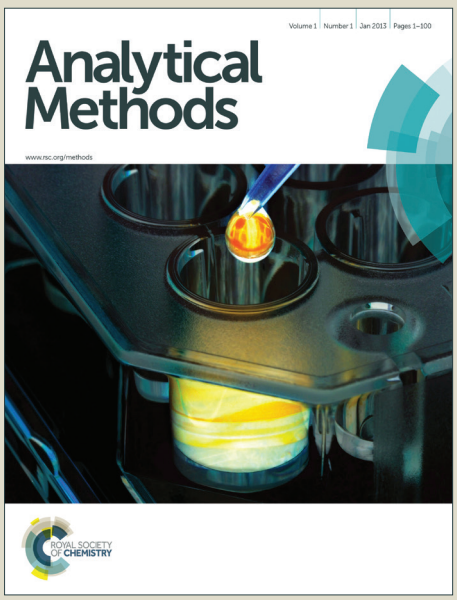
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Analytical Methods

PAPER

Discrimination of normal and malignant mouse ovarian surface epithelial cells *in vitro* using Raman microspectroscopy

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

S. Borel,[a†] E. A. Prikryl,[b†] N. H. Vuong,[c] J. Jonkman,[a] B. Vanderhyden,[c] B. C. Wilson[a,d] and S. Murugkar[b,*]

Raman microspectroscopy in conjunction with multivariate statistical analysis is a powerful technique for label-free classification of live cells based on their molecular composition, which can be correlated to variations in protein, DNA/RNA, and lipid macromolecules. We apply this technique *in vitro*, to discriminate between normal mouse ovarian surface epithelial (MOSE) cells and spontaneously-transformed ovarian surface epithelial (STOSE) cells, which are derived from the MOSE cells and are a model for high-grade serous ovarian cancer. The Raman spectra collected from individual cells undergo initial preprocessing (background subtraction, normalization and noise reduction) to yield true Raman spectra representative of the cells for subsequent statistical analysis. Using Principal Component Analysis (PCA) followed by Linear Discriminant Analysis (LDA) yields a separation of the cells into the two groups (MOSE and STOSE). This classification model has a sensitivity and specificity of 92% and 85%, respectively, after treatment of the cells to bring them into cell cycle synchrony. The main source of this separation is correlated with the increased nucleic acid content in the malignant OSE cells. As expected, a lower accuracy of 72% is obtained with asynchronous MOSE and STOSE cell populations. These results are expected to have a positive impact on the future development of improved strategies for early detection and therapeutics related to ovarian cancer.

Introduction

Ovarian cancer is the most lethal of gynecological cancers and is the fifth leading cause of cancer death among women in North America.[1] It is a prime example of a disease that is very difficult to detect, due to the lack of overt symptoms. More than 90% of ovarian cancer is of the epithelial type[2] and the majority of cases are diagnosed at a late stage when the metastatic disease is advanced and resists treatment, resulting in a dismal 5-year survival rate of less than 30%. Despite the many advances intended to enhance response to treatment, this survival rate has only marginally improved in the past few decades. Immense progress could be made if very early neoplastic changes in the Ovarian Surface Epithelium (OSE) cells and tissue could be detected and if factors that cause the disease could be better understood.[1,2]

Optical imaging offers several distinct approaches to early disease diagnostics due to high resolution, sensitivity and specificity. Several modalities, such as optical coherence tomography (OCT)[3], confocal fluorescence microscopy[4] and multiphoton microscopy, including two-photon excitation fluorescence and second harmonic generation imaging[5], have been recently investigated for early detection of changes to ovarian tissue. Although promising for accessing morphological and functional properties of tissue, these techniques do not provide the biochemical information that is important for the robust identification of early neoplastic changes, such as is required in ovarian cancer detection.

Raman spectroscopy, combined with statistical multivariate analysis, is a powerful technique for analysis of live cells and tissue.[6,7] It is based on the inelastic scattering of light due to vibrations of common molecular bonds and does not rely on exogenous agents to provide the biochemical contrast. It provides a "molecular fingerprint" and enables extraction of subtle biochemical differences. By implementing this technique on an optical microscope platform, Raman microspectroscopy has been successfully applied in the analysis of single cells, e.g. in classifying human breast epithelial cells[8], urological cells[9] and leukaemia cells[10] and for differentiating between stem cells and their differentiated progeny.[11] Raman microspectroscopy, coupled with multivariate analysis, has accurate classification between cell and tissue[12,13] types, cell cycle stage and response to drug exposure.[14]

It is, therefore, hypothesized that Raman micro-spectroscopy could be developed into a powerful diagnostic tool in ovarian cancer. Contrary to expectations, only a limited number of Raman-based studies of ovarian cancer have been conducted to date.[15] One such study from 2005 by Krishna et al[16] found significant differences between spectral profiles of

[a] *Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada.*
[b] *Department of Physics, Carleton University, Ottawa, ON, Canada.*
[c] *Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada.*
[d] *Princess Margaret Cancer Centre, Toronto, ON, Canada*
† *These authors contributed equally to this work*
* email: smurugkar@physics.carleton.ca

normal and malignant formalin-fixed ovarian tissues. This was attributed to an excess of DNA and lipids and decreased amount of proteins in cancerous tissue. The same group performed a small pilot study in 2008[17] showing the potential clinical usefulness of the Raman-based approach. Other than this earlier work, the application of Raman spectroscopy to investigations involving OSE cells and tissue has been largely unexplored. In this initial *in vitro* study involving live mouse OSE cells, we demonstrate that Raman microspectroscopy combined with multivariate analysis permits the differentiation of normal and malignant OSE cells for the first time. We show that the main source of this separation is correlated with the increased nucleic acid content in the malignant OSE cells.

Materials and Methods

Cell Culture

The live OSE cell samples comprised both normal mouse ovarian surface epithelial (MOSE: M0505) cells and spontaneously-transformed ovarian surface epithelial (STOSE) cells, which are derived from the former and are a known model for high-grade serous ovarian cancer. MOSE and STOSE cells were cultured in MOSE media according to the protocol described in Gamwell et al.[18] After the cells reached confluence, TrypLe (Life Technologies, Burlington, ON) was used to detach them from the tissue culture dish. They were then plated at low density on a thin 25.4 mm diameter quartz coverslip (01019T-AB SPI Supplies, West Chester, PA) contained in a 60 mm tissue culture dish.

It has been demonstrated in earlier work[19,20] that both cell cycle progression and changes in cell culture confluency, are major sources of spectral variability in Raman spectra of single cells measured from the same cell culture *in vitro*. Since the objective of this work was to find the true biochemical differences between MOSE and STOSE cells, appropriate steps were implemented as follows, to minimize the effect of the confluency and the cell cycle on the measured spectral variability. Both MOSE and STOSE cells were propagated until they were close to confluent and a cell cycle inhibitor Roscovitine (CST, Whitby, ON, Canada) was used to arrest the cell cycle. A solution of Roscovitine was added to the cell media to a final concentration of 20 µM in 0.1% DMSO. This concentration of Roscovitine was determined from a set of preliminary dose-response experiments to identify the maximally effective concentration that inhibited cell proliferation. This is a concentration that we have used previously[21] without apparent cytotoxicity. The same amount of 0.1% DMSO was added to the untreated control cells for which the cell cycle was not arrested. This amount of DMSO is expected to have no impact on cell viability.

Imaging was done 24h after Roscovitine treatment application. The quartz coverslip containing the adhered cells was transferred to a Chamlide magnetic chamber (Live Cell Instrument, Seoul, Korea) and filled with fresh media containing either Roscovitine in DMSO or DMSO only. A Chamlide stage top incubator (Live Cell Instrument, Seoul, Korea) was used to maintain the cells under optimum conditions of 5% CO2, 50% humidity and 37°C during the acquisition of Raman spectra.

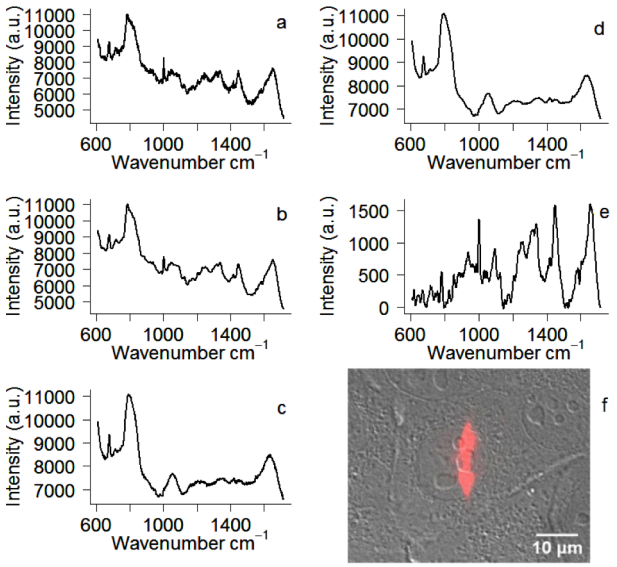


Fig. 1: Pre-processing steps for a sample Raman cell nucleus spectrum. (a) Raw cell spectrum, (b) Savitzky-Golay smoothing applied to cell spectrum, (c) average of raw background spectra, (d) Savitzky-Golay smoothing applied to background spectra, (e) correction for the background contributions using the SMIRF background removal algorithm, and (f) The 785nm laser focus spot superimposed on a differential interference contrast (DIC) image of the cell monolayer, acquired with a 63x water immersion objective. The laser spot is centred on the nucleolus of the cell and spans the nucleus.

Raman Spectra Collection

Raman measurements were made on a confocal Raman system (Renishaw inVia, Hoffman Estates, IL, USA) configured to an inverted microscope base (Leica DMI6000B, Concord, ON, Canada) fitted with a 63X water-immersion objective lens. 785 nm excitation light was used at a power of 70mW at the sample. Each spectrum was acquired with a total integration time of 1 min (6 scans each of 10s).

Spectra were acquired from 40 randomly selected cells adhered to a quartz substrate for each of the four groups (MOSE control, MOSE treated, STOSE control, and STOSE treated) for a total of 160 spectra. **Fig. 1**f is an example of a typical image of MOSE cells in the field of view of the Raman microscope obtained using Differential Interference Contrast (DIC). The inset illustrates a laser spot superimposed on one MOSE cell such that the sampling area of 2 µm by 15 µm is centred on the nucleolus and spans the nucleus and some of the cytoplasm. Background spectra were acquired in four different locations of the quartz substrate containing only the cell media with or without Roscovitine in DMSO/ DMSO.

Data processing

A Savitzky-Golay smoothing filter (15 points, second-order polynomial) was first applied to each spectrum. An effective background subtraction method was then used to yield true Raman spectra for analysis and reduce or eliminate the varying levels of background fluorescence signal from the cells, media, DMSO, Roscovitine and the Raman signal from the quartz substrate. These background contributions were removed using the Spectrum-based Method for Iterative Removal of Fluorescence (SMIRF) Background Removal algorithm developed in MATLAB® by Beier and Berger,[22] for which the average of the background spectra collected from the cell media was used as the contaminant spectrum and a 5th-order polynomial was used to model the broad and slowly-varying background fluorescence. Each spectrum was then normalized to the total area (**Fig. 1**). This normalization method makes no assumptions about the cell biochemistry, unlike other methods such as normalization to a single Raman peak at 1450 cm^{-1}, which has the added minor disadvantage of not being able to assess variation of this peak in the spectral analysis. The Raman shifts were mean-centred for subsequent processing.

The quartz Raman contribution to the background spectra was significant and varied with location. Hence, the background spectra were stronger than the cell spectra in certain spectral ranges, most notably in the 800 cm^{-1} region. In addition, a sharp peak at 670 cm^{-1} from DMSO[23] may not have been removed consistently and could result in artificial contributions in this region. DMSO also produces a small and broader peak at 710 cm^{-1} [24] but this peak has less impact on the background subtraction.

Principal Component Analysis and Linear Discriminant Analysis

Several chemometric methods exist for analyzing biomedical spectroscopy data[25,26]. Principal component analysis (PCA) is an unsupervised technique that is commonly used for dimensional reduction of the spectroscopic data set. It produces a new set of orthogonal variables, called principal components (PCs), along which the data set's variance is maximized. In particular, the first principal component accounts for the greatest variance, and each subsequent principal component accounts for successively less variance. In this manner, important information in the data for the characterization and classification of spectra is retained, while redundancy in the data is reduced.[27] In contrast, Linear discriminant analysis (LDA) is a supervised technique that aims to optimize the separation between groups of data, but minimizes the separation within groups. It takes the cell labels (for example, MOSE-control, STOSE-control, MOSE-treated, or STOSE-treated in this study) as input, and computes the component, called the linear discriminant, along which the separation between the two groups is greatest.[28] The technique known as PCA-LDA uses the PCs as input variables instead of the actual Raman shifts in wavenumbers, to prevent over-fitting of the LDA model. An important assumption in LDA is that the data, in this case the PC scores, are normally distributed.[29] PCA combined with LDA has been widely used to classify and analyze Raman spectra.[9,12,13,30,31-35]

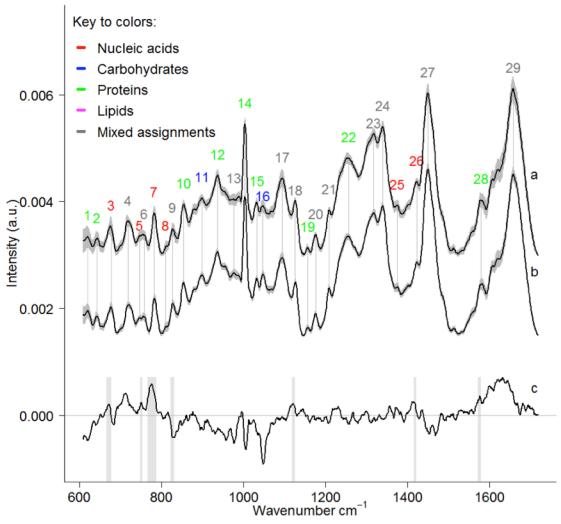


Fig. 2: (a) Average STOSE spectrum, (b) average MOSE spectrum, and (c) difference spectrum generated by subtracting the average MOSE spectrum from the average STOSE spectrum. The grey curves are the standard deviation envelopes. Spectral peaks of MOSE and STOSE spectra are highlighted by grey lines and annotated with numbers assigned as per Table 1. Peaks in the difference spectrum associated with nucleic acid contributions are highlighted with grey bands.

In this study PCA was performed on the cell spectra of the control and treatment MOSE and STOSE groups using the STATS package in the R programming language.[36] LDA was applied to the output of PCA using the MASS package in R.[37] The choice of the number of PCs to use in the LDA is determined by examining the location of the elbow of the scree plot and keeping the principal components that precede the elbow. Only the first two PCs were found to be the dominant contributors to the LDA as explained in the sections below. The specificity and sensitivity of the classification for the confusion matrix were computed taking the STOSE classification as the positive and the MOSE classification as the negative. This was based on the leave-one-out-classification validation (LOOCV) method.[28] The loadings of the first principal component and the linear discriminant were examined to identify the Raman shifts that accounted for the biomolecular differences between the cell types.

Results and Discussion

Control (untreated) samples

Fig. 2 shows the average of 40 STOSE spectra (a) and the average of 40 MOSE spectra (b) of the control (untreated) samples. Spectral differences between the control MOSE and STOSE cells were determined from the difference spectrum (c) obtained by subtracting the average MOSE spectrum from the average STOSE spectrum. These spectra contain many contributions from proteins, lipids and nucleic acids, as well as a few contributions that may be from polysaccharides. Table 1

provides a summary of the most likely molecular assignments of peaks in the average MOSE and STOSE spectra based on literature values. The peaks corresponding to nucleic acid content are indicated in bold. The wavenumber identifications for all of these peaks should be considered in terms of narrow bands of wavenumber ranges instead of a single wavenumber, since the peaks have non-zero widths. However, it is often the case that the bands overlap. Thus, several molecular assignments can correspond to adjacent wavenumbers that all lie within a given band, resulting in mixed assignments for several of the bands. The peaks in the two cell types are the same, but their intensities vary.

Examination of the peaks in the difference spectrum in Fig. 2(c) yielded information about variations in molecular content. The positive Raman bands labelled in Fig. 2 suggest greater nucleic acid content in STOSE cells than in MOSE cells. Specifically, the band at 666-678 cm^{-1} is associated with G, T ring breathing modes in DNA/RNA bases and a C-S stretching mode in cytosine (C).[38,39] The band at 766-786 cm^{-1} is associated with U, T, C ring breathing modes in DNA/RNA bases as well as O-P-O backbone stretching. There are also several bands with mixed assignments that include possible nucleic acid contributions. In particular, the band at 748-753 cm^{-1} may be associated with DNA, but there are also assignments to the symmetric breathing of the amino acid tryptophan that can be made in this range.[38] The band at 1118-1125 cm^{-1} may be due to C-O stretching of ribose, which could be a marker for RNA, but there are also many assignments to C-N, C-C and C-O-C stretching due to proteins, lipids, and carbohydrates in this region.[38] The band at 1415-1421 cm^{-1} could have DNA contributions as well as protein, lipid and carbohydrate contributions through C-H deformations.[38] The band at 1573-1580 cm^{-1} could be due to A,G ring breathing of DNA/RNA or due to tryptophan.[38] However, a negative band was observed at 822-831 cm^{-1} which has possible contributions from the O-P-O stretching of DNA as well as from the amino acids proline and tyrosine.[38]

Table 1: Tentative molecular assignments for peaks in average MOSE and STOSE spectra from control group[a]. **Bold** indicates nucleic acid-related peaks.

Band	Raman Shift (cm^{-1})	Molecular Assignment(s) with Raman Shifts from Literature	Ref.
1	618 621	C-C twisting (p) C-C twisting mode of phenylalanine (p)	38 38
2	640 643	C-S stretching & C-C twisting of proteins - tyrosine (p) C-C twisting mode of tyrosine (p)	38 38,43
3	678	**G ring breathing (d)**	38
4	717-719	Choline (l), C-N (membrane phospholipids head) (l), **A (d)**	38
5	746	**T (ring breathing mode of DNA/RNA bases) (d)**	38
6	755 759	Symmetric breathing of tryptophan (p) Tryptophan (p), ethanolamine group (l), or phosphatidylethanolamine (l)	38 38
7	782	**U,C,T ring breathing (d)**	38
8	810	**phosphodiester (d)**	38
9	827	Out-of-plane ring breathing, tyrosine (p), **O-P-O stretch (d)**	38,43,20
10	853	Ring breathing mode of tyrosine & C-C stretch of proline ring (p)	38,43, 20
11	898	Monosaccharides (beta-glucose) and disaccharides, (C-O-C) skeletal mode (c)	38
12	936 937	C-C symmetric stretch backbone, α-helix (p) C-C backbone (p)	20,38
13	974 980 980	**Ribose vibration (d)** C-C stretching beta-sheet (p) =CH bending (l)	38 38 38
14	1000/1/3	Phenylalanine (p)	38,43, 20
15	1030 to 1040	C-H bending phenylalanine (p)	44
16	1048 1049	glycogen (c) C-O stretch in carbohydrates (c)	38 45
17	1093 1094 1094	**Symmetric O-P-O stretch of DNA backbone (d)**, or C-N stretch (p) **(d)** C-N stretch (p), or chain C-C stretch (l)	38 38 20
18	1125	C-N stretching (p); C-C stretching (l)	44
19	1158	C-C/C-N stretch (p)	38, 45
20	1175/6 1176	**C, G (d)** C-H bending tyrosine (p)	38 38
21	1208	tryptophan, phenylalanyne (p), or **A, T ring-breathing (d)**	38
22	1254	C-N in-plane stretch (p)	38
23	1320 1320	**G (d)** C-H deformation (p)	38 38, 46, 45
24	1336-1345	**A, G (d)**, C-H deformations (p), CH2 twist (l)	44
25	1373 1374	**T, A, G (ring breathing of DNA/RNA) (d)** **T (d)**	38 20
26	1421 1422	**A, G (d)** **Deoxyribose (d)**	38 38
27	1450	CH def (p, l)	38, 20
28	1582	phenylalanine (p)	38
29	1656 1655-1680	amide (p) or C=C stretch (l) Amide I (p)	20, 38

[a] Abbreviations: (d) nucleic acid, (c) carbohydrate, (p) protein, (l) lipid (G) guanine, (C) cytosine, (A) adenine, (T) thymine, (U) uracil

The search for biomolecular differences through analysis of the difference spectrum and PCA-LDA was guided by the paper on STOSE cells by McCloskey et al.[40] STOSE cells are characterized by their rapid growth, having a doubling time of 13 h that is nearly four times faster than MOSE cells.[40] STOSE cells also have a high degree of aneuploidy;[40] that is, they have an abnormal (higher) number of chromosomes in some of the chromosomal sets. Moreover, the majority of the STOSE cells are near-triploid and a smaller portion are polyploid.[40] Although there is also some degree of aneuploidy in MOSE cells (2/5 are near-tetraploid cells, while 3/5 are near-diploid), this agrees with the increased DNA content observed in cell cycle analysis of these cells, and is not to the same extent as the aneuploidy in the STOSE cells.[40] Thus, a higher DNA content was expected in the spectra of the STOSE cells than the MOSE cells. The above results are consistent with this,

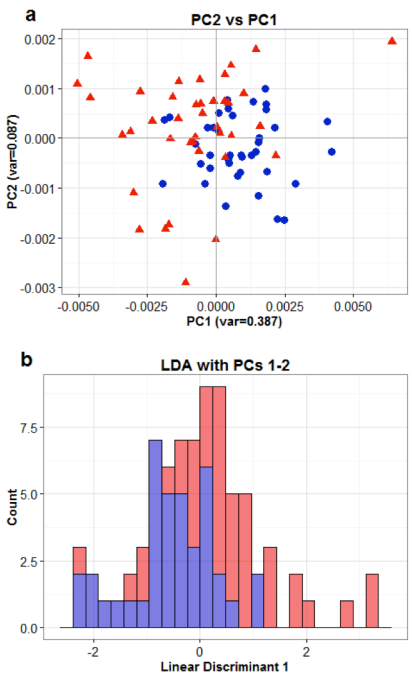


Fig. 3: PCA score plots (a) and LDA histograms (b) for control MOSE (blue) and STOSE (red) cells.

albeit not conclusively because of the large number of mixed assignments.

Classification of control sample cells (PCA-LDA)

PCA was performed on the control sample set of MOSE and STOSE cells. Fig. 3a shows a plot of PC scores for the control cell samples after removing two outlier spectra (one MOSE and one STOSE). The PC1 scores produce some separation between the two groups, but it also accounts for some spectral variation within the groups, while PC2 accounts primarily for variation within the groups, as can be seen by the spread of the points within the group along that variable (Fig. 3a). LDA was performed on the first two PCs (accounting for 38.7% and 8.7% of the variance, respectively), yielding a poor separation of the two groups (Fig. 3b). The choice of the number of PCs to use in the LDA was determined by examining the location of the elbow of the scree plot (not shown) and keeping the principal components that precede the elbow. If more PCs are used in the LDA, a greater separation is produced; however, this greater separation is sample-dependent because, after the first few PCs, the variance that subsequent PCs account for is likely due to subtle cellular variations that would become less important with greater sample size, and not to consistently identifiable intergroup variation. PCA demonstrated that the spectral variation among the control group cells was due not only to variation between groups, but also to variation within groups, resulting at least in part from the cells being in various phases of the cell cycle.

Classification of treated samples (PCA-LDA)

The cells were treated with Roscovitine to arrest them in the G1/G0 or G2/M phases.[41,42] This was expected to minimize cell cycle variability as the source of spectral variation within groups as explained earlier.[20] The average of the Raman spectra of 40 treated MOSE and 40 treated STOSE cells is shown in Fig. 4a. PCA performed on the treated cells revealed that PC1 and PC2 accounted for 52.4% and 6.2% of the variance, respectively, with a decreasing contribution from the rest of the PCs. Hence only PC1 and PC2 are considered for the LDA classification. Fig. 4b shows the difference spectrum, PC1 loadings plot, and the LD loadings plot.

The PC scores plot for the treated MOSE and STOSE cells is presented in Fig. 4c. It clearly shows that the clustering of the two treated groups is improved compared to that seen with the untreated control group of samples (Fig. 3). This enhanced group separation is also observed in the LDA in Fig. 4d. These results are summarized in Table 2 by means of the confusion matrix for LOOCV on PCA-LDA of the treated and control (untreated) groups of cells. The PC-LDA model separated the treated groups of MOSE and STOSE cells into two clusters with 92% sensitivity and 85% specificity (Table 2). In contrast, the specificity and sensitivity of classification of control (untreated) cells were both found to be 72%.

The loadings plots of the first principal component and the linear discriminant were examined for the treated samples to identify the Raman shifts that accounted for the biomolecular differences between the cell types. It can be seen from Fig. 4b that these two plots are very similar to the difference spectrum. This serves to confirm the significance of the peaks in the difference spectrum to account for intergroup spectral variation.[20] Table 3 summarizes the tentative Raman band assignments for the difference spectrum of the treated samples, with the nucleic acid contributions in bold. The positive features in Fig. 4b suggest an overall increase in the nucleic acid and protein content (including some mixed assignment) in the STOSE spectra. There is also a relative

Table 2: Confusion Matrix for LOOCV on PCA-LDA of Treated and Control Cells

TREATED	Predicted		
Actual		MOSE	STOSE
	MOSE	34	6
	STOSE	3	37
	Sensitivity	92%	
	Specificity	85%	
	Accuracy	89%	
CONTROL	Predicted		
Actual		MOSE	STOSE
	MOSE	28	11
	STOSE	11	28
	Sensitivity	72%	
	Specificity	72%	
	Accuracy	72%	

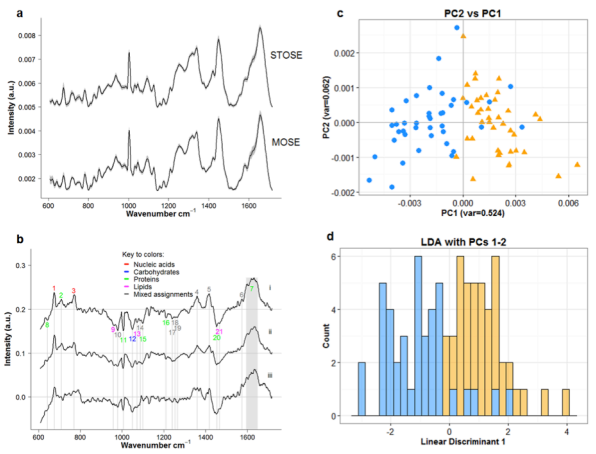


Fig. 4: a) Average STOSE and MOSE spectra from treated group. (b) comparison of: (i) difference spectrum (STOSE - MOSE), (ii) PC1 loadings plot, and (iii) LD loadings plot for treated spectra. Spectral peaks are numbered and highlighted by grey lines, as per Table 3. (c) PC scores plots and (d) LDA histogram for treated MOSE (blue) and STOSE (orange) cells.

decrease in the lipid and protein content in the treated STOSE spectra compared to the treated MOSE spectra. At first glance, this seems to contradict earlier reports [47,48] of the observation of increased lipid metabolism in cancer cells. This was investigated further by comparing control and treated samples to determine whether the effect of Roscovitine treatment could have resulted in this observation.

Effect of Roscovitine treatment within group

Fig. 5 shows the average Raman spectrum of the treated and control cell samples for the MOSE and STOSE groups. An overall decrease in the lipid and to some extent the protein content is observed in the treated cells relative to the control (untreated) cells in both the MOSE and STOSE groups. This is suggested by the negative features in the difference spectrum in Fig. 5 and is based on the band assignments in Table 3. A relatively strong increase is seen in the protein band around 1660 cm^{-1} in the treated samples compared to control (untreated) samples of both the MOSE and STOSE groups. The presence of the sharp feature at 675 cm^{-1} could be attributed to nucleic acid content or could be an artifact of background subtraction as mentioned earlier.

Fig. 5 also shows good clustering of the control and treated cells in the PC scores plots for each of the MOSE and STOSE group of cells. This suggests that there are significant spectral differences between the treated and control cells and that Roscovitine is eliminating variation in the cell cycle to a large degree but not with 100% effectiveness. Although Roscovitine is known to arrest cells in either the G1/G0 or the G2/M phases, the majority of the cells were expected to be arrested in the G1/G0 phases because cells do not generally stay in the G2/M phases for long periods of time, although both results

Table 3: Provisional molecular assignments for peaks in the difference spectrum from the treatment group[a]. **Bold** indicates nucleic acid-related peaks.

Band	Wavenumber (cm^{-1})	Molecular Assignment(s)	Ref.
1	653-673 678	**G (d)** **G ring breathing (d)**	39 38
2	700-45	C-S trans vibration (aminoacid methionine) (p)	38
3	766	**Pyrimidine ring breathing (d)**	38
4	1355/7 1359 1360 1361/2/3/5	**G (d)** **Tryptophan (p)** **Tryptophan (p)** **G (d)**	38
5	1417 1420 1420/1	C=C stretch in quinoid ring CH2 (p,l) **Deoxyribose (d)**	38
6	1573 1575	**A, G (d),** tryptophan (p) **A, G ring-breathing (d)**	38
7	1600-1800 1640-1680	(p) Amide I band (p)	38
8	643	C-C twisting mode of tyrosine (p)	38,43
9	957	Cholesterol (l)	38,43
10	977 980	Phenylalanine (p) C-C stretching of beta-sheet (p), =CH bending (l)	38 38
11	1005	Phenylalanine (p)	38
12	1048	Glycogen (c)	38
13	1073	Triglycerides (l)	38
14	1087-1090	C-C stretch, phosphate stretch (**d**,l)	38
15	1099	C-N stretch (p)	38
16	1209 1210	Tryptophan, phenylalanine (p) Tyrosine, phenylalanine (p)	38
17	1239 1240 1241	Amide III (p) **RNA peak, or asymmetric O-P-O stretching modes (d),** or collagen (p) **Asymmetric O-P-O Stretching modes (d)**	38
18	1250/2 1254	**G, C (NH_2 mode) (d)** C-N in plane stretching (p)	38
19	1264 1265/6	Triglycerides (l) Amide III, α-helix (p)	38
20	1453	(p)	38
21	1465	(l)	38

[a] Abbreviations: (d) nucleic acid, (c) carbohydrate, (p) protein, (l) lipid (G) guanine, (C) cytosine, (A) adenine, (T) thymine, (U) uracil

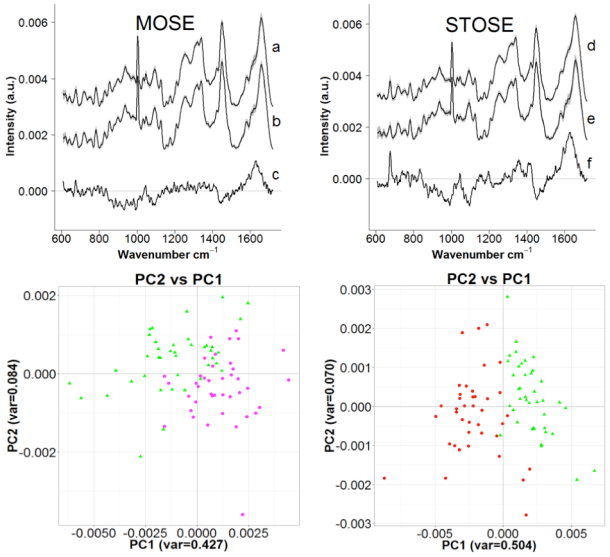


Fig. 5: Left column: Average MOSE (a) treated, (b) control and (c) difference spectra (treated-control), and PC scores plot for MOSE control and treatment spectra. Right column: Average STOSE (d) control and (e) treatment, and (f) difference spectrum, and PC scores plot for STOSE control and treatment spectra.

have been observed in previous studies using Roscovitine.[41,42] It does serve the main purpose in these experiments of reducing variability from DNA replication in S phase and so allows better comparison of baseline DNA levels in the MOSE and STOSE cells.

Discrimination of MOSE and STOSE cells

Fig. 6 presents a comparison of the difference spectra representing the difference between the average spectrum of STOSE and the average spectrum of MOSE, for both the control and treated groups. A higher overall nucleic acid content is clearly visible in the STOSE cells compared to the MOSE cells in both the control groups and the treated groups in Fig. 6. In addition, lower lipid (and mixed protein assignment) content is observed in the STOSE compared to the MOSE cells, in addition to a significantly higher protein contribution around 1660 cm^{-1}. This result that is related to the lower lipid and protein contribution is attributed to the effect of Roscovitine on the STOSE cells, as discussed above. Hence, it can be concluded that the main source of the biochemical differences between the STOSE and MOSE cells is the increased nucleic acid content in the STOSE cells. As found from Table 2, PCA-LDA of the treated group provides improved classification based on these biochemical differences, compared to the classification of the control cells.

Raman microspectroscopy from localized regions such as the nuclei as performed here is a useful approach for discriminating between cell types. However additional Raman spectra from the rest of the cell could further improve the basis for the PCA-LDA classification[49]. Moreover, alternative statistical techniques such as biochemical components analysis[50] as well as other methods for classification[30] such as boosted trees algorithms[51] may yield more precise information regarding biochemical differences and classification.

Conclusions

In this initial study involving live mouse ovarian surface epithelial cells, we have demonstrated that Raman microspectroscopy combined with multivariate analysis is a powerful analytical tool that enables effective discrimination between normal and malignant mouse OSE cells for the first time. This investigation of the near-infrared Raman spectral characteristics of normal and transformed mouse ovarian surface epithelial cells has demonstrated that there are significant differences between these, most likely associated with nuclear changes, as might be expected from the known alterations in the nuclear size and degree of aneuploidy. This is significantly, consistent with the hyperploidy (increased number of chromosomes) in the nuclei of STOSE cells recently reported by some of us using standard techniques.[40] In terms of the capability of Raman microspectroscopy to correctly classify OSE cells as normal or malignant, the results are encouraging, in that PCA-LDA analysis gives sensitivity and specificity of 92% and 85%, respectively, after treatment to bring the cells into synchrony. Further *in vitro* work will be required to consolidate these findings, including more detailed analysis of the spectral assignments.

This initial work suggests a promising path for constructing multivariate classification models in the future using calibration Raman spectra of OSE cells of known identity such as MOSE vs STOSE, normal ovarian stem cells vs cancer stem cells, as well as malignant ovarian cells that are sensitive or

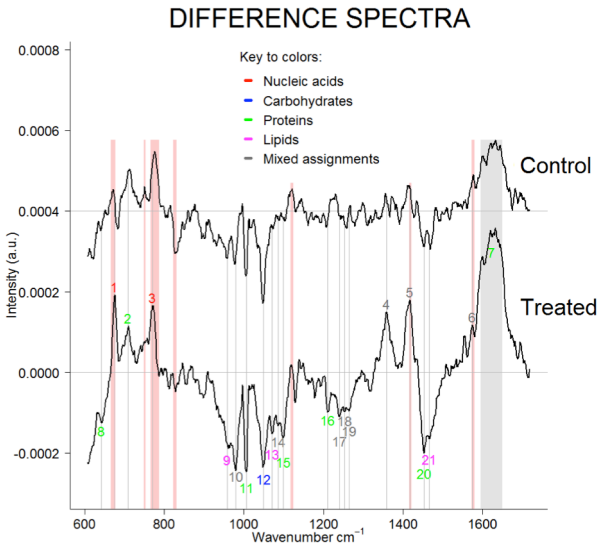


Fig. 6: Difference spectra (STOSE - MOSE) for control and treated groups. Red bands are the bands with nucleic acid contributions identified in the control difference spectrum from Fig. 2, and features that are numbered and highlighted in grey are for the treated difference spectrum from Fig. 4b.

resistant to treatment. This could potentially lead to i) early detection of ovarian cancer, ii) an efficient method for isolating stem cells as therapeutic targets in ovarian cancer and iii) early assessment of treatment response in ovarian cancer.

In terms of translation to potential clinical impact, there have been a number of studies demonstrating that point Raman spectroscopy can be implemented endoscopically[52] or intraoperatively[51], using fiber-optic probes specifically designed for this purpose. This approach has been demonstrated in several different clinical settings and for different cancers, both to detect and stage lesions[53] and to identify tumor margins for surgical guidance.[51] Both approaches would be highly relevant to the management of ovarian cancer patients during laparoscopy. For these applications, it will be critical to demonstrate that high sensitivity and specificity for tumor identification can also be achieved in tissue *in vivo*, where clearly there will be heterogeneity in cell cycle as well as variations in the tumor microenvironment that may adversely affect the results. Nevertheless, we are encouraged by the present positive results to take this approach to the next stage of investigation in animal models of ovarian cancer and in human tissues, firstly *ex vivo* and then *in vivo*.

Acknowledgements

The authors gratefully acknowledge funding from Carleton University Start-up funds and NSERC Discovery Grant (SM) and Canadian Institutes for Health Research grant (BV) and (BCW/ #RMF11162). We would also like to thank Dr. Brooke Beier, Dr. Yelena Ilin, Dr. Quinn Matthews, Prof. Andrew Jirasek and Prof. G. D. Sockalingum, for helpful advice and discussions.

Notes and references

[1] R. C. Bast, B. Hennessy, and G. B. Mills, Nat Rev Cancer 2009, 9(6), 415.
[2] P.E.Colombo, M. Fabbro, C. Theillet, F. Bibeau, P. Rouanet, I. Ray-Coquard, Crit. Rev. in Oncology/Hematology, 2009, 89, 207–216.
[3] L. P. Hariri et al., Gynecol. Oncol., 2009, 114(2), 188–194.
[4] A. A. Tanbakuchi et al., Am. J. Obstet. Gynecol., 2010 202(90), e1–e9.
[5] R. M. Williams et al., Transl. Oncol., 2010, 3(3), 181–194.
[6] C. Kendall, N. Stone et al, Analyst , 2009, 134, 1029–1045.
[7] D. Ellis, R. Goodacre et al, Analyst, 2013, 138, 3871.
[8] C. Yu et al, Cancer Detection and Prevention, 2006, 30, 515–522.
[9] T. J. Harvey et al, J. Biophoton., 2009, 2(1–2), 47–69.
[10] J. W. Chan et al, Anal. Chem., 2008, 80, 2180.
[11] H. G. Schulze, R. F. B. Turner et al, Anal. Chem., 2010, 82, 5020–5027.
[12] M. Larraona-Puy et al., J. Biomed. Optics, 2009, 14(5), 054031-054031-10.
[13] M. Fullwood et al., Anal. Methods, 2014, 6, 3948-3961.
[14] I. W. Schie and T. Huser, Appl. Spectroscopy, 2013, vol 67, 813.
[15] C. Mallidis, V. Sanchez, J. Wistuba, F. Wuebbeling, M. Burger, C. Fallnich and S. Schlatt, Human Reproduction Update, 2013,0 pp. 1–12.
[16] C. M. Krishna, G. D. Sockalingum, L. Venteo, R. A. Bhat, P. Kushtagi, M. Pluot and M. Manfait, Biopolymers 2005,79, 269–276.
[17] K. Maheedhar, R. A. Bhat, R. Malini, N.B. Prathima, K. Patil, P. Kushtagi and C. M. Krishna, Photomedicine and Laser Surgery, 2008, 26, 83.
[18] L.F. Gamwell, O. Collins, B.C. Vanderhyden, Biology of Reproduction, 2012, 87, 80.
[19] R. Swain et al, Journal of Cellular Biochemistry, 2009, 104, 1427–1438.
[20] Q. Matthews, A. Jirasek, J. Lum, X. Duan and A. G. Brolo, J. Appl. Spectrosc., 2010, 64, 871-887.
[21] F. H. Thomas, J. F. Ethier, S. Shimasaki, B. C. Vanderhyden, Endocrinology, 2005, 146(2):941-949.
[22] B.D. Beier and A.J. Berger, Analyst, 2009, 134, 1198-1202.
[23] C. W. Freudiger, W. Min, B. G. Saar, S. Lu, G. R. Holtom, C. He, J. C. Tsai, J. X. Kang and X. S. Xie, Science, 2008, 322, 1857-1861.
[24] D. C. Johnson, M. D. Nicholson and F. C. Haigh, IPC Technical Paper Series, 1975, 5, pp. 1- 30.
[25] R. K. Reddy and R. Bhargava, in Emerging Raman applications and techniques in biomedical and pharmaceutical fields," Eds. Matousek and Morris, (2010).
[26] J. Trevisan, P. P. Angelov, P. L. Carmichael, A. D. Scottc and F. L. Martin, Analyst, 2012, 137(14), 3202-3215.
[27] J. Shlens, Systems Neurobiology Laboratory, University of California at San Diego, 2005.
[28] P. Cunningham in Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval, ed. M. Cord and P. Cunningham, Springer, 2008, pp. 97-98.
[29] M. Pohar, M. Blas and S. Turk, Metodološki zvezki, 2004, 1, 143-161.
[30] K. M. G. Lima, K. Gajjar, G. Valasoulis, M. Nasioutziki, M. Kyrgiou, P. Karakitsos, E. Paraskevaidis, P. L. Martin Hirsch and F. L. Martin, Anal. Methods, 2014, 6, 9643-9652.
[31] A. Molckovsky, L.-M. Wong Kee Song, M. G. Shim, N. E. Marcon, B. C. Wilson, Gastrointestinal Endoscopy¸2003, 57, 396-402.
[32] J. L. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-García, R. Huerta-Franco, J. L. González-Trujillo, C. A. Ramírez-Alvarado, G. Gutiérrez-Juárez, C. Medina-Gutiérrez, Lasers in Medical Science, 2007, 22, 229-236.
[33] S. K. Teh, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, and Z. Huang, British Journal of Cancer, 2008, 98, 457-465.
[34] Christoph Krafft, Gerald Steiner, Claudia Beleites, Reiner Salzer, Journal of Biophotonics, 2009, 2, 13-28.
[35] P. Crow, A. Molckovsky, N. Stone, J. Uff, B. Wilson and L.-M. Wong Kee Song, Urology, 2005, 65, 1126-1130.
[36] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2000, URL http://www.R-project.org/.
[37] W. N. Venables and B. D. Ripley, Modern Applied Statistics with S. Fourth Edition. Springer, New York, 2002, ISBN 0-387-95457-0
[38] Z. Movasaghi, S. Rehman, I. U. Rehman, Applied Spectroscopy Reviews, 2007, 42, 493-541.
[39] Kurt W. Short, Susan Carpenter, James P. Freyer, Judith R. Mourant, Biophysical Journal, 2005, 88, 4274-4288.
[40] C.W. McCloskey, R.L. Goldberg, L.E. Carter, L.F. Gamwell, E.M. Al-Hujaily, O. Collins, E.A. Macdonald, K. Garson, M. Daneshmand, E. Carmona and B.C.Vanderhyden, Frontiers in Oncology, 2014, 4, DOI: 10.3389/fonc.2014.00053.
[41] Pei-Chang Wu, Ming-Hong Tai, Dan-Ning Hu, Chien-Hsiung Lai, Yi-Hao Chen, Yi-Chen Wu, Chia-Ling Tsai, Shyi-Jang Shin, Hsi-Kung Kuo, Journal of Ocular Pharmacology and Therapeutics, 2008, 24, 25-33.
[42] S. R. Cho, S. A. Ock, J. G. Yoo, B. Mohana kumar, S. Y. Choe, G. J. Rho, Reproduction in Domestic Animals, 2005, 40, 171-176.

[43] N. Stone, C. Kendall, J. Smith, P. Crow and H. Barr, Faraday Discussions, 2004, 126, 141-157.
[44] I. Delfino, G. Perna, M. Lasalvia, V. Capozzi, L. Manti, C. Camerlingo, M. Lepore, Journal of Biomedical Optics, 2015, 20, 35003.
[45] I. Notingher, S. Verrier, S. Haque, J. M. Polak and L. L. Hench, Biopolymers, 2003, 72, 230-240.
[46] I. Notingher, I. Bisson, A. E. Bishop, W. L. Randle, J. M. P. Polak, L. L. Hench, Analytical Chemistry, 2004, 76, 3185-3193.
[47] V. W. Daniëls, K. Smans, I. Royaux, M. Chypre, J. V. Swinnen, N. Zaidi, PLoS ONE, 2014, 9, e106913.
[48] A. Mukherjee, J. Wu, S. Barbour, X. Fang , Journal of Biological Chemistry, 2012, 287, 24990-25000.
[49] I. W. Schie, L. Alber, A. L. Gryshuk, J. W. Chan Analyst. 2014 Jun 7;139(11):2726-33.
[50] Y. Hong Ong, M. Lim, Q. Liu, Optics Express, 2012, 20, 22158-22171.
[51] M. Jermyn, K. Mok, J. Mercier, J. Desroches, J. Pichette, K. Saint-Arnaud, L. Bernstein, M. C. Guiot, K. Petrecca and F. Leblond, Sci Transl Med., 2015, 7, 274ra19.
[52] M. G. Shim and B. C. Wilson, J Raman Spectr 1997, 28, 131-142.
[53] A. Molckovsky, L. M. Song, M. G. Shim, N. E. Marcon and B. C. Wilson, Gastrointestinal Endoscopy, 2003, 57, 396-402.