

Journal of Biomedical Optics

BiomedicalOptics.SPIEDigitalLibrary.org

***Ex vivo* Raman spectroscopic study of breast metastatic lesions in lungs in animal models**

Tanmoy Bhattacharjee
Sneha Tawde
Rasika Hudlikar
Manoj Mahimkar
Girish Maru
Arvind Ingle
C. Murali Krishna

Ex vivo Raman spectroscopic study of breast metastatic lesions in lungs in animal models

Tanmoy Bhattacharjee, Sneha Tawde, Rasika Hudlikar, Manoj Mahimkar, Girish Maru, Arvind Ingle, and C. Murali Krishna*

Advanced Centre for Treatment Research and Education in Cancer (ACTREC), Tata Memorial Centre (TMC), Kharghar, Navi Mumbai 410210, India

Abstract. The lung is one of the most common sites of metastases, with approximately 50% of patients with extrathoracic cancer exhibiting pulmonary metastases. Correct identification of the metastatic status of a lung lesion is vital to therapeutic planning and better prognosis. However, currently available diagnostic techniques, such as conventional radiography and low dose computed tomography (LDCT), may fail to identify metastatic lesions. Alternative techniques such as Raman spectroscopy (RS) are hence being extensively explored for correct diagnosis of metastasis. The current *ex vivo* study aims to evaluate the ability of a fiber optic-based Raman system to distinguish breast cancer metastasis in lung from primary breast and lung tumor in animal models. In this study, spectra were acquired from normal breast, primary breast tumor, normal lung, primary lung tumor, and breast cancer metastasis in lung tissues and analyzed using principal component analysis and principal component-linear discriminant analysis. Breast cancer metastasis in lung could be classified with 71% classification efficiency. Approximately 6% breast metastasis spectra were misclassified with breast tumor, probably due to the presence of breast cancer cells in metastasized lungs. Test prediction results show 64% correct prediction of breast metastasis, while 13% breast metastasis spectra were wrongly predicted as breast tumor, suggesting the possible influence of breast cancer cells. Thus, findings of this study, the first of such explorations, demonstrate the potential of RS in classifying breast metastasis in lungs from primary lung and primary breast tumor. Prospective evaluation on a larger cohort with better multivariate analysis, combined with LDCT and recently developed real-time *in vivo* probes, RS can play a significant role in nonsurgical screening of lesions, which can lead to individualized therapeutic regimes and improved prognoses. © 2015 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JBO.20.8.085006](https://doi.org/10.1117/1.JBO.20.8.085006)]

Keywords: breast cancer metastasis; primary lung tumor; primary breast tumor; Raman spectroscopy; principal component analysis; principal component-linear discriminant analysis.

Paper 150233RR received Apr. 9, 2015; accepted for publication Aug. 4, 2015; published online Aug. 21, 2015.

1 Introduction

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 such as 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 The latest report estimated 2,21,200 and 1,58,040 new cases and deaths, respectively, in the United States during 2015.² Most of the lung cancer cases are detected at advanced stages of the disease, resulting in a 5-year survival rate as low as 16%. Studies have shown better prognosis with the early detection of lung cancer.³ Lung cancer screening using sputum cytology and chest radiography did not result in a reduction of advanced lung cancer cases or deaths.^{4,5} The current US Preventive Services Task Force guidelines recommend annual screening for lung cancer with low-dose computed tomography (LDCT) in adults.⁶ Randomized studies have significantly indicated fewer lung cancer deaths in a cohort screened using LDCT compared to the control group.⁷ One of the disadvantages of LDCT is the difficulty in a confident diagnosis of pulmonary metastasis.⁸ Conventional radiography

also fails to distinguish primary from metastatic lesions in two-third of the cases.⁹ Differential diagnosis between primary lung lesion and breast metastatic lesion in lung has been reported to be especially difficult.¹⁰⁻²¹ 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-1, a known strategy to identify lung malignancy and on the other hand, some breast cancers do not stain for estrogen receptor, an indicator for the 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 a differential diagnosis is vital for effective therapeutic intervention and favorable prognosis. Hence, there is a need for sensitive, rapid, objective, and cost-effective alternate tools for diagnosis of metastasis.

Raman spectroscopy (RS), a sensitive, rapid, objective technique, has shown promising results in the diagnosis of cervix, gastrointestinal, brain, oral, skin, lung, breast, and several other cancers.²²⁻³⁴ The technique has also been explored to study metastasis. Oliveira et al. have shown the feasibility of differentiating primary and metastatic cutaneous melanoma, while Terentis et al. have demonstrated discrimination of live human

*Address all correspondence to: C. Murali Krishna, E-mail: mchilakapati@actrec.gov.in, pittu1043@gmail.com

metastatic melanoma cells from skin fibroblasts using Raman microspectroscopy.^{27,35} Another Raman microspectroscopic study has shown a distinction between metastatic and nonmetastatic cell lines.^{36,37} A similar approach has been employed by Fullwood et al. to study metastatic brain tumors and has explored the possibility of identifying primary sites of origin.³⁸ Stone et al. have studied lymph node metastasis in breast cancer using fiber-optic and Raman microscopic techniques.^{33,39,40}

Recently, Short et al. have reported the development of a probe to collect real-time *in vivo* lung spectra and have successfully acquired spectra of lungs.⁴¹ This may prove to be an invaluable nonsurgical adjunct to LDCT, wherein lesions detected by LDCT may be categorized into primary lung cancer and pulmonary metastases using RS. In light of this, the current *ex vivo* study aims to evaluate the ability of a fiber optic-based Raman system to distinguish metastatic lesions from primary lung tumors in animal models. Breast cancer metastasis to lungs is the most common of several exthoracic cancers that metastatize to lungs. Moreover, as discussed earlier, differential diagnosis between lung cancer and breast metastasis in lung has been reported to be very difficult, but very necessary from a therapeutic and prognostic point of view. Therefore, in this study, experimental breast cancer metastasis in the lungs was implemented. To the best of our knowledge, this is the first study to attempt classification of metastatic breast lesions in lung from primary lung as well as breast tumors using fiber-optic based-RS. Spectra were acquired from breast metastatic lesions in lung, primary lung tumor, and normal lung. For robust analysis, spectra from normal breast and primary breast tumor were also incorporated in the study. Spectra from all five groups were analyzed using principal component analysis (PCA) and principal component-linear discriminant analysis (PC-LDA). The findings of the study are reported in this manuscript.

2 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 ($n = 4$, where “ n ” is the number of animals). Use of these models has been extensively reported.^{42,43} Briefly, the female mice of this strain contain MMTV virus in their milk, which gets transferred into pups. The viral DNA gets integrated into their breast cells and acts as an oncogene resulting in spontaneous tumors at a later age. 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.⁴⁴ 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 (4×10^6 cells) into new 8 week old C3H Jax mice. After 3 weeks, the mice were sacrificed by cervical dislocation, the lungs were harvested ($n = 8$) and used for spectroscopy. 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, Government of India guidelines. All animals were housed under standard laboratory conditions, fed a diet of in-house-prepared pellets containing natural ingredients such as wheat, roasted bengal gram, casein, skimmed milk powder, vitamins, minerals, probiotics, and ground nut oil and provided with water *ad libitum*.

2.1 Spectra Acquisition

Approximately 8 to 10 spectra were acquired from normal breast, normal lung, primary breast tumor, and primary lung tumor tissues; while approximately 16 to 22 spectra were recorded from metastatic breast lesions. All spectra were acquired using Raman spectrometer as described elsewhere.²⁶ Briefly, this system consists of a diode laser (PI-ECL-785-300-FC, Process Instruments) of 785-nm wavelength as the excitation source, a high efficiency spectrograph (HE-785,

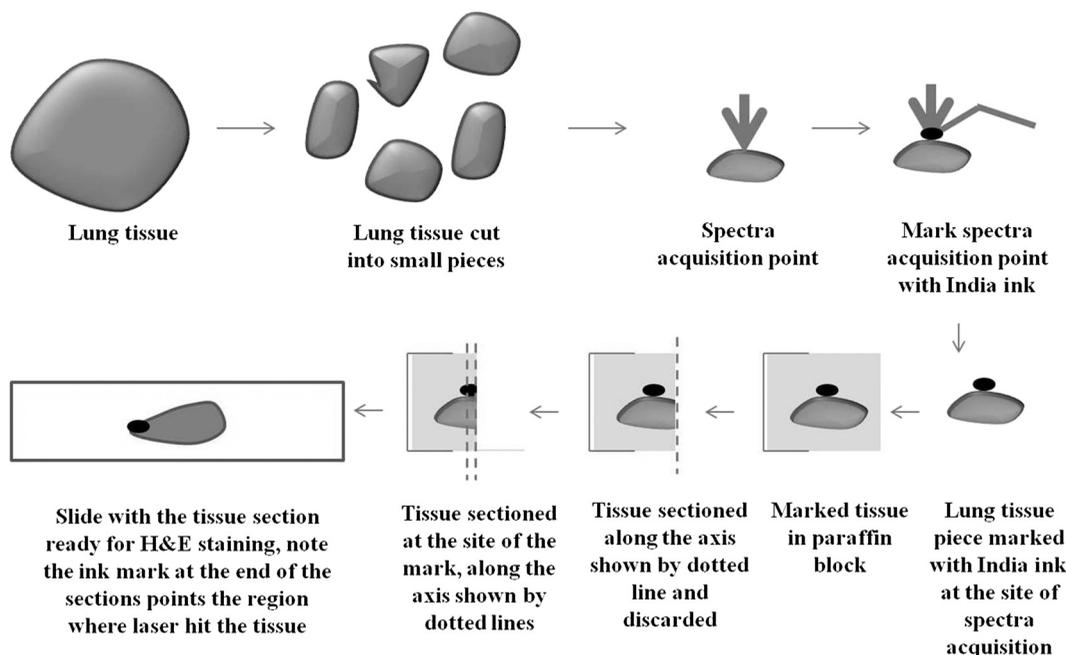


Fig. 1 Methodology for tissue sectioning and H&E staining.

Jobin-Yvon-Horiba, France) with a fixed 950 gr/mm grating coupled with a charge-coupled device (CCD) (CCD-1024X256-BIDD-SYN, Synapse). The spectrograph has no movable parts and the spectral resolution is $\sim 4 \text{ cm}^{-1}$. A commercial RamanProbe (RPS 785/ 12-5, In Photonics Inc., Downy Street) consisting of an excitation and a collection fiber (NA-0.40) of diameters 105 and 200 μm , respectively, was used to couple the excitation source and detection system. This probe utilizes a backscattering ($\theta = 180 \text{ deg}$) sampling geometry. The estimated spot size and depth of penetration as per the manufacturer's specifications is 105 μm and 1 mm, respectively. Spectral acquisition parameters were: $\lambda_{\text{ex}} = 785 \text{ nm}$, laser power-80 mW, spectra were integrated for 15 s and averaged over three accumulations.

2.2 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 the laser hits the tissue was marked with India ink and fixed with 2% glacial acetic acid. Paraffin embedded blocks were 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 the laser interacted with the tissue, the pathology and spectra can be directly correlated (Fig. 1).

2.3 Spectral Preprocessing

Raman spectra from all samples were corrected for CCD response with a National Institute of Science and Technology certified Standard Reference Material 2241 (SRM 2241) followed by the subtraction of background signals from optical elements and substrate. Background spectra were acquired by removing the sample and keeping everything else constant. These signals were acquired before collecting spectra from each sample. The subtraction was performed in LabSpec. To remove interference of the slow moving background, first derivatives of spectra (Savitzky-Golay method and window size 3) were computed.²⁶ Spectra were interpolated in 1200 to 1800 cm^{-1} region, vector-normalized, and used as input for multivariate analysis. The utility of this spectral range in distinguishing normal from malignant tissues of oral,²⁶ cervix⁴⁵ and breast⁴⁶ cancers, discrimination of anatomical sites,⁴⁶ study of physiological conditions such as pregnancy, lactation⁴⁷ and ageing,⁴⁸ as well as prediction of tumor appearance⁴⁹ has been demonstrated earlier. The range also helps to avoid fiber spectral artifacts.

2.4 Multivariate Analysis

First derivative, vector normalized spectra were subjected to multivariate unsupervised PCA and supervised PC-LDA.

First derivatives of preprocessed spectra were subjected to supervised PC-LDA. PCA is a 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 the mean-centered data. Because of their orthogonal characteristics, the first few PCs are enough to represent maximum data variance. For visual

discrimination, we project each of the spectra in the newly formed co-ordinate space of these selected PCs. Although 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 the development of PCs, 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

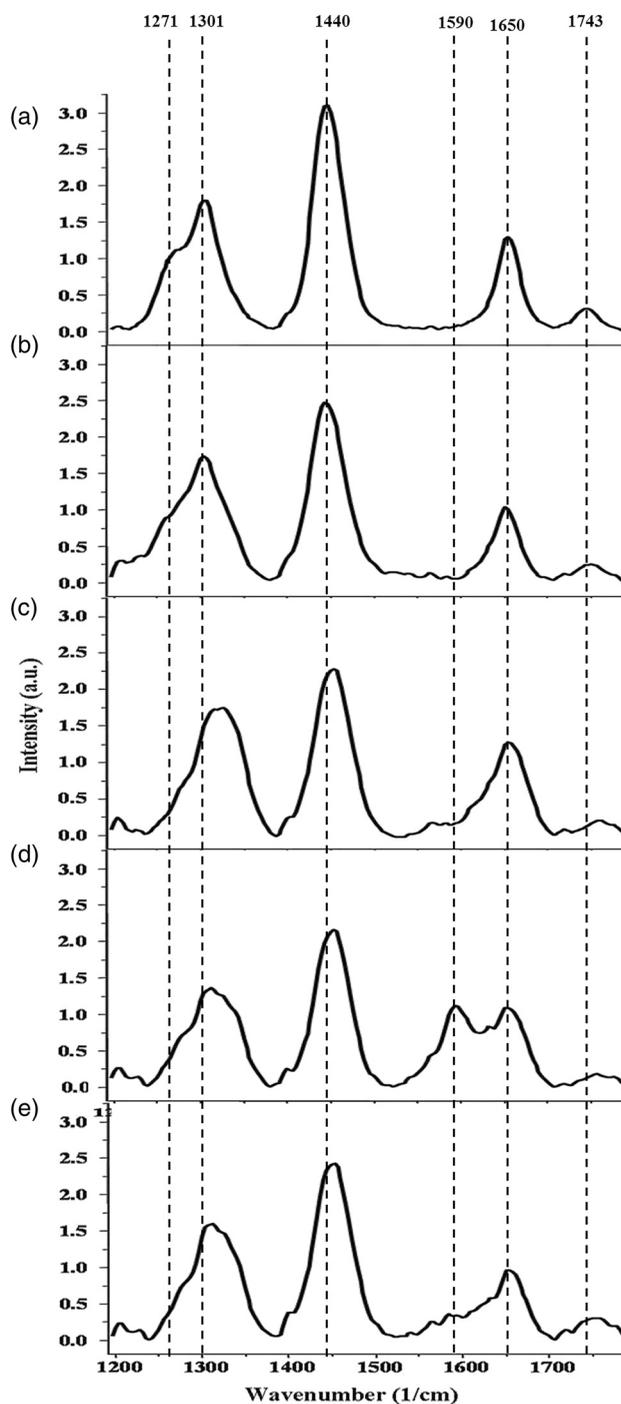


Fig. 2 Mean spectra interpolated in 1200 to 1800 cm^{-1} region of (a) breast control, (b) primary breast tumor, (c) breast cancer metastasis in lung (d) lung control, and (e) primary lung tumor.

further identified as an 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 the Euclidean/RMS or Mahalanobis distance of transformed test spectra and the mean of the transformed input dataset. In this study we have employed Mahalanobis distance for class prediction, since it handles nonlinearity well.²² 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 a maximum variance among 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).

LOOCV is a type of rotation estimation mainly used 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. 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 is correctly classified.

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 the 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 gives a visual understanding of the classification pattern in the data.^{50–52} Algorithms for these analyses were implemented in MATLAB® (Mathworks Inc.) based software using in-house codes.²⁶

Mean spectra were computed from the background subtracted spectra prior to derivatization for each class by averaging *Y*-axis variations and keeping *X*-axis constant for each class, and were baseline corrected by fitting a fifth-order polynomial function. These baselines corrected spectra were vector normalized and then used for computing the mean spectra. Difference spectra were also calculated by subtracting the mean spectra of breast and lung tumors from breast control and lung control, respectively, and by subtracting the mean breast metastasis spectrum from mean spectra of control breast, breast tumor, control lung, and lung tumor.

3 Results and Discussion

3.1 Spectral Analysis

Mean spectra: the spectral features of the mean control breast spectrum [Fig. 2(a)]—1743 cm^{-1} (C=O ester); 1653 cm^{-1} (amide I); 1440 cm^{-1} (δ CH₂); 1301 cm^{-1} (τ CH₂); and 1271 cm^{-1} —can be attributed to lipids. The mean breast tumor spectrum [Fig. 2(b)]

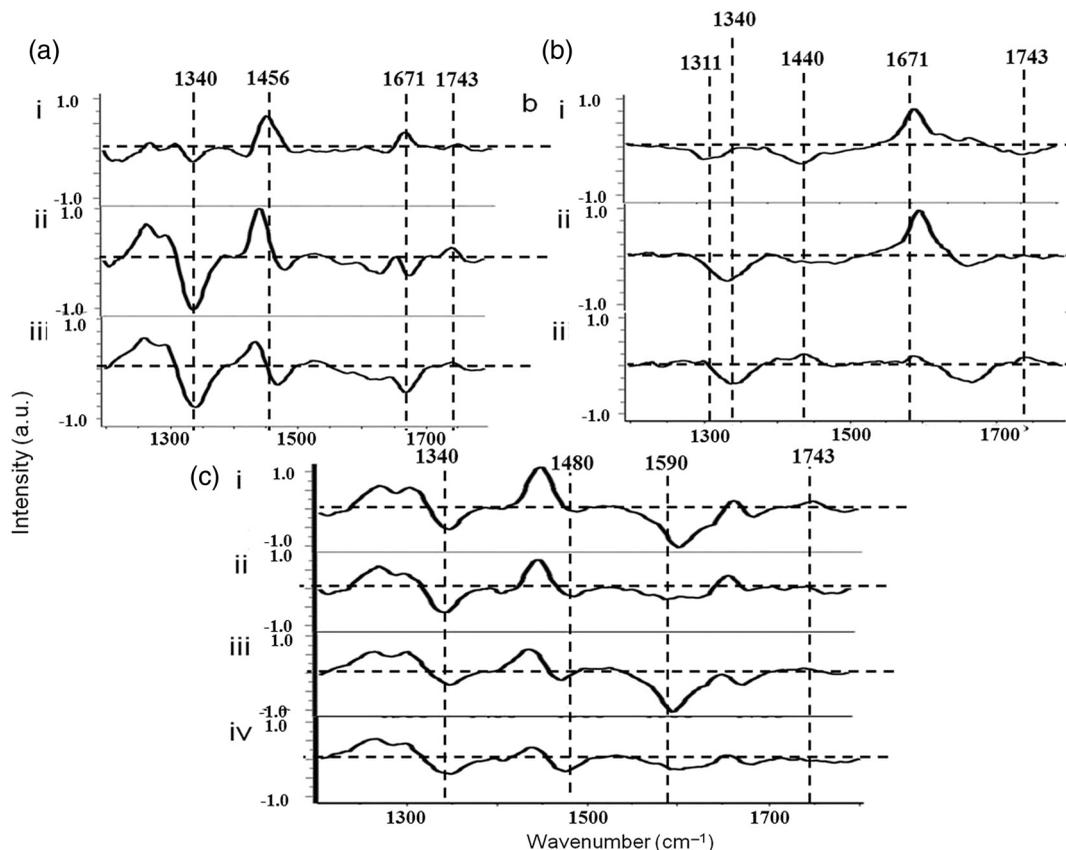


Fig. 3 Difference spectra (a) i—breast control—breast tumor, ii—breast control—breast metastasis, iii—breast tumor—breast metastasis, (b) i—lung control—lung tumor, ii—lung control—breast metastasis, iii—lung tumor—breast metastasis, and (c) i—breast control—lung control, ii—breast control—lung tumor, iii—breast tumor—lung control, and iv—breast tumor—lung tumor.

shows broad amide I and features in the 1200 to 1400 cm^{-1} region, suggesting dominance of proteins and DNA in tumor. Normal breast consists of mammary epithelium and lipid rich mammary fat pad.⁵³ This explains lipid dominance in normal breast. Tumor is characterized by changes in protein profiles, breast architecture, and increase in cell proliferation.⁵⁴ This may explain variation in protein, increase in DNA, and loss of lipids in tumor. These findings corroborate well with earlier studies.^{55,56} The mean control lung spectrum [Fig. 2(d)] exhibits features at 1650 cm^{-1} (amide I), 1311 cm^{-1} , 1335 cm^{-1} , 1450 cm^{-1} (δ CH₂), 1301 cm^{-1} (τ CH₂), and 1590 cm^{-1} as reported earlier.⁵⁷ The 1590 cm^{-1} band has been suspected to be carbon particles since the mice were sacrificed using CO₂ asphyxiation in the reported study. However, since in this study, cervical dislocation was used to sacrifice mice, the 1590 cm^{-1} along with 1311 cm^{-1} may be attributable to cytochrome.⁵⁸ Cytochromes are abundantly present in lungs since the organ is involved with the oxygen transfer process. Mean lung adenoma spectrum [Fig. 2(e)] shows a loss of 1590 cm^{-1} band with respect to the control. Broad amide I and amide III with respect to control breast, breast tumor, control lung, and lung tumor mean spectrum [Fig. 2(c)] is observed in mean breast metastasis spectra.

Difference spectra: to elucidate the spectral variations among groups, difference spectra were computed. Subtraction of mean spectra is one of the conventional ways of looking at spectral differences. It provides differences over a selected spectral range, thus understanding of the moieties that may have been

modified is facilitated. The prominent bands of the difference spectra have been used to infer from the difference spectra.

The breast control–breast tumor [Fig. 3(a), i] were computed by subtracting the breast tumor mean spectrum from the breast control mean spectrum. In this difference spectrum, the positive peaks are due to breast control, whereas the negative peaks are due to breast tumor. Thus, a positive 1740 cm^{-1} indicates higher lipid content in the control compared to tumor, while a negative 1340 cm^{-1} indicates a lower DNA content in the control compared to tumor. Thus, breast control–breast tumor [Fig. 3(a), i] suggest a lower amount of proteins (negative peaks at 1671, 1456, and 1471 cm^{-1}) and DNA (negative peaks at 1480, 1340 cm^{-1}) and higher lipid content (positive peaks at 1743, 1440 cm^{-1}) in breast control compared to tumors, which corroborate previous reports.^{32,46} Control breast–breast metastasis difference spectra [Fig. 3(a), ii] suggest a higher lipid content (positive peaks at 1740, 1440 cm^{-1}) and lower DNA content (negative peaks at 1340, 1480 cm^{-1}) in control with respect to metastasis. Breast tumor–breast metastasis difference spectra [Fig. 3(a), iii] suggest decreased DNA (negative peaks at 1340, 1470 cm^{-1}) in breast tumor compared to metastasis.

Lung control–lung tumor difference spectra [Fig. 3(b), i] suggest lower lipid content (negative peaks at 1740, 1440, and 1301 cm^{-1}) in lung control compared to lung tumor. The same has been demonstrated in other studies.⁵⁷ The loss of the 1311 cm^{-1} band is another characteristic of lung tumor spectra. Lung control–breast metastasis [Fig. 3(b), ii] and lung

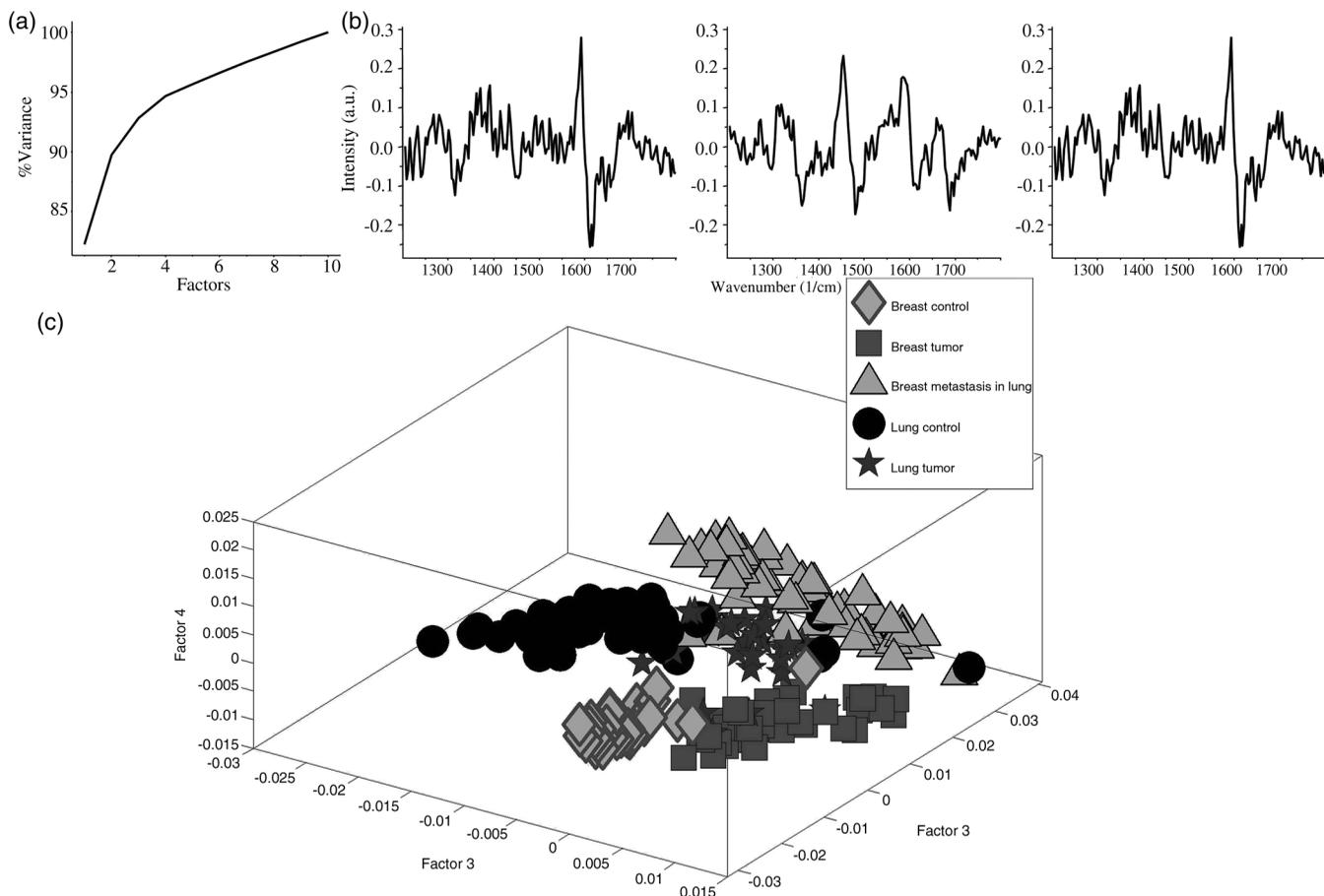


Fig. 4 Principal component analysis (PCA): (a) variance plot, (b) loading factors 1, 3, and 4, and (c) scatter plot.

tumor–breast metastasis difference spectra [Fig. 3(b), iii] suggest lower DNA content (negative peaks at 1340, 1470 cm^{-1}) in lung control and lung tumor compared to metastasis.

Breast control–lung control difference spectra [Fig. 3(c), i] suggest lipid dominance in breast control, while 1590 cm^{-1} characterizes the lung control. Breast control–lung tumor difference spectra [Fig. 3(c), ii] suggest decreased DNA content in breast control compared to lung tumor (negative peaks at 1480, 1340 cm^{-1}). Breast control–lung control difference spectra [Fig. 3(c), iii] also highlight the characteristic 1590 cm^{-1} band of lungs. Breast tumor–lung tumor spectra [Fig. 3(c), iv] suggest decreased DNA content in breast tumor compared to lung tumor. The spectral assignments are based on the available literature.⁵⁸

Overall, protein and DNA content are 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 cm^{-1} band is characteristic of control lung, but it disappears in lung tumor.

3.2 Multivariate Analysis

Preprocessed interpolated spectra in the 1200 to 1800 cm^{-1} range were subjected to PCA for delineating trends in the dataset. PCA variance plot and loadings are shown in Figs 4(a) and 4(b). As seen in Fig. 4(a), the cumulative variance covered by factors 1, 3, and 4 are 82.3%, 92.7, and 94.6% respectively. A scatter plot of the PCA factors [Fig. 4(c)] shows distinct clusters

of control breast, breast tumor, control lung, and breast metastasis. The lung tumor cluster lies in the center and is close to the control lung cluster and the breast metastasis cluster. The ability of RS to distinguish normal breast and breast cancer has been reported earlier.^{55,56} Earlier studies in mouse model have shown that *ex vivo* spectra of lung and breast differ considerably.⁵⁷ The distinct spectral identity of breast from several anatomical sites *in vivo* has also been demonstrated.⁴⁶ 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 RS. Thus, PCA suggests the 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, three factors⁵⁰ contributing ~85% of correct classification were used [Fig. 5(a)]. The plot of PC-LDA factors 1, 2, and 3 [Fig. 5(b)] shows a clustering pattern similar to PCA. The results of PC-LDA can be 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 the contribution of all factors selected for analysis. The confusion matrix for PC-LDA model building is shown in Table 1(a). In this analysis, 59 out of 62 spectra were correctly classified as breast control, while 1/62 and 2/62 spectra were misclassified with breast and lung tumors, respectively. 38/40 breast tumor spectra were correctly classified, while 2 were misclassified as breast control. In the case of breast metastasis in lungs, 45/63

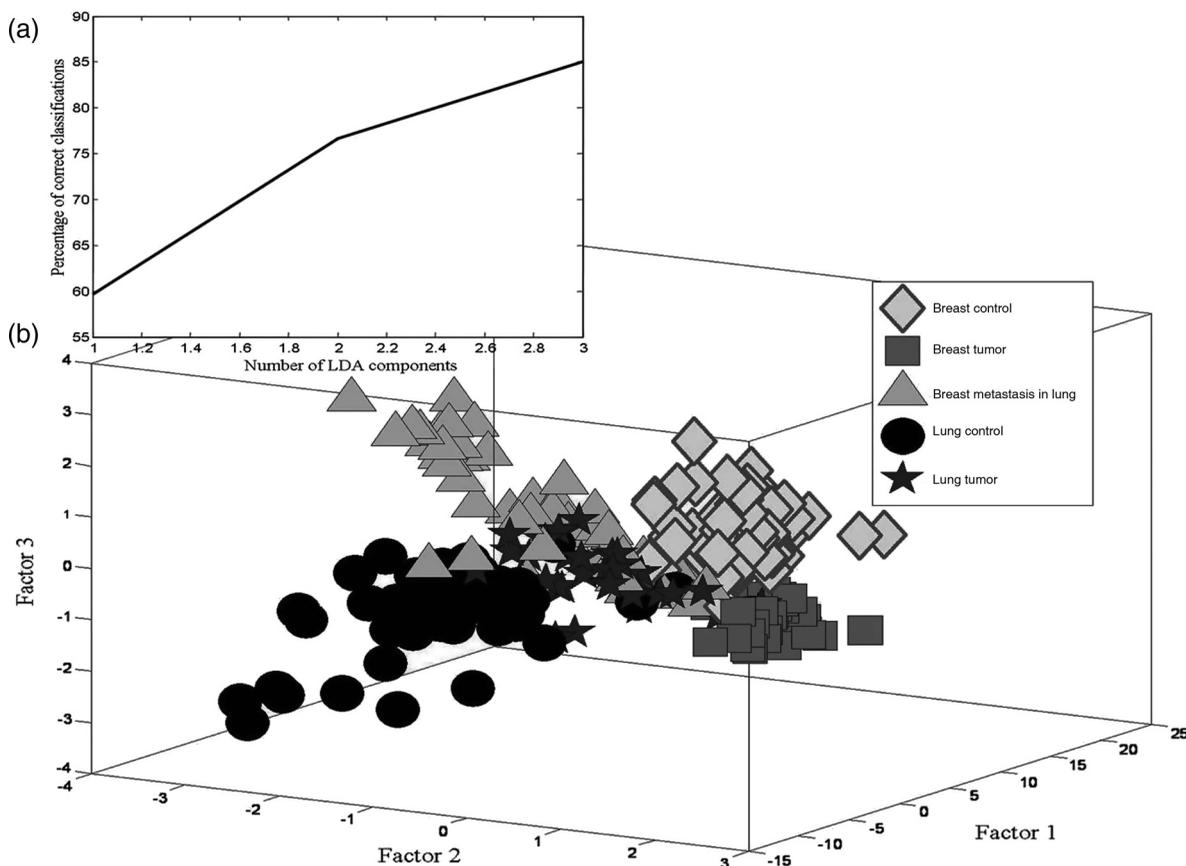


Fig. 5 Principal component-linear discriminant analysis (PC-LDA): (a) scree plot and (b) scatter plot.

spectra were correctly classified, 3/63 were misclassified with breast tumor, and 15/63 spectra were misclassified with lung tumor. 44/49 lung control spectra were correctly classified as lung control, whereas 4/49 were misclassified with lung tumor and 1/49 were misclassified with breast tumor. 25/34 lung tumor spectra were correctly classified, while 2/34 were misclassified with lung control, 3/34 were misclassified with

breast cancer metastasis, and 4/34 were misclassified with breast tumor.

LOOCV was carried out to evaluate the results obtained by PC-LDA. In analysis of LOOCV as shown in Table 1(b); 59 out of 62 spectra were correctly classified as breast control, while 1/62 and 2/62 spectra were misclassified with breast and lung tumors, respectively. Correct classification of breast spectra suggests the distinctness of breast spectra. Misclassification of breast control and breast tumor may be due to the 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 were correctly classified, while 2 were misclassified as breast control. The heterogeneity of tumors may explain the misclassification with normal breast. In the case of breast metastasis in lungs, 45/63 spectra were correctly classified, 4/63 were misclassified with breast tumor, and 14/63 spectra were misclassified 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 the architectural similarity of primary and metastatic tumor. 44/49 lung control spectra were correctly classified as lung control, whereas 4/49 were misclassified with lung tumor and 1/49 were misclassified with breast tumor. Misclassification between lung control and lung tumor may be attributed to the 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 were misclassified with lung control, 4/34 were misclassified with breast cancer metastasis, and 4/34 were misclassified with breast tumor. The heterogeneity of lung tumor may explain misclassification with lung control. The architectural similarity among 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 1(c)].

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 1(d). 56 out of 88 spectra are correctly predicted as breast metastasis in lungs, while 21/88 and 11/88 were wrongly predicted as lung and breast tumors, respectively. As mentioned earlier, prediction as lung tumor may be due to the architectural similarity between lung tumor and breast metastasis in lungs. Prediction as breast tumor may be due to the presence of breast tumor cells that have metastasized into lungs.

4 Conclusion

Differential diagnosis between primary and metastatic lesions in lung is vital for effective therapeutic intervention and favorable prognosis. However, this is especially difficult between primary lung lesions and metastatic breast lesions, as reported by several studies, due to the morphological, radiological, and pathological similarities. Therefore, the current *ex vivo* study aims to evaluate the ability of a fiber optic-based Raman system to distinguish metastatic lesions on lung from primary lung tumors in animal models. To subject multivariate analysis (PCA and PC-LDA) to more complex scenario, spectra were acquired from control breast and breast tumor along with control lung, primary

Table 1 PC-LDA Confusion matrix for (a) model building, (b) Leave-one-out cross, (Diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions), (c) LOOCV-sensitivity and specificity, and (d) independent test prediction.

(a) MODEL (Number of animals, Number 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 (Number of animals, Number of spectra)	Breast control	Breast tumor	Breast cancer metastasis in lungs	Lung control	Lung tumor
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 (Number of animals, Number of spectra)	Breast control	Breast tumor	Breast metastasis in lungs	Lung control	Lung tumor

lung tumor, and breast metastasis in lung tissues. Breast metastatic lesions in lung could be classified with 71% efficiency. Approximately 6% and 22% breast metastasis spectra were misclassified as breast and lung tumors, respectively, probably due to the presence of breast cells in metastatic lesions. Test prediction results show a 64% correct prediction of breast metastasis, while 13% and 24% breast metastasis spectra were wrongly predicted as breast and lung tumors, respectively. Thus, the findings of the study, the first of such investigations, demonstrate the potential of classifying breast metastasis in lungs from primary lung and primary breast tumor. Prospective evaluation on a larger cohort with multivariate analysis, combined with LDCT and recently developed real-time *in vivo* probes, may help non-surgical screening of lesions. The technique can then aid in treatment planning and improved prognosis.

References

- J. Ferlay et al., "GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide", 2013, International Agency for Research on Cancer, Lyon, France, <http://globocan.iarc.fr> (February 2015).
- R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA Cancer J. Clin.* **65**(1), 5–29 (2015).
- N. A. Howlander et al., Eds, *SEER Cancer Statistics Review, 1975–2011*, National Cancer Institute, Bethesda, Maryland (2013).
- P. W. Payne et al., "Sputum screening by quantitative microscopy a re-examination of a portion of the National Cancer Institute cooperative early lung cancer study," *Mayo Clin. Proc.* **72**, 697–704 (1997).
- M. R. Melamed, "Lung cancer screening results in the National Cancer Institute New York study," *Cancer* **89**(S11), 2356–2362 (2000).
- A. M. Davis and A. S. Cifu, "Lung cancer screening," *J. Am. Med. Assoc.* **312**(12), 1248–1249 (2014).
- P. B. Bach et al., "Benefits and harms of CT screening for lung cancer: a systematic review," *J. Am. Med. Assoc.* **307**(22), 2418–2429 (2012).
- J. B. Seo et al., "Atypical pulmonary metastases: spectrum of radiologic findings 1," *Radiographics* **21**(2), 403–417 (2001).
- N. Pavlidis et al., "Diagnostic and therapeutic management of cancer of an unknown primary," *Eur. J. Cancer* **39**(14), 1990–2005 (2003).
- S. Raab et al., "Adenocarcinoma in the lung in patients with breast cancer. A prospective analysis of the discriminatory value of immunohistochemistry," *Am. J. Clin. Pathol.* **100**(1), 27–35 (1993).
- S. Raab, C. Sturgis, and M. Wick, "Metastatic tumors in the lung: a practical approach to diagnosis," Chapter 17 in *Practical Pulmonary Pathology: A Diagnostic Approach*, K. O. Leslie and M. R. Wick, Eds., pp 597–639, Churchill Livingstone, Philadelphia (2005).
- W. G. Cahan and E. Castro, "Significance of a solitary lung shadow in patients with breast cancer," *Ann. Surg.* **181**(2), 137 (1975).
- J. J. Casey et al., "The solitary pulmonary nodule in the patient with breast cancer," *Surgery* **96**(4), 801–805 (1984).
- D. J. Dabbs et al., "Detection of estrogen receptor by immunohistochemistry in pulmonary adenocarcinoma," *Ann. Thor. Surg.* **73**(2), 403–406 (2002).
- E. Y. Chang et al., "The evaluation and treatment implications of isolated pulmonary nodules in patients with a recent history of breast cancer," *Am. J. Surg.* **191**(5), 641–645 (2006).
- P. L. Wagner et al., "Immunohistochemical expression of estrogen and progesterone receptors in primary pulmonary neuroendocrine tumors," *Arch. Pathol. Lab. Med.* **132**(12), 1889 (2008).
- 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 Pathol.* **20**(2), 208–214 (2007).
- 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," *Arch. Pathol. Lab. Med.* **132**(2), 239 (2008).
- 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," *J. Surg. Oncol.* **98**(7), 510–514 (2008).
- J. Herbst et al., "Evidence-based criteria to help distinguish metastatic breast cancer from primary lung adenocarcinoma on thoracic frozen section," *Am. J. Clin. Pathol.* **131**(1), 122–128 (2009).
- R. T. Vollmer, "Primary lung cancer vs. metastatic breast cancer a probabilistic approach," *Am. J. Clin. Pathol.* **132**(3), 391–395 (2009).
- D. Evers et al., "Optical spectroscopy: current advances and future applications in cancer diagnostics and therapy," *Future Oncol.* **8**(3), 307–320 (2012).
- U. Utzinger et al., "Near-infrared Raman spectroscopy for *in vivo* detection of cervical precancers," *Appl. Spectrosc.* **55**(8), 955–959 (2001).
- S. K. Teh et al., "Diagnostic potential of near-infrared Raman spectroscopy in the stomach: differentiating dysplasia from normal tissue," *Br. J. Cancer* **98**(2), 457–465 (2008).
- S. Koljenović et al., "Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy," *Lab. Invest.* **82**(10), 1265–1277 (2002).
- S. P. Singh et al., "In vivo Raman spectroscopy of oral buccal mucosa: a study on malignancy associated changes (MAC)/cancer field effects (CFE)," *Analyst* **138**(14), 4175–4182 (2013).
- A. C. Terentis et al., "Confocal Raman microspectroscopy discriminates live human metastatic melanoma and skin fibroblast cells," *J. Raman Spectrosc.* **44**(9), 1205–1216 (2013).
- Z. Huang et al., "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer," *Int. J. Cancer* **107**(6), 1047–1052 (2003).
- S. K. Teh et al., "Near-infrared Raman spectroscopy for gastric precancer diagnosis," *J. Raman Spectrosc.* **40**(8), 908–914 (2009).
- H. Abramczyk and B. Brozek-Pluska, "Raman imaging in biochemical and biomedical applications. Diagnosis and treatment of breast cancer," *Chem. Rev.* **113**(8), 5766–5781 (2013).
- R. Alfano et al., "Human breast tissues studied by IR Fourier transform Raman spectroscopy," *Lasers Life Sci.* **4**(1), 23–28 (1991).
- A. S. Haka et al., "Diagnosing breast cancer by using Raman spectroscopy," *Proc. Natl. Acad. Sci. U. S. A.* **102**(35), 12371–12376 (2005).
- J. Horsnell et al., "Raman spectroscopy—a new method for the intra-operative assessment of axillary lymph nodes," *Analyst* **135**(12), 3042–3047 (2010).
- N. Stone et al., "Subsurface probing of calcifications with spatially offset Raman spectroscopy (SORS): future possibilities for the diagnosis of breast cancer," *Analyst* **132**(9), 899–905 (2007).
- A. F. d. Oliveira et al., "Differential diagnosis in primary and metastatic cutaneous melanoma by FT-Raman spectroscopy," *Acta Cir. Bras.* **25**(5), 434–439 (2010).
- 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).
- M. Marro et al., "Molecular monitoring of epithelial-to-mesenchymal transition in breast cancer cells by means of Raman spectroscopy," *Biochim. et Biophys. Acta* **1843**(9), 1785–1795 (2014).
- 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," *Anal. Methods* **6**(12), 3948–3961 (2014).
- M. Sattler et al., "Investigation of support vector machines and Raman spectroscopy for lymph node diagnostics," *Analyst* **135**(5), 895–901 (2010).
- J. Smith et al., "Raman spectral mapping in the assessment of axillary lymph nodes in breast cancer," *Technol. Cancer Res. Treat.* **2**(4), 327–332 (2003).
- M. A. Short et al., "Development and preliminary results of an endoscopic Raman probe for potential *in vivo* diagnosis of lung cancers," *Opt. Lett.* **33**(7), 711–713 (2008).
- F. Mustafa, M. Lozano, and J. P. Dudley, "C3H mouse mammary tumor virus superantigen function requires a splice donor site in the envelope gene," *J. Virol.* **74**(20), 9431–9440 (2000).
- L. M. Hook et al., "Genetics of mouse mammary tumor virus-induced mammary tumors: linkage of tumor induction to the gagGene," *J. Virol.* **74**(19), 8876–8883 (2000).
- 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).

45. S. Rubina et al., "In vivo Raman spectroscopy of cervix cancers," *Proc. SPIE* **8940**, 89400E (2014).
46. 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).
47. T. Bhattacharjee et al., "Transcutaneous *in vivo* Raman spectroscopic studies in a mouse model: evaluation of changes in the breast associated with pregnancy and lactation," *J. Biomed. Opt.* **18**(4), 047004 (2013).
48. T. Bhattacharjee et al., "Transcutaneous *in vivo* Raman spectroscopy: detection of age-related changes in mouse breast," *Vib. Spectrosc.* **67**, 80–86 (2013).
49. T. Bhattacharjee et al., "Transcutaneous Raman spectroscopy of breast tumors and pretumors," *J. Raman Spectrosc.* (2015).
50. J. G. 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).
51. D. Pena, *Multivariate Data Analysis*, McGraw Hill-Interamericana de Espana SAU (2002).
52. R. Eisenbeis and R. Avery, *Discriminant Analysis and Classification Procedures: Theory and Applications*, DC Heath, Lexington, Massachusetts (1972).
53. M. M. Richert et al., "An atlas of mouse mammary gland development," *J. Mammary Gland Biol. Neoplasia* **5**(2), 227–241 (2000).
54. P. P. Rosen et al., *Rosen's Breast Pathology*, 2nd ed, Lippincott Williams & Wilkins, Philadelphia (2001).
55. 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).
56. M. V. Chowdary et al., "Discrimination of normal, benign, and malignant breast tissues by Raman spectroscopy," *Biopolymers* **83**(5), 556–569 (2006).
57. N. Huang et al., "Full range characterization of the Raman spectra of organs in a murine model," *Opt. Express* **19**(23), 22892–22909 (2011).
58. F. S. Parker, *Applications of Infrared, Raman, and Resonance Raman Spectroscopy in Biochemistry*, Plenum Press, New York (1983).

Biographies for the authors are not available.