter plot of factors 1, 3 and 4.

**Table 1** Confusion matrix for PC-LDA of control, pre-surgical and post-surgical groups: model building (a) and LOOCV (b) (Diagonal elements are true positive predictions and Ex-diagonal elements are false positive predictions).

Model (No. of animals, no. of spectra)	Control	Pre-surgical	Post-surgical
Control (5, 40)	33	4	3
Pre-surgical (5, 39)	3	36	0
Post surgical (6, 48)	4	0	44
LOOCV			
Control (5, 40)	33 (82.5%)	4	3
Pre-surgical (5, 39)	5	36 (87.2%)	0
Post surgical (6, 48)	4	0	44 (91.6%)

cal' group and 'post surgical' group and absence of misclassification amongst these groups suggest a possible role of tumor-associated factors in classification. Spectral identification of tumor-associated factors may help determine risk of recurrence.

#### 4. Conclusion

The aim of the study was to evaluate the feasibility of distinguishing 'pre surgery' serum from 'post surgery'. 5–6 animals were used per group-control, pre surgery and post surgery; to obtain serum. Spectra acquired from each group could be classified with 82%, 87% and 91% classification efficiency, suggesting possibility of distinguishing pre and post surgery serum. However, investigations on effect of tissue loss, minor, moderate and major surgeries on serum and further study with larger sample size are imperative to validate results of the current study. Additionally, changes in serum post tumor resection might suggest possibility of tumor associated or tumor mediated host response factors in play. Therefore, rigorous experiments need to be performed to link tumor associated factors with classification of 'post surgery' serum from 'pre surgery'. Prospectively, the possibility of distinguishing 'pre' from 'post' surgery opens up interesting avenues such as exploration of tumor recurrence prediction, treatment efficacy, etc. using RS.

**Author biographies** Please see Supporting Information online.

#### References

- [1] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, *CA, A Cancer Journal for Clinicians* **61**(2), 69–90 (2011).
- [2] A. M. Gonzalez-Angulo, F. Morales-Vasquez, and G. N. Hortobagyi, *Adv. Exp. Med. Biol.* **608**, 1–22 (2007).
- [3] K. D. Voduc, M. C. Cheang, S. Tyldesley, K. Gelmon, T. O. Nielsen, and H. Kennecke, *J. Clin. Oncol.* **28**, 1684–1691 (2010).
- [4] M. T. Weigel and M. Dowsett, *Endocr. Relat. Cancer* **17**, R2 45–R2 62 (2010).
- [5] F. van der Leij, P. H. Elkhuisen, H. Bartelink, and M. J. van de Vijver, *Semin. Radiat. Oncol.* **22**, 100–107 (2012).
- [6] F. J. Esteva and G. N. Hortobagyi, *Breast Cancer Res.* **6**, 109–118 (2004).
- [7] M. J. Duffy, *Clinical Chemistry* **52**(3), 345–351 (2006).
- [8] M. T. Weigel and M. Dowsett, *Endocrine-Related Cancer* **17**, R245–R262 (2010).

- [9] Y. Baskin and T. Yiğitbaşı, *Curr. Genomics* **11**(7), 528–536 (2010).
- [10] C. V. Raman and K. S. Krishnan, *Nature* **121**, 501–502 (1928).
- [11] V. Chang, P. Cartwright, S. Bean, G. Palmer, R. Bentleyand, and N. Ramanujam, *Neoplasia* **11**(4), 325–332 (2009).
- [12] H. Yamazaki, S. Kaminaka, E. Kohda, M. Mukai, and H. Hamaguchi, *Radiat. Med.* **21**(1), 1–6 (2003).
- [13] S. Teh, W. Zheng, K. Ho, M. Teh, K. Yeoh, and Z. Huang, *Br. J. Cancer* **98**(2), 457–465 (2008).
- [14] Y. Zhou, C. H. Liu, Y. Sun, Y. Pu, S. Boydston-White, Y. Liu, and R. R. Alfano, *J. Biomed. Opt.* **17**(11), 116021 (2012).
- [15] S. P. Singh, A. Deshmukh, P. Chaturvedi, and C. M. Krishna, *J. Biomed. Opt.* **17**(10), 105002 (2012).
- [16] M. Gniadecka, H. C. Wulf, N. N. Mortensen, O. F. Nielsen, and D. H. Christensen, *J. Raman. Spectrosc.* **28**, 125–129 (1997).
- [17] A. Molckovsky, L. Song, M. G. Shim, N. E. Marcon, and B. C. Wilson, *Gastrointest. Endosc.* **57**, 396–402 (2003).
- [18] D. J. Evers, B. H. W. Hendriks, G. W. Lucassen, and T. J. M. Ruers, *Future Oncol.* **8**(3), 307–320 (2012).
- [19] R. R. Alfano, C. H. Liu, W. L. Sha, H. R. Zhu, D. L. Akins, J. Cleary, R. Prudente, and E. Celmer, *Laser Life Sci.* **4**(1), 23–28 (1991).
- [20] A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari, and M. S. Feld, *Proc. Natl. Acad. Sci. USA* **102**(35), 12371–12376 (2005).
- [21] M. V. Chowdary, K. K. Kumar, S. Mathew, L. Rao, C. M. Krishna, and J. Kurien, *Biopolymers* **91**, 539–546 (2009).
- [22] M. D. Keller, E. Vargis, N. Granja, R. H. Wilson, M. A. Mycek, M. C. Kelley, and A. Mahadevan-Jansen, *J. Biomed. Opt.* **16**(7), 07 7006 (2011).
- [23] N. Stone, R. Baker, K. Rogers, A. W. Parker, and P. Matousek, *Analyst.* **132**, 899–905 (2007).
- [24] N. Stone and P. Matousek, *Cancer. Res.* **68**(11), 4424–4430 (2008).
- [25] M. V. Chowdary, K. K. Kumar, J. Kurien, P. S. Mathew, and C. M. Krishna, *Biopolymers* **83**(5), 556–569 (2006).
- [26] R. E. Kast, G. K. Serhatkulu, A. Cao, A. K. Pandya, H. Dai, J. S. Thakur, V. M. Naik, R. Naik, M. D. Klein, G. W. Auner, and R. Rabah, *Biopolymers* **89** (3), 235–241(2008).
- [27] L. Raniero, R. A. Canevari, L. N. Ramalho, F. S. Ramalho, E. A. dos Santos, R. A. Bitar, K. J. Jalkanen, H. S. Martinho, and A. A. Martin, *Theor. Chem. Acc.* **130**(4), 1239–1247 (2011).
- [28] C. Nieva, M. Marro, N. Santana-Codina, S. Rao, and D. Petrov, *PLoS ONE* **7**(10), e46456 (2012).
- [29] S. E. Clare, H. Neubauer, C. Schuetz, R. Kurek, E. Solomayer, D. R. Wallwiene, D. F. Dye, J. M. Zaleski, R. J. Goulet, and T. Fehm, *J. Clin. Oncol.* **24**(18S), 10619 (2006).
- [30] P. N. Goel, S. P. Singh, C. M. Krishna, and R. P. Gude, *J. Innov. Opt. Health Sci.*, in press
- [31] O. Matthews, A. G. Brolo, J. Lum, X. Duan, and A. Jirasek, *Phys. Med. Biol.* **56**, 19 (2011).

- [32] M. S. Vidyasagar, K. Maheedhar, B. M. Vadhiraaja, D. J. Fernandes, V. B. Kartha, and C. M. Krishna, *Biopolymers* **89**(6), 530–537 (2008).
- [33] N. D. Magee, J. R. Beattie, C. Carland, R. Davis, K. McManus, I. Bradbury, D. A. Fennell, P. W. Hamilton, M. Ennis, J. J. McGarvey, and J. S. Elborn, *J. Biomed. Opt.* **15**(2), 026015 (2010).
- [34] A. S. Haka, Z. Volynskaya, J. A. Gardecki, J. Nazemi, J. Lyons, D. Hicks, M. Fitzmaurice, R. R. Dasari, J. P. Crowe, and M. S. Feld, *Cancer. Res.* **15**(6), 3317–3322 (2006).
- [35] A. Sahu, S. Sawant, H. Mamgain, and C. M. Krishna, *Analyst.* **138**(14), 4161–4174 (2013).
- [36] J. L. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-García, R. Huerta-Franco, J. L. González-Trujillo, C. A. Ramírez-Alvarado, G. Gutiérrez-Juárez, and C. Medina-Gutiérrez, *Lasers in Medical Science* **22** (4), 229–236 (2007).
- [37] S. Feng, R. Chen, J. Lin, J. Pan, G. Chen, Y. Li, M. Cheng, Z. Huang, J. Chen, and H. Zeng, *Biosens. Bioelectron.* **25**, 2414–2419 (2010).
- [38] J. L. González-Solís, J. Rodríguez-López, J. C. Martínez-Espinosa, C. Frausto-Reyes, L. F. Jave-Suárez, A. C. Aguilar-Lemarroy, H. Vargas-Rodríguez, and E. Martínez-Cano, *AIP, Conf. Proc.* **1226**, 91–95 (2010).
- [39] X. Li, T. Yang, and S. Li, *Appl. Opt.* **51**, 5038–5043 (2012).
- [40] A. Sahu, K. Dalal, S. Naglot, P. Aggarwal, and C. Murali Krishna, *PLoS One* **8**(11), e78921 (2013).
- [41] J. Desborough, *British Journal of Anaesthesia* **85**(1), 109–117 (2000).
- [42] R. Lakshmi, V. Kartha, C. Murali Krishna, J. Solomon, G. Ulhas, and P. Devi, *Radiat. Res.* **157**, 175–182 (2002).
- [43] B. Chernow, H. R. Alexander, R. C. Smallridge, W. R. Thompson, D. Cook, D. Beardsley, M. P. Fink, C. R. Lake, and J. R. Fletcher, *Arch. Intern. Med.* **147**(7), 1273–1278 (1987).
- [44] A. Lindh, M. Lindholm, L. Holmquist, and L. A. Carlsson, *JPEN J. Parenter. Enteral. Nutr.* **10**(3), 265–273 (1986).
- [45] <http://www.mims.com/India/drug/info/povidone%20iodine/?type=full> 18–210
- [46] A. Nijssen, K. Maquelin, P. J. Caspers, T. C. B. Schut, M. H. A. Neumann, J. C. den Hollander, L. F. Santos, and G. J. Puppels, *J. Biomed. Opt.* **12**, 034004–1–034004–7 (2007).
- [47] S. Koljenovic, L. P. Choo-Smith, T. C. B. Schut, J. M. Kros, H. J. Van den Berge, and G. J. Puppels, *Lab. Invest.* **2**, 1265–1277 (2002).
- [48] T. J. Harvey, E. Gazi, A. Henderson, R. D. Snook, N. W. Clarke, M. Brown, and P. Gardner, *Analyst.* **134**, 1083–1091 (2009).
- [49] A. D. Ghanate, S. Kothiwale, S. P. Singh, D. Bertrand, and C. M. Krishna, *J. Biomed. Opt.* **16**, 025003–1–025003–9 (2011).
- [50] H. S. Lai, J. C. Lee, P. H. Lee, S. T. Wang, and W. J. Chen, *Semin. Cancer Biol.* **15**, 267–276 (2005).
- [51] Y. Miyagi, M. Higashiyama, A. Gochi, M. Akaike, T. Ishikawa, T. Miura, N. Saruki, E. Bando, H. Kimura, F. Imamura, M. Moriyama, I. Ikeda, A. Chiba, F. Oshita, A. Imaizumi, H. Yamamoto, H. Miyano, K. Horimoto, O. Tochikubo, T. Mitsushima, M. Yamakado, and N. Okamoto, *PLoS One* **6**(9), e24143 (2011).
- [52] J. Kelly, J. Trevisan, A. Scott, P. Crmichael, H. Pollock, P. Martin-Hirsch, and F. Martin, *J. Proteome. Res.* **10**, 1437–1448 (2011).
- [53] B. K. Kim, J. W. Lee, P. J. Park, Y. S. Shin, W. Y. Lee, K. A. Lee, S. Ye, H. Hyun, K. N. Kang, D. Yeo, Y. Kim, S. Y. Ohn, D. Y. Noh, and C. W. Kim, *Breast Cancer Research* **11**, R22 (2009).
- [54] M. C. Gast, C. H. V. Gils, L. F. Wessels, N. Harris, J. M. Bonfrer, E. J. Rutgers, J. H. Schellens, and J. H. Beijnen, *Oncol. Rep.* **22**(1), 205–213 (2009).
- [55] D. Nelson and M. Cox, *Lehninger Principles of Biochemistry*, 4th edition, W. H. Freeman, (2008).
- [56] F. A. Bettelheim and W. H. Brown, 8th edition, Cengage learning (2007).
- [57] L. G. Perez-Rivas, J. M. Jerez, C. E. Fernandez-De Sousa, V. de Luque, C. Quero, B. Pajares, L. Franco, A. Sanchez-Muñoz, N. Ribelles, and E. Alba, *Int. J. Oncol.* **41**(6), 2200–2206 (2012).



CrossMark  
click for updates

Cite this: *Analyst*, 2015, 140, 456

# A preliminary Raman spectroscopic study of urine: diagnosis of breast cancer in animal models

T. Bhattacharjee,<sup>†a</sup> A. Khan,<sup>†a</sup> G. Maru,<sup>b</sup> A. Ingle<sup>c</sup> and C. Murali Krishna<sup>\*a</sup>

Prognosis of breast cancer, the most common cancer in females worldwide, has been shown to improve with early detection. Owing to disadvantages like low sensitivity, specificity, tedious sample preparation, long output times and inter-observer variance of currently available screening/diagnostic tools, rapid and objective alternatives such as Raman spectroscopy (RS) are being extensively explored. Body fluid (serum and saliva) based RS assays have shown promising results in diagnosis of oral, lung and nasopharyngeal cancers. The current study aims to explore the feasibility of breast cancer diagnosis using urine based RS. In this study, spectra were acquired from unprocessed as well as concentrated urine of controls (C) and breast tumor bearing (T) rats and analyzed using Principal Component Analysis (PCA) and Principal Component-Linear Discriminant Analysis (PC-LDA). Classification efficiencies of 80% and 72% using unprocessed urine and 78% and 91% using concentrated urine for C and T rats were achieved. Thus, results suggest the possibility of breast cancer diagnosis using urine based RS. Further, spectra were also acquired from concentrated urine samples collected prior to breast tumor development (TT) in rats and from rats that did not develop tumors despite carcinogen treatment (NTT). Concentrated urine of NTT rats could be classified as 'normal' (C or NTT) with ~83% efficiency whereas concentrated urine from visibly and palpably normal rats that eventually developed tumor (TT rats) could be classified as 'abnormal' (TT or T) with ~72.5% efficiency using PC-LDA. These results suggest the possibility of detecting biochemical changes occurring prior to tumor development using urine based RS.

Received 18th September 2014  
Accepted 13th November 2014

DOI: 10.1039/c4an01703j

www.rsc.org/analyst

## 1. Introduction

Breast cancer is the most common and leading cause of cancer deaths amongst females worldwide.<sup>1</sup> In 2012, 226 870 new invasive breast cancer cases and 39 510 breast cancer deaths were estimated to occur in the US.<sup>2</sup> In developing countries like India, incidence rates as high as 39.5 per 100 000 women have been reported.<sup>3</sup> The literature suggests improved prognosis with early detection of breast cancer.<sup>4</sup> *In lieu* of this, efforts have been directed towards development of screening techniques for early detection of this cancer. Clinical breast exam (CBE) and mammography are the two most widely used screening tools.<sup>5,6</sup> However, the sensitivity of CBE is low whereas mammography suffers from disadvantages like low positive predictive values (only 25%),<sup>6</sup> unsuitability for women with dense breast, radiation exposure, *etc.* Alternatives like ultrasonography, thermography, Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET) have low sensitivity, cannot detect

small tumors and are expensive. The gold standard for diagnosis of breast cancer, histopathology, also suffers from several disadvantages like tedious sample preparation, long output times and inter-observer variance.<sup>7,8</sup> Rapid, objective and preferably non-invasive alternate screening/diagnostic techniques are hence being extensively explored.

Raman spectroscopy (RS), a rapid and objective tool with potential for non-invasive/minimally invasive applications, has shown promising results in the diagnosis of cervix, lung, gastrointestinal, brain, oral, skin, colon<sup>9-13</sup> and several other cancers<sup>14</sup> including breast cancers.<sup>15-22</sup> RS is based on an inelastic scattering process where the energy of photons scattered by the sample is different from that of the incident photons due to transfer of energy to or from vibrational modes of molecules in the sample. Since the bands of the Raman spectrum are characteristic of specific molecular vibrations unique to a molecule, RS can provide a chemical fingerprint/biochemical profile of a sample. The ability of this technique to classify normal breast tissues from benign and malignant tissues,<sup>16,17</sup> metastatic from non-metastatic cell lines,<sup>23</sup> and invasive carcinoma from ductal *in situ* carcinoma using cryopreserved sections<sup>24</sup> has been reported. Detection of constituents deep inside breast tissue phantoms<sup>20,21</sup> as well as transcutaneous detection of breast tumors in rats has been established.<sup>25,26</sup> Tumor margin assessment using *in vivo* RS

<sup>a</sup>Chilakapati Lab, Advanced Center for Treatment Research and Education in Cancer (ACTREC), Tata Memorial Center (TMC), Kharghar, Sector '22', Navi-Mumbai, 410210, India. E-mail: mchilakapati@actrec.gov.in; pittu1043@gmail.com; Tel: +91-22-2740-5039

<sup>b</sup>Maru Lab, ACTREC, Kharghar, Navi-Mumbai, 410210, India

<sup>c</sup>Laboratory Animal Facility, ACTREC, Kharghar, Navi-Mumbai, 410210, India

<sup>†</sup> Authors have contributed equally to this work.

during partial mastectomy surgery has also been demonstrated.<sup>22</sup> However, the requirement of invasive procedures for spectroscopy in human subjects remains a problem. The use of body fluid (serum, saliva, and urine) based RS may circumvent this problem. Apart from minimal invasiveness, body fluid based tests have several advantages like accessibility, multiple sampling, easy handling, storage and transportation. Serum based RS to classify normal from cancer serum and monitor asthma<sup>27–32</sup> and saliva based surface enhanced RS assay for diagnosis of oral, nasopharyngeal and lung cancer<sup>33–35</sup> have been reported. Diagnosis of bladder cancer using cells voided into urine has also been demonstrated.<sup>36</sup> However, the feasibility of using urine for breast cancer diagnosis is yet to be explored.

The current study aims to explore the feasibility of breast cancer diagnosis using urine based RS. In this study, spectra were acquired from unprocessed urine of controls and breast tumor bearing rats. To check the possibility of better classification between normal and tumor cells, spectra were also acquired after concentrating (by dehydration in a vacuum) urine of control and tumor bearing rats. In order to further explore the sensitivity of urine based RS in breast cancer diagnosis and the possibility of early detection, spectra were also acquired from urine samples collected prior to breast tumor development in rats. The data were analyzed using Principal Component Analysis (PCA) and Principal Component-Linear Discriminant Analysis (PC-LDA). The results of the study are reported in the manuscript.

## 2. Materials and methods

### 2.1. Animals

A total of 42 Sprague-Dawley (SD) rats were used in this study. 25 fifty day old SD rats were administered with 65 mg kg<sup>-1</sup> 7,12-dimethyl benzantracene (DMBA, Sigma-Aldrich, purity 95%) dissolved in groundnut oil (Dhara, India) intragastrically by gavage. 20 rats developed breast tumors (histopathologically confirmed adenocarcinoma of breast) approximately six months post-carcinogen treatment. 17 fifty day old SD rats were administered with oil (control). None of the control rats developed breast tumors. The study was approved by the Institutional Animal Ethics Committee, ACTREC endorsed by the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA), Government of India. All animals were housed under standard laboratory conditions, fed a diet of in-house-prepared pellets and provided with water *ad libitum*.

### 2.2. Urine collection

The rats were restrained; airlifted and voided urine was collected in sterile Petri dishes. The urine was then transferred to sterile Eppendorf tubes using a micropipette. A separate Petri dish and Eppendorf were used for each rat urine sample. Approximately 150–200  $\mu$ l urine was collected per rat. The urine samples were snap frozen immediately after collection and

stored at  $-80$  °C. Using this procedure, urine was taken from six groups:

(a) Unprocessed control urine ( $n = 9$ ): urine samples of control rats were thawed and spectra were acquired from these samples.

(b) Unprocessed tumor urine ( $n = 9$ ): urine samples of tumor bearing rats were thawed and spectra were acquired from these samples.

(c) Concentrated control urine ( $n = 8$ ): urine samples of control were thawed, dehydrated in a vacuum using a Speed Vac and rehydrated with 40  $\mu$ l normal saline before spectra acquisition.

(d) Concentrated tumor urine ( $n = 7$ ): urine samples of tumor bearing rats were thawed, dehydrated in a vacuum using a Speed Vac and rehydrated with 40  $\mu$ l normal saline before spectral acquisition.

(e) Concentrated TT urine ( $n = 4$ ): urine samples were collected 5 months post-carcinogen (DMBA) treatment from visibly and palpably normal rats. These rats were palpated every two weeks after urine collection. Approximately 1 month post-urine collection ( $\sim 6$  months post-carcinogen treatment), these rats developed breast tumors. Biopsy followed by histopathology confirmed the tumors to be adenocarcinoma. The urine samples collected from these rats were labeled 'Tumor Treated' and will henceforth be referred to as 'TT'. Before spectral acquisition, these samples were dehydrated and rehydrated as described above.

(f) Concentrated NTT urine ( $n = 5$ ): urine samples were collected 5 months post-carcinogen (DMBA) treatment from visibly and palpably normal rats. However, these rats failed to develop tumor even 8 months post-carcinogen treatment. Urine samples from these rats were labeled 'No Tumor Treated' and henceforth are referred to as 'NTT'. The urine was processed in the same way before spectral acquisition.

Urine samples from control, TT, NTT and tumor bearing rats were collected at the same time. Thus, all samples were collected from age matched rats.

### 2.3. Spectral acquisition

After passive thawing/rehydrating, samples were subjected to Raman spectroscopy by placing a 40  $\mu$ l sample on a calcium fluoride (CaF<sub>2</sub>) window and spectra were recorded using a Fiber Optic Raman microprobe (Horiba-Jobin-Yvon, France). This Raman system consists of a laser (785 nm, Process Instruments) as an excitation source and a HE 785 spectrograph (Horiba-Jobin-Yvon, France) coupled with a CCD (Synapse, Horiba-Jobin-Yvon) as dispersion and detection elements, respectively. Optical filtering of the laser line and Rayleigh signals is accomplished through the '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. The Raman microprobe was assembled by coupling a 40 $\times$  microscopic objective (Nikon, Japan) to the superhead. Spectral acquisition details were: excitation wavelength ( $\lambda_{\text{ex}}$ ) = 785 nm and

laser power = 40 mW. Spectra were integrated for 10 seconds and averaged over 6 accumulations. On average, 8 spectra were recorded from each sample to generate a total of 355 spectra under 6 groups, 81 spectra from unprocessed urine of control rats, 82 from unprocessed urine of tumor bearing rats, 64 spectra from concentrated urine of control rats, 56 from concentrated urine of tumor bearing rats, 40 spectra from concentrated urine of TT rats and 32 spectra from concentrated urine of NTT rats.

#### 2.4. Spectral pre-processing

Raman spectra from all urine samples were corrected for instrument response with a National Institute of Science and Technology (NIST) certified Standard Reference Material 2241 (SRM 2241) followed by the subtraction of background signals from optical elements and the substrate. To remove the interference of the low frequency background, first derivatives of spectra (Savitzky-Golay method and window size 3) were computed.<sup>37–39</sup> Spectra were interpolated in the 600–1800  $\text{cm}^{-1}$  region, vector-normalized and used as an input for multivariate analysis.

#### 2.5. Multivariate analysis

First derivative, vector normalized spectra were subjected to multivariate unsupervised Principal Component Analysis (PCA) and supervised Principal Component-Linear Discriminant Analysis (PC-LDA). PCA is a routinely used method for data compression and visualization while LDA provides data classification based on an optimized criterion which is aimed for better class separation. LDA can be used in combination with PCA to increase the efficiency of classification. For this, PCA scores obtained using a set of significant 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. LDA models were validated by Leave-One-Out Cross-Validation (LOOCV). LOOCV is a type of rotation estimation used mainly for smaller datasets *i.e.* a technique useful for assessing the performance of a predictive model with a hypothetical validation set when an explicit validation set is not available. Algorithms for these analyses were implemented in MATLAB (Mathworks Inc., USA) based software using in-house codes.<sup>40</sup>

Mean spectra were computed from the background subtracted spectra prior to derivatization for each class, by averaging Y-axis variations for each class, and baseline corrected by fitting a fifth order polynomial function. These baseline corrected spectra were vector normalized and then used for computing mean spectra. Standard deviation was computed using background subtracted baseline corrected spectra. Difference spectra were also calculated by subtracting mean spectra of the control group from the tumor, TT and NTT groups.

### 3. Results and discussion

#### 3.1. Spectral analysis

(a) Unprocessed control and tumor urine: the vector-normalized average spectrum (Fig. 1a(i)) of control rat urine exhibits urea peaks at 1004  $\text{cm}^{-1}$  (symmetrical C–N stretch) and 1161  $\text{cm}^{-1}$  (attributed to NH<sub>2</sub> modes) and creatinine peaks at 680  $\text{cm}^{-1}$  (C–NH<sub>2</sub> and C=O stretching, ring vibrations) and 850  $\text{cm}^{-1}$  (C–NH<sub>2</sub> deformation and ring vibrations), as reported elsewhere.<sup>41</sup> Mean tumor bearing rat urine spectra (Fig. 1a(ii)) showed differences in the intensities of several peaks, indicating the difference in the concentration of the urine's biochemical components. Differences were seen in the intensity of specific peaks such as increased intensity of the peaks of urea (1006  $\text{cm}^{-1}$ ) and creatinine (680  $\text{cm}^{-1}$ ) in the cancer group compared to control. To elucidate the spectral variations amongst groups, difference spectra were computed by subtracting the mean control spectrum from the mean tumor spectrum. Subtraction of mean spectra is one of the conventional ways of looking at spectral differences, it provides differences over a selected spectral range and thus understanding of the moieties that may have been modified is facilitated. The positive peaks of the difference spectrum are from the mean tumor spectrum while negative peaks are from the mean control spectrum. Tumor – control difference spectra (Fig. 1a(ii)) also show a prominent positive urea peak at 1006  $\text{cm}^{-1}$  suggesting increased urea concentration in urine during cancer.

(b) Concentrated control and tumor urine: the mean concentrated control rat urine spectrum has features similar to unprocessed urine with additional peaks at 653, 756, 781, 885 and 925  $\text{cm}^{-1}$  (Fig. 1b(i)). The mean tumor spectrum (Fig. 1b(ii)) shows a difference in the intensity of urea and creatinine peaks with respect to control. Tumor – control difference spectra show a prominent positive urea peak at 1006  $\text{cm}^{-1}$  suggesting increased urea concentration in urine during cancer (Fig. 1b(ii)).

(c) Concentrated NTT and TT urine: mean concentrated NTT (Fig. 3a(i)) and TT (Fig. 3a(ii)) rat urine show a difference in the intensity of the urea peak. The TT mean spectrum exhibits the highest intensity compared to control and tumor while the NTT mean spectrum shows the lowest concentration compared to all. TT – control (Fig. 2b(i)) and control – NTT (Fig. 2b(iii)) difference spectra suggest higher urea concentration in TT compared to control and higher urea concentration in control compared to NTT. TT – NTT difference spectra (Fig. 2b(ii)) also suggest increased urea concentration in TT compared to NTT.

Standard deviations for each group, unprocessed urine control (Fig. 3a), unprocessed urine tumor (Fig. 3b), concentrated urine control (Fig. 3c), concentrated urine tumor (Fig. 3d), concentrated urine NTT (Fig. 3e) and concentrated urine TT (Fig. 3f) were calculated to assess the reproducibility of data processing and sample-sample variation.

Although differences in the urine biochemistry of control and tumor rats are observed, the link between breast cancer and urine is not yet completely known. Studies have indicated

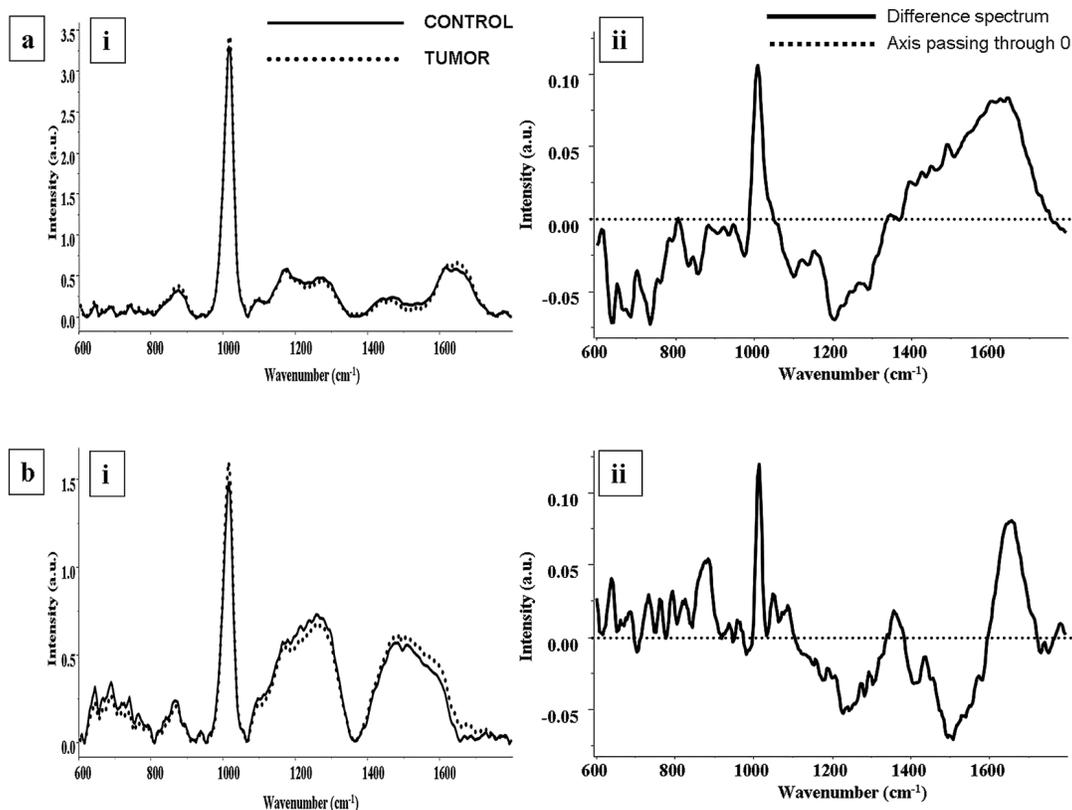


Fig. 1 (a) (i) Mean spectra of unprocessed urine interpolated in the 600–1800  $\text{cm}^{-1}$  region from control and tumor bearing rats and (ii) tumor – control difference spectrum; (b) mean spectra of concentrated urine interpolated in the 600–1800  $\text{cm}^{-1}$  region from control and tumor bearing rats, and (ii) tumor – control difference spectrum.

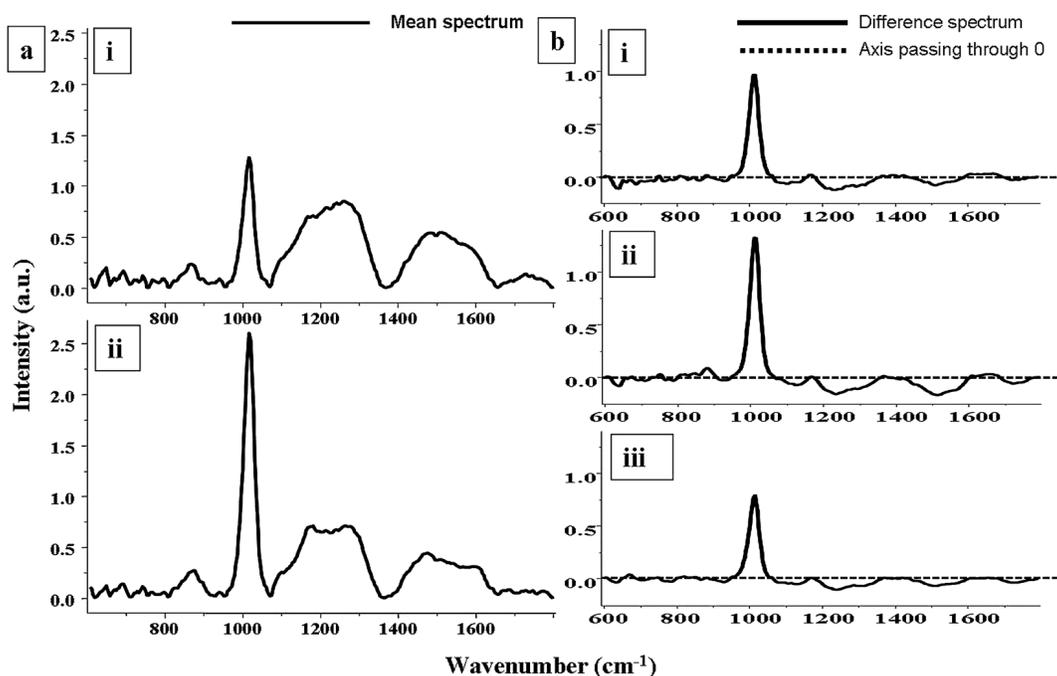


Fig. 2 (a) Mean spectra of concentrated urine interpolated in the 600–1800  $\text{cm}^{-1}$  region from (i) NTT and (ii) TT rats and (b) (i) TT – control, (ii) TT – NTT, and (iii) control – NTT difference concentrated urine spectra.

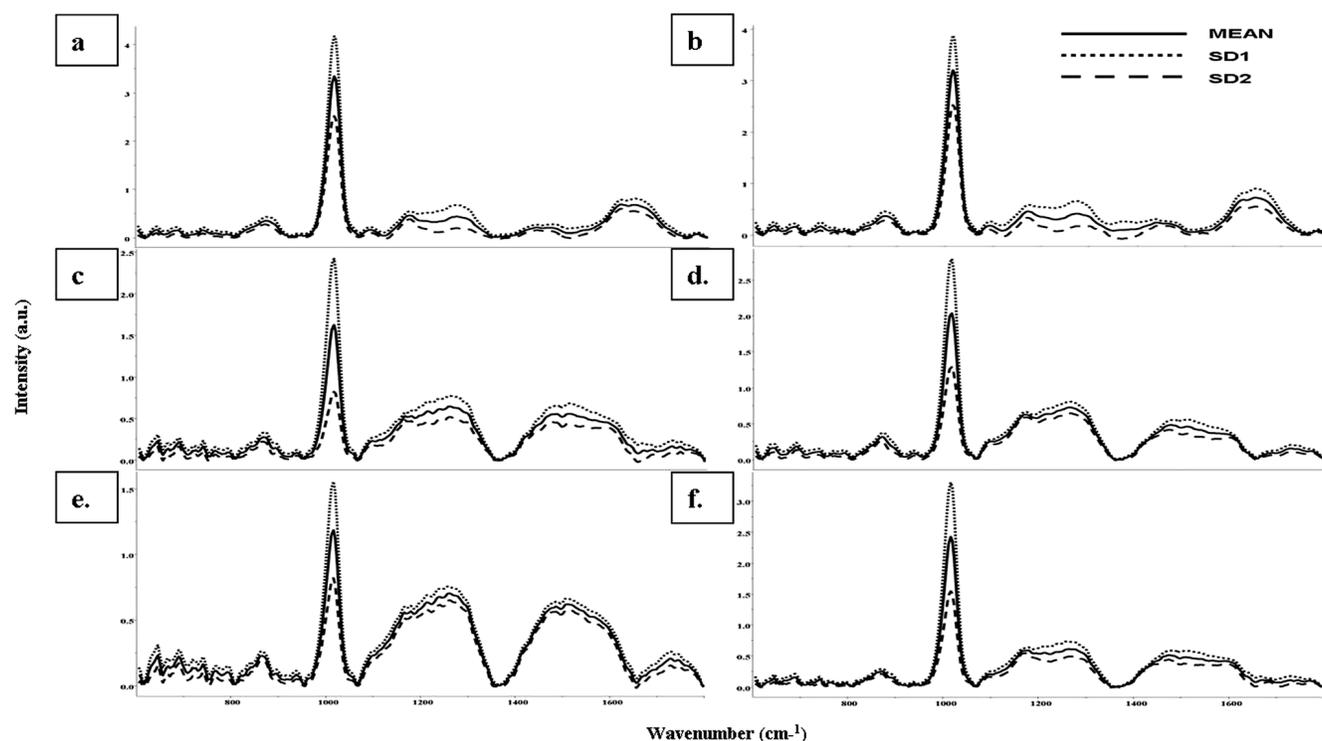


Fig. 3 Mean and standard deviations for each group, (a) unprocessed urine control, (b) unprocessed urine tumor, (c) concentrated urine control, (d) concentrated urine tumor, (e) concentrated urine NTT and (f) concentrated urine TT.

excretion of androgen, corticoid and estrogen metabolites in urine to be associated with breast cancer.<sup>42,43</sup> However, further studies are needed to establish an underlying connection between urine and breast cancer.

Further, it is unclear why TT and NTT have a huge difference while the control and tumor spectral difference is comparatively smaller. A probable explanation is as follows. TT is the pre-tumor condition. Intense biochemical changes may be expected for establishing a tumor. In the NTT group, a successful immune response resulting in aborted carcinogenesis may alter metabolism. Altered metabolism is linked with changed excretion profiles. Since the metabolic and biochemical alterations in TT and NTT are intense and different, a large difference is probably observed. Compared to this, in control there is no tumor and hence additional biochemical changes are not expected. In the tumor group, cancer is established. Hence, minor changes and secretions from the tumor are expected. Hence, the difference between these groups is probably less. Extensive studies in this direction are warranted.

### 3.2. Multivariate analysis

(a) Unprocessed control and tumor urine: preprocessed spectra interpolated in the 600–1800  $\text{cm}^{-1}$  range were subjected to PCA for delineating trends in the dataset. The PCA variance plot and loadings are shown in Fig. 4a and b. As can be seen in Fig. 4a, cumulative variances covered by factors 2 and 3 are 81% and 84%, respectively. The scatter plot of PCA factors (Fig. 4c) shows a tendency towards classification of

control and tumor bearing rat unprocessed urine. To explore the feasibility of classifying the above groups, PC-LDA was used. To avoid over fitting, 9 factors<sup>44</sup> contributing ~80% percent of correct classification were used (Fig. 5a). The plot of PC-LDA factors 1, 2 and 3 (Fig. 5b) shows clusters of control and tumor unprocessed urine spectra. The confusion matrix for PC-LDA model building shows that 69/81 control spectra are correctly classified as control while 12/81 are misclassified as tumor; whereas 63/82 spectra are correctly classified as tumor while 19/82 spectra are misclassified as control. Leave-One-spectrum-Out-Cross-Validation (LOSOVC) was carried out to evaluate the results obtained by PC-LDA. LOSOVC builds a model based on all observations but one, and tests the left out observation against the model built; this is repeated until all observations are left out once. The performance is estimated in terms of classification efficiency, which is the percentage of spectra from each group that is correctly classified. In analysis of LOSOVC as shown in Table 1a, 65/81 control spectra are correctly classified as control while 16/81 are misclassified as tumor; whereas 59/82 spectra are correctly classified as tumor while 23/82 spectra are misclassified as control.

Urine is a complex colloidal solution consisting mainly of urea, creatinine, salts and colloids made of glycoprotein, proteins and mucopolysaccharides.<sup>45</sup> Their concentration ranges from 9.3  $\text{g L}^{-1}$  (urea) to 0.67  $\text{g L}^{-1}$  (creatinine). The meager quantity present enhances the possibility of irregular distribution of the components mentioned. Further, as breast cancer progresses, minute concentration changes occur in the

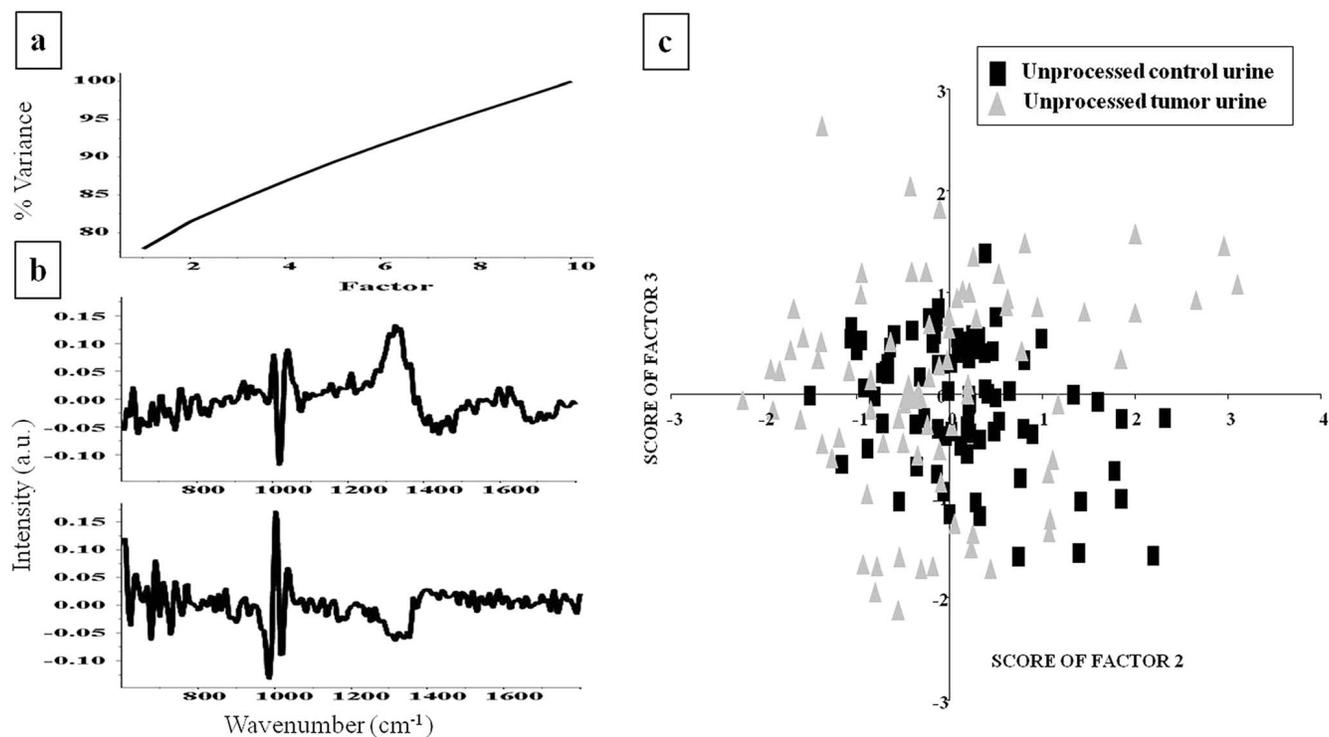


Fig. 4 PCA of unprocessed urine from control and tumor bearing rats: (a) variance plot, (b) loading factors 2 and 3, and (c) scatter plot.

limited number of urine components while the concentration of all other components of urine remains unchanged. These factors may contribute greatly to the misclassification observed

amongst groups. Overall, the classification efficiency of control and tumor (using unprocessed urine samples) groups was 80% and 72%, respectively.

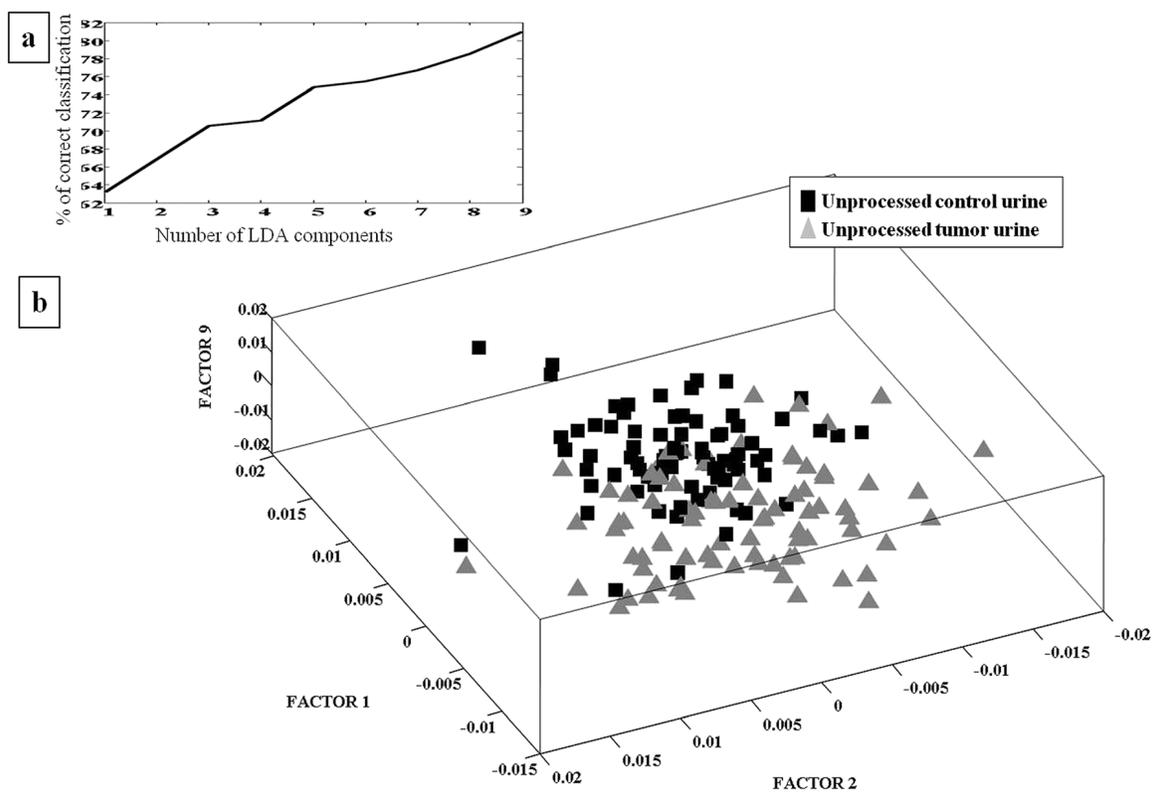


Fig. 5 PC-LDA of unprocessed urine from control and tumor bearing rats: (a) scree plot and (b) scatter plot.

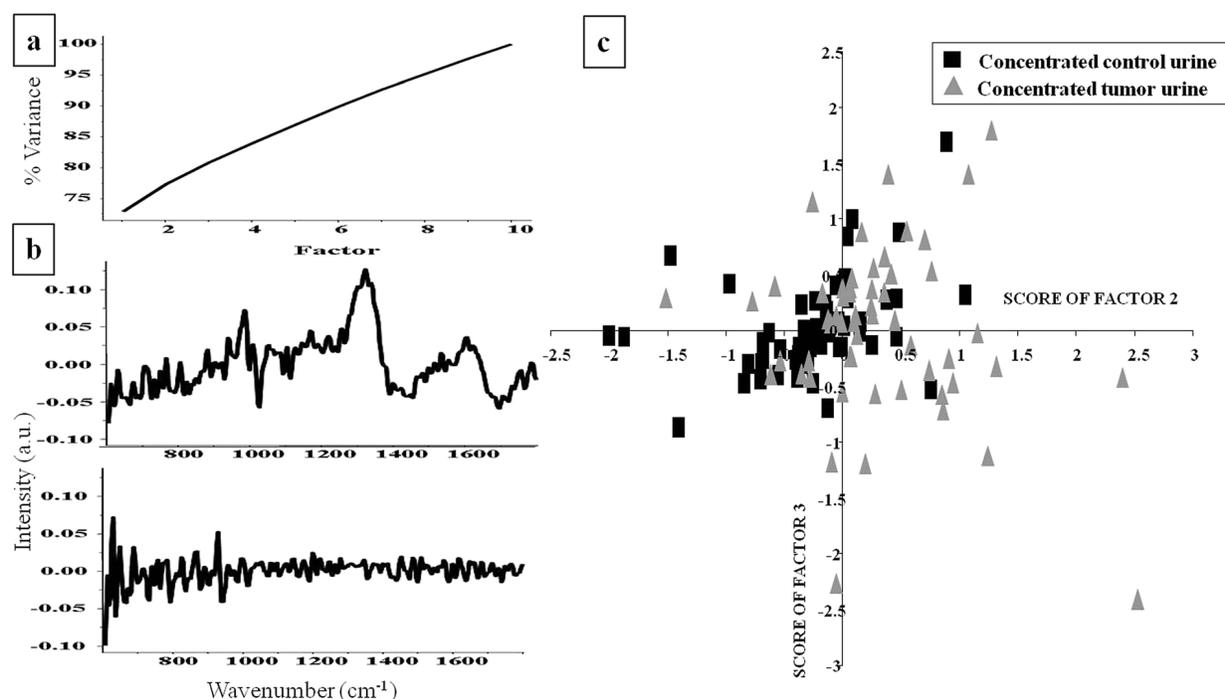
**Table 1** PC-LDA confusion matrix of (a) Leave-One-Out Cross-Validation (LOOCV) of unprocessed control and tumor bearing rat urine and (b) LOOCV of concentrated control and tumor bearing rat urine (ex-diagonal elements are false positive predictions)

(a) LOOCV (no. of animals, no. of spectra)	Unprocessed urine control	Unprocessed urine tumor
Unprocessed urine control (9, 81)	65 (80.24%)	16
Unprocessed urine tumor (9, 82)	23	59 (71.95%)
(b) LOOCV (no. of animals, no. of spectra)	Concentrated urine control	Concentrated urine tumor
Concentrated urine control (8, 64)	50 (78.12%)	14
Concentrated urine tumor (7, 56)	5	51 (91.07%)

(b) Concentrated control and tumor urine: spectra of concentrated urine of control and tumor bearing rats, interpolated in the 600–1800  $\text{cm}^{-1}$  range, were also subjected to PCA. The PCA variance plot and loadings are shown in Fig. 6a and b, respectively. As can be seen in Fig. 6a, cumulative variances covered by factors 2 and 3 are 82% and 84%, respectively. The scatter plot of PCA factors (Fig. 6c) shows clusters of concentrated control and tumor bearing rat urine. To explore the feasibility of classifying the above groups from control, PC-LDA was used. 4 factors contributing  $\sim 85\%$  of correct classification were applied (Fig. 7a). The plot of PC-LDA factors 1, 2 and 3 (Fig. 7b) shows well-separated clusters of control and tumor spectra. The confusion matrix for PC-LDA model building shows that 50/64 control spectra are correctly classified as control while 14/64 are misclassified as tumor; whereas 51/56 spectra are correctly classified as tumor while 5/56 spectra are

misclassified as control. A model is built based on LOOCV results (Table 1b); 50/64 control spectra are correctly classified as control while 14/64 are misclassified as tumor; whereas 51/56 spectra are correctly classified as tumor while 5/56 spectra are misclassified as control. As discussed previously, the limiting concentration of urine components and their irregular distribution may explain the observed misclassification.

Although, in this case, the samples are concentrated, the total amount of components present in the sample analyzed is still very low. Since 150–200  $\mu\text{l}$  samples were concentrated and used for spectral acquisition, the total quantity of major component urea expected in one urine sample will be  $\sim 2 \mu\text{g}$ . Thus, concentrated samples have higher quantity and probably more regular distribution of components compared to unprocessed urine, but the quantities being analyzed are meager and possibly results in the misclassification observed. Overall, the



**Fig. 6** PCA of concentrated urine from control and tumor bearing rats: (a) variance plot, (b) loading factors 2 and 3, and (c) scatter plot.

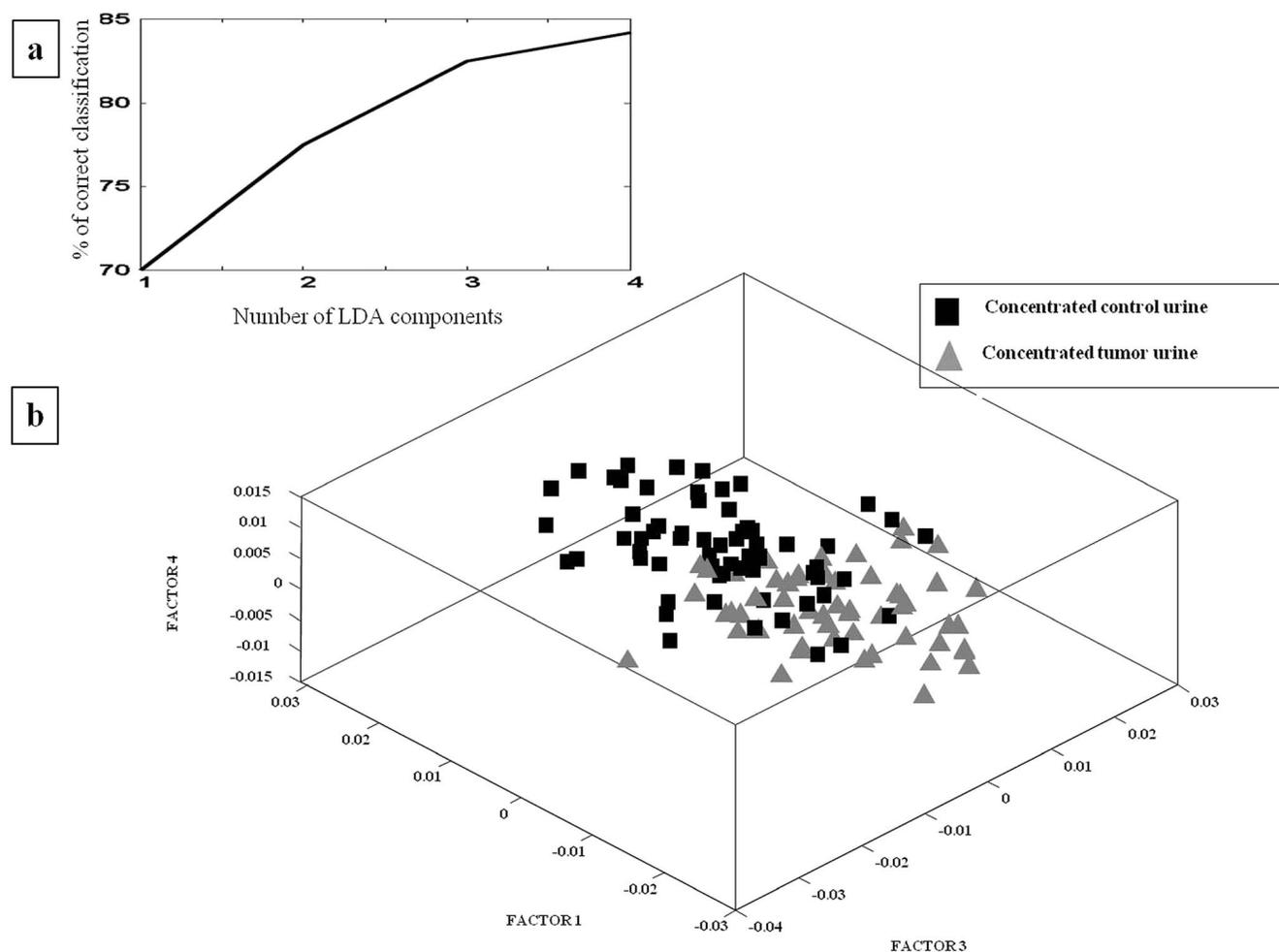


Fig. 7 PC-LDA of concentrated urine from control and tumor bearing rats: (a) scree plot and (b) scatter plot.

classification efficiency of the control and tumor (unprocessed urine samples) group was 78% and 91%, respectively. While the classification efficiency of the control group in the case of both unprocessed and concentrated control urine samples is  $\sim 80\%$ , classification efficiency of the tumor group in the case of concentrated urine is higher (91%) compared to unprocessed urine (72%). Therefore, further studies were conducted using concentrated urine.

(c) Concentrated NTT and TT urine: to further explore the sensitivity of urine based RS in diagnosis of breast cancer and the possibility of early detection, as mentioned previously, urine samples were also collected prior to tumor development. Spectra acquired from concentrated urine of control, tumor bearing, NTT and TT rats were preprocessed, interpolated in the  $600\text{--}1800\text{ cm}^{-1}$  range and were subjected to PCA and PC-LDA. The PCA variance plot and loading factors 1 and 3 are shown in Fig. 8a and b, respectively. The TT spectra in the PCA scatter plot (Fig. 8c) show a tendency of classification. The PC-LDA scatter plot (Fig. 9b) of factors 1, 2 and 3 shows overlapping clusters of control, tumor, TT and NTT. It is however noteworthy that control and NTT populate the left side of the plot whereas TT and tumor lie on the right side. The results of PC-LDA in the form of the confusion matrix

shows that 34/64 spectra are correctly classified as control, while 9/64 misclassified as NTT, 11/64 misclassified as TT and 10/64 misclassified as tumor. 19/40 NTT spectra were correctly classified as NTT, while 15/40, 2/40 and 4/40 misclassified as control, TT and tumor, respectively. In the case of TT, 23/32 were correctly classified whereas 7/32 and 2/32 were misclassified as control and tumor, respectively. 30/56 tumor spectra were classified correctly while 3/56, 11/56 and 12/56 were misclassified as control, NTT and TT. The results of LOOCV are shown in Table 2a. As can be seen, 34/64 spectra are correctly classified as control, while 9/64 are misclassified as NTT, 11/64 are misclassified as TT and 10/64 are misclassified as tumor. 19/40 NTT spectra were correctly classified as NTT, while 16/40, 2/40 and 5/40 were misclassified as control, TT and tumor, respectively. In the case of TT, 22/32 were correctly classified whereas 8/32 and 2/32 were misclassified as control and tumor, respectively. 30/56 tumor spectra were classified correctly while 3/56, 11/56 and 12/56 were misclassified as control, NTT and TT.

Despite misclassification amongst groups, 67.2% control spectra are classified as either control or NTT, while 82.5% NTT spectra are classified as either control or NTT (Table 2a). Control rats were not treated with carcinogen whereas NTT rats

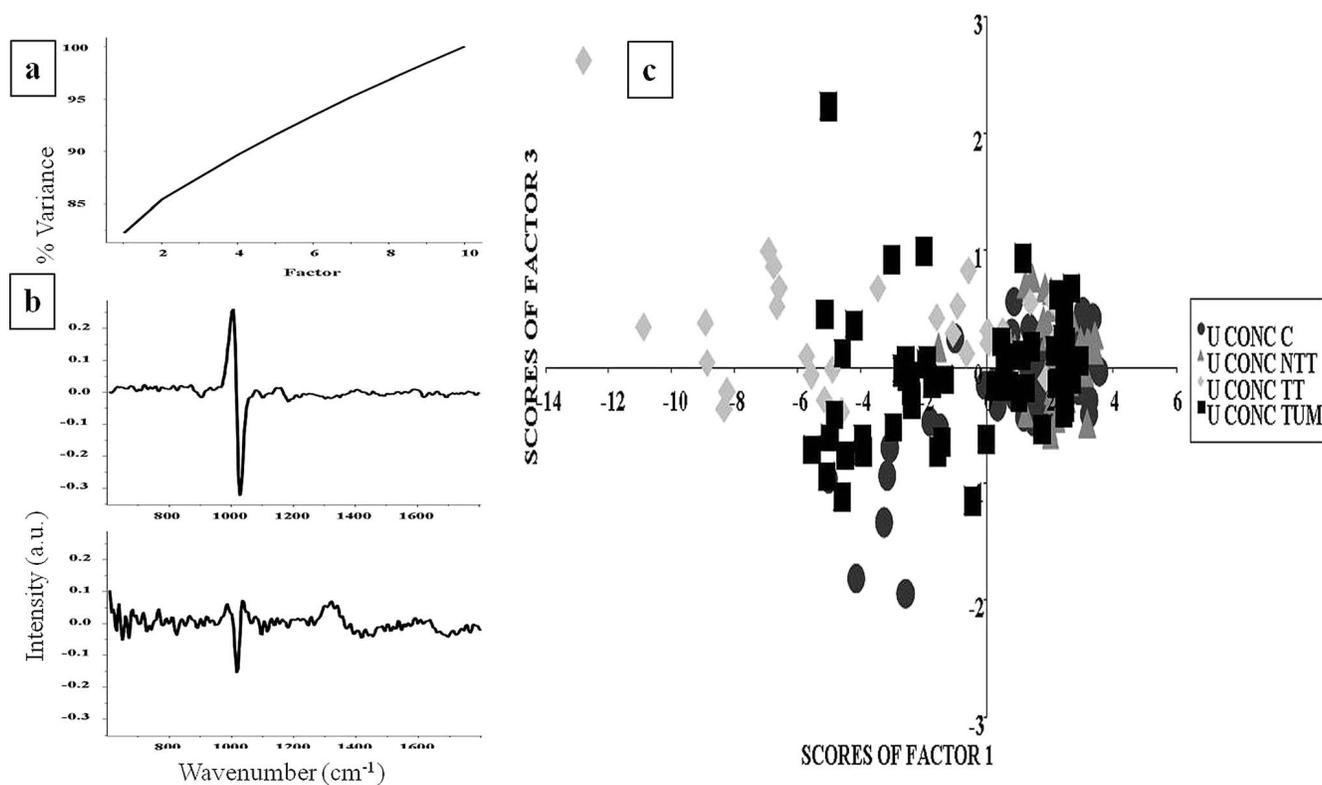


Fig. 8 PCA of concentrated urine from control, NTT, TT and tumor bearing rats: (a) variance plot, (b) loading factors 1 and 3, and (c) scatter plot.

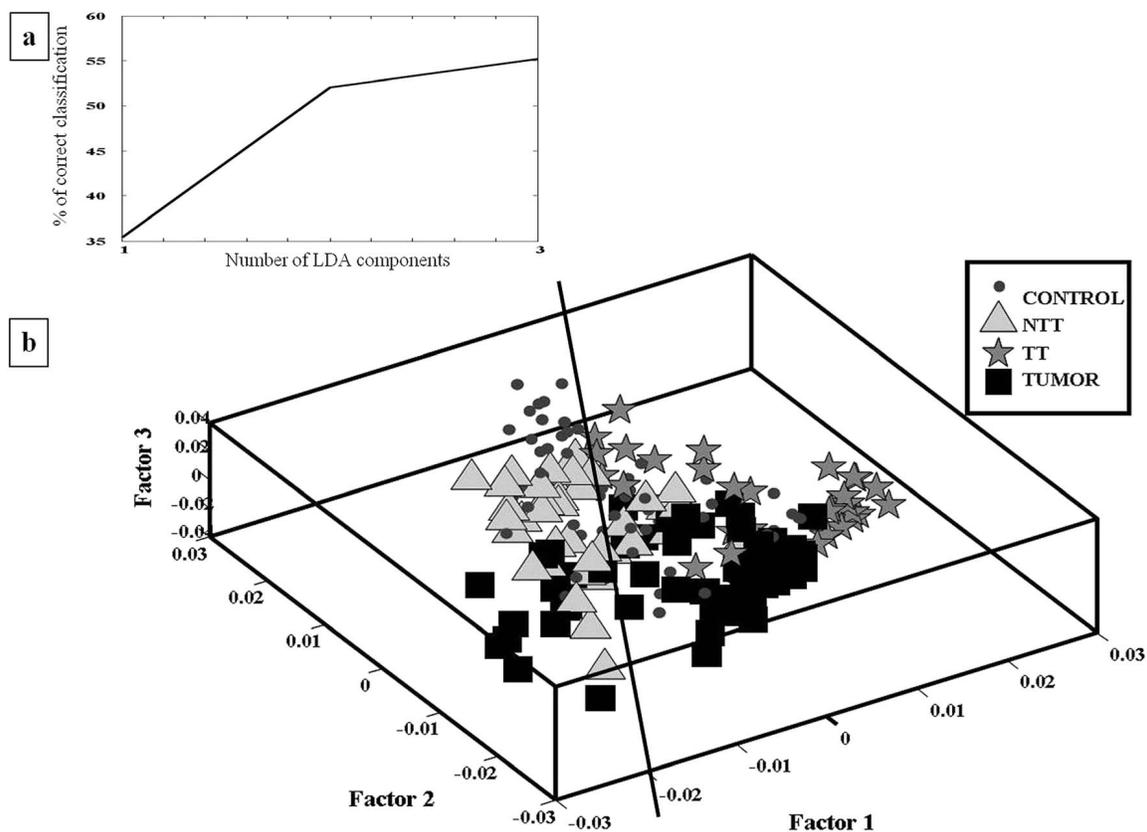


Fig. 9 PC-LDA of concentrated urine from control, NTT, TT and tumor bearing rats: (a) scree plot and (b) scatter plot.

**Table 2** PC-LDA confusion matrix of C, NTT, TT and T concentrated urine for (a) Leave-One-spectrum-Out Cross-Validation and (b) Leave-One-rat-Out Cross-Validation (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions)

(a) Group (no. of animals, no. of spectra)	Concentrated urine control	Concentrated urine NTT	Concentrated urine TT	Concentrated urine tumor
Concentrated urine control (8, 64)	30 (46.8%)	17 (26.6%)	8 (12.5%)	9 (14.1%)
Concentrated urine NTT (5, 40)	20 (50%)	14 (35%)	0	6 (15%)
Concentrated urine TT (4, 32)	8 (25%)	1 (3.1%)	19 (59%)	4 (12.5%)
Concentrated urine TUMOR (7, 56)	2 (3.6%)	15 (26.8%)	17 (30.4%)	22 (39%)
(b) Group (no. of animals, no. of spectra)	Concentrated urine control	Concentrated urine NTT	Concentrated urine TT	Concentrated urine tumor
Concentrated urine control (8, 64)	34 (53.12%)	9 (14.1%)	11 (17.2%)	10 (15.6%)
Concentrated urine NTT (5, 40)	16 (40%)	17 (42.5%)	2 (5%)	5 (12.5%)
Concentrated urine TT (4, 32)	8 (25%)	0	22 (62.5%)	2 (6.3%)
Concentrated urine tumor (7, 56)	3 (5.3%)	11 (19.6%)	12 (21.4%)	30 (51.78%)

did not develop tumors in spite of carcinogen treatment. Therefore, control and NTT urine spectra represent 'normal' (non-cancerous) conditions. TT rats were rats that eventually developed tumors whereas tumor group rats had breast tumors at the time of urine collection. Thus, urine spectra of TT and tumor group rats represent 'abnormal' (cancerous) conditions. As observed in Table 2a, 75% TT and 75% tumor spectra are correctly classified as abnormal (TT/tumor).

The results of Leave-One-rat-Out-Cross-Validation (LOR-OCV) implemented using 2 PC-LDA factors are shown in Table 2b. As can be seen, 30/64 spectra are correctly classified as control, while 17/64 are misclassified as NTT, 8/64 are misclassified as TT and 9/64 are misclassified as tumor. 14/40 NTT spectra were correctly classified as NTT, while 20/40, 0/40 and 6/40 were misclassified as control, TT and tumor, respectively. In the case of TT, 19/32 were correctly classified whereas 8/32, 1/32 and 4/32 were misclassified as control, NTT and tumor, respectively. 22/56 tumor spectra were classified correctly while 2/56, 15/56 and 17/56 were misclassified as control, NTT and TT. Thus, 73.4% and 85% control and NTT spectra, respectively were classified as control/NTT while 72% and 70% spectra from TT and tumor, respectively were classified as TT/tumor. The results corroborate the outcome of LOOCV.

In a nutshell, results suggest that rats that did not develop tumors could be classified as 'normal' (with ~83% efficiency) even though these rats were treated with carcinogen and had a high probability of developing tumors, while rats that did develop tumors after carcinogen treatment were classified as 'abnormal' (with 72.5% efficiency) using urine collected prior to any visible or palpable abnormality.

## 4. Conclusion

The aim of the study was to explore the feasibility of breast cancer diagnosis using urine based RS. Using unprocessed urine, control and tumor bearing rats could be classified with ~80% and ~72% efficiency, respectively. Using concentrated

urine, control and tumor groups could be classified with ~80% and 91% efficiency. These results suggest the possibility of cancer diagnosis using urine based RS. Further, concentrated urine of rats that did not develop tumors even after the carcinogen challenge could be classified as 'normal' with 83% efficiency whereas concentrated urine from visibly and palpably normal rats that eventually developed tumors could be classified as 'abnormal' with 72.5% efficiency. These results suggest the possibility of detecting biochemical changes occurring prior to tumor development using RS. Further studies in this direction may help to develop urine based RS as an early breast cancer detection tool.

## Abbreviations used

RS	Raman Spectroscopy
PC-LDA	Principal Component Linear Discriminant Analysis
LOOCV	Leave One Out Cross-Validation

## References

- 1 A. Jemal, F. Bray, M. Center, J. Ferlay, E. Ward and D. Forman, *Ca-Cancer J. Clin.*, 2011, **61**(2), 69–90.
- 2 R. Siegel, D. Naishadham and A. Jemal, *Ca-Cancer J. Clin.*, 2012, **62**, 10–29.
- 3 A. Nandakumar, T. Ramnath and M. Chaturvedi, *Indian J. Surg. Oncol.*, 2010, **1**(1), 8–9.
- 4 *SEER Cancer Statistics Review*, ed. Howlader N., Noone A. M., Krapcho M., Neyman N., Aminou R., Altekruse S. F., Kosary C. L., Ruhl J., Tatalovich Z., Cho H., Mariotto A., Eisner M. P., Lewis D. R., Chen H. S., Feuer E. J. and Cronin K. A., National Cancer Institute, Bethesda, MD, 2001, pp. 1975–2009.
- 5 G. I. Andreea, R. Pegza, L. Lascu, S. Bondari, Z. Stoica and A. Bondari, *Curr. Health Sci. J.*, 2011, **37**(2), 55–61.
- 6 D. Kennedy, T. Lee and D. Seely, *Integr. Cancer Ther.*, 2009, **8**(1), 9–16.

- 7 S. M. Ismail, A. B. Coleclough, J. S. Dinnen, D. Eakins, D. M. Evans, E. Gradwell, J. P. O'Sullivan, J. M. Summerell and R. G. Newcombe, *BMJ*, 1989, **298**, 707–710.
- 8 J. B. Sørensen, F. R. Hirsch, A. Gazdar and J. E. Olsen, *Cancer*, 1993, **71**, 2971–2976.
- 9 V. Chang, P. Cartwright, S. Bean, G. Palmer, R. Bentley and N. Ramanujam, *Neoplasia*, 2009, **11**(4), 325–332.
- 10 H. Yamazaki, S. Kaminaka, E. Kohda, M. Mukai and H. Hamaguchi, *Radiat. Med.*, 2003, **21**(1), 1–6.
- 11 S. Teh, W. Zheng, K. Ho, M. Teh, K. Yeoh and Z. Huang, *Br. J. Cancer*, 2008, **98**(2), 457–465.
- 12 Y. Zhou, C. H. Liu, Y. Sun, Y. Pu, S. Boydston-White, Y. Liu and R. R. Alfano, *J. Biomed. Opt.*, 2012, **17**(11), 116021.
- 13 S. P. Singh, A. Deshmukh, P. Chaturvedi and C. M. Krishna, *J. Biomed. Opt.*, 2012, **17**(10), 105002.
- 14 D. J. Evers, B. H. Hendriks, G. W. Lucassen and T. J. Ruers, *Future Oncol.*, 2012, **8**(3), 307–320.
- 15 R. R. Alfano, C. H. Liu, W. L. Sha, H. R. Zhu, D. L. Akins, J. Cleary, R. Prudente and E. Celmer, *Laser Life Sci.*, 1991, **4**(1), 23–28.
- 16 S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari and M. S. Feld, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**(35), 12371–12376.
- 17 M. V. Chowdary, K. K. Kumar, S. Mathew, L. Rao, C. Murali Krishna and J. Kurien, *Biopolymers*, 2009, **91**(7), 539–546.
- 18 M. V. Chowdary, K. K. Kumar, J. Kurien, P. Mathew and C. M. Krishna, *Biopolymers*, 2006, **83**(5), 556–569.
- 19 M. D. Keller, E. Vargis, N. Granja, R. H. Wilson, M. A. Mycek, M. C. Kelley and A. Mahadevan-Jansen, *J. Biomed. Opt.*, 2011, **16**(7), 077006.
- 20 N. Stone, R. Baker, K. Rogers, A. W. Parker and P. Matousek, *Analyst*, 2007, **132**, 899–905.
- 21 N. Stone and P. Matousek, *Cancer Res.*, 2008, **68**(11), 4424–4430.
- 22 A. Haka, Z. Volynskaya, J. A. Gardecki, J. Nazemi, J. Lyons, D. Hicks, M. Fitzmaurice, R. R. Dasari, J. P. Crowe and M. S. Feld, *Cancer Res.*, 2006, **66**(6), 3317–3322.
- 23 C. Nieva, M. Marro, N. Santana-Codina, S. Rao and D. Petrov, *PLoS One*, 2012, **7**(10), e46456.
- 24 S. E. Clare, H. Neubauer, C. Schuetz, R. Kurek, E. Solomayer, D. Wallwiener, D. F. Dye, J. M. Zaleski, R. J. Goulet and T. Fehm, *J. Clin. Oncol.*, 2006, **24**(18S), 10619.
- 25 L. Raniero, R. A. Canevari, L. N. Ramalho, F. S. Ramalho, E. A. dos Santos, R. A. Bitar, K. J. Jalkanen, H. S. Martinho and A. A. Martin, *Theor. Chem. Acc.*, 2011, **130**(4), 1239–1247.
- 26 T. Bhattacharjee, P. Kumar, G. Maru, A. Ingle and C. Murali Krishna, *Laser Med. Sci.*, 2014, **29**(1), 325–333.
- 27 A. Sahu, S. Sawant, H. Mamgain and C. M. Krishna, *Analyst*, 2013, **138**(14), 4161–4174.
- 28 J. L. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-García, R. Huerta-Franco, J. L. González-Trujillo, C. A. Ramírez-Alvarado, G. Gutiérrez-Juárez and C. Medina-Gutiérrez, *Laser Med. Sci.*, 2007, **22**(4), 229–236.
- 29 S. Feng, R. Chen, J. Lin, J. Pan, G. Chen, Y. Li, M. Cheng, Z. Huang, J. Chen and H. Zeng, *Biosens. Bioelectron.*, 2010, **25**, 2414–2419.
- 30 J. L. González-Solís, J. Rodríguez-Lóopez, J. C. Martínez-Espinosa, C. Frausto-Reyes, L. F. Jave-Suárez, A. C. Aguilar-Lemarroy, H. Vargas-Rodríguez and E. Martínez-Cano, *AIP Conf. Proc.*, 2010, **1226**, 91–95.
- 31 X. Li, T. Yang and S. Li, *Appl. Opt.*, 2012, **51**, 5038–5043.
- 32 A. Sahu, K. Dalal, S. Naglot, P. Aggarwal and C. Murali Krishna, *PLoS One*, 2013, **8**(11), e78921.
- 33 J. C. Kah, K. W. Kho, C. G. Lee, C. J. Sheppard, Z. X. Shen, K. C. Soo and M. C. Olivo, *Int. J. Nanomed.*, 2007, **2**(4), 785–798.
- 34 S. Feng, D. Lin, J. Lin, Z. Huang, G. Chen, Y. Li, S. Huang, J. Zhao, R. Chen and H. Zeng, *Appl. Phys. Lett.*, 2014, **104**, 073702.
- 35 X. Li, T. Yang and J. Lin, *J. Biomed. Opt.*, 2012, **17**(3), 037003.
- 36 A. Shapiro, O. N. Gofrit, G. Pizov, J. K. Cohen and J. Maier, *Eur. Urol.*, 2011, **59**(1), 106–112.
- 37 A. Nijssen, K. Maquelin, P. J. Caspers, T. C. B. Schut, M. H. A. Neumann, J. C. den Hollander, L. F. Santos and G. J. Puppels, *J. Biomed. Opt.*, 2007, **12**, 034004.
- 38 S. Koljenovic, L. P. Choo-Smith, T. C. B. Schut, J. M. Kros, H. J. Van den Berge and G. J. Puppels, *Lab. Invest.*, 2002, **2**, 1265–1277.
- 39 T. J. Harvey, E. Gazi, A. Henderson, R. D. Snook, N. W. Clarke, M. Brown and P. Gardner, *Analyst*, 2009, **134**, 1083–1091.
- 40 A. D. Ghanate, S. Kothiwale, S. P. Singh, D. Bertrand and C. M. Krishna, *J. Biomed. Opt.*, 2011, **16**, 025003.
- 41 J. A. Bispo, E. E. Vieira, L. Silveira and A. B. Fernandes, *J. J. Biomed. Opt.*, 2013, **18**(8), 087004.
- 42 R. D. Bulbrook, *et al.*, *Lancet*, 1971, **298**(7721), 395–398.
- 43 G. C. Kabat, *et al.*, *Cancer Epidemiol., Biomarkers Prev.*, 1997, **6**, 505–509.
- 44 J. Kelly, J. Trevisan, A. Scott, P. Crmichael, H. Pollock, P. Martin-Hirsch and F. Martin, *J. Proteome Res.*, 2011, **10**, 1437–1448.
- 45 A. J. Butt, E. A. Hauser and J. Seifter, *Calif. Med.*, 1952, **76**(3), 123–129.

# Transcutaneous *in vivo* Raman spectroscopy of breast tumors and pretumors

T. Bhattacharjee, G. Maru, A. Ingle and C. Murali Krishna\*



Breast cancer is the most common cancer amongst women worldwide. Early detection of this cancer results in better prognosis. Owing to the disadvantages of currently available screening tools for early detection of this cancer, rapid and sensitive alternatives such as optical spectroscopic techniques are being extensively explored. Detection of premalignant lesions using these techniques has been reported. However, premalignant lesions are risk indicators and may not be true predictors of tumor development. Therefore, the current study aims at correlation between spectral changes and tumor appearance. In this context, transcutaneous *in vivo* spectra were acquired from same carcinogen-induced rats immediately before carcinogen treatment, 3, 8–10, and 12–14 weeks after carcinogen treatment and from frank tumors. These were analyzed using multivariate statistical tools principal component analysis and principal component linear discriminant analysis. Further, a complex test data set consisting of spectra from rats of varying ages, tumor appearance times, and tumor induction protocols was used to test the feasibility of correctly identifying controls and pretumors using Raman spectroscopy. Results suggest feasibility of distinguishing pretumor spectra from controls. Taking into consideration the heterogeneity of afflicted breast, rat-wise analysis was performed wherein a rat was declared 'will develop tumor', even if one spectrum was found abnormal. Using this criterion, *in vivo* Raman spectroscopy could predict tumor appearance with 82% sensitivity and 95% specificity. Prospectively, combined with emerging technologies like deep Raman spectroscopy and fiber-probe-based whole sample imaging, Raman spectroscopy may prove as an invaluable adjunct to currently available breast cancer screening tools. Copyright © 2015 John Wiley & Sons, Ltd.

Additional supporting information may be found in the online version of this article at the publisher's website.

**Keywords:** Raman spectroscopy; early detection; screening; breast cancer; transcutaneous *in vivo*; animal model

## Introduction

Early detection of breast cancer, the most common cancer amongst women worldwide,<sup>[1,2]</sup> can substantially increase prognosis and disease-free survival.<sup>[3]</sup> The currently available breast cancer screening tools – mammography, magnetic resonance imaging, ultrasonography, positron emission tomography, tomosynthesis, and thermography – suffer from several disadvantages such as low sensitivity, specificity, high cost, high-output times, tedious sample preparation, and inter-observer variations.<sup>[4,5]</sup> This underlines the need for development of rapid, objective, potentially non-invasive screening tools. Raman spectroscopy (RS), a technique based on the phenomenon of inelastic scattering that can provide chemical fingerprint of a sample, is a rapid, objective tool that has been shown to be effective in the diagnosis of several cancers including breast cancers.<sup>[6–11]</sup> Ability of RS to monitor treatment using serum,<sup>[12]</sup> classify malignancy using exfoliated cells,<sup>[13]</sup> and identify cancer field effects/malignancy-associated changes,<sup>[14]</sup> pre-cancer lesions, hyperplastic, and dysplastic patches<sup>[15,16]</sup> demonstrates the sensitivity of the technique to subtle biochemical changes. Therefore, development of RS as an 'early cancer detection' tool is the focus of current biomedical optics-based research.

In this context, the present study aims to test the feasibility of distinguishing pretumor spectra from controls. Because most of the patients present at advanced stages, such studies are not possible in human subjects. Hence, animal models were employed. In addition, experimental carcinogenesis in animal models provides unique opportunities of following cancer progression in a

controlled environment with minimum variables. Several studies have reported animal-model-based breast cancer studies using RS. Feasibility of distinguishing physiological conditions such as pregnancy, lactation, and aging as well as pathological conditions using transcutaneous *in vivo* spectroscopy has been demonstrated.<sup>[17–20]</sup> Study of pre-cancer conditions using animal models can be performed in many ways. Typically, spectral changes are correlated with histopathologically confirmed premalignant conditions such as atypical ductal hyperplasia (ADH) and ductal carcinoma *in situ* (DCIS). However, literature suggests that premalignant conditions are indicators for risk assessment, that is, the risk of developing breast cancer in women presenting premalignant lesions ADH and DCIS are four to five and eight to ten times higher, respectively than those that do not exhibit such abnormalities.<sup>[4]</sup> Thus, detection of a premalignant condition may not be a true predictor of prospective tumor development. Hence, the objective of this study is to find an association between spectral changes and tumor appearance. In the present study, breast tumor was induced by subcutaneous injection of carcinogen in breast. Spectra were acquired 0, 3, 8–10, 12–14 and 20 weeks post-carcinogen treatment from treated rats and post-

\* Correspondence to: C. Murali Krishna, Chilakapati Laboratory, Advanced Center for Treatment Research and Education in Cancer (ACTREC), Tata Memorial Center (TMC), Kharghar, Sector 22, Navi-Mumbai 410210, India.  
E-mail: mchilakapati@actrec.gov.in, pittu1043@gmail.com

ACTREC, Kharghar, Navi-Mumbai 410210, India

oil treatment from controls. The spectra were analyzed using principal component analysis (PCA) and principal component-linear discriminant analysis (PC-LDA). PC-LDA models were also validated by independent test data. The test data consists of spectra acquired at varying rat ages (9–38 weeks old rats), from rats with varying times of tumor appearance (8–22 weeks post-carcinogen treatment) and different protocols for inducing carcinogenesis (intragastric administration and subcutaneous dusting of carcinogen), for rigorous evaluation of the model. Results of the study are discussed in the manuscript.

## Materials and methods

### Animals

A total of 81 Sprague–Dawley (SD) rats were used in this study. Thirty four 47 days old rats were administered oil (controls), while forty seven 47 days old rats were administered 7, 12 dimethyl benzanthracene (DMBA, Sigma-Aldrich, purity 95%). The study was approved by Institutional Animal Ethics Committee, ACTREC endorsed by the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA), Government of India guidelines. All animals were housed under standard laboratory conditions, fed a diet of in-house-prepared pellets, and provided with water *ad libitum*.

### Carcinogenesis

The most commonly used protocol for inducing mammary tumors is intragastric administration of carcinogen DMBA. However, using this protocol, tumor may appear in any of the 12 mammary glands. Scanning 12 glands is time consuming and impractical. To circumvent this, carcinogen delivery using dusting and subcutaneous injection for site specific tumorigenesis was explored along with intragastric carcinogen delivery.

- (a) Subcutaneous injection: A DMBA solution of 15 mg/ml concentration was prepared. An insulin syringe fitted with 16G needle was used to inject 80  $\mu$ l of DMBA solution. The needle was inserted in the third left inguinal breast, and the solution was dispersed subcutaneously between the first and second left inguinal breast nipples (Supplementary Figure 1a). The same protocol was used for controls except oil was injected instead of DMBA. Twenty six SD rats were treated using this protocol, and 20 rats were kept as control.
- (b) Dusting: 1 mg DMBA was mixed with 2 mg cholesterol, cholesterol being the vehicle for delivery of DMBA. The rats were anesthetized using ketamine and xylazine. A single incision was made using a surgical blade between the first and second left inguinal breast nipples (Supplementary

Figure 1b). The skin was lifted and DMBA–cholesterol mixture was dusted over the mammary gland. The wound was sealed by Vetbond tissue adhesive. The control rats were dusted with cholesterol in the same manner. Four SD rats were dusted with DMBA, and five rats were kept as control.

- (c) Intragastric delivery (Gavage): Seventeen SD rats were administered 65 mg/kg DMBA dissolved in groundnut oil (Dhara, India) intragastrically by gavage (Supplementary Figure 1c). Nine rats were administered oil (Control).

### Raman spectroscopy

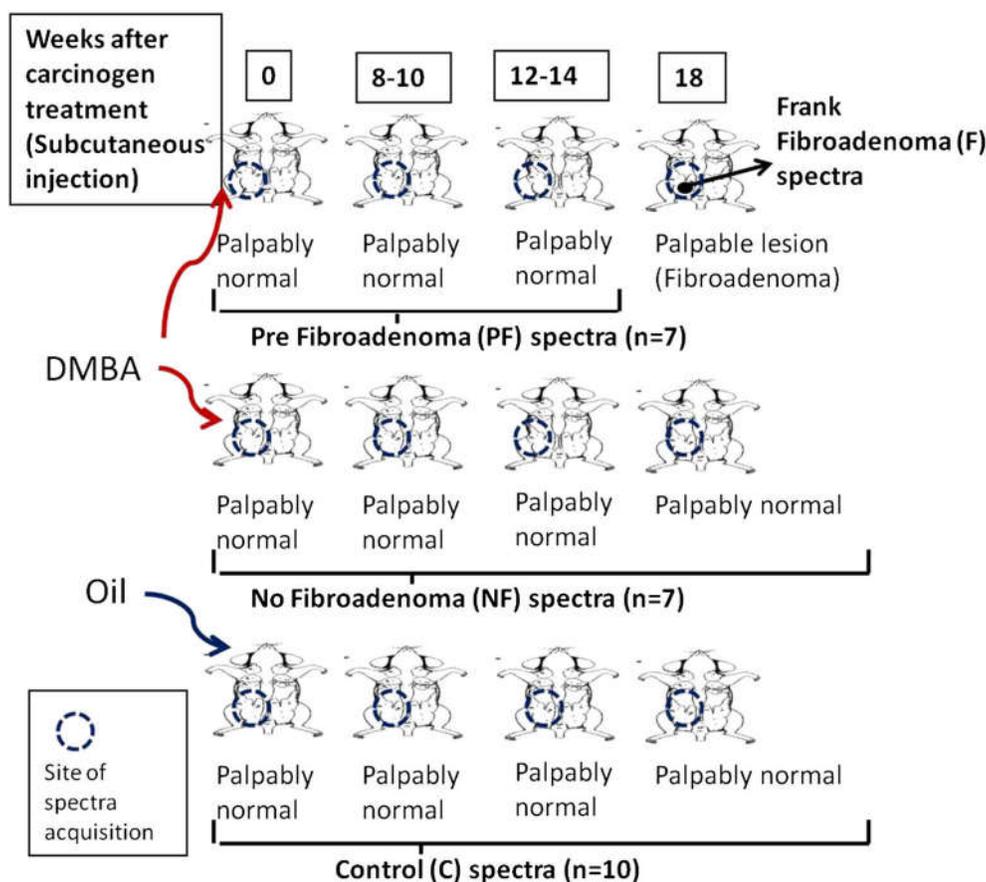
All spectra were acquired transcutaneously using Raman spectrometer described elsewhere.<sup>[17]</sup> This system consist of a diode laser (PI-ECL-785-300-FC, Process Instruments) of 785 nm wavelength as excitation source, a high-efficiency spectrograph (HE-785, Jobin-Yvon-Horiba, France) with fixed 950 g/mm grating coupled with a Charge Coupled Device (CCD) (CCD-1024X256-BIDD-SYN, Synapse). The spectrograph has no movable parts. The spectral resolution is  $\sim 4 \text{ cm}^{-1}$ . Commercial RamanProbe (RPS 785/12-5, In Photonics Inc, Downy St. USA) consisting of an excitation and collection fiber (NA-0.40) of diameters 105 and 200  $\mu$ m, respectively, was used to couple excitation source to the detection system. The probe utilizes a backscattering ( $\theta = 180^\circ$ ) sampling with the estimated spot size and depth of penetration as per the manufacturer's specifications 105  $\mu$ m and 1 mm, respectively. Spectral acquisition parameters used were  $\lambda_{\text{ex}} = 785 \text{ nm}$ , laser power=80 mW, spectra were integrated for 15 s and averaged over three accumulations.

Transcutaneous *in vivo* spectra were acquired from left inguinal mammary gland of rats as summarized in Table 1 and elaborated in the following paragraph:

- (a) Subcutaneous injection:
  - (a) Batch I: Spectra were acquired from control and treated rats 0, 3, 8–10, and 12–14 weeks posttreatment. The rat breast during these time points of spectra acquisition was visibly and palpably normal. Seven treated rats developed breast tumor (benign fibroadenoma) 18 weeks after carcinogen delivery (Fig. 1). Because spectra were acquired from visibly and palpably normal breast prior tumor development (pre-tumor), the 0, 3, 8–10, and 12–14 weeks spectra of these seven rats were labeled pre-fibroadenoma (PF). Spectra acquired from rats that did not develop tumor even after 30 weeks posttreatment were labeled no-fibroadenoma (NF). Spectra were also acquired from frank fibroadenoma, labeled F, of seven rats. The F spectra are spectra acquired from tumors of PF rats 20 weeks post-carcinogen treatment. The control and F spectra were used as training data set. The PF and NF spectra were used as test data set.

**Table 1.** Details of protocols, number of animals and time of spectra acquisition

Protocol	Batch	Control (no. of rats)	Treated (no. of rats)	Weeks of spectra acquisition	Tumor status (time of appearance post-carcinogen treatment)
Subcutaneous injection	I	10	14	0, 3, 8–10, 12–14, 20	18 weeks
	II	6	7	3, 7, 11	22 weeks
	III	4	5	3, 7, 11	14 weeks
Dusting		5	4	3, 7, 12	14 weeks
Gavage		16	17	0, 3, 8–10, 12–14, 20	None



**Figure 1.** Protocol for fibroadenoma progression study using rats treated by subcutaneous injection: DMBA treated rats shown by red arrow, oil treated rats shown by blue arrow, weeks after carcinogen/ oil treatment shown in black outlined text boxes.

- (b) **Batch II:** Spectra were acquired from control and treated rats 3, 7, and 11 weeks post oil/DMBA treatment. The treated rats developed fibroadenoma 22 weeks post-treatment. The 3, 7, and 11 week spectra of rats that developed tumor were labeled PF. The PF and control spectra were used as test data set.
- (c) **Batch III:** Spectra were acquired from control and treated rats 3, 7, and 11 weeks post-oil/DMBA treatment. All treated rats developed fibroadenoma 14 weeks posttreatment. Because the treated rats eventually developed fibroadenoma, spectra from these rats were also grouped as PF. Both control and PF spectra were used as test data set.
- (b) **Dusting:** Spectra were acquired from control and treated rats 3, 7, and 12 weeks post-carcinogen treatment. During spectra acquisition, these rats were palpably and visibly normal. Although normal during spectra acquisition weeks, the treated rats developed fibroadenoma 14 weeks post-carcinogen treatment and hence were labeled PF. These spectra as well as those from control rats were used as test data set.
- (c) **Gavage:** Spectra were acquired from control and treated rats 0, 3, 8–10, 12–14, and 20 weeks post-carcinogen/oil treatment. The treated rats did not develop fibroadenoma at the location scanned despite carcinogen treatment and hence were labeled NF. Both control and treated rat spectra were used as test data set.

The spectra were acquired at specific points with respect to the first inguinal nipple (Supplementary Fig. 1d) for all rats. This ensures

that the same locations in breast are probed throughout the study. Different batches of rats were administered carcinogen on different days. Weeks post-carcinogen treatments are calculated with respect to the day of treatment, which is referred to as the week 0. Spectra of a specific batch of rats were acquired on the same day.

Because this is a follow-up study wherein spectra were acquired from the same animal from 0 week (day of carcinogen delivery) till 14–22 week (time of tumor appearance), biopsy and subsequent histopathology of intermediate weeks could not be carried out. Histopathology was carried out at the end of the study to confirm tumors. Post-mortem surgery also revealed that all tumors were subcutaneous tumors, that is, the tumors were immediately below the skin.

#### Data analysis

As described elsewhere,<sup>[17]</sup> spectra were preprocessed by correcting for CCD response with a NIST certified SRM 2241 material and subtracting background signals from optical elements. To remove the interference of low frequency background variations, first derivative of the corrected Raman spectra were calculated (Savitzky-Golay, window size 3), interpolated in 1200–1800  $\text{cm}^{-1}$  range (Raman fingerprint region), and vector normalized. Analysis of preprocessed spectra was carried out using multivariate analysis tool PC-LDA implemented in MATLAB-based in-house software (Mathwork Inc.). PCA is routinely used method for data compression and visualization, while LDA provides data classification based on an optimized criterion for more class separability. LDA can be

used in conjunction with PCA (PC-LDA) to further increase the performance efficiency of classification. PC-LDA models were cross validated using leave one spectrum out cross validation (LOOCV). LOOCV builds a model based on all spectra but one, and tests the left out 'spectrum' against the model built. This is repeated until all spectra have been left out once. In this mode of cross validation, each spectrum is treated as an individual sample.

Average spectra were computed from the background-subtracted spectra (without derivatization) for each class and baseline corrected by fitting a fifth order polynomial function. These baseline-corrected, vector-normalized spectra were used for spectral comparisons and for computing difference spectra.

## Results and discussion

The study aims to test the feasibility of distinguishing pretumor spectra from controls. In this context, transcutaneous breast spectra were acquired from carcinogen-treated and control rats. In this study, breast tumor was induced by subcutaneous injection of carcinogen in breast of 14 rats. For controls, ten rats were subcutaneously injected with oil. Spectra were acquired before carcinogen treatment (0 weeks), 3 weeks after treatment and 8–10 and 12–14 weeks after treatment with carcinogen/oil from treated and control rats. Tumors appeared in seven treated rats approximately around 18 weeks post-carcinogen treatment while no abnormalities were observed in the remaining seven rats and control rats. Spectra were also acquired from frank tumors of rats that developed tumors and breast of control rats and treated rats that did not develop tumor 20 weeks after carcinogen treatment. The spectra were used for multivariate statistical analysis. For independent test prediction, rats induced using intragastric carcinogen administration and carcinogen dusting protocols were used along with rats induced by subcutaneous injection protocol.

### Spectral analysis

For spectral analysis, spectra acquired from subcutaneous injection batch I rats were used. Mean spectrum of control rats 0 week post-carcinogen treatment ( $C_0$ ) was used as standard to compare mean spectra of control rats 3 week ( $C_3$ ), 8–10 week ( $C_{8-10}$ ), 12–14 week ( $C_{12-14}$ ), and 20 week ( $C_{20}$ ) post-oil treatment, NF rats 3 week ( $NF_3$ ), 8–10 week ( $NF_{8-10}$ ), 12–14 week ( $NF_{12-14}$ ), and 20 week ( $NF_{20}$ ) post-carcinogen treatment, PF rats 3 week ( $PF_3$ ), 8–10 week ( $PF_{8-10}$ ), 12–14 week ( $PF_{12-14}$ ), and 20 week ( $PF_{20}$ ) post-carcinogen treatment and F (frank fibroadenoma). Difference spectra were calculated by subtracting  $C_0$  from  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ ,  $C_{20}$ ,  $NF_3$ ,  $NF_{8-10}$ ,  $NF_{12-14}$ ,  $NF_{20}$ ,  $PF_3$ ,  $PF_{8-10}$ ,  $PF_{12-14}$ ,  $NF_{20}$ , and F spectra.

$C_0$  spectrum (Fig. 2 a.1) exhibits lipid features – 1745  $\text{cm}^{-1}$ , 1653  $\text{cm}^{-1}$ , 1445  $\text{cm}^{-1}$ , 1301  $\text{cm}^{-1}$ , 1271  $\text{cm}^{-1}$ , and 1337  $\text{cm}^{-1}$  DNA band and broad amide I.  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ , and  $C_{20}$  (Fig. 2 a.2–a.5) show loss of 1337  $\text{cm}^{-1}$  band, sharper lipid (1653  $\text{cm}^{-1}$ , 1445  $\text{cm}^{-1}$ , and 1301  $\text{cm}^{-1}$ ), and amide I bands. Several changes take place in breast as the rat ages. At birth, the mammary glands consist of stroma – connective tissue, fibroblasts, the mammary fat pad, and epithelial cords – a small, branched ductal network of mammary epithelium that invades from nipple into fat pad. During puberty (from ~35–60 days after birth), mammary epithelial cells divide rapidly and invade into the fat pad resulting in branched ducts throughout the breasts. After initial intense proliferation, the cell proliferation becomes constant throughout the reproductive phase (~2–12 months after birth) at a rate lower

than puberty. Cell division declines during perimenopause and menopause; breast tissue decreases (1.2–1.5 years after birth).<sup>[21]</sup> Because  $C_0$  (day of carcinogen treatment) corresponds to 47th day after birth, the rat is in puberty phase. Thus, rapid cell proliferation during puberty may explain DNA band and broad amide I in  $C_0$  mean spectrum, while decrease in the cell proliferation during reproductive phase compared with puberty in probably explains loss of DNA band, sharper amide I, and lipid bands  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ , and  $C_{20}$ . Different spectra also support this finding. The  $C_3$ – $C_0$  different spectrum (Fig. 3 a.1) exhibit positive protein bands at 1660 and 1450 and negative DNA band at 1340  $\text{cm}^{-1}$ . This suggests slight change in DNA and protein profile.  $C_{8-10}$ – $C_0$  (Fig. 3 a.2) show positive lipid bands at 1740, 1440, 1301, and 1262  $\text{cm}^{-1}$  negative DNA bands at 1337 and 1480  $\text{cm}^{-1}$ .  $C_{12-14}$ – $C_0$  and  $C_{20}$ – $C_0$  (Fig. 3 a.3–a.4) difference spectra show profiles similar to  $C_{8-10}$ – $C_0$ , suggesting increased lipid content and decreased DNA content during  $C_{8-10}$ ,  $C_{12-14}$ , and  $C_{20}$  compared to  $C_0$ . The profile of  $C_3$  is slightly different probably because it is a phase where puberty ends and reproductive phase begins.

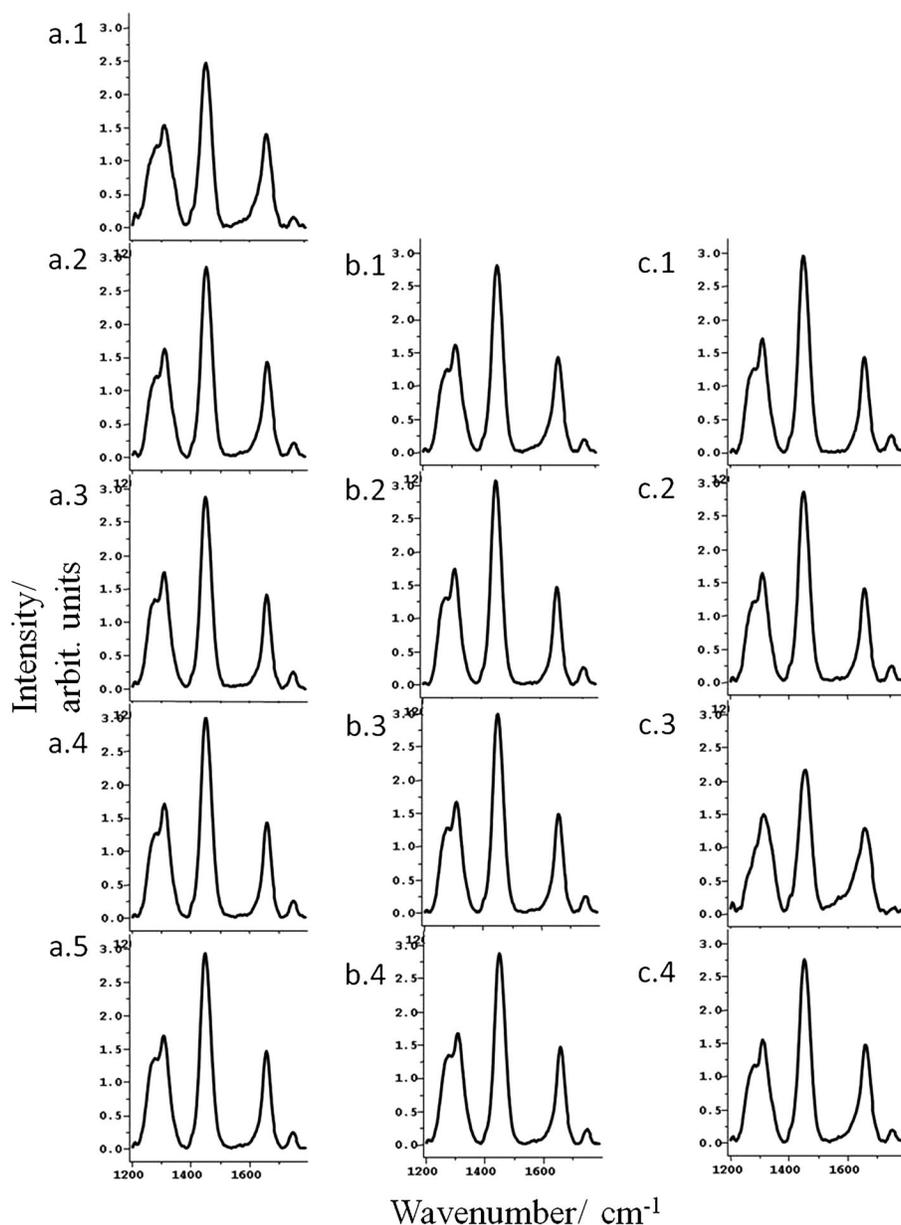
The mean NF spectra,  $NF_3$ ,  $NF_{8-10}$ ,  $NF_{12-14}$ , and  $NF_{20}$  (Fig. 2 b.1–b.4), have characteristics similar to  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ , and  $C_{20}$ , respectively. Loss of 1337  $\text{cm}^{-1}$  DNA band and sharper amide I in  $NF_3$ – $NF_{20}$  mean spectra compared with  $C_0$  indicates lower cell proliferation in these groups.  $NF_3$ – $C_0$ ,  $NF_{8-10}$ – $C_0$ , and  $NF_{12-14}$ – $C_0$  (Fig. 3 b.1–b.4) also have profiles similar to  $C_3$ – $C_0$ ,  $C_{8-10}$ – $C_0$ , and  $C_{12-14}$ – $C_0$ , respectively. These spectra suggest increased lipid and decreased DNA compared with  $C_0$ . The similarity between controls and NF can be explained because NF rats do not develop fibroadenoma.

The mean PF spectra,  $PF_3$ ,  $PF_{8-10}$ , and  $PF_{12-14}$  (Fig. 2 c.1–c.4), show an additional shoulder band at 1337  $\text{cm}^{-1}$  (DNA) and broader amide I compared with their respective week control and NF rat mean spectra. This suggests higher cell proliferation in PF compared with control and NF rats.  $PF_3$ – $C_0$ ,  $PF_{8-10}$ – $C_0$ , and  $PF_{12-14}$ – $C_0$  (Fig. 3 c.1–c.3) exhibit less intense DNA band compared with  $C_3$ – $C_0$ ,  $C_{8-10}$ – $C_0$ , and  $C_{12-14}$ – $C_0$  and  $NF_3$ – $C_0$ ,  $NF_{8-10}$ – $C_0$ , and  $NF_{12-14}$ – $C_0$ , respectively. This probably suggests higher cell proliferation in  $PF_3$ ,  $PF_{8-10}$ , and  $PF_{12-14}$  compared with  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ , and  $C_{20}$  and  $NF_3$ ,  $NF_{8-10}$ ,  $NF_{12-14}$ , and  $NF_{20}$ – $C_0$ , respectively. The high cell proliferation may indicate changes preceding fibroadenoma development.

The mean F spectrum (Fig. 3 c.4) show broad amide I,  $\delta$   $\text{CH}_2$  band at 1450  $\text{cm}^{-1}$ , and change in features in 1200–1400  $\text{cm}^{-1}$  region. F– $C_0$  difference spectra show positive DNA bands at 1480  $\text{cm}^{-1}$  and 1340  $\text{cm}^{-1}$  and negative lipid bands at 1260  $\text{cm}^{-1}$ , 1440  $\text{cm}^{-1}$ , and 1743  $\text{cm}^{-1}$ . These features suggest low lipid content and high DNA content with respect to  $C_0$ . This suggests dominance of proteins and DNA and thus, higher proliferation compared with  $C_0$ . Normal breast consists of mammary epithelium supported by mammary fat pad rich in lipids. Whereas, fibroadenoma is characterized by changes in protein profiles, increase in cell proliferation, and changes in breast architecture. This explains lipid dominance in control and variation in protein, increase in DNA, and loss of lipids in fibroadenoma. These findings corroborate well with earlier studies. The tentative spectral assignments are based on available literature.<sup>[22]</sup>

### Multivariate analysis

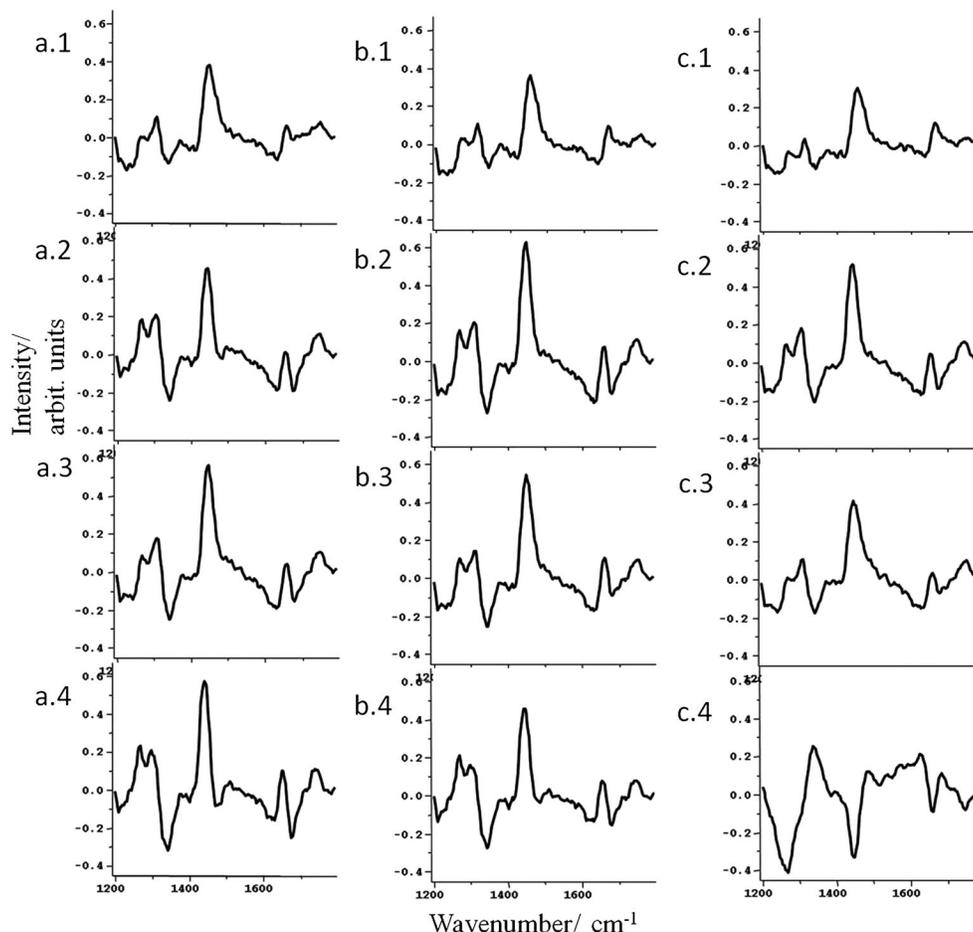
- PCA: Preprocessed spectra interpolated in 1200–1800  $\text{cm}^{-1}$  range were subjected to PCA for delineating trends in the data set. PCA of initial weeks (Fig. 4a) of all three groups –  $C_3$ ,  $NF_3$ , and  $PF_3$  – show no clustering. This indicates that  $C_3$ ,  $NF_3$ ,



**Figure 2.** Mean spectra of (a.1)  $C_0$ – $C_3$ , (a.2)  $C_{8-10}$ , (a.3)  $C_{12-14}$ , and (a.4)  $C_{20}$ ; (b.1)  $NF_3$ , (b.2)  $NF_{8-10}$ , (b.3)  $NF_{12-14}$ , and (b.4)  $NF_{20}$ ; (c.1)  $PF_3$ , (c.2)  $PF_{8-10}$ , (c.3)  $PF_{12-14}$ , (c.4) F. - C: control, NF: No Fibroadenoma, PF: Pre Fibroadenoma, F: Fibroadenoma, weeks post carcinogen/oil treatment shown in subscripts.

and  $PF_3$  spectra have similar features. PCA of final weeks (Fig. 4b),  $C_{20}$ ,  $PF_{20}$ , and F show three clusters. F cluster is distinct, while  $C_{20}$  and  $NF_{20}$  overlap. This suggests that frank fibroadenoma spectra can be distinguished from C and NF rat spectra. PCA of  $C_0$ ,  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ , and  $C_{20}$  (Fig. 4c) show three overlapping clusters –  $C_3$ ,  $C_{8-14}$ , and  $C_{20}$ . As explained earlier,  $C_0$  is phase of rapid cell proliferation, while  $C_3$  is the beginning of reproductive phase, wherein cell proliferation remains constant. However, in biological systems, phase boundaries – puberty to reproductive phase – are not sharp. Hence some cell proliferation similar to  $C_0$  is expected, which may explain the cluster of  $C_3$ .  $C_{8-14}$  belongs to the reproductive phase, hence no clustering is observed amongst this group.  $C_{20}$ , although part of the reproductive phase, may have lower cell proliferation, because cell proliferation slowly declines as rat approaches perimenopause

phase. This may explain  $C_{20}$  cluster. PCA of  $NF_0$ ,  $NF_3$ ,  $NF_{8-10}$ ,  $NF_{12-14}$ , and  $NF_{20}$  (Fig. 4d) shows clustering similar to controls and suggests spectral similarity between controls and NF. PCA of  $PF_0$ ,  $PF_3$ ,  $PF_{8-10}$ ,  $PF_{12-14}$ ,  $PF_{20}$ , and F (Fig. 4e) show four overlapping clusters –  $PF_3$ ,  $PF_{8-10}$ ,  $PF_{12-14}$ , and F. This may be caused by change in cell proliferation patterns that lead to fibroadenoma development. It is also observed that many  $PF_{8-10}$  and  $PF_{12-14}$  spectra overlap with F cluster, probably indicating spectral features similar to F in  $PF_{8-10}$  and  $PF_{12-14}$  groups. It is important to note that  $PF_3$ ,  $PF_{8-10}$ , and  $PF_{12-14}$  rats were visibly and palpably normal. Overlap of these groups with F group may thus be indicative of future fibroadenoma development. This possibility is supported by PCA of all groups (Fig. 4f) –  $C_0$ ,  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ ,  $C_{20}$ ,  $NF_0$ ,  $NF_3$ ,  $NF_{8-10}$ ,  $NF_{12-14}$ ,  $NF_{20}$ ,  $PF_0$ ,  $PF_3$ ,  $PF_{8-10}$ ,  $PF_{12-14}$ ,  $PF_{20}$ , and F. As seen in this figure, C and NF groups cluster away



**Figure 3.** Difference spectra of (a.1)  $C_3-C_0$ , (a.2)  $C_{8-10}-C_0$ , (a.3)  $C_{12-14}-C_0$ , and (a.4)  $C_{20}-C_0$ ; (b.1)  $NF_3-C_0$ , (b.2)  $NF_{8-10}-C_0$ , (b.3)  $NF_{12-14}-C_0$ , and (b.4)  $NF_{20}-C_0$ ; (c.1)  $PF_3-C_0$ , (c.2)  $PF_{8-10}-C_0$ , (c.3)  $PF_{12-14}-C_0$ , and (c.4)  $F-C_0$ . - C: control, NF: No Fibroadenoma, PF: Pre Fibroadenoma, F: Fibroadenoma, weeks post carcinogen/oil treatment shown in subscripts.

from F, while PF cluster overlap with F. Thus, results suggest possibility of distinguishing pretumor spectra from normal.

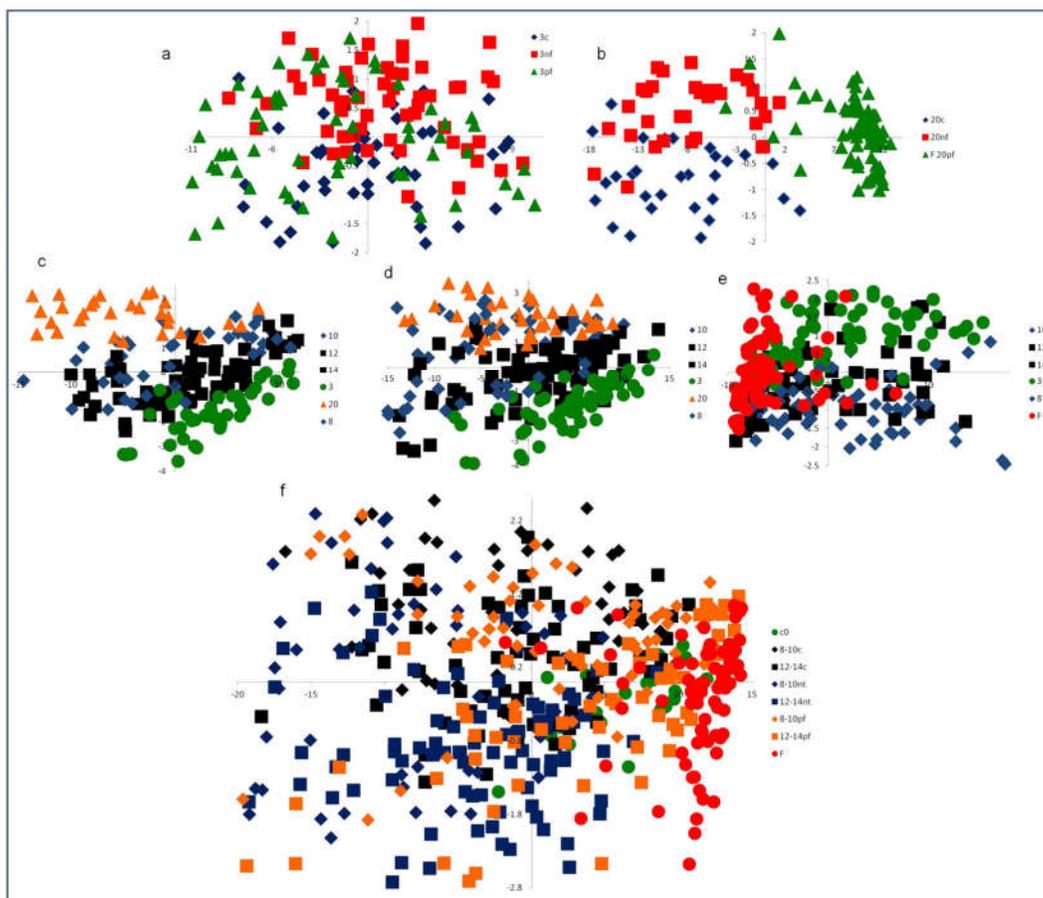
- (b) PC-LDA: to further explore the possibility of distinguishing pretumor spectra from control, PC-LDA was implemented. As mentioned earlier, several changes occur in breast during puberty and reproductive phase. Previous studies have shown that these age-related changes affect breast spectra.<sup>[19]</sup> In view of this, PC-LDA model was trained using spectra from control rats of different age groups –  $C_0$ ,  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ , and frank fibroadenoma. Only subcutaneous injection batch I spectra were used as training set, while the remaining was used as test data set. The PC-LDA scatter plot of  $C_0$ ,  $C_3$ ,  $C_{8-10}$ ,  $C_{12-14}$ , and F (Fig. 5) shows two distinct clusters – C and F. Some  $C_0$  spectra overlap with F. As mentioned earlier, high rate of cell proliferation during  $C_0$  and during fibroadenoma development may explain this. The confusion matrix for PC-LDA model LOOCV is shown in Table 2. LOOCV was carried out to evaluate the results obtained by PC-LDA. LOOCV builds a model based on all observations but one and tests the left out observation against the model built; this is repeated until all observations are left out once. The performance is estimated in terms of classification efficiency, which is percentage of spectra from each group that are correctly classified. In this table, 3/22  $C_0$  misclassify with F whereas 16/56 F misclassify with  $C_0$ . Misclassification between F and  $C_0$  may be explained based on rapid cell

proliferation characteristic of these groups. Misclassification between different control age groups can be attributed to age-related changes reported earlier. Overall, results suggest that F can be classified from controls with 71% efficiency.

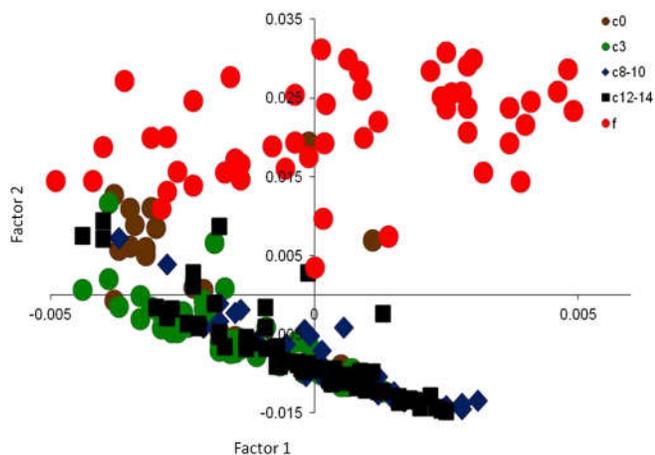
In the previous section, PCA results suggest that control and NF groups cluster away from F, while PF cluster overlap with F. To further support this observation, the PC-LDA model was evaluated with independent test data. As seen in Table 3a, none of the spectra from gavage  $C_0$ ,  $C_3$ ,  $C_8$ ,  $C_{12-14}$ , and  $C_{18}$  are predicted as F. One spectrum out of 20 spectra from subcutaneous injection batch II  $C_7$  is predicted as F, while no spectra from  $C_3$  and  $C_{11}$  are predicted as F (Table 3b). None of the spectra from subcutaneous injection batch III  $C_3$ ,  $C_7$ , and  $C_{11}$  and dusting  $C_3$  and  $C_7$  are predicted as F (Table 3c). These results support the PCA findings that control spectra are different from F spectra.

In case of NF (Table 4), 2/108 spectra of gavage  $NF_3$  are predicted as F, while 1/275 spectra of gavage  $NF_{12-14}$  is predicted as F. None of the spectra from gavage  $NF_7$ ,  $NF_{20}$ , subcutaneous injection batch II  $NF_3$ ,  $NF_{8-10}$ , and  $NF_{12-14}$  are predicted as F. Thus, NF group spectra differ from F group spectra, as observed in PCA.

In case of PF (Table 5), 7/63, 5/81, and 8/60 spectra of subcutaneous injection batch I  $PF_3$ ,  $PF_{8-10}$ , and  $PF_{12-14}$  are predicted as F while 6/64, 3/63, and 7/55 spectra from subcutaneous injection batch II are predicted as F. Spectra 3/49 and 2/34 from subcutaneous injection batch III  $PF_7$  and  $PF_{11}$ , respectively, 4/36, 3/36, and 7/35



**Figure 4.** Principal component analysis of (a) initial weeks, C<sub>3</sub> (blue), NF<sub>3</sub>(red), and PF<sub>3</sub> (green); (b) final weeks, C<sub>20</sub> (blue), NF<sub>20</sub> (red), and F (green); (c) C<sub>3</sub> (green), C<sub>8-10</sub> (blue), C<sub>12-14</sub> (black), and C<sub>20</sub> (orange); (d) NF<sub>3</sub> (green), NF<sub>8-10</sub> (blue), NF<sub>12-14</sub> (black), NF<sub>20</sub> (orange); (e) PF<sub>3</sub> (green), PF<sub>8-10</sub> (blue), PF<sub>12-14</sub> (black), and F (red); and (f) C<sub>0</sub> (green), C<sub>3</sub> (black), C<sub>8-10</sub> (black), C<sub>12-14</sub> (black), C<sub>20</sub> (black), NF<sub>3</sub> (blue), NF<sub>8-10</sub> (blue), NF<sub>12-14</sub> (blue), NF<sub>20</sub> (blue), PF<sub>3</sub> (orange), PF<sub>8-10</sub> (orange), PF<sub>12-14</sub> (orange), NF<sub>20</sub> (orange), and F (red) (x-axis, score of factor 1; y-axis, score of factor 2). - C: control, NF: No Fibroadenoma, PF: Pre Fibroadenoma, F: Fibroadenoma, weeks post carcinogen/oil treatment shown in subscripts, colors in brackets correspond to legend colors.



**Figure 5.** Principal component-linear discriminant analysis scatter plot of C<sub>3</sub>, C<sub>8-10</sub>, C<sub>12-14</sub>, C<sub>20</sub>, and F. - C: control, NF: No Fibroadenoma, PF: Pre Fibroadenoma, F: Fibroadenoma, weeks post carcinogen/oil treatment shown in subscripts.

spectra from dusting PF<sub>3</sub>, PF<sub>7</sub>, and PF<sub>12</sub>, respectively, were predicted as F. Thus, PF spectra are predicted as F more frequently compared with control and NF. Thus, results corroborate the outcome of PCA. Overall, results suggest possibility of distinguishing pretumor spectra from controls.

**Table 2.** Principal component-linear discriminant analysis confusion matrix scatter plot of C<sub>3</sub>, C<sub>8-10</sub>, C<sub>12-14</sub>, C<sub>20</sub>, and F for leave one spectrum out cross validation (LOOCV) (diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions). Sample size is shown in brackets

LOOCV (no. of animals used, no. of spectra)	C <sub>0</sub>	C <sub>3</sub>	C <sub>8-10</sub>	C <sub>12-14</sub>	F
C <sub>0</sub> (10, 22)	12	6	1	0	3
C <sub>3</sub> (10, 49)	10	22	7	10	0
C <sub>8-10</sub> (10, 44)	2	6	25	11	0
C <sub>12-14</sub> (10, 69)	8	15	30	16	0
F (7, 56)	16	0	0	0	40

Although the results show that PF rat spectra are predicted as F, it should be noted that very few PF spectra are predicted as F, while majority are predicted as C/NF. This is probably due to heterogeneity of the sample. Breast is a large organ, but tumor occurs in a small region. Thus, an afflicted breast has largely normal regions with a small abnormal region. Because spectra were acquired from the several sites, they represent heterogeneity. As discussed by Malini *et al.*<sup>[23]</sup> and Stone *et al.*,<sup>[24]</sup> a heterogeneous sample can be considered abnormal even if one spectrum is abnormal. Similar practice is used in histopathological assessment, wherein even if a single region of a single slide is abnormal, the whole sample is

**Table 3.** Evaluation of principal component–linear discriminant analysis model with control test data set (a) subcutaneous injection batch II control, (b) subcutaneous injection batch III controls, (c) dusting controls, and (d) gavage controls

Test data (no. of animals used, no. of spectra)	C <sub>0</sub>	C <sub>3</sub>	C <sub>8–10</sub>	C <sub>12–14</sub>	C <sub>20</sub>	F
(a)						
C <sub>3</sub> (5, 20)	4	9	3	4	0	0
C <sub>7</sub> (5, 20)	3	7	5	4	0	1
C <sub>11</sub> (5, 24)	5	12	3	4	0	0
(b)						
C <sub>3</sub> (5, 26)	6	8	6	6	0	0
C <sub>7</sub> (4, 20)	8	5	2	5	0	0
C <sub>11</sub> (2, 10)	1	1	7	1	0	0
(c)						
C <sub>3</sub> (4, 20)	6	1	8	5	0	0
C <sub>7</sub> (5, 25)	3	10	5	7	0	0
(d)						
C <sub>3</sub> (9, 41)	6	9	16	10	0	0
C <sub>8–10</sub> (4, 20)	0	3	14	3	0	0
C <sub>12–14</sub> (6, 36)	5	2	22	7	0	0
C <sub>20</sub> (5, 20)	0	2	14	4	0	0

**Table 5.** Evaluation of principal component–linear discriminant analysis model with pre-fibroadenoma (PF) test data set (a) subcutaneous injection batch I PF, (b) subcutaneous injection batch II PF, (c) subcutaneous injection batch III PF, and (d) dusting PF

Test data (no. of animals used, no. of spectra)	C <sub>0</sub>	C <sub>3</sub>	C <sub>8–10</sub>	C <sub>12–14</sub>	C <sub>20</sub>	F
(a)						
PF <sub>3</sub> (7, 63)	23	14	7	12	0	7
PF <sub>8–10</sub> (6, 61)	20	6	21	10	0	4
PF <sub>12–14</sub> (6, 60)	20	9	12	11	0	8
(b)						
PF <sub>3</sub> (7, 64)	48	7	0	3	0	6
PF <sub>7</sub> (7, 63)	45	12	0	3	0	3
PF <sub>11</sub> (6, 55)	30	13	2	3	0	7
(c)						
PF <sub>3</sub> (5, 46)	6	12	11	17	0	0
PF <sub>7</sub> (5, 49)	26	14	2	4	0	3
PF <sub>11</sub> (4, 34)	6	5	17	4	0	2
(d)						
PF <sub>3</sub> (4, 36)	8	4	12	8	0	4
PF <sub>7</sub> (4, 36)	28	4	0	1	0	3
PF <sub>12</sub> (4, 35)	19	6	1	2	0	7

declared abnormal. Taking this into consideration, a rat wise analysis was performed for the current data. All spectra from each rat irrespective of time of spectra acquisition were considered, and even if one spectrum was found to be abnormal (predicted as F), the rat was declared abnormal, that is, will develop tumor. Using this criterion, 82% rats could be correctly predicted to develop tumor in the future. Only 5% rats that did not develop were predicted wrongly as abnormal. Thus, sensitivity and specificity of RS to predict tumor occurrence in this study was 82% and 95%, respectively. This further strengthens the evidence for feasibility of breast cancer screening using RS.

Early detection of breast cancers results in improved prognosis, but the currently available screening tools have several disadvantages. A search for better screening techniques has instigated investigation in several diverse fields such as genomics, proteomics, and optical spectroscopy.<sup>[25,26]</sup> Optical spectroscopic techniques have an edge over others as screening tools because these techniques are rapid, objective, and amenable to *in vivo* applications. Several studies have shown the potential of different optical spectroscopic techniques in identification of premalignant lesions.<sup>[26–29]</sup>

**Table 4.** Evaluation of principal component–linear discriminant analysis model with no-fibroadenoma (NF) test data set (a) subcutaneous injection batch I NF and (b) gavage NF

Test data (no. of animals used, no. of spectra)	C <sub>0</sub>	C <sub>3</sub>	C <sub>8–10</sub>	C <sub>12–14</sub>	C <sub>20</sub>	F
(a)						
NF <sub>3</sub> (7, 63)	12	29	10	12	0	0
NF <sub>8–10</sub> (6, 64)	1	7	44	12	0	0
NF <sub>12–14</sub> (7, 84)	5	26	39	14	0	0
(b)						
NF <sub>3</sub> (7, 108)	44	29	14	19	0	2
NF <sub>8–10</sub> (5, 81)	11	27	22	21	0	0
NF <sub>12–14</sub> (17, 275)	53	63	104	54	0	1
NF <sub>20</sub> (8, 126)	12	28	65	21	0	0

But, such premalignant lesions can only help risk estimation and cannot predict actual tumor development.<sup>[41]</sup> Therefore, the current study aims to analyze spectral data based on outcome – success or failure to develop tumor and establish a correlation between spectral changes and tumor appearance. In order to achieve this, rats were injected subcutaneously with carcinogen DMBA in their left inguinal mammary gland. The advantage of this methodology is that tumor appears approximately around the site of injection. Spectra were acquired from the whole left inguinal breast 0, 3, 8–10, and 12–14 weeks after carcinogen treatment from both treated rats and their corresponding controls. Seven treated rats developed tumor approximately 18 weeks posttreatment (PF), while seven rats did not develop any abnormality despite carcinogen treatment throughout the study (NF). The control rats also did not exhibit any abnormality throughout the study (C). Spectra were also acquired from frank tumor (F). PCA of all spectra acquired showed that C and NF formed clusters distinct from F and that, while majority of PF spectra overlapped with C/NF, several PF spectra overlapped with F.

To further evaluate these findings, a PC-LDA model was trained using C and F spectra. PC-LDA showed that F can be classified from C with 71% efficiency. The model was validated using independent test data. The test data consisted of spectra acquired at rat ages different from that used for training the model, varying time points from time of tumor appearance, and different protocols for inducing carcinogenesis. Despite being subjected to a complex test data set, the model could correctly predict all controls as controls, showing high specificity of RS. Several PF spectra were predicted as F corroborating with the results of PCA. Taking into consideration that tumor bearing breast is heterogeneous, which is largely normal with a small region harboring the tumor showing abnormality, a rat-wise analysis was performed, wherein a rat was declared to ‘develop tumor in the future’, even if one spectrum irrespective of time of spectra acquisition was predicted as F. Using this criterion, the specificity and sensitivity of RS in predicting tumor was found to be 95% and 82%, respectively. These results suggest the possibility of detecting tumor early.

Combined with developments in deep RS,<sup>[11,30–32]</sup> it may be possible to identify biochemical changes indicating tumor development at different depths. Moreover, technologies are now emerging that may allow quick imaging of the whole breast, replacing the tedious step by step spectra acquisition procedures. Recently, Schmälzlin *et al.*<sup>[33]</sup> have reported Raman imaging with a fiber – coupled multichannel spectrograph that allows capture of entire Raman image with one single exposure and chemical mapping without the need for scanning procedure. A Raman chemical map of the whole breast can be obtained using the fiber-coupled multichannel spectrograph system, and abnormal spectra (that are predicted as F) can be identified. Images of breasts that are not likely to develop tumor will show normal map, whereas those that are in the future going to develop tumor will show a map with largely normal areas harboring small areas of abnormality corresponding to the site of future tumor appearance. Thus, emerging technologies combined with sensitivity of RS to pretumor changes may allow identification of precancerous changes in the whole breast volume. Further studies in this area may help develop this technique as an alternative/adjunct breast cancer screening tool.

### Acknowledgements

The authors would like to acknowledge Mr Pramod N. Tawde for his help in animal-handling and spectra acquisition. The Raman spectrometer employed in the study was procured from 'Development of in vivo laser Raman spectroscopy methods for diagnosis of oral precancerous and cancerous conditions' (BT/PRI11282/MED/32/83/2008), Department of Biotechnology, Government of India.

### References

- [1] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, DM. Parkin, D. Forman, F. Bray, GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; **2013**. Available from: <http://globocan.iarc.fr>, accessed on 04/06/2015
- [2] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, D. Forman, *CA Cancer J. Clin.* **2011**, *61*, 69.
- [3] N. N. A. Howlader, M. Krapcho, J. Garshell, D. Miller, S. F. Altekruse, C. L. Kosary, M. Yu, J. Ruhl, Z. Tatalovich, A. Mariotto, D. R. Lewis, H. S. Chen, E. J. Feuer, K. A. Cronin (eds) National Cancer Institute. *Bethesda, MD* **2013**.
- [4] C. B. Matsen, L. A. Neumayer, *JAMA Surg.* **2013**, *148*, 971.
- [5] A. Qaseem, V. Snow, K. Sherif, M. Aronson, K. B. Weiss, D. K. Owens, *Ann. Intern. Med.* **2007**, *146*, 511.
- [6] D. Evers, B. Hendriks, G. Lucassen, T. Ruers, *Future Oncol.* **2012**, *8*, 307.
- [7] A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari, M. S. Feld, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12371.
- [8] A. S. Haka, Z. Volynskaya, J. A. Gardecki, J. Nazemi, J. Lyons, D. Hicks, M. Fitzmaurice, R. R. Dasari, J. P. Crowe, M. S. Feld, *Cancer Res.* **2006**, *66*, 3317.
- [9] M. Chowdary, K. Kalyan Kumar, S. Mathew, L. Rao, C. M. Krishna, J. Kurien, *Biopolymers* **2009**, *91*, 539.
- [10] J. Horsnell, P. Stonelake, J. Christie-Brown, G. Shetty, J. Hutchings, C. Kendall, N. Stone, *Analyst* **2010**, *135*, 3042.
- [11] N. Stone, R. Baker, K. Rogers, A. W. Parker, P. Matousek, *Analyst* **2007**, *132*, 899.
- [12] A. Sahu, K. Dalal, S. Naglot, P. Aggarwal, C. M. Krishna, *PLoS One* **2013**, *8*, e78921.
- [13] S. Rubina, M. Amita, R. Bharat, C. M. Krishna, *Vib. Spectrosc.* **2013**, *68*, 115.
- [14] S. P. Singh, A. Sahu, A. Deshmukh, P. Chaturvedi, C. M. Krishna, *Analyst* **2013**, *138*, 4175.
- [15] S. K. Teh, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, Z. Huang, *Br. J. Cancer* **2008**, *98*, 457.
- [16] C. Zheng, L. Liang, S. Xu, H. Zhang, C. Hu, L. Bi, Z. Fan, B. Han, W. Xu, *Anal. Bioanal. Chem.* **2014**, *406*, 5425.
- [17] T. Bhattacharjee, P. Kumar, G. Maru, A. Ingle, C. M. Krishna, *Lasers Med. Sci.* **2014**, *29*, 325.
- [18] T. Bhattacharjee, G. Maru, A. Ingle, C. M. Krishna, *J. Biomed. Opt.* **2013**, *18*, 047004.
- [19] T. Bhattacharjee, G. Maru, A. Ingle, C. M. Krishna, *Vib. Spectrosc.* **2013**, *67*, 80.
- [20] L. Raniero, R. Canevari, L. Ramalho, F. Ramalho, E. dos Santos, R. Bitar, K. J. Jalkanen, H. Martinho, A. Martin, *Theor. Chem. Acc.* **2011**, *130*, 1239.
- [21] M. M. Richert, K. L. Schwertfeger, J. W. Ryder, S. M. Anderson, *J. Mammary Gland Biol. Neoplasia* **2000**, *5*, 227.
- [22] F. S. Parker, *Applications of Infrared, Raman, and Resonance Raman Spectroscopy in Biochemistry*, Plenum Press, New York, **1983**.
- [23] R. Malini, K. Venkatakrishna, J. Kurien, M. K. Pai, L. Rao, V. Kartha, C. M. Krishna, *Biopolymers* **2006**, *81*, 179.
- [24] M. Sattlecker, C. Bessant, J. Smith, N. Stone, *Analyst* **2010**, *135*, 895.
- [25] P. R. Srinivas, S. Srivastava, S. Hanash, G. L. Wright, *Clin. Chem.* **2001**, *47*, 1901.
- [26] M. M. Stephen, J. L. Jayanthi, N. G. Unni, P. E. Kolady, V. T. Beena, P. Jeemon, N. Subhash, *BMC Cancer* **2013**, *13*, 278.
- [27] L. Yang, W.-T. Liu, H. Wu, C. Wang, B. Ping, D.-R. Shi, *BioMed Res. Int.* **2011**, *2011*, 214781.
- [28] I. J. Bigio In LEOS Annual Meeting Conference Proceedings, 2009. *LEOS'09. IEEE; IEEE: 2009*, p 268.
- [29] T. Denkçeken, T. Simsek, G. Erdogan, E. Pestereli, S. Karaveli, D. Ozel, U. Bilge, M. Canpolat, *Biomed. Eng., IEEE Trans* **2013**, *60*, 123.
- [30] R. Baker, P. Matousek, K. L. Ronayne, A. W. Parker, K. Rogers, N. Stone, *Analyst* **2007**, *132*, 48.
- [31] P. Matousek, N. Stone, *J. Biomed. Opt.* **2007**, *12*, 024008.
- [32] N. Stone, P. Matousek, *Cancer Res.* **2008**, *68*, 4424.
- [33] E. Schmälzlin, B. Moralejo, M. Rutowska, A. Monreal-Ibero, C. Sandin, N. Tarcea, J. Popp, M. M. Roth, *Sensors* **2014**, *14*, 21968.

### Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website.