

Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

Differentiation of cancer cells in two-dimensional and three-dimensional breast cancer models by Raman spectroscopy

Nur P. Damayanti
Yi Fang
Mukti R. Parikh
Ana Paula Craig
Julia Kirshner
Joseph Irudayaraj

Differentiation of cancer cells in two-dimensional and three-dimensional breast cancer models by Raman spectroscopy

Nur P. Damayanti,^{a,c*} Yi Fang,^{a,c*} Mukti R. Parikh,^{b*} Ana Paula Craig,^{c*} Julia Kirshner,^b and Joseph Irudayaraj^{a,c}

^aPurdue University, Department of Agricultural and Biological Engineering, 225 S. University Street, West Lafayette, Indiana 47907

^bPurdue University, Department of Biological Sciences, 915 W. State Street, West Lafayette, Indiana 47907

^cPurdue University, Bindley Bioscience Center, West Lafayette, Indiana 47907

Abstract. We demonstrate the first application of Raman spectroscopy in diagnosing nonmalignant, premalignant, malignant, and metastatic stages of breast cancer in a three-dimensional (3-D) cell culture model that closely mimics an *in vivo* environment. Comprehensive study comparing classification in two-dimensional (2-D) and 3-D cell models was performed using statistical methods composed of principal component analysis for exploratory analysis and outlier removal, partial least squares discriminant analysis, and elastic net regularized regression for classification. Our results show that Raman spectroscopy with an appropriate classification tool has excellent resolution to discriminate the four stages of breast cancer progression, with a near 100% accuracy for both 2-D and 3-D cell models. The diversity in chemical groups related to nucleic acids, proteins, and lipids, among other chemicals, were identified by appropriate peaks in the Raman spectra that correspond to the correct classification of the different stages of tumorigenesis model comprising of MCF10A, MCF10AneoT, MCF10CA1h, and MCF10CA1a cell lines. An explicit relationship between wavenumber and the stages of cancer progression was identified by the elastic net variable selection. © 2013 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.18.11.117008]

Keywords: Raman spectroscopy; three-dimensional cell culture; breast cancer staging; elastic net analysis; diagnosis.

Paper 130383RR received Jun. 3, 2013; revised manuscript received Oct. 10, 2013; accepted for publication Oct. 15, 2013; published online Nov. 18, 2013.

1 Introduction

Raman spectroscopy is an analytical technique based on inelastic scattering of light, capable of providing chemical specificity with a suitable enhancement down to a single molecule resolution.^{1–5} The presence of specific chemical bond diversity on cell surfaces relates to the diversity of cell surface composition consisting of carbohydrates, proteins, and lipids among others.^{3,5,6} With the surface-enhanced Raman spectroscopy, subcellular fingerprinting (e.g., nucleic acids and amino acids) has been possible, enabling Raman spectroscopy as a viable method for the classification of complex *in vivo* biomolecular analysis and monitoring.⁴ Raman spectroscopy has been shown to be accurate in differentiating nonmalignant and malignant entities, such as in skin lesions,^{5,7} epithelial breast cancer cells,⁶ gastrointestinal, colorectal, and lung cancer cells.^{8–11} The focus of our study is on breast cancer using a three-dimensional (3-D) cell culture model system that mimics an *in vivo* condition.

Breast cancer is a lethal heterogeneous disease¹² with diversity within and between tumors and also among patients.¹³ The complexity of this disease poses a challenge in cancer diagnosis, prognosis, and in the assessment of populations that have developed a therapeutic resistance. Therefore, there is a need for robust diagnostic and classification tools that are reproducible and have clinical potential.¹²

Although Raman spectroscopy has been used for *ex vivo* and *in vivo* classification of breast cancer,^{4,14–19} most of the past work predominantly focused only on the differentiation of nontumorigenic and tumorigenic cells, while few address stage-based classification (from nonmalignant, premalignant, malignant, to metastatic). Furthermore, most of the past work using Raman spectroscopy for breast cancer classification used two-dimensional (2-D) cell culture models, where the extracellular matrix, a critical factor in cell phenotype regulation, is unaccounted for.²⁰ Under 2-D conditions, cells tend to lose part of their context due to the absence of an extracellular environment and other characteristics necessary for cell-to-cell communication, growth, and differentiation.

The work presented here is the first to evaluate Raman spectroscopy as a potential tool for cancer staging using 3-D breast cancer model with elastic net regularized regression analysis to reveal the relationship between characteristic Raman fingerprint and the stage of cancer. We tested our hypothesis that Raman spectroscopy is a viable tool for breast cancer classification using 2-D and 3-D cell culture models in conjunction with the application of principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), and elastic net analysis. Our work addresses the: (1) potential of Raman spectroscopy to differentiate cancer cells at different stages of tumorigenesis; (2) classification efficacy to differentiate cells at different stages of the disease; and (3) determination of key chemical components that account for the differences observed in the cells at different stages of malignant transformation. Spectral analysis from elastic net is supported by statistical methods, PCA, and PLS-DA.

*These authors contributed equally.

Address all correspondence to: Joseph Irudayaraj, Purdue University, Department of Agricultural and Biological Engineering, 225 S. University Street, West Lafayette, Indiana 47907. Tel: +1-765-494-0388; Fax: (765) 496-1115; E-mail: josphi@purdue.edu

2 Materials and Methods

2.1 Cancer Cell Lines

The MCF10 model is a panel of isogenic human breast epithelial cell lines, namely MCF10A (nonmalignant), MCF10AneoT (pre-malignant), MCF10CA1h (malignant, but not metastatic), and MCF10CA1a (metastatic), each at a distinct stage of cancer progression.²¹ Spontaneously immortalized, nonmalignant breast epithelial MCF10A cells were developed from a human benign fibrocystic breast disease specimen. Introduction of the H-Ras oncogene in MCF10A cells gave rise to the pre-malignant MCF10AneoT cells which were serially passaged in immunocompromised mice resulting in the accumulation of tumor-promoting mutations. This led to the development of human tumor xenografts with varying degree of malignancy, ultimately giving rise to various MCF10CA1 clones. MCF10CA1h cell line consists of well-differentiated tumor cells that are malignant, but not metastatic, whereas MCF10CA1a line consists of poorly differentiated cells that are metastatic in immunocompromised mice.^{21–26} These cell lines have been extensively characterized by cytogenetic and microarray analysis and have been used to study key signaling pathways and molecular and genetic events in tumorigenesis.^{27–29} Since the cell lines in the MCF10 series are isogenic and represent the major stages of breast cancer progression, they were an ideal choice for this study. All the cell lines were obtained from Barbara Ann Karmanos Cancer Institute (Detroit, Michigan).

2.2 Cell Lines Growth Conditions

MCF10A and MCF10AneoT cells were cultured and maintained in Dulbecco's modified Eagle's medium nutrient mixture F-12 HAM [with 15 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl] ethanesulfonic acid (HEPES), NaHCO₃, pyridoxine, and L-glutamine; Sigma] supplemented with 5% horse serum (Sigma), 100 ng/mL cholera toxin (Sigma), 20 ng/mL EGF (SAFC Biosciences, St. Louis, Missouri), 10 μg/mL insulin (Sigma, St. Louis, Missouri), and 0.5 μg/mL hydrocortisone (Sigma). MCF10CA1h and MCF10CA1a cells were grown in Dulbecco's modified Eagle's medium nutrient mixture F-12 HAM (with 15 mM HEPES, NaHCO₃, pyridoxine, and L-glutamine; Sigma) supplemented with 5% horse serum. All the cells were cultured in 5% CO₂ incubator at 37°C.

2.3 Sample Preparation

For 2-D culture, the cells were cultured in a 25-cm² flask (Nunc, Rochester, New York) to reach 80% confluency. The cells were trypsinized and 0.5 mL of the cell suspension was added to 35-mm tissue culture dishes (Falcon, Franklin lakes, New Jersey) with 1.5 cm × 1.5 cm gold slides placed inside the dish. Subsequently, 2.5 mL of culture medium was added and the cells were cultured to reach 80% confluency. By using this method, the cells were grown both on the surface of the gold slide and in the tissue culture dish. The slides were subsequently taken out and placed into a new tissue culture dish. The cells on the gold slide were rinsed twice with Gibco phosphate-buffered saline (PBS) at 37°C (pH 7.4, Life Technologies, Carlsbad, California) to remove the remaining culture medium and placed in 1.3 mL PBS to prevent the cells from drying out during measurement. The main steps involved in sample preparation are schematically illustrated in Fig. 1.

In Fig. 1, the 2-D cell culture and the two common types of 3-D cell culture models are shown: Fig. 1(b) shows, an “embedded model” in which a mixture of cell suspension with liquefied Matrigel is prepared as a first step followed by the addition of culture media after the solidification of the Matrigel. The second method for 3-D cell culture preparation is the “on-top model,” Fig. 1(c), where solidified Matrigel is used as a substratum before adding the mixture of culture media and cell suspension. Here, the cells adhere to the surface of the Matrigel and initiate the formation of clusters on the surface. In this work, in order to keep the cell clusters alive during measurement, the embedded model was adopted as the surrounding Matrigel protects the specimen from drying out due to the evaporation of the liquid medium during measurement.

For 3-D cell culture, 2×10^5 cells were resuspended in 10 μL of growth medium and added to 150 μL Matrigel basement membrane matrix (BD Biosciences, San Jose, California) on ice. The mixture of cells and Matrigel was evenly layered in a 24-well tissue culture treated plate (BD Falcon, Franklin lakes, California) and incubated in a 5% CO₂ incubator at 37°C for 30 min. After 30 min, 1.5 mL of prewarmed growth medium was gently layered from the side of the well onto the matrix/cell mixture. The cells were maintained in the 3-D culture at 37°C with 5% CO₂ for 5 days with the medium replaced every 2–3 days. The cells were allowed to form 3-D spheroids and then imaged using the Zeiss AxioObserver microscope with Axiovision software.

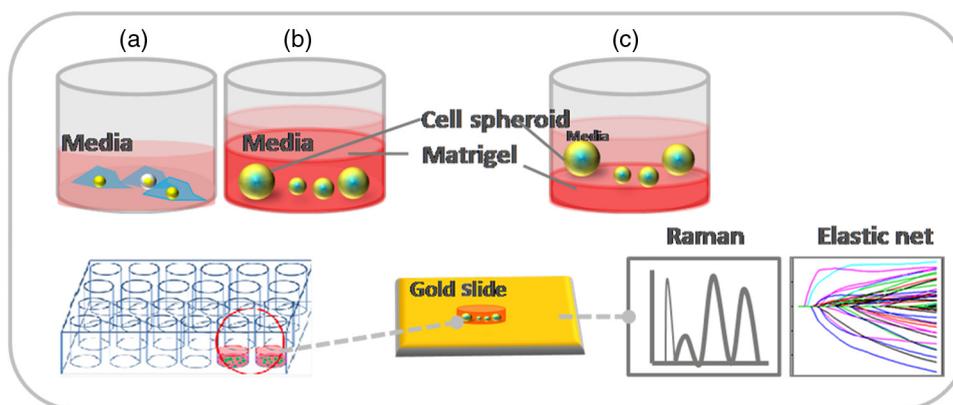


Fig. 1 Schematic of a two-dimensional (2-D) and three-dimensional (3-D) cell culture systems: (a) 2-D; (b) and (c) 3-D culture models (top) and 3-D cell sample preparation (bottom). 3-D cell lines were first cultured in a 24-well plate in an embedded model, washed, and a piece of gel was then taken, placed onto the gold slide for Raman measurement that followed.

The cells were analyzed after 5 days in the 3-D culture. The culture medium was removed and the cells were washed twice with PBS to remove the remaining culture medium; PBS in the well was subsequently removed and a section of the Matrigel with 3-D cell spheroids was removed with a spatula. The Matrigel was placed on the gold slide for the spectral analysis, and the remaining Matrigel was kept in the incubator for further measurements.

2.4 Data Collection by Raman Spectrometer

The Senterra Confocal Raman Spectrometer (Bruker Optics Inc., Billerica, Massachusetts), fitted with a 50× objective was used. The 50 μm pinhole was activated to remove stray light from the out-of-focus plane. Cells grown in 2-D and 3-D cultures are shown in Figs. 2(a) and 2(b), respectively. A 633 nm He-Ne laser with 20 mW power was used for excitation and an integration time of 45 s was used for Raman spectra acquisition. During measurement, the cells attached to the gold slide displayed a spindle form, thus no tissue damage was observed under the laser conditions used. This can be

explained by the fact that cells gradually display a round shape instead of the spindle shape when dying. The schematic of the optical layout is illustrated in Fig. 2(c).³⁰⁻³⁴

For both 2-D and 3-D cell samples, three independent biological replicates were performed. In the 2-D cell cultures, 10 cells were measured in each cell, five measurements were made by choosing different spots (one is at the center) to ensure that all components of the cell were probed. For the 3-D cell sample, since the cells formed clusters, called spheroids, the samples were studied as spheroids rather than as single cells. In each biological replicate, four spheroids were chosen for analysis and 15 different spots were chosen across the spheroid (one center point and four spots surrounding this spot and 10 spots at the exterior) for measurement in each sample. The data collection and analysis for both 2-D and 3-D cultures are noted below.

2.5 Data Pretreatment

Data analysis was done using R software. The raw spectra were first processed to accommodate the 500–1800 cm⁻¹

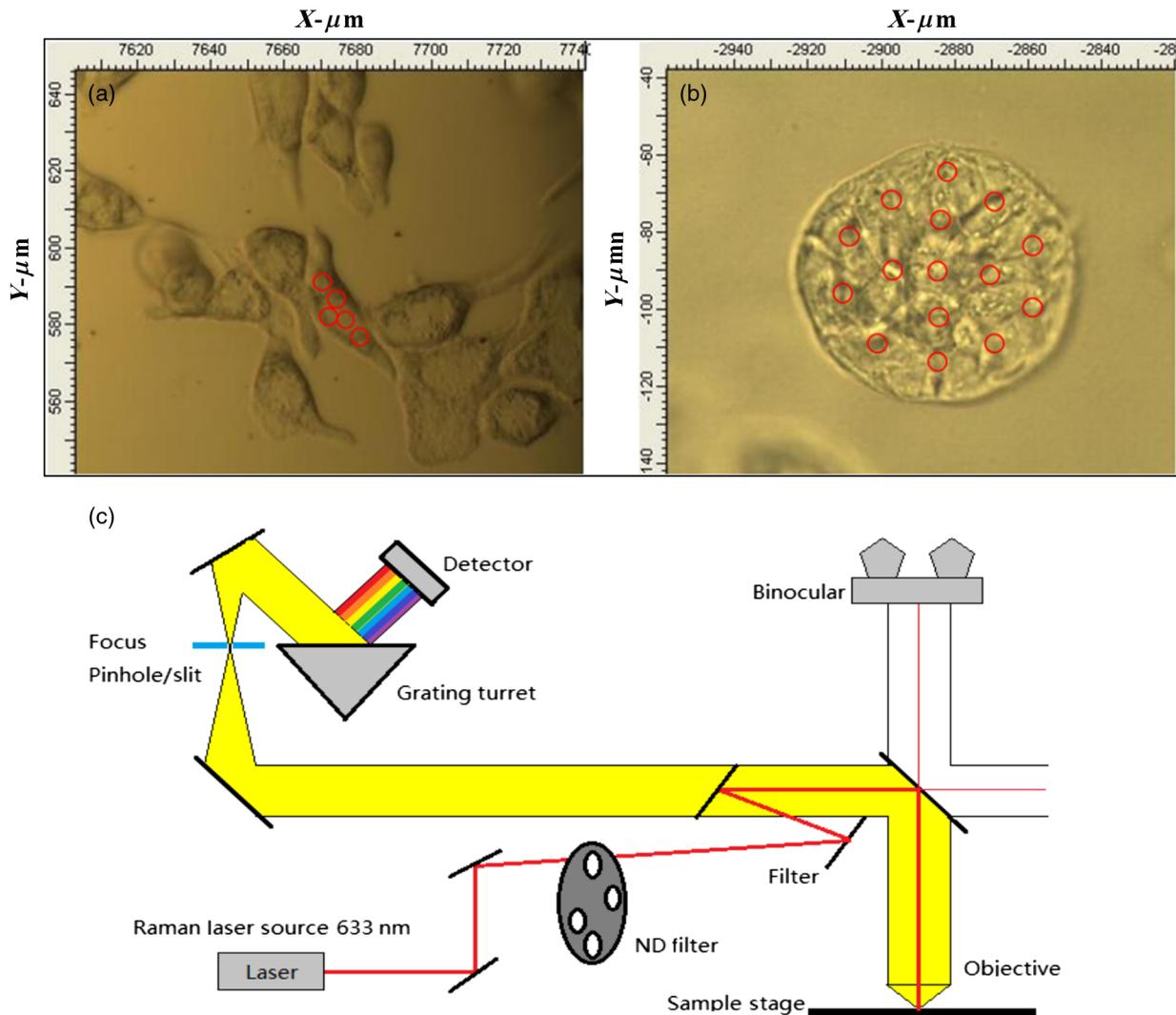


Fig. 2 Schematic of sampling for 2-D and 3-D culture systems: (a) five spots were chosen for data collection in each 2-D cell, (b) 15 spots distributed throughout the spheroid were chosen in the 3-D spheroid, and (c) schematic of the optical layout of the Senterra Confocal Raman spectrometer: the laser is directed through a neutral density filter to excite the sample through the objective and the scattered photons are collected by the same objective. A pinhole is used to detect photons from the plane of focus.

range of the Raman shift which contains most of the biological fingerprints.¹⁷⁻¹⁹ The spectra were baseline corrected by fitting polynomials using the R software to remove the effects of fluorescence and allow for the acquisition of spectra with band edges up to the theoretical baseline to visualize minor peaks. The spectra obtained were normalized to ensure that the area under the curve was unity, for further evaluation.

2.6 Statistical Analysis

The statistical analysis of this study was conducted in three steps. First, PCA was applied to assess data variability and to detect outliers. These outliers were removed from the data set and consequently not included in the subsequent statistical techniques. In sequence, PLS-DA, a well-established classification technique was applied to the data to actually classify the samples based on their cell lines. Beyond the development of classification models, in a third step, an interpretation of the main differences between cells in different stages of cancer was attempted by elastic net,³⁵ providing an interesting visualization of specific compounds associated with this progression. Elastic net is a version of penalized least squares that combines both Ridge and Lasso regressions. In this technique, the wavenumbers that reveal an explicit relationship with the different stages of tumor progression are selected with their respective coefficient estimates that represent the extent of contribution of each of these wavenumbers to the correct classification of each cell line.

2.6.1 Exploratory analysis and outlier removal

PCA is an exploratory multivariate analysis technique that projects the data matrix to a lower-dimensional space spanned by the loading vectors. The loading vectors corresponding to the k largest eigenvalues are retained to optimally capture the variance of the data and to minimize the effect of random noise.³⁶ The goodness-of-fit between the data and the model can be calculated using the residual matrix and Q statistics that measure the distance of a sample from the new space of the PCA model.³⁶ Hotelling's T^2 statistics indicates how far the estimated sample by the PCA model is from the multivariate mean of the data, thus these statistics provide an indication of variability within the normal subspace.³⁷ The combination of Q and T^2 tests is used to detect the remaining abnormal observations. Given the level of significance for the Q and T^2 statistics, measurements with Q or T^2 values over the threshold are classified as outliers.³⁷ After the elimination of the outlier spectra from the model, the procedure was continually repeated until no outliers could be identified. The software MATLAB (The MathWorks, Co., Natick, Massachusetts) and the computational package PLS_Toolbox (Eigenvector Research, Inc., Wenatchee, Washington) were employed. All data were mean centered prior to PCA.

2.6.2 Classification of cells in different stages of cancer

PLS-DA is a parametric and linear model and one of the most applied techniques for the classification of spectral data. The basics of PLS-DA consist of the application of a partial least squares regression model with variables which are indicators of groups. The second step of PLS-DA is to classify observations from the results of PLS regression on the indicator variables.³⁸ MATLAB (The MathWorks, Co., Natick, Massachusetts) and the computational package PLS_Toolbox

(Eigenvector Research, Inc.) were employed for PLS-DA analysis. In both 2-D and 3-D models, spectral data were divided into calibration (75%) and validation (25%) data sets. Cross-validation (comprising of leave-one-out analysis) was applied to estimate the performance of the models and to choose the optimal number of latent variables. The predictability of the resulting models was evaluated based on the classification error for the validation set. Sensitivity is defined as the proportion of true positives that were classified as positive, and specificity represents the proportion of true negatives that were classified as negative. The later parameters were calculated as follows:

$$\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$\text{Specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$

2.6.3 Variable selection

Elastic net³⁵ was applied using the glmnet package of the R software that fits generalized linear models via penalized maximum

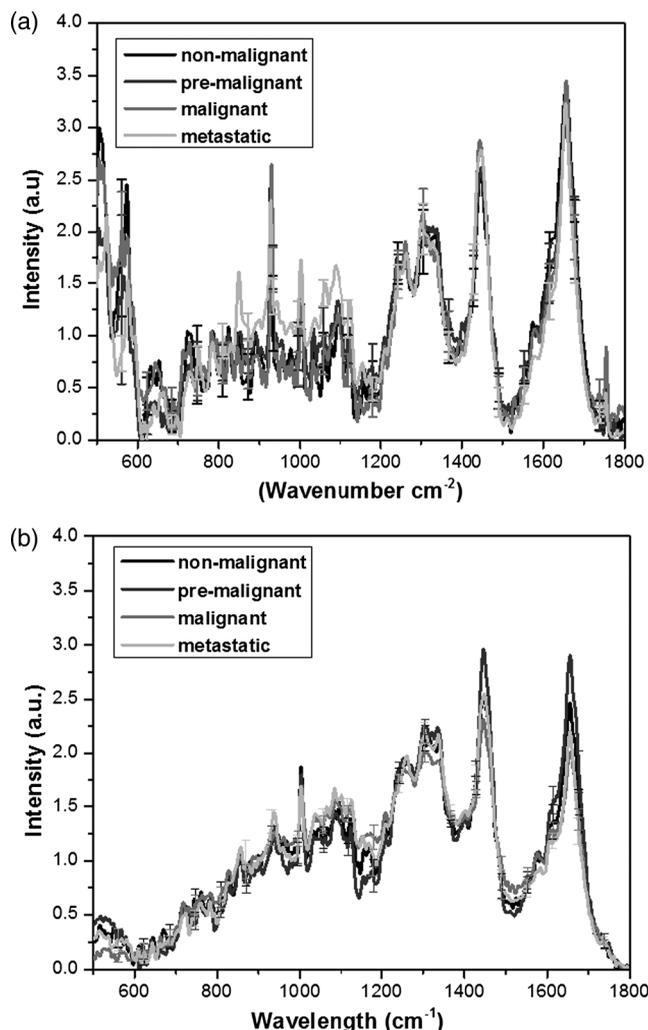


Fig. 3 Mean Raman spectra of (a) 2-D and (b) 3-D cell cultures of benign, MCF10A (black), premalignant, MCF10AneoT (dark gray), malignant, MCF10CA1h (medium gray), and metastatic MCF10CA1h (light gray). Standard deviation is provided.

likelihood. Samples were randomly separated into training (75%) and validation (25%) data sets. The regularization parameter lambda causes coefficient shrinkage, minimizing the residual sum of squares. In order to obtain the lambda value that gives a minimum cross-validated error, leave-one-out cross-validation was performed. In sequence, multinomial logistic models were fitted with the training data set at α values ranging from 0 to 1, in steps of 0.25. The parameter α controls the mixing between Ridge and Lasso regressions. Ridge regression ($\alpha = 0$) imposes an L_2 -penalty to the model resulting in a coefficient shrinkage, while Lasso regression ($\alpha = 1$) imposes an L_1 -penalty which expects many predictors to be close to zero and a small subset to be nonzero, providing automatic variable selection.³⁹ Elastic net ($0 < \alpha < 1$) provides both coefficient shrinkage and variable selection. The developed models were used to predict the degree of classification as well as validation using the validation data set. The α -value that provided the higher predictability based on the lowest classification errors was considered as the best-fit model. The discriminant wave-numbers from the best-fit model and their respective coefficient estimates were then extracted.

3 Results

3.1 Raman Spectra

Figure 3 shows the average Raman spectra from a total of 150 spectra for the 2-D [Fig. 3(a)] and 180 spectra for the 3-D [Fig. 3(b)] cell culture model in the range

500 – 1800 cm^{-1} . Each spectrum was previously baseline corrected and normalized as described in Sec. 2. Standard deviation information is also provided.

Despite the similarities between the 2-D and the 3-D spectra, especially in the region 800–1150, 1200–1350, and 1400–1700 cm^{-1} , differences could be noted in the spectral window 500 – 700 cm^{-1} , and peaks in the 2-D cell culture samples merged into a broad shoulder in the 3-D cell systems.

Spectra from the metastatic (MCF10CA1a) cell lines had a higher intensity than other cell lines in the region 850–1150 cm^{-1} . This trend was observed in both the 2-D and 3-D cell culture samples, although it is more evident in the 2-D cells. In the 3-D system, it is evident that premalignant samples show higher intensity than other cell lines in the 1451 and 1650 cm^{-1} peaks, and nonmalignant samples show a higher intensity than malignant and metastatic at the 1650 cm^{-1} peak. An opposite trend was observed in nonmalignant and premalignant at the 1620 cm^{-1} shoulder, where the 2-D samples exhibited intensity in the following order; nonmalignant > premalignant > malignant > metastatic, while in the 3-D system the premalignant samples show higher intensity than nonmalignant. However, based only on the visual observation of the spectra, it is not possible to arrive at major conclusions, especially when considering the standard deviation between each cell line. In the 2-D model the minimum–maximum values of the standard deviation in the nonmalignant, premalignant, malignant, and metastatic cell lines were 0.06–1.22, 0.05–1.04, 0.07–0.57, and 0.05–0.57, respectively; in the 3-D models, these values

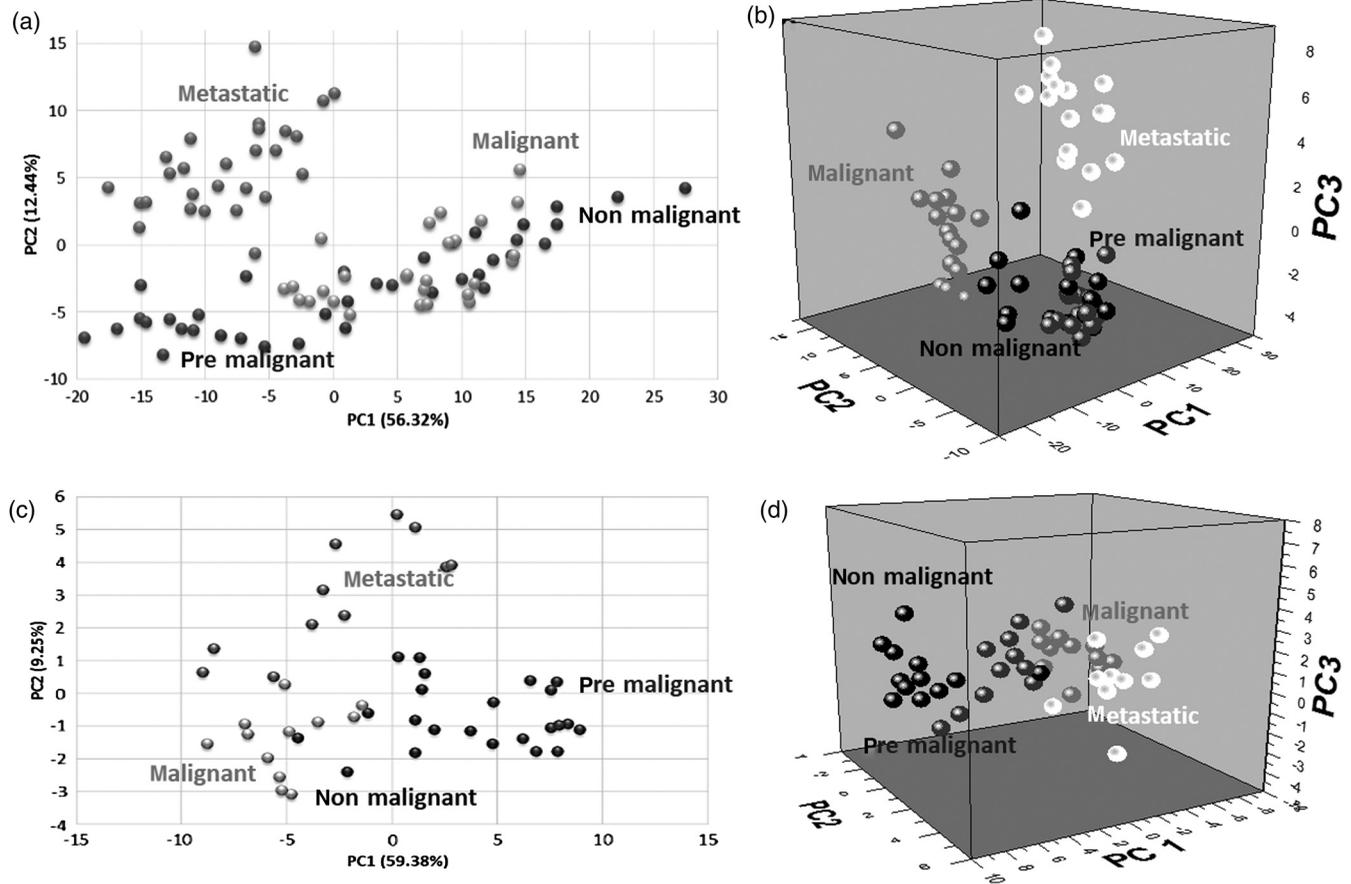


Fig. 4 Three-dimensional cell culture model better discriminates the four stages of breast cancer; PC scores scatter plots: (a), (c), PC1 versus PC2 of 2-D (a) and 3-D (c) cell models, (b), (d), PC1 versus PC2 versus PC3 of (b) 2-D, and (d) 3-D models.

Table 1 Summary of PLS-DA prediction results of the model developed for classification of cancer cell lines.

	MCF10A	MCF10AneoT	MCF10CA1a	MCF10CA1h
Two-dimensional				
Sensitivity (Cal)	1	0.941	1	1
Specificity (Cal)	1	1	1	1
Sensitivity (CV)	1	0.941	1	1
Specificity (CV)	1	1	1	1
Sensitivity (Val)	0.6	1	1	1
Specificity (Val)	1	0.895	1	0.938
Class. Err (Cal)	0	0.029	0	0
Class. Err (CV)	0	0.029	0	0
Class. Err (Val)	0.2	0.053	0	0.031
Three-dimensional				
Sensitivity (Cal)	1	1	1	1
Specificity (Cal)	1	1	1	1
Sensitivity (CV)	1	0.889	1	1
Specificity (CV)	0.923	1	0.963	0.923
Sensitivity (Val)	1	1	1	1
Specificity (Val)	0.889	1	1	1
Class. Err (Cal)	0	0	0	0
Class. Err (CV)	0.038	0.055	0.018	0.038
Class. Err (Val)	0.056	0	0	0

were 0.03–0.21, 0.02–0.24, 0.02–0.18, and 0.02–0.31, respectively. A discussion on the main differences between the spectra of each cell line will be conducted in the next sections supported by statistical analysis.

3.2 Exploratory Analysis and Outlier Removal by PCA

Using PCA, the original high-dimensional spectral data set containing chemical and biological characteristics of the cancer cell lines was projected to a low-dimensional space where the first principal components (PCs) explained most of the variance between the samples. Samples with Q or $T2$ statistical values over the threshold were classified as outliers and removed from the data set. The procedure was repeated until no outliers were identified. For the 2-D cell culture models, 19 observations from the original data set containing 110 observations were identified as outliers. On the other hand, two from the original 48 observations from the data set of the 3-D cell culture models were identified as outliers and removed from the data set.

The resulting scatter plots obtained by PCA analysis are shown in Fig. 4. For the 2-D cell culture, the first two PCs clearly separated metastatic cells from the other cell lines

[Fig. 4(a)]. The inclusion of the third PC provides a 3-D scatter plot that also discriminates malignant, but nonmetastatic from the rest of the group, as shown in Fig. 4(b). Thus, based on the variance of the spectral data from the 2-D cell culture, the two stages of cancer could be separated into two clusters while nonmalignant and premalignant were clustered together. In the case of the 3-D cell culture, the first two PCs displayed four clusters of the cell lines, although some overlap was observed [Figs. 4(c) and 4(d)].

3.3 Classification Based on PLS-DA

Results from our exploratory analysis indicate that Raman spectra can provide sufficient information to develop classification models for different stages of cancer cells. Thus, PLS-DA was further applied to the spectral data with the purpose of obtaining classification models for assigning samples to categories. Based on leave-one-out cross-validation, five and six latent variables were chosen and employed in the development of the 2-D and 3-D cell culture models. Table 1 summarizes the classification results obtained. Accurate classification was obtained for both the 2-D and 3D cell culture models with the accuracy up to 100% for malignant and premalignant in both models.

The sensitivity in the validation set for nonmalignant cells is lower in 2D models compare to the sensitivity of nonmalignant cells in 3D models. The overall classification error for the 3-D model is considered to be lower than the 2-D model, especially for the classification of nonmalignant cells.

Due to the fact that relatively small number of samples was used in the development of the 3-D model, a lower number of latent variables would be preferred in order to guarantee the stability and robustness of the model. However, the number of latent variables employed reflects the complexity of the samples being analyzed, and a reduction in this number leads to an

accuracy decay. Nevertheless, Table 2 shows the classification results for the 2-D and 3-D models constructed with one to three latent variables, respectively. Satisfactory results were obtained with only three latent variables for both the 2-D and 3-D models.

3.4 Variable Selection by Elastic Net

Elastic net was applied in this study for visualization of the discriminating wavenumbers and to know what extent these wavenumbers contributed to the correct classification of each cell line. Multinomial logistic models were developed

Table 2 PLS-DA prediction results of the models developed for the classification of 2-D and 3-D cancer cell lines (number of latent variables are included).

LV	Cell Line	Sens. (Cal)	Specif. (Cal)	Sens. (CV)	Specif. (CV)	Sens. (Val)	Specif. (Val)	Class. Err (Cal)	Class. Err (CV)	Class. Err (Val)
Two-dimensional										
1	MCF10A	0.727	0.491	0.727	0.491	0.8000	0.667	0.391	0.391	0.267
	MCF10AneoT	0.706	0.078	0.647	0.098	0.000	0.211	0.608	0.627	0.895
	MCF10CA1a	0.15	0.396	0.15	0.396	0.000	0.500	0.727	0.727	0.750
	MCF10CA1h	0.9	0.292	0.85	0.292	1.000	0.563	0.404	0.429	0.219
2	MCF10A	0.818	0.474	0.818	0.474	1.000	0.611	0.354	0.354	0.194
	MCF10AneoT	0.353	0.725	0.471	0.549	0.000	0.632	0.461	0.490	0.684
	MCF10CA1a	1	0.854	1	0.854	1.000	0.750	0.073	0.073	0.125
	MCF10CA1h	0.85	0.688	0.85	0.688	0.857	0.688	0.231	0.231	0.228
3	MCF10A	0.818	0.702	0.818	0.702	0.000	0.722	0.240	0.240	0.639
	MCF10AneoT	0.941	0.784	0.824	0.784	1.000	0.947	0.137	0.196	0.026
	MCF10CA1a	0.95	1	0.95	0.958	1.000	1.000	0.025	0.046	0.000
	MCF10CA1h	0.8	0.792	0.75	0.708	0.714	0.688	0.204	0.271	0.299
Three-dimensional										
1	MCF10A	0.111	0.654	0.333	0.654	0	0.556	0.618	0.506	0.722
	MCF10AneoT	0.889	0.769	0.889	0.769	1	0.75	0.171	0.171	0.125
	MCF10CA1a	0.375	0.593	0.5	0.667	0.667	0.625	0.516	0.417	0.354
	MCF10CA1h	0.222	0.231	0.222	0.231	0	0	0.774	0.774	1.000
2	MCF10A	0.889	0.462	0.889	0.423	1	0.333	0.325	0.344	0.333
	MCF10AneoT	0.889	0.923	0.889	0.923	1	1	0.094	0.094	0.000
	MCF10CA1a	0.625	0.556	0.625	0.556	1	0.625	0.410	0.410	0.188
	MCF10CA1h	0.889	0.769	0.778	0.769	1	0.75	0.171	0.226	0.125
3	MCF10A	0.889	0.385	0.889	0.385	1	0.444	0.363	0.363	0.278
	MCF10AneoT	0.889	0.923	0.889	0.846	1	1	0.094	0.132	0.000
	MCF10CA1a	1	1	1	0.963	1	1	0.000	0.019	0.000
	MCF10CA1h	0.889	0.846	0.889	0.808	1	1	0.132	0.152	0.000

by varying the parameter α . Perfect statistical classification was achieved for an α value of 0.5 in both the 2-D and 3-D systems. The coefficient estimates from the developed models were then extracted.³³

As demonstrated in Fig. 5, each stage of breast cancer was identified by specific coefficient estimates in both the 2-D and 3-D models; overall, 3-D cells exhibited fewer coefficients for each stage. Premalignant cells were identified by the prominent positive peak at 1228 – 1234 cm^{-1} and a negative peak at 1774 cm^{-1} in both 2-D and 3-D, but only in the 3-D model a prominent positive peak appeared at 612 – 614 cm^{-1} . High negative coefficients for nonmalignant cells were observed at 1365 cm^{-1} in the 3-D model. Premalignant, malignant, and metastatic stages were identified by different coefficient estimates in the 2-D and 3-D cell cultures, although the premalignant state yielded a negative peak at 604 – 612 cm^{-1} . A prominent negative peak for the premalignant cell line was observed at 1651 cm^{-1} in the 2-D model. Malignant cells were identified by a strong positive peak at 688 cm^{-1} in the 2-D model and strong negative peaks at 555 and 889 cm^{-1} in the 3-D and 2-D models, respectively. Higher peaks associated with the metastatic cell line were mostly observed in the 3-D model, with a strong positive peak at 1349 – 1358 cm^{-1} and negative peaks in the following regions: 732, 1006, 1236–1240, and 1578 cm^{-1} .

We examined each of the nonzero coefficient estimates and, based on the existing work, tentative assignment of these coefficients to chemical compounds were made that relate to specific regions in the Raman spectrum (Table 3).

4 Discussion

4.1 Raman Spectroscopy for Breast Cancer Classification

Based on our results (Fig. 4 and Table 1), the methods and analysis procedures used for classifying 3-D models using PCA and PLS-DA has a similar trend as the 2-D classification, providing evidence that the methods developed are robust.³⁸ Based on PLS-DA, excellent results were achieved in the classification of metastatic and malignant cells (specificity 100%, sensitivity 100%); however, lower sensitivity was observed when classifying nonmalignant cells in the 2-D model compared to that of the 3-D system. Low sensitivity (60%) and high specificity (100%) for nonmalignant cells indicate the possibility of nonmalignant cells to be classified as premalignant cells in the 2-D model; however, malignant and metastatic stages would not be classified as nonmalignant. Higher sensitivity was achieved for nonmalignant cells in the 3-D model, which suggests that the change in the extracellular matrix environment and interaction

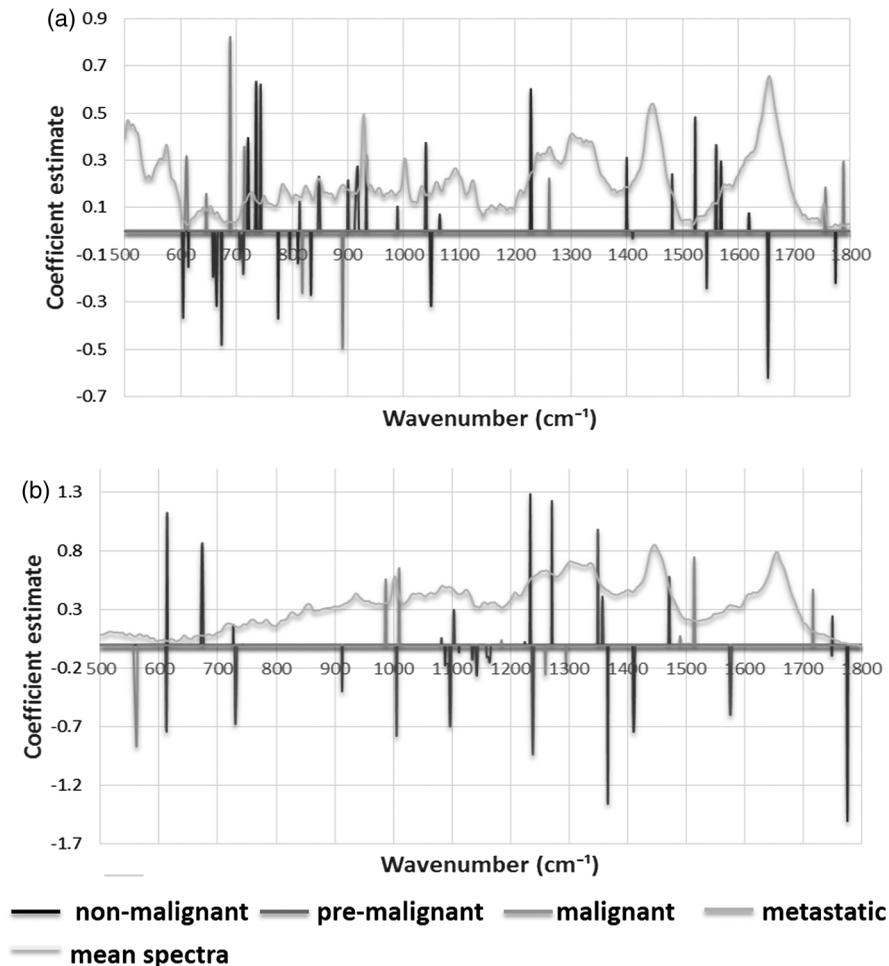


Fig. 5 Elastic net coefficient estimates for $\alpha = 0.5$ for (a) 2-D and (b) 3-D cell culture models. A peak indicates that the correct classification of spectra is associated with the corresponding spectral region. A positive peak indicates higher intensity than other cell lines; a negative peak indicates lower intensity.

Table 3 Coefficient estimates obtained from elastic net and its tentative peak assignment from the literature (Refs. 32–40).

Wavenumber (cm ⁻¹)	Two-dimensional		Three-dimensional		Lipid assignment	Protein assignment	Others
	MCF10A	MCF10AneoT	MCF10CA1h	MCF10CA1a			
556–562					–	S-S stretch, gauche and trans conformation ⁷	
602–605	–						
607–612		+			(C=C) vibration		
611–614	–		+	–			
644–646		+				Tyrosine (C=C) twist, aromatic amino acid ⁵	
656–665	–						
671–675	–			+			
687–690			+				
704–713		–					
712–715			+		C=N stretch phospholipids		
719–722	+						
726–727				+			
729–733				–			
733–736	+						
742–744		+					Ring breathing mode of DNA/RNA thymine ⁶
774–796	–						Symmetric breathing tryptophan, phenylalanine ⁷
809–811	–						C=C stretch of proline in collagen
812–813		+					RNA (backbone of P-O-P stretching) ⁸
817–819			–				
832–834	–						DNA–RNA backbone (P-O-P) ⁸

Table 3 (Continued).

Wavenumber (cm ⁻¹)	Two-dimensional		Three-dimensional		Lipid assignment	Protein assignment	Others
	MCF10A	MCF10AneoT	MCF10CA1a	MCF10CA1h			
846-850		+				Proline ⁶ Tyrosine ring breathing ⁸	
889-891			-				CH deformation, Beta linkage from glycosaminogly can ^{1,3}
898-919		+					
912-913			-				
932-933		+			Skeletal mode polysaccharide	Glycan, polysaccharide ³	
985-988				+			
988-989		+					
1004-1006			-		Symm ring breathing mode of phe ¹	Aromatic amino acid, protein ^{1,2}	
1009-1011				+			
1038-1041		+					
1047-1050			-				
1063-1065		+			C=C skeleton, trans ¹	ν (C=C) aromatic ³	
1081-1082				+			
1088-1089			-				DNA backbone ³
1095-1098					O-PO symmetry ³		
1102-1104				+			
1134-1166			-		lipids(C=Cstr) ³		
1164-1166			-			(C-H) bending tyrosine ⁶ Amide III ⁷	ν (PO ₂), nucleic acid
1224-1225			+				
1226-1229		+					
1233-1235			+				
1236-1240			-				

Table 3 (Continued).

Wavenumber (cm ⁻¹)	Two-dimensional		Three-dimensional		Lipid assignment	Protein assignment	Others
	MCF10A	MCF10AneoT	MCF10CA1h	MCF10CA1a			
1259–1260							
1260–1265		+				V(CN), d(NH) amide III, α-helix collagen Tryptophan ⁷	
1294–1295							
1335–1358			+			CH ₃ CH ₂ wagging collagen	CH ₃ CH ₂ wagging nucleic acids
1366–1367							
1399–1401	+						
1409–1412		-					
1470–1472			+				
1481–1482	+						
1489–1515					+		Carotene ⁹
1522–1523	+						
1542–1544		-					
1559–1561	+						
1568–1570	+						
1571–1578			+			ν(C=C) phenylalanine aromatic amino acid ⁴ ν(C=C), tryptophan ⁴	
1618–1620	+						ν(C=O) Phorphyrin ⁴
1652–1655	-					ν(C=O) amide I, α-helix, collagen, elastin ⁵	
1716–1718							
1749–1750			+			ν(C=O) phospholipids ⁷	
1755–1758		+					
1774–1777	-						
1788–1790		+					

plays a key role in the development of premalignancy in breast cancer.⁴⁰ Malignant stages are accurately classified in both the 2-D and 3-D models; suggesting that the malignant characteristic is preserved in both the 2-D and 3-D cell culture models.

4.2 Variable Selection by Elastic Net

Although PLS-DA alone can provide satisfactory classification, this technique cannot reveal sparse variables that contribute to the classification of each cell line, since each of the latent variables used in the development of the PLS-DA models is a linear combination of several original variables, representing a problem in terms of variable selection. Elastic net was used in this study to solve the aforementioned problem, making the data more succinct and simpler, and providing a good interpretation of the model.⁴¹ More importantly, elastic net reveals the relationship between breast cancer stages and discriminating wavenumbers that give the most contribution to the correct classification of each cell line. This knowledge will be useful in understanding the underlying biomolecular processes in the breast cancer tumorigenesis. Based on our results, extracellular matrix plays an important role in the genesis of malignancy. Prominent coefficients indicate that nonmalignant cells exhibit a high content of polysaccharide and the lowest aromatic amino acid content (Table 3),^{42,43–51} which is in agreement with the previous report which implicates lipid genesis in neoplastic transformation^{52,50} and characteristic aromatic amino acid content during malignancy development.^{53,51} High lipid and nucleic acid content are characteristics of metastatic cells, while high aromatic content and lower polysaccharide composition are characteristic of malignant cells.^{42–53} These results suggest that lipid, polysaccharide, and aromatic amino acid constituents can serve as potential biomarkers in monitoring tumor progression.

5 Conclusion

Our study presents the first application of Raman spectroscopy to classify 3-D breast cancer cell culture models and to discriminate between the nonmalignant, premalignant, malignant, and metastatic stages of breast cancer. The 3-D model is unique because it accounts for the extracellular matrix and is a better representative of an *in vivo* environment than the 2-D cell cultures. Hence, our work presents an important step in translation. Analysis of Raman spectra using elastic net indicates that the tools developed can accurately discriminate the premalignant from nonmalignant with the specificity and sensitivity nearing 100%. We also show that Raman spectroscopy can differentiate the two forms of malignant cancer; metastatic, and nonmetastatic. Evaluation of the spectra indicates that the polysaccharide, lipid, nucleic acid, aromatic amino acid, and extracellular matrix components are involved in the staging of breast cancer. Our methods suggest that the incorporation of Raman spectroscopy and statistical techniques, such as PLS-DA and elastic net, makes the classification of cells within the tissues possible with a high degree of accuracy.

Acknowledgments

Partial funding from the Purdue University Center for Cancer Research innovation grant and the Clinical Translational Sciences Institute (CTSI) of Indiana is acknowledged. Financial support from the Brazilian Government Agency CNPq for AC is appreciated. The authors thank Sarah Voisin and Kate Stephen, former students in Irudayaraj group for assistance in the R software.

References

1. K. Larsson and R. P. Rand, "Detection of changes in the environment of hydrocarbon chains by Raman spectroscopy and its application to lipid-protein systems," *Biochim. Biophys. Acta* **326**(2), 245–255 (1973).
2. W. E. Moerner and L. Kador, "Optical detection and spectroscopy of single molecules in a solid," *Phys. Rev. Lett.* **62**(21), 2535–2538 (1989).
3. E. C. Le Ru and P. G. Etchgoïn, "Single-molecule surface-enhanced Raman spectroscopy," *Ann. Rev. Phys. Chem.* **63**, 65–87 (2012).
4. K. Kneipp et al., "Single molecule detection using surface-enhanced Raman scattering (SERS)," *Phys. Rev. Lett.* **78**(9), 1667–1670 (1997).
5. R. S. Das and Y. K. Agrawal, "Raman spectroscopy: recent advancements, techniques and applications," *Vib. Spectrosc.* **57**(2), 163–176 (2011).
6. C. Yu et al., "Characterization of human breast epithelial cells by confocal Raman microspectroscopy," *Cancer Detect. Prevent.* **30**(6), 515–522 (2006).
7. M. Gniadecka et al., "Distinctive molecular abnormalities in benign and malignant skin lesions: studies by Raman spectroscopy," *Photochem. Photobiol.* **66**(4), 418–423 (1997).
8. K. Chen et al., "Diagnosis of colorectal cancer using Raman spectroscopy of laser-trapped single living epithelial cells," *Opt. Lett.* **31**(13), 2015–2017 (2006).
9. J. R. Mourant et al., "Biochemical differences in tumorigenic and non-tumorigenic cells measured by Raman and infrared spectroscopy," *J. Biomed. Opt.* **10**(3), 031106 (2005).
10. X. L. Yan et al., "Raman spectra of single cell from gastrointestinal cancer patient," *World J. Gastroenterol.* **11**(21), 3290–3292 (2005).
11. K. Chen et al., "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer," *Int. J. Cancer* **107**(6), 1047–1105 (2003).
12. L. Hutchinson, "Breast cancer: challenges, controversies, breakthroughs," *Nat. Rev. Clin. Oncol.* **7**(12), 669–670 (2010).
13. K. Polyak, "Heterogeneity in breast cancer," *J. Clin. Invest.* **121**(10), 3786–3788 (2011).
14. A. S. Haka et al., "Diagnosing breast cancer by using Raman spectroscopy," *Proc. Natl. Acad. Sci. U. S. A.* **102**(35), 12371–12376 (2005).
15. B. Brožek Pluska et al., "Breast cancer diagnostics by Raman spectroscopy," *J. Mol. Liq.* **141**(3), 145–148 (2008).
16. C. Matthäus et al., "New ways of imaging uptake and intracellular fate of liposomal drug carrier systems inside individual cells, based on Raman microscopy," *Mol. Pharm.* **5**(2), 287–293 (2008).
17. Z. Movasaghi, S. Rehman, and D. I. U. Rehman, "Raman spectroscopy of biological tissue," *Appl. Spectrosc. Rev.* **42**(5), 493–541 (2007).
18. D. Zhang et al., "Gold nanoparticles can induce the formation of protein-based aggregates at physiological pH," *Nano Lett.* **9**(2), 666–671 (2009).
19. N. Stone and P. Matousek, "Advanced transmission Raman spectroscopy: a promising tool for breast disease diagnosis," *Cancer Res.* **68**(11), 4424–4430 (2008).
20. M. J. Bissell, D. C. Radisky, and A. Rizki, "The organizing principle: microenvironmental influences in the normal and malignant breast," *Differentiation* **70**(9–10), 537–546 (2002).
21. S. J. Santner et al., "Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells," *Breast Cancer Res. Treat.* **65**(2), 101–110 (2001).
22. F. R. Miller, Models of progression spanning preneoplasia and metastasis: the human MCF10AneoT, TGN series and a panel of mouse mammary tumor subpopulations, in *Mammary Tumor Cell Cycle, Differentiation, and Metastasis*, pp. 243–263, Kluwer academic publisher, Dordrecht, the Netherlands (1996).
23. F. R. Miller et al., "Xenograft model of progressive human proliferative breast disease," *J. Nat. Cancer Inst.* **85**(21), 1725–1732 (1993).
24. H. D. Soule et al., "Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10," *Cancer Res.* **50**(18), 6075–6086 (1990).
25. L. B. Strickland et al., "Progression of premalignant MCF10AT generates heterogeneous malignant variants with characteristic histologic types and immunohistochemical markers," *Breast Cancer Res. Treat.* **64**(3), 235–240 (2000).
26. M. Kadota et al., "Delineating genetic alterations for tumor progression in the MCF10A series of breast cancer cell lines," *PLoS One* **5**, e9201 (2010).

27. N. V. Marella et al., "Cytogenetic and cDNA microarray expression analysis of MCF10 human breast cancer progression cell lines," *Cancer Res.* **69**(14), 5946–5953 (2009).
28. D. K. Rhee, S. H. Park, and Y. K. Jang, "Molecular signatures associated with transformation and progression to breast cancer in the isogenic MCF10 model," *Genomics* **92**(6), 419–428 (2008).
29. J. Y. So et al., "Differential expression of key signaling proteins in MCF10 cell lines, a human breast cancer progression model," *Mol. Cell. Pharmacol.* **4**(1), 31–40 (2012).
30. A. Shamsaie, J. Heim, and J. Irudayaraj, "Quantification of MBA in living cell by surface enhanced raman spectroscopy," *Chem. Phys. Lett.* **461**, 131–135 (2008).
31. A. Shamsaie and J. Irudayaraj, "Intracellularly grown gold nanoparticles (IGAuN) as potential SERS probes," *J. Biomed. Opt.* **12**(2), 020502 (2007).
32. S. Ravindranath et al., "Surface-enhanced Raman imaging of intracellular bioreduction of chromate in *Shewanella oneidensis*," *PLoS One* **6**(2), e16634 (2011).
33. K. Stephen, C. Nakatsu, and J. Irudayaraj, "Surface enhanced Raman spectroscopy (SERS) for the discrimination of *Arthrobacter* strains based on variations in cell surface composition," *Analyst* **137**(18), 4280–4286 (2012).
34. K. Lee and J. Irudayaraj, "Correct spectral conversion between surface-enhanced raman and plasmon resonance scattering from nanoparticle dimers for single molecule detection," *Small J.* **9**(7), 1106–1115 (2013).
35. H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc., Ser. B* **67**, 301–320 (2005).
36. N. K. Shah and P. J. Gemperline, "A program for calculating Mahalanobis distances using principal component analysis," *TrAC Trends Anal. Chem.* **8**(10), 357–361 (1989).
37. P. J. de Groot et al., "Application of principal component analysis to detect outliers and spectral deviations in near-field surface-enhanced Raman spectra," *Anal. Chim. Acta* **446**(1–2), 71–83 (2001).
38. S. Chevallier et al., "Application of PLS-DA in multivariate image analysis," *J. Chemometr.* **20**(5), 221–229 (2006).
39. M. Zanon et al., "Assessment of linear regression techniques for modeling multisensor data for non-invasive continuous glucose monitoring," in *Conf. Proc.: Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Boston, MA, pp. 2538–2541 (2011).
40. M. P. V. Shekhar, R. Pauley, and G. Heppner, "Host microenvironment in breast cancer development: extracellular matrix-stromal cell contribution to neoplastic phenotype of epithelial cells in the breast," *Breast Cancer Res.* **5**(3), 130–135 (2003).
41. T. Zhou, D. Tao, and X. Wu, "Manifold elastic net: a unified framework for sparse dimension reduction," *Data Min. Knowl. Discovery* **22**(3), 340–371 (2011).
42. I. A. Shumova, "A cytochemical investigation of the nucleic acids, proteins and polysaccharides in malignant tumors of the uterine cervix and breast in humans," *Bull. Exp. Biol. Med.* **48**(1), 860–864 (1959).
43. H. F. Zhu et al., "Hepatocellular carcinoma cells Raman spectra with gold and silver colloid as SERS substrate," *Proc. SPIE* **8311**, 83112M1 (2011).
44. A. R. Peck et al., "Low levels of Stat5a protein in breast cancer are associated with tumor progression and unfavorable clinical outcomes," *Breast Cancer Res.* **14**(5), R130 (2012).
45. G. D. McEwen et al., "Subcellular spectroscopic markers, topography and nanomechanics of human lung cancer and breast cancer cells examined by combined confocal Raman microspectroscopy and atomic force microscopy," *Analyst* **138**(3), 787–797 (2013).
46. Z. Huang et al., "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer," *Int. J. Cancer* **107**(6), 1047–1052 (2003).
47. D. H. Kim et al., "Raman chemical mapping reveals site of action of HIV protease inhibitors in HPV16 E6 expressing cervical carcinoma cells," *Anal. Bioanal. Chem.* **398**(7–8), 3051–3061 (2010).
48. D.-W. Sun, *Modern Techniques for Food Authentication*, pp. 1, online resource (xxiv, 689 p.), Elsevier/Academic Press, Amsterdam, Boston (2008).
49. F. Severcan and P. L. Haris, *Vibrational Spectroscopy for Diagnosing and Screening*, pp. 30 online resource, IOS Press Inc., Amsterdam (2012).
50. H. G. Schulze et al., "Label-free imaging of mammalian cell nucleoli by Raman microspectroscopy," *Analyst* **138**(12), 3416–3423 (2013).
51. V. E. Baracos and M. L. Mackenzie, "Investigations of branched-chain amino acids and their metabolites in animal models of cancer," *J. Nutr.* **138**(1), 237S–242S (2006).
52. F. Zhang and G. Du, "Dysregulated lipid metabolism in cancer," *World J. Biol. Chem.* **3**(8), 167–174 (2012).
53. K. Deng et al., "High levels of aromatic amino acids in gastric juice during the early stages of gastric cancer progression," *PLoS One* **7**(11), e49434 (2012).