Toward improving fine needle aspiration cytology by applying Raman microspectroscopy

Melanie Becker-Putsche
Thomas Bocklitz
Joachim Clement
Petra Rösch
Jürgen Popp

Toward improving fine needle aspiration cytology by applying Raman microspectroscopy

Melanie Becker-Putsche,[a] **Thomas Bocklitz,**[a] **Joachim Clement,**[b] **Petra Rösch,**[a] **and Jürgen Popp**[a,c]

[a]University of Jena, Institute of Physical Chemistry and Abbe Center of Photonics, Helmholtzweg 4, 07743 Jena, Germany
[b]University Hospital Jena, Department of Hematology and Oncology, Clinic for Internal Medicine II, Erlanger Allee 101, 07740 Jena, Germany
[c]Institute of Photonic Technology, Albert-Einstein-Strasse 9, 07745 Jena, Germany

Abstract. Medical diagnosis of biopsies performed by fine needle aspiration has to be very reliable. Therefore, pathologists/cytologists need additional biochemical information on single cancer cells for an accurate diagnosis. Accordingly, we applied three different classification models for discriminating various features of six breast cancer cell lines by analyzing Raman microspectroscopic data. The statistical evaluations are implemented by linear discriminant analysis (LDA) and support vector machines (SVM). For the first model, a total of 61,580 Raman spectra from 110 single cells are discriminated at the cell-line level with an accuracy of 99.52% using an SVM. The LDA classification based on Raman data achieved an accuracy of 94.04% by discriminating cell lines by their origin (solid tumor versus pleural effusion). In the third model, Raman cell spectra are classified by their cancer subtypes. LDA results show an accuracy of 97.45% and specificities of 97.78%, 99.11%, and 98.97% for the subtypes basal-like, HER2 + /ER−, and luminal, respectively. These subtypes are confirmed by gene expression patterns, which are important prognostic features in diagnosis. This work shows the applicability of Raman spectroscopy and statistical data handling in analyzing cancer-relevant biochemical information for advanced medical diagnosis on the single-cell level. © *2013 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JBO.18.4.047001]

1 Introduction

With 1.3 million new cases and 458,000 deaths worldwide in 2008, breast cancer is one of the most common cancer in females.[1] Increasing incidences of male breast cancer[2] should not be underestimated, with about 13,000 new cases worldwide annually.[3] Hence, the very early detection and quick sampling of palpable breast masses combined with an unambiguous medical diagnosis is of utmost importance to reduce mortality rates of breast cancer patients. Conventional diagnostic techniques for tissue and cell collection are surgically implemented biopsy, punch biopsy, and fine needle aspiration biopsy (FNAB), followed by extensive cytological evaluation. In comparison to punch biopsies, FNAB is known to be a minimally invasive, very fast,[4,5] and cost-effective technique,[4–6] accompanied by a low traumatization of breast tissue because of the very small needle diameter.

Rosa[4] has outlined developments in cell collection techniques from 1847 to date. She mentioned that needles with a smaller diameter (improved from 18-gauge to 22-gauge needles) were employed, and the experience of clinicians and pathologists is increased. Further, cytological examinations were advanced by sophisticated staining techniques. Hence, the increasing diagnostic accuracy has a large share in today´s popularity of FNAB. At this point it is necessary to mention that the rates of patients with false-positive results who undergo a biopsy after 10 mammograms or 10 clinical breast examinations are enormous, with 18.6% and 6.2%, respectively.[7]

Ariga et al.[8] illustrated the precision in estimating palpable breast lesions by means of FNAB without image guidance. They accomplished examinations of 1158 fine-needle aspirates over almost 20 years. Here, malignant FNA results have been diagnosed with an accuracy of 99%. The otherwise suspicious FNA results needed to be examined by adjuvant clinical diagnostic techniques because they bear ambiguous features, which complicate an accurate diagnosis by pathologists and cytologists.

Also, Alkuwari et al.[9] described FNA as a sensitive and specific method for detecting breast cancer metastasis. In all 115 cases of axillary lymph node FNAs, they achieved an overall sensitivity of 65%, and the specificity was 100%.

Further studies by Mansoor and Jamal[10] in 2002 achieved an overall diagnostic accuracy of 93% on 72 FNA cytologies. They compared their results with 27 other studies between 1975 and 2002, and revealed diagnostic results of malignancy for FNAs with sensitivities between 66% and 100%, and specificities between 82% and 99%.

A recent study[6] worked out that the differences in grading results of palpable breast lesions applying FNAB or core-needle biopsy (CNB) with image guidance are comparably low, with an accuracy of 96% for both cases. They have achieved sensitivities of 89% and 100% for FNAB and CNB, respectively, while the specificity for FNAB and CNB accounted for 98% and 90%, respectively.

All these studies demonstrate that there is a broad range of sensitivities and specificities for diagnostic FNA results. Suspicious FNAs are also commonly occurring because of unavoidable limitations of FNA, like a limited cellular amount of the tumor and problems during histological examination.[10]

Address all correspondence to: Jürgen Popp, University of Jena, Institute of Physical Chemistry and Abbe Center of Photonics, Helmholtzweg 4, 07743 Jena, Germany, and Institute of Photonic Technology, Albert-Einstein-Strasse 9, 07745 Jena, Germany. Tel: +49 3641 948320; Fax: +49 3641 948302; E-mail: Juergen.Popp@uni-jena.de

Here, it is complicated to obtain an accurate differentiation between malignant and benign breast lesions. Accordingly, the diagnosis of suspicious FNAs requires additional and cost-intensive surgical biopsies[5] to clarify the breast lesions. These facts imply that the diagnostic accuracy is extremely dependent on the clinician's experience taking samples and the expert knowledge of pathologists and cytologists. Hence, improving diagnostic techniques associated with higher diagnostic accuracy would make such additional surgical operations dispensable. Biochemical information on the molecular level of protein, lipid, and nucleic acid structures[11] would enhance the cytologist´s and pathologist´s knowledge on breast cancer lesions.

Aside from commonly used staining techniques like hematoxylin and eosin stain, papanicolaou stain, or May-Grunwald Giemsa stain, Raman microspectroscopy is a valuable technique for collecting supplementary biochemical information without staining. Raman microspectroscopy is a noninvasive and label-free technique that enables the investigation of biological systems, e.g., eukaryotic[12] and prokaryotic cells.[13–15] Hence, Raman spectra can be recorded from single cells just after FNAB because only a minimal sample preparation is needed, and Raman signals are unimpaired by aqueous ingredients. This fact allows a following application of cytological staining techniques on the same sample. The combination of chemometric evaluation methods and Raman spectroscopy has been successfully applied for bacterial identification[16–18] and differentiation of benign and malignant cells.[11,19–21] Already, in 1995, Frank et al.[22] characterized pathological alterations in breast tissue by applying Raman spectroscopy. They highlighted the diagnostic potential of the biochemical information provided by Raman spectra. Thereupon, Haka et al.[23] studied an algorithm that describes precise chemical alterations that are associated with breast diseases. They emphasized that the biochemical information based on the contribution of lipid and collagen are diagnostically relevant. Accordingly, the results showed an increased amount of collagen in pathological tissues, whereas normal tissue reveals a larger amount of fat. The FT-Raman spectroscopic investigations on normal and tumoral human breast tissue by Bitar et al.[24] also demonstrated that Raman spectra provide diagnostically relevant information. They distinguished between normal tissue and six pathological subtypes of breast tissue by utilizing the biochemical Raman information on the lipid content, alterations in the collagen amount, and variations in the DNA content. Further studies by Moreno et al.[25] achieved a correct identification of pathologically altered tissues with 98.5% accuracy by using principal component analyses. For instance, they found different amounts of phenylalanine in normal tissue compared to altered tissue. Kelly et al.[26] already mentioned the potential of biospectroscopic tools like IR spectroscopy and Raman spectroscopy combined with multivariate analysis for supporting the diagnosis of FNA.

In this contribution, we are aiming to improve cancer diagnosis of FNAB with very poor cellular amounts of breast tumors in order to avoid insufficient diagnostic results. Therefore, Raman microspectroscopy is utilized to study six breast cancer cell lines (MCF-7, JIMT-1, T47-D, MT-3, BT-20, and HCC-1143) on a single-cell level. Here, chemometric evaluation procedures are implemented to establish classification models for the obtained Raman data. We used an SVM to discriminate between the different breast cancer cell lines. To classify based on the origin of extraction and on the grade of the breast cancer subtypes, we used LDA.

2 Materials and Methods

2.1 Cell Cultures and Preparation

The cell lines MT-3, MCF-7, BT-20, T47-D, JIMT-1, and HCC-1143 were obtained from the Department of Hematology and Oncology at Jena University Hospital, Germany. BT-20, MCF-7, and JIMT-1 were grown in 90% Dulbecco's modified eagle medium (Dulbecco's MEM, cell culture medium) with 10% fetal calf serum (FCS). MT-3 and T47-D were grown in 90% Roswell Park Memorial Institute medium (RPMI) 1640 with 10% FCS. HCC-1143 was grown in 80% RPMI 1640 with 20% FCS.

Cells were cultivated in cell culture flasks with an optimal incubation temperature of 37°C and a constant humified atmosphere (90% humidity) of 5% carbon dioxide (CO_2) in air. After cultivation in cell culture flasks, cells were removed from the flasks surface by applying trypsin. The resuspended cells were then cultivated on fused silica slides (Frank Optic Products GmbH, Germany) in petri dishes. For Raman analysis, fused silica slides were removed from culture media and air-dried.

2.2 Raman Microspectroscopy

Raman measurements were carried out using the confocal Raman microscope (CRM),[27] CRM 300 (WITec, Germany). The output of a 785-nm diode laser was focused with a 50× numerical aperture (NA) 0.95 objective (Zeiss, Germany) onto the sample. The back-scattered light was spectrally dispersed with a monochromator of 300-mm focus length and equipped with 600 lines/mm grating. An electron multiplying charge-coupled device (EM-CCD) camera with 1024×128 pixels cooled down to $-75°C$ was used for detection.

The scan mode was carried out for single cell measurements, where a given number of line scans with a given number of points was performed. The number of lines and points were chosen to achieve a step size of 0.5 μm. An acquisition time of 10 s and a laser power of about 50 mW on the sample was applied for each spectrum.

Raman data from two independently cultivated batches were recorded at different times. The first data set comprised single cell maps of complete cells including cell-free areas. Here, for each cell map, a rectangular grid with approximately 50×70 points was defined covering the spherically formed sample (single cell). For the second data set, one complete cell map and a large amount of small Raman cell maps were recorded. These small Raman cell maps were collected in the middle of the cell with a square base of 10×10 points. This measurement method was chosen to keep the measurement time for a single cell sufficiently short. Accordingly, approximately 15 cells were measured in the same amount of time (approximately 12 h), which allowed us a better insight into the cell-line diversity by studying more Raman data.

2.3 Chemometric Evaluation

The spectral analysis of Raman spectra starts with a preprocessing routine, which corrects various corrupting effects, like fluorescence phenomena, CCD baselines, noise, and cosmic spikes. First, all spikes are removed from the scans with the upper bound method,[28] and a wavenumber calibration is performed.[29] Thereafter, a polynomial of degree five is subtracted[30] in order to

reduce the influence of fluorescence from the sample. At the end of the preprocessing procedure, a principal component analysis (PCA)[31] was implemented by doing singular value decomposition (SVD) on the covariance matrix. The number of principal components (PCs) was determined by prestudies.[11,32] The optimal number of PCs depends on the classifier and number of classes. Therefore, 20 scores with 99.42% variance were used for analysis by the SVM with six classes (breast cancer cell lines). Thirty scores with 99.57% variance were utilized for evaluation by LDA of the cell line origin with two classes. Fifty scores with 99.72% variance were determined for the LDA of the cancer subtypes (three classes).

Before starting final data analysis techniques, a further preselection of Raman spectra was necessary, since Raman data of single cell measurements (cell scans) contain Raman spectra with substrate as well as Raman spectra with vibrational (i.e., molecular) information of measured cancer cells. To achieve a suitable preselection of Raman data, the cell value was evaluated for each Raman spectrum. The cell value defines the amount of biological information, and thus, the applicability of each Raman cell spectrum for investigation.

$$\frac{\max[I(1426 - 1467 \text{ cm}^{-1})]}{\sum\limits_{737-877 \text{ cm}^{-1}} I} > 0.013 \qquad (1)$$

In Eq. (1), the maximum peak intensity of the wavenumber region between 1426 and 1467 cm^{-1}, referring to biological information (CH$_2$), was divided by the sum of all Raman intensities between approximately 737 and 876 cm^{-1}, including Raman signals from DNA (phosphate backbone) and amino acids, as well as fused silica information in background. All Raman spectra exhibiting a larger cell value than 0.013 were assigned as cell Raman spectra, since they present an optimal amount of biologic information. Verification was done by comparing white light images with visualized cell values over the measured area. Thus, the area above the threshold of 0.013 was accordant with the cell shape in white light images. Figure 1 shows the fingerprint region of such mean Raman spectra of all breast cancer cell lines containing a useful amount of biological information. These Raman spectra were then used for further chemometric evaluations.

LDA[33] was utilized to analyze the Raman spectra. Here, classification models were constructed to differentiate the cancer cell line origin as well as the breast cancer subtypes. This is a supervised classification technique, which creates a linear model for separating the data into given classes. For a binary classification task, only two classes exist (group 1 and group 2), and the model can be written with the LD scaling vector $\vec{s}$ by

$$\vec{s} = \Sigma^{-1} \cdot (\vec{\mu}_1 - \vec{\mu}_2). \qquad (2)$$

In this equation, $\vec{\mu}_1$ and $\vec{\mu}_2$ stand for both group means, and Σ^{-1} represents the inverse covariance matrix. The dot product of a Raman spectrum S with $\vec{s}$ is called the LD value, and is converted by a threshold to a classification decision. The LD value can be interpreted as a class membership property. In order to evaluate the prediction properties of such a model a leave-one-out cross-validation (LOOCV) was used, and the result is arranged in a matrix called a confusion table.[34]

Another supervised classification method is the SVM.[35] The SVM seeks to construct a hyper-plane, which separates two groups from each other, and the classification function is then
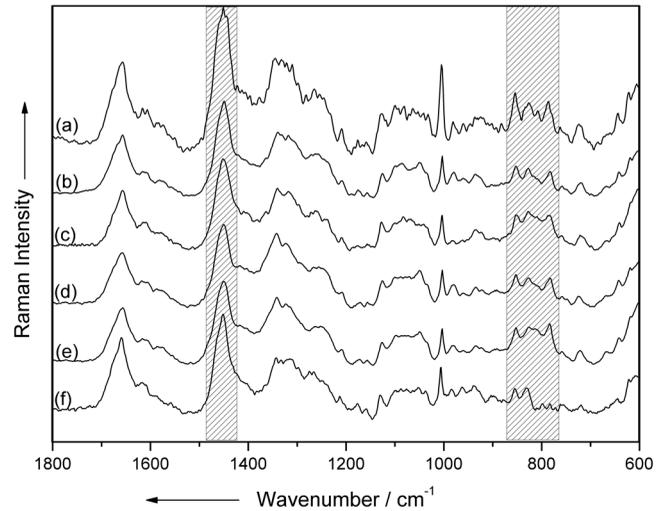
Fig. 1 Selected and preprocessed Raman mean spectra in the fingerprint region of the breast cancer cell lines (a) MCF-7, (b) T47-D, (c) BT-20, (d) HCC-1143, (e) JIMT-1, and (f) MT-3. The light-gray highlighted wavenumber regions were chosen for preselecting cancer cell Raman spectra from the complete cell scan data. Therefore, the maximum Raman intensity of the Raman band at approximately 1449 cm^{-1} was divided by the sum of Raman intensities of the wavenumber region between 737 and 876 cm^{-1}. The Raman spectra exhibiting a cell value larger than 0.013 were assigned as Raman spectra containing significant cell information [see Eq. (1)].

$$\text{sgn}\left(\sum_{i=1}^{N} y_i \cdot \alpha_i \cdot \vec{s}^{(i)} \cdot \vec{S} - \text{b}\right). \qquad (3)$$

In this equation, $\vec{s}^{(i)}$ are the N-support vectors, y_i are their class values, and α_i are weights. A new Raman spectrum S is then classified to the group -1 or $+1$ depending of the sign of the equation above. This large margin classifier is implemented using the library, libSVM. The parameters cost $= 1$ and gamma $= 1/$number of spectra[36] were used. These are the default parameters from the libSVM library and can be further optimized. The SVM builds up a multiclass classifier by the one-against-one scheme. In doing so, a binary SVM was constructed for each breast cancer cell line combination, and voting based on the prediction for each binary SVM provides the results for cell line classification. Accordingly, an SVM based on the one-against-one scheme is more suitable over other schemes for classification tasks on several groups. Therefore, SVM was utilized for the classification of the cell lines itself. For the evaluation of prediction properties of the SVM, a 10-fold cross-validation was used, where the results are demonstrated in a confusion table.

3 Results and Discussion

Single cell measurements were performed on a Raman microspectroscopic setup with an excitation wavelength of 785 nm, and Raman maps were generated using a linear scanning mode. Altogether, 110 Raman cell maps were collected. These maps include 19, 23, 26, 16, 5, and 19 single cell maps of the breast cancer cell lines BT-20, HCC-1143, JIMT-1, MCF-7, MT-3, and T47-D, respectively.

After a preselection based on the calculated cell value as mentioned in Sec. 2, a total of 61,580 Raman cell spectra (correlating with 65.7% of all Raman data) were finally evaluated. The enormous data volume, and the crucial but often minute

Table 1 Confusion table with classification results for six breast cancer cell lines, T47-D, MT-3, MCF-7, JIMT-1, HCC-1143, and BT-20, generated by SVM plus corresponding specificities and sensitivities.

True labels	Predicted labels						Sensitivity (%)
	T47-D	MT-3	MCF-7	JIMT-1	HCC-1143	BT-20	
T47-D	**13205**	1	0	39	15	1	99.58
MT-3	0	**8246**	0	0	1	0	99.98
MCF-7	2	0	**4362**	0	5	80	98.05
JIMT-1	38	1	0	**12952**	45	3	99.33
HCC-1143	9	0	0	21	**11762**	1	99.74
BT-20	7	0	25	0	4	**10763**	99.67
Specificity	99.96%	99.99%	99.96%	99.88%	99.86%	99.83%	

spectral differences in Raman spectra of cancer cells, requires a statistically based data processing method.

Figure 1 shows selected Raman mean spectra in the fingerprint region of all breast cancer cell lines, demonstrating the impossibility to distinguish such small spectral differences for several thousand Raman spectra by eye. Thus, we used supervised classification methods like SVM and LDA for data evaluation. The classifier SVM with the one-against-one voting scheme was used for a stable evaluation of multiclass problems, while the LDA was used for tasks with only few group members.

In this contribution, a classification model was created for discriminating all Raman cell spectra by their breast cancer cell line through application of SVM. Table 1 shows the SVM results for all Raman cell spectra. This classification model revealed an almost exact separation of all Raman data, with an overall accuracy of 99.52%. Thus, the classification model achieved specificities (true-negative rate) of 99.96%, 99.99%, 99.96%, 99.88%, 99.86%, and 99.83% for breast cancer cell lines T47-D, MT-3, MCF-7, JIMT-1, HCC-1143, and BT-20, respectively. Also, the sensitivities (true-positive rate) are between 98.05% and 99.98% for cell lines MT-3 and MCF-7, respectively. An additional correction of the internal correlation within the Raman cell scans was implemented. In doing so, we studied the influence of similar Raman spectra within one cell scan on the overall classification capability of the model. Here, the SVM yields an accuracy of 97.22%, which is in the range of the result obtained from the correction-free model. We refrained from using this computationally expensive correction technique in further calculations, since this evaluation allows only the prediction of single-cell scans (by LOOCV) and not of single Raman spectra.

A further classification model was carried out to separate all Raman cell spectra according to their cell line origin. Here, cell lines MT-3, BT-20, and HCC-1143 had their origin in solid tumors (group "solid tumor"), while cell lines MCF-7, JIMT-1, and T47-D were extracted out of pleural effusions (group "pleural effusion"). For this issue, the supervised classification method LDA was utilized for separating all Raman data by their physiological origin.

Out of a total of 30,839 Raman spectra from the solid tumor group, 28,896 Raman data were assigned correctly and 1943

Raman cell spectra were misclassified. For Raman data from the pleural effusions group, 29,022 Raman spectra were assigned correctly and 1727 Raman data were misclassified out of a total of 30,749 cell spectra.

The LDA model achieved a classification accuracy of 94.04% with specificities of 94.4% and 93.7% for group "solid tumor" and group "pleural effusion," respectively. The separation experiment of various cell lines by their origin achieved no exact classification results for both groups. This can be seen in Fig. 2 as an overlap in the histogram (right-hand side) and in detail in the distribution of the single spectra (left-hand side). An explanation for these results is that the cell lines undergo enormous changes during their establishment and maintenance.
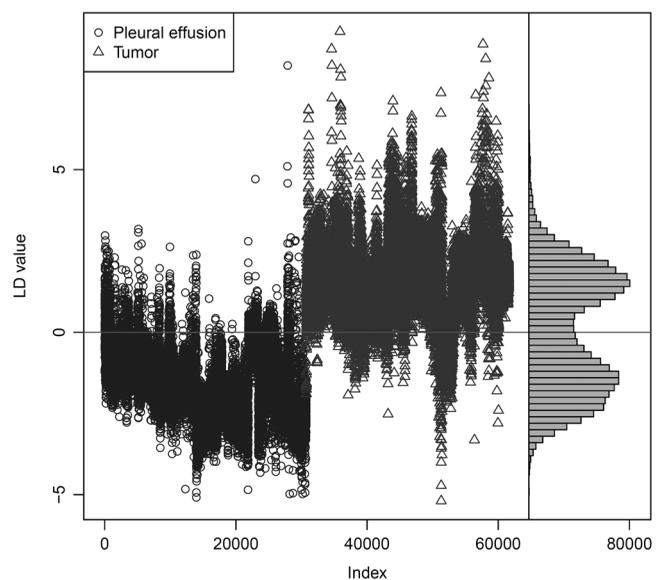
Fig. 2 Linear discriminant analysis (LDA) results for classification of breast cancer cell lines by their origin of extraction (group "pleural effusion" versus group "solid tumor"). The scatter plot and the histogram show the distribution of single Raman spectra for the group "solid tumor" containing Raman data of MT-3, BT-20, and HCC-1143 (triangle) and for the group "pleural effusion" comprising Raman data of MCF-7, JIMT-1, and T47-D (circle).

Table 2 Subtype grading for each breast cancer cell line with corresponding subtype features used for characterization.

Cell line	Subtype	ER	PR	ERBB2/HER2	Source	Tumor type	Reference(s)
MT-3	Basal	−	−	−	Primary tumor	Adenocarcinoma	39–41
BT-20	Basal A	−	−	−	Primary tumor	Adenocarcinoma	37, 42
HCC-1143	Basal A	−	−	−	Primary tumor	Ductal carcinoma	37, 43
MCF-7	Luminal A	+	+	−	Pleural effusion	Metastatic adenocarcinoma	37, 38, 42, 43
T47-D	Luminal A	+	+	−	Pleural effusion	Invasive ductal carcinoma	37, 38, 42, 43
JIMT-1	Her2 + /ER−	−	−	+	Pleural effusion	Ductal carcinoma	44
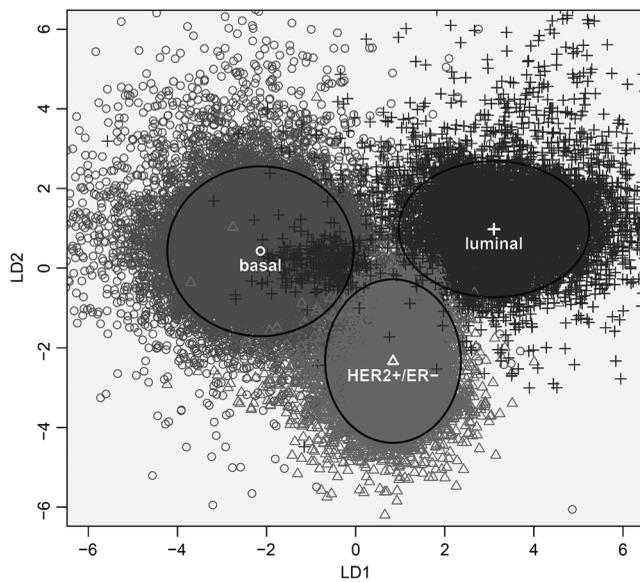
Fig. 3 LDA (linear discriminant analysis) subtype classification results for breast cancer Raman spectra. All Raman data were classified into the groups "basal-like" (circle), "HER2 + /ER−" (triangle), or "luminal" (+). A detailed breast cancer cell line assignment to respective subtypes is given in Table 2.

More interesting than the origin of cell line development is the information on the molecular profiles of each cell line. Accordingly, the cancer cells had been characterized by dominant differences in their gene expression pattern as subtypes namely, basal-like, HER2 + /ER−, and luminal A and B.[37,38] Such classifications are possible because breast cancer is established molecularly as a very heterogeneous disease. Especially the markers like estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (ERBB2/HER2) were used for subtype specification.

Table 2 summarizes molecular information corresponding to distinct subtypes for all breast cancer cell lines used here. Hence, breast cancer cell lines MT-3, BT-20, and HCC-1143 were assigned to the basal-like subtype (group "basal-like"), since their status is negative for ER, PR, and ERBB2/HER2. The luminal A subtype corresponding cell lines are T47-D and MCF-7 (group "luminal"), as their status is positive for ER and PR, while for ERBB2/HER2 the status is negative. The subtype grading for cancer cell line JIMT-1 (group "HER2 + /ER−") indicates an assignment to HER2 + /ER− subtype, because of a positive ERBB2/HER2 status and a negative status for ER and PR.

A LDA was carried out to discriminate Raman cell spectra by their molecular profiles. The scatter plot in Fig. 3 shows the classification result of the LDA. Here, 95.4% of the respective Raman cell spectra for each subtype goup were localized within each cluster area marked with a black circle. These data correlate with the double standard deviation (SD) for all Raman data of each subtype group.

Table 3 displays the discrimination result for each Raman spectrum. Consequently, out of a total of 30,839 basal-like subtype Raman spectra, 30,304 were classified correctly into group "basal-like"; 12,461 Raman spectra were assigned correctly into group "HER2 + /ER−" out of a total of 13,039 Raman data for the HER2 + /ER− subtype; and 17,253 out of 17,710 luminal subtype Raman data were accurately classified into group "luminal." Accordingly, the assignment of Raman cell data into several subtypes achieved a classification accuracy

Table 3 Confusion table with discrimination results for Raman cell spectra separated by their gene expression pattern (breast cancer subtype) plus corresponding specificities and sensitivities.

	Predicted labels			
True labels	Basal-like subtype	HER2 + /ER − subtype	Luminal subtype	Sensitivity (%)
Basal-like subtype	**30304**	371	164	98.27
HER2 + /ER − subtype	288	**12461**	290	95.57
Luminal subtype	394	63	**17253**	97.42
Specificity	97.78%	99.11%	98.97%	

of 97.45%. Here, the sensitivities and specificity are 98.27% and 97.78% for group "basal-like," 95.57% and 99.11% for group "HER2 + /ER−," and 97.42% and 98.97% for group "luminal."

These results demonstrate the high potential of Raman spectroscopy to extract diagnostic relevant information from single cancer cells without the use of time-consuming staining techniques.

4 Conclusions

In conclusion, we showed that the biochemical information in Raman cell spectra allow for classification experiments under various aspects on single breast cancer cells. In this proof-of-principle study, Raman microspectroscopy was combined with multivariate analysis, namely SVM and LDA.

The Raman data of each cell line contain significant biochemical information so that the SVM results exhibited a differentiation between the breast cancer cell lines T47-D, MT-3, MCF-7, JIMT-1, HCC-1143, and BT-20 with an accuracy of 99.51%. Furthermore, all Raman cell spectra were assigned according to their respective origin (solid tumor or pleural effusion) with 94.04% classification accuracy by applying LDA.

A diagnostically relevant step is the differentiation of breast cancer cell lines by their gene expression pattern. Accordingly, LDA revealed a classification accuracy of 97.45% with high specificities of 97.78%, 99.11%, and 98.97% for the discrimination between the breast cancer subtypes basal-like, HER2 + /ER−, and luminal, respectively. These results reveal that biochemical information in Raman cell spectra have a high potential to acquire diagnostically relevant information on single cells. Hence, the application of Raman microspectroscopy combined with chemometric evaluation can omit extensive cytological evaluations, and thus, will improve clinical diagnosis.

In future research, the current data should be enhanced with more breast cancer cell lines as well as patients' samples, because including information on biochemical diversity, e.g., in age, sex, ethnic background, progress of the disease, molecular profiling, diagnosis, and therapy, would optimize the statistical model. Identification experiments on unknown samples are planned in future studies. Exact identification of real samples on the single-cell level by means of Raman microspectroscopy can support pathologists and cytologists in clinical diagnosis on samples with suspicious FNA results.

References

1. J. Ferlay, D. M. Parkin, and E. Steliarova-Foucher, "Estimates of cancer incidence and mortality in Europe in 2008," *Eur. J. Cancer* **46**(4), 765–781 (2010).
2. J. White et al., "Male breast carcinoma: increased awareness needed," *Breast Cancer Res.* **13**(5), 219–225 (2011).
3. V. Speirs, "Not just for women," *Nature* **485**(7400), S66 (2012).
4. M. Rosa, "Fine-needle aspiration biopsy: a historical overview," *Diagn. Cytopathol.* **36**(11), 773–775 (2008).
5. P. D. Britton, "Fine needle aspiration or core biopsy," *Breast* **8**(1), 1–4 (1999).
6. S. Nagar et al., "An analysis of fine needle aspiration versus core needle biopsy in clinically palpable breast lesions: a report on the predictive values and a cost comparison," *Am. J. Surg.* **204**(2), 193–198 (2012).
7. J. G. Elmore et al., "Ten-year risk of false positive screening mammograms and clinical breast examinations," *New Eng. J. Med.* **338**(16), 1089–1096 (1998).
8. R. Ariga et al., "Fine-needle aspiration of clinically suspicious palpable breast masses with histopathologic correlation," *Am. J. Surg.* **184**(5), 410–413 (2002).
9. E. Alkuwari and M. Auger, "Accuracy of fine-needle aspiration cytology of axillary lymph nodes in breast cancer patients," *Cancer Pathol.* **114**, 89–93 (2008).
10. I. Mansoor and A. A. Jamal, "Role of fine needle aspiration in diagnosing breast lesions," *Saudi. Med. J.* **23**(8), 915–920 (2002).
11. T. Bocklitz et al., "A comprehensive study of classification methods for medical diagnosis," *J. Raman Spectrosc.* **40**(12), 1759–1765 (2009).
12. I. Notingher and L. L. Hench, "Raman microspectroscopy: a noninvasive tool for studies of individual living cells in vitro," *Expert. Rev. Med. Dev.* **3**(2), 215–234 (2006).
13. U. Schmid et al., "Gaussian mixture discriminant analysis for the single-cell differentiation of bacteria using micro-Raman spectroscopy," *Chemom. Intell. Lab. Syst.* **96**(2), 159–171 (2009).
14. M. Krause et al., "The investigation of single bacteria by means of fluorescence staining and Raman spectroscopy," *J. Raman Spectrosc.* **38**(4), 369–372 (2007).
15. M. Harz et al., "Direct analysis of clinical relevant single bacterial cells from cerebrospinal fluid during bacterial meningitis by means of micro-Raman spectroscopy," *J. Biophoton.* **2**(1–2), 70–80 (2009).
16. A. Walter et al., "Raman spectroscopic detection of physiology changes in plasmid-bearing Escherichia coli with and without antibiotic treatment," *Anal. Bioanal. Chem.* **400**(9), 2763–2773 (2011).
17. S. Meisel et al., "Raman spectroscopy as a potential tool for detection of Brucella spp. in milk," *AEM* **78**(16), 5575–5583 (2012).
18. D. F. M. Willemse-Erix et al., "Optical fingerprinting in bacterial epidemiology: Raman spectroscopy as a real-time typing method," *JCM* **47**(3), 652–659 (2009).
19. 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).
20. M. V. P. Chowdary et al., "Discrimination of normal, benign, and malignant breast tissues by Raman spectroscopy," *Biopolymers* **83**(5), 556–569 (2006).
21. C. Kendall et al., "Vibrational spectroscopy: a clinical tool for cancer diagnostics," *Analyst* **134**(6), 1029–1045 (2009).
22. C. J. Frank, R. L. McCreery, and D. C. Redd, "Raman spectroscopy of normal and diseased human breast tissues," *Anal. Chem.* **67**(5), 777–83 (1995).
23. A. S. Haka et al., "Diagnosing breast cancer by using Raman spectroscopy," *PNAS* **102**(35), 12371–12376 (2005).
24. R. A. Bitar et al., "Biochemical analysis of human breast tissues using Fourier-transform Raman spectroscopy," *J. Biomed. Opt.* **11**, 054001 (2006).
25. M. Moreno et al., "Raman spectroscopy study of breast disease," *Theor. Chem. Acc.* **125**(3–6), 329–334 (2010).
26. J. G. Kelly et al., "A biospectroscopic interrogation of fine needle aspirates points towards segregation between graded categories: an initial study towards diagnostic screening," *Anal. Bioanal. Chem.* **401**(3), 957–967 (2011).
27. L. Baia et al., "Confocal Raman investigations on hybrid polymer coatings," *Vib. Spectrosc.* **29**(1–2), 245–249 (2002).
28. D. Zhang and D. Ben-Amotz, "Removal of cosmic spikes from hyperspectral images using a hybrid upper-bound spectrum method," *Appl. Spectrosc.* **56**(1), 91–98 (2002).
29. T. Dörfer et al., "Checking and improving calibration of Raman spectra using chemometric approaches," *Zeitschrift Physikal. Chem.* **225**(6–7), 753–764 (2011).
30. C. A. Lieber and A. Mahadevan-Jansen, "Automated method for subtraction of fluorescence from biological Raman spectra," *Appl. Spectrosc.* **57**(11), 1363–1367 (2003).
31. K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philos. Mag.* **2**(11), 559–572 (1901).

32. K. Hartmann et al., "A study of docetaxel-induced effects in MCF-7 cells by means of Raman microspectroscopy," *Anal. Bioanal. Chem.* **403**, 745–753 (2012).

33. R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals Eugen.* **7**(2), 179–188 (1936).

34. W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed., Springer, New York (2002).

35. O. Ivanciuc, "Application of support vector machines in chemistry," in *Reviews on Computational Chemistry*, K. B. Lipkowitz and T. R. Cundari, Eds., pp. 291–400, Wiley & Sons, Hoboken, NJ (2007).

36. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Sys. Technol.* **2**, 27:1–27:27 (2011).

37. J. Kao et al., "Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery," *PLoS ONE* **4**(7), e6146 (2009).

38. C. M. Perou et al., "Molecular portraits of human breast tumours," *Nature* **406**(6797), 747–752 (2000).

39. M. Lacroix and G. Leclercq, "Relevance of breast cancer cell lines as models for breast tumours: an update," *Breast Cancer Res. Treat.* **83**(3), 249–289 (2004).

40. N. Prang et al., "Cellular and complement-dependent cytotoxicity of Ep-CAM-specific monoclonal antibody mt201 against breast cancer cell lines," *Br. J. Cancer* **92**(2), 342–349 (2005).

41. R. J. Hambly et al., "Establishment and characterisation of new cell lines from human breast tumours initially established as tumour xenografts in NMRI nude mice," *Breast Cancer Res. Treat.* **43**(3), 247–258 (1997).

42. L. W. Engel and N. A. Young, "Human breast carcinoma cells in continuous culture: a review," *Cancer Res.* **38**(11 Pt. 2), 4327–4339 (1978).

43. R. M. Neve et al., "A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes," *Cancer Cell* **10**(6), 515–527 (2006).

44. M. Tanner et al., "Characterization of a novel cell line established from a patient with herceptin-resistant breast cancer," *Mol. Cancer Ther.* **3**(12), 1585–1592 (2004).