Analyst

Cite this: DOI: 10.1039/c0an00910e

www.rsc.org/analyst

PAPER

Correlation of p16^{INK4A} expression and HPV copy number with cellular FTIR spectroscopic signatures of cervical cancer cells

Kamila M. Ostrowska,[a] Amaya Garcia,[a] Aidan D. Meade,[a] Alison Malkin,[b] Ifeoluwapo Okewumi,[b] John J. O'Leary,[c] Cara Martin,[c] Hugh J. Byrne[d] and Fiona M. Lyng[a]

Received 13th November 2010, Accepted 6th January 2011
DOI: 10.1039/c0an00910e

Cervical cancer, a potentially preventable disease, has its main aetiology in infection by high risk human papillomavirus (HR-HPV). Approaches to improving cervical cancer screening and diagnostic methodologies include molecular biological analysis, targeting of biomarker proteins, but also exploration and implementation of new techniques such as vibrational spectroscopy. This study correlates the biomarker protein p16^{INK4A} expression levels dependent on HPV copy number with the infrared absorption spectral signatures of the cervical cancer cell lines, HPV negative C33A, HPV-16 positive SiHa and CaSki and HPV-18 positive HeLa. Confocal fluorescence microscopy demonstrated that p16^{INK4A} is expressed in all investigated cell lines in both nuclear and cytoplasmic regions, although predominantly in the cytoplasm. Flow cytometry was used to quantify the p16^{INK4A} expression levels and demonstrated a correlation, albeit nonlinear, between the reported number of integrated HPV copies and p16^{INK4A} expression levels. CaSki cells were found to have the highest level of expression, HeLa intermediate levels, and SiHa and C33A the lowest levels. FTIR spectra revealed differences in nucleic acid, lipid and protein signatures between the cell lines with varying HPV copy number. Peak intensities exhibited increasing tendency in nucleic acid levels and decreasing tendency in lipid levels with increasing HPV copy number, and although they were found to be nonlinearly correlated with the HPV copy number, their dependence on p16^{INK4A} levels was found to be close to linear. Principal Component Analysis (PCA) of the infrared absorption spectra revealed differences between nuclear and cytoplasmic spectroscopic signatures for all cell lines, and furthermore clearly differentiated the groups of spectra representing each cell line. Finally, Partial Least Squares (PLS) analysis was employed to construct a model which can predict the p16^{INK4A} expression level based on a spectral fingerprint of a cell line, demonstrating the diagnostic potential of spectroscopic techniques.

Introduction

Cervical cancer is a potentially preventable disease and it remains the second most common malignancy in women worldwide.[1] The incidence rates range from less than 15 per 100 000 in Europe up to 33.5 per 100 000 in Latin America.[2] Many aetiological factors are associated with cervical cancer, such as diet, cigarette smoking, multiple sexual partners, multiple pregnancies, contraceptive pills, sexually transmitted diseases (Chlamydia, HIV) or aging. However, the most important factor identified is infection with high risk human papillomavirus (HR-HPV).[3] There are 15 identified HR-HPV types[4] and 70% of all cervical cancers are associated with either HPV-16 or HPV-18.

Differentially expressed proteins in cancer have potential utility as biomarkers. As the cell cycle is often disrupted in a cancerous cell, proteins associated with it are often candidate biomarkers. Putative biomarkers of cervical cancer that are currently under study include proteins such as CDC 6 (DNA licensing protein), minichromosome maintenance proteins (MCM 2, MCM 5), p53 or p16^{INK4A}.[5] These biomarkers have been used to detect the presence of abnormal cells, based upon immunocytochemical methods.

p16^{INK4A} regulates the levels of active cyclin D/CDK in the cell, providing a feedback mechanism that regulates the levels of MCM (minichromosome maintenance proteins), PCNA (proliferating-cell nuclear antigen) and cyclin E. Overexpression of p16^{INK4A}, which is considered a marker of elevated E7 expression, can be detected in some CIN1 (Cervical Intra-epithelial Neoplasia grade 1) lesions, as well as in CIN2 and CIN3 lesions that show evidence of integration.[6,7] Integration of

[a]*RESC, Focas Research Institute, Dublin Institute of Technology, Kevin Street, Dublin, 8, Ireland. E-mail: kamila.m.ostrowska@gmail.com*
[b]*School of Biological Sciences, Dublin Institute of Technology, Kevin Street, Dublin, 8, Ireland*
[c]*Department of Pathology, Coombe Women & Infants University Hospital, Dolphin's Barn Street, Dublin, 8, Ireland*
[d]*Focas Research Institute, Dublin Institute of Technology, Kevin Street, Dublin, 8, Ireland*

the HPV genome into the host cell chromosome is a critical event in the development of most cervical cancers.[8] Overexpression of $p16^{INK4A}$ has been demonstrated in cervical cancers as a result of functional inactivation of pRb by the HPV E7 protein.[9] This overexpression highlights the potential of $p16^{INK4A}$ as a marker for CIN and cervical cancer. HPV positivity and $p16^{INK4A}$ positivity have shown a correlation, even though $p16^{INK4A}$ expression was also seen in a limited number of HPV negative biopsy samples.[10] Also, a correlation between $p16^{INK4A}$ expression and cervical lesion grade and HR-HPV positivity has been documented.[11] $p16^{INK4A}$ has been proven to be the most reliable marker of cervical dysplasia and was found to mark all grades of squamous and glandular lesions of the cervix. The use of $p16^{INK4A}$ immunocytochemical analysis as a complement to conventional screening programmes could potentially aid in the reduction of false positive and false negative results.[12]

In the first part of this study, expression of $p16^{INK4A}$ was analysed in cervical cancer cell lines using immunocytochemical staining and both confocal fluorescence microscopy and flow cytometry. Confocal fluorescence microscopy is an optical imaging technique that allows for high-resolution structural imaging at the cellular level at varying depths in the sample while flow cytometry is a well established technique for examining and sorting fluorochrome-labeled cells, simultaneously providing assessment of a multitude of characteristics of individual cells.

Four cervical cell lines were used in this study: HPV negative C33A, HPV-18 positive HeLa containing 20–50 integrated HPV copies per cell, HPV-16 positive CaSki containing 60–600 integrated HPV copies per cell and SiHa with 1–2 integrated HPV-16 copies per cell.

The second part of the study utilised Fourier Transformed Infrared (FTIR) Spectroscopy to spectroscopically fingerprint and differentiate the cervical cancer cell lines. FTIR, along with Raman spectroscopy, is a vibrational spectroscopic technique which provides unique information about the chemical composition of a sample. A wide range of biological applications, particularly cancer detection, has been reported to date.[13–19] The development of applications of vibrational spectroscopy to medical diagnostics has recently been reviewed by Diem *et al.*[20] Significant activity in the area of the potential application of FTIR spectroscopy to cervical cancer detection has been reported over the last decade.[17,21–43] Understanding and processing the biochemical information delivered by vibrational spectroscopic techniques and its correlation with the existing screening and diagnostic methods is key to further development of realistic applications of Raman and FTIR spectroscopy to cervical cancer detection.

It has recently been demonstrated that both Raman and FTIR absorption spectroscopy can effectively discriminate between cell lines with varying HPV infection levels.[44] The current study explores the potential correlation of spectroscopic features with HPV infection levels characteristic of each cell line as well as the resultant expression levels of $p16^{INK4A}$ biomarker.

Materials and methods

Cell culture and immunocytochemistry

Cell lines were obtained from the ATCC cell culture collection. All cell lines were grown in RPMI 1640 medium (Sigma-Aldrich, Ireland) supplemented with 10% Foetal Bovine Serum (FBS, BioWhittaker®, Lonza, Ireland), 1% penicillin–streptomycin (BioWhittaker®, Lonza, Ireland), 1% L-glutamine (Sigma-Aldrich, Ireland), and 0.2% hydrocortisone (Sigma-Aldrich, Ireland). Cells were incubated at 37 °C in 5% CO_2 and maintained up to 70–80% confluency.

Immunocytochemical staining for $p16^{INK4A}$ was performed on each of the cell lines and appropriate negative controls. Cells were fixed using 4% paraformaldehyde for 30 min, washed 3 times with phosphate buffered saline (PBS) and permeabilised with cold methanol for 10 min. Cells were blocked by incubating with a blocking buffer (1% bovine serum albumin (BSA) in PBS) for 30 min at room temperature with agitation. After decanting the serum, cells were incubated for one hour at 37 °C with the primary antibody, purified mouse anti-human $p16^{INK4A}$ (BD Pharmingen™, Ireland) in dilution buffer (1% BSA in PBS). Following incubation, cells were washed 5 times with PBS and were incubated with the secondary antibody, goat polyclonal to mouse Immunoglobulin G (IgG) (Abcam, Ireland) labeled with fluorescein isothiocyanate (FITC) in dilution buffer. This was carried out in darkness for 45 min at 37 °C. The cells were additionally incubated with propidium iodide (PI) staining solution (dilution 1 : 250) (BD Pharmingen™, Ireland) for 15 min and washed 3 times with PBS. A negative control was also prepared for each cultured cell line according to the same protocol but with no primary antibody (NC1) and no primary or secondary antibody (NC2).

Sample preparation

Confocal microscopy. For confocal microscopy, cells were grown on glass coverslips in 6 well plates and stained as described above. Each coverslip with the stained cells was removed from the multi-well plate and mounted upside down on a pre-labelled glass slide using Sigma diagnostic mounting medium (Sigma-Aldrich, Ireland). Samples were left to dry in a dark room and thereafter stored at 4 °C until they were examined under the confocal microscope.

Flow cytometry. Flow cytometry samples were stained as a pellet in Eppendorf tubes according to the staining protocol described above. Pellets were examined directly after the staining procedure.

FTIR spectroscopy. Cervical cell lines for FTIR study were grown for 24 hours on CaF_2 windows. Cells prepared for FTIR study were not fixed to avoid contribution of the fixative agent to the cellular spectra.[45] Before FTIR measurements, cells were washed twice in PBS and dried to minimise contributions from water.

Measurements and data handling

Confocal fluorescence microscopy. An inverted confocal laser scanning fluorescence microscope (LSM 510 META, Carl Zeiss) was used to record fluorescence images of the cells. The argon ion laser was chosen to excite the FITC fluorochromes at 488nm and the HeNe laser was used to excite the PI fluorochromes at 543 nm. A 63× oil immersion objective lens was used for

recording. Five to six images were collected for each sample from different regions of the slides.

Flow cytometry. A CyFlow® Space Flow cytometer System (Partec) was used for the cervical cell analysis. Measurements were performed for each cell line, represented by 2–4 samples, prepared as stained cellular pellets. Fluorescence intensity of the staining was calculated based on analysis of approx. 2×10^4 cells for each sample. Analysis of the dataset was performed using FloMax 3.0 software. Gating R1 selected a well defined population of cells excluding any possible debris present in the samples. Gating R2 was performed to exclude doublets and triplets of cells analysed by the instrument. The subsequent step included applying gating G1 (logical gate of R1 and R2) to the investigated fluorescence signal and choosing the range of interest within which fluorescence intensity calculations were performed.

Infrared spectroscopy. Fourier Transformed Infrared (FTIR) spectroscopic measurements were performed using a Perkin Elmer Spotlight 400 microscope system in transmission mode. The time frame between washing the FTIR samples with PBS and start of acquiring spectra was 30–40 minutes. Spectra were collected as an average of 64 scans with a resolution of 2 cm^{-1} using Image Mode. FTIR signals were accumulated over a spectral range of 750 to 4000 cm^{-1}. Signals were accumulated with the maximal imaging resolution of the system of 6.25 μm per pixel. The sampling included investigation of two cell culture passages prepared on separate slides. Four to five cells per slide were analysed.

Spectroscopic data analysis was carried out in Matlab, version R2006 (Mathworks, CA, USA) according to protocols developed and routinely used in DIT. Processing of the FTIR spectra included an EMSC resonant Mie scattering correction according to the algorithm developed by Bassan *et al.*[46] with the Matrigel spectrum supplied with the algorithm used as reference. In the original work, it was demonstrated using simulated spectra that the algorithm, which iteratively evolves the reference spectrum to fit and correct for both the resonant and nonresonant scattering contributions, reproduces the ideal spectra even using a "non-ideal" reference and that for example the true position of the biologically significant amide I band can be obtained with the RMieS-EMSC algorithm.[46] After pre-processing the spectra, Principal Component Analysis (PCA) and Partial Least Squares (PLS) analysis were performed on the dataset. It should be noted that although the spectra were recorded over the range 750–4000 cm^{-1}, the Mie scattering correction was only executable over the maximum range of the reference spectrum provided with the algorithm, namely 1000–4000 cm^{-1}, and thus only this region is described.

Results and discussion

Confocal microscopy

Fluorescence images of cells were collected using the confocal fluorescence microscopy system. Representative images of cells are presented in Fig. 1, where A, C, E and G show only green fluorescence staining (FITC), while B, D, F and H show the overlay of green (FITC) and red (PI) fluorescence signals. The FITC fluorescence intensity is associated with the $p16^{INK4A}$ expression level, while the PI dye was used to visualise the position of nuclei within the cells. The fluorescence intensities for the defined FITC fluorescence emission wavelengths for the samples stained with primary and secondary antibody ($p16^{INK4A}$ staining) and NC1 and NC2 samples were compared. The results revealed a strong emission peak for samples stained with primary and secondary antibody and negligible emission for the negative control samples (NC1 and NC2) (data not shown). Based on these observations, immunostaining was found to be specific to $p16^{INK4A}$. Fig. 1 shows that purified mouse monoclonal anti-human $p16^{INK4A}$ stains both cytoplasmic and nuclear regions of the cells, although staining is predominantly in the cytoplasm. The presence of $p16^{INK4A}$ in the cytoplasmic region may be a result of post-transcriptional modification or overexpression of the protein, which forces its transfer into the cytoplasm.[12] These findings support previous studies confirming the theory that $p16^{INK4A}$ is overexpressed in dysplastic cells of the cervix.[47] The observations are in very good agreement with the work of Klaes *et al.*, who demonstrated that the $p16^{INK4A}$ specific immunohistochemical staining allows for sensitive and specific identification of dysplastic cervical cells in cervical cancer tissue sections and cervical cancer cell lines.[48] The results also support the findings of Murphy *et al.* that $p16^{INK4A}$ marks dysplastic squamous and glandular cells of the cervix with a sensitivity of 99.9% and specificity of 100%.[49] The negative HPV cell line C33A was $p16^{INK4A}$ positive (Fig. 1A and B), but exhibited the lowest staining intensity of the investigated cell lines. This finding confirms the previous study of Milde-Langosch *et al.* in which $p16^{INK4A}$ overexpression was reported in 41% of HPV negative adenocarcinomas and it was suggested that a non-HPV E7 mediated mechanism of $p16^{INK4A}$ up regulation may also exist.[50]

Quantitative analysis of the fluorescence intensity of FITC visible in the confocal images, albeit on a limited number of individual cells, indicated that there is a correlation between the number of integrated HPV copies per cell and $p16^{INK4A}$ expression characteristic to each cell line. CaSki cells were found to have the highest intensity of staining, with decreasing tendency for HeLa and SiHa and C33A expressing the lowest intensity (data not shown).

Flow cytometry

To more precisely quantify expression of $p16^{INK4A}$ protein in cervical cancer cell lines, flow cytometry measurements were performed to enable a larger statistical basis. For all the investigated cell lines, the populations were well defined, allowing for a precise investigation of the cells of interest.

Mean fluorescence intensity levels related to $p16^{INK4A}$ expression are presented in Table 1 together with their standard deviations. Values of fluorescence intensity for negative control (NC1) samples prepared for the investigated cell lines according to the immunostaining protocol but with no primary antibody are also presented in Table 1. Values of fluorescence intensities for NC1 were lower than the standard deviations for fluorescence intensities related to $p16^{INK4A}$ expression, confirming that the flow cytometry analysis was specific to $p16^{INK4A}$. A degree of correlation between the number of HPV copies per cell line and

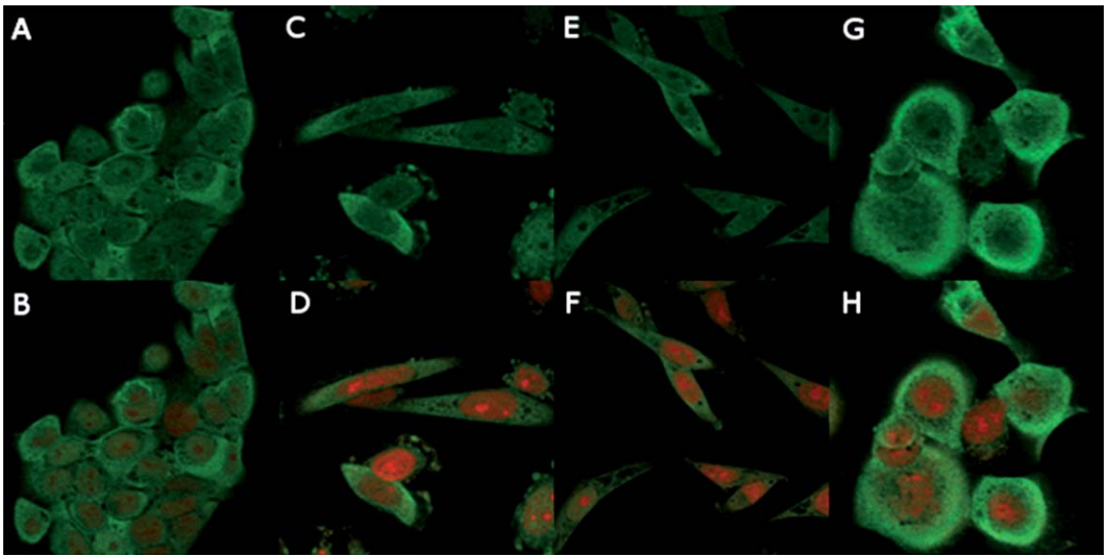

Fig. 1 Confocal microscopy images of cervical cell lines: (A–B) C33A cells, (C–D) SiHa cells, (E–F) HeLa cells, (G–H) CaSki cells. Images (A), (C), (E) and (G) show FITC fluorescence staining (related to $p16^{INK4A}$ expression), while images (B), (D), (F) and (H) present FITC and PI staining.

the fluorescence intensity related to $p16^{INK4A}$ expression was observed for the flow cytometry data (Fig. 2). The highest level of $p16^{INK4A}$ expression (intensity of staining) was observed for CaSki cells, followed by HeLa, SiHa and the lowest for C33A. The relationship between $p16^{INK4A}$ expression and HPV copy number illustrated by Fig. 2 is supported by a study conducted by Klaes *et al.*, wherein a correlation between increasing grade of cervical lesion and staining intensity of $p16^{INK4A}$ was observed.[48] Similarly, Agoff *et al.* showed that $p16^{INK4A}$ expression correlates with an increasing severity of cervical disease.[51] Murphy *et al.* showed a strong correlation between HR-HPV positivity and $p16^{INK4A}$ staining pattern.[12,52,53] In the study conducted by Wang *et al.* a correlation between $p16^{INK4A}$ immunostaining and cervical disease severity stratified by HPV status was observed.[54] The results in the present study demonstrate a correlation between $p16^{INK4A}$ immunostaining and the presence of HR-HPV in cervical cancer cell lines supporting the previous studies and implies that $p16^{INK4A}$ is a highly sensitive marker of HR-HPV in cervical cancer cell lines. The hyperbolic-like relationship between the levels of $p16^{INK4A}$ and HPV copy number is typical of the response of a cell to the action of an agonist[56] and in Fig. 2 is approximated by a fit with the following function $y = a + b\sqrt{x} + cx$, where $a = 62.4$, $b = 3.5$ and $c = -0.1$. It should be noted that, in Fig. 2, the HPV copy number is represented by the average of the range quoted in literature[57,58] and so error margins in the horizontal axis are potentially very large. However, the sublinear nature of the plot indicates that $p16^{INK4A}$ expression levels are particularly sensitive for low HPV copy number.

FTIR spectroscopy

FTIR maps were recorded for C33A, SiHa, HeLa and CaSki cells (2–4 maps for each cell line) using the Perkin Elmer Spotlight 400 system. Representative images are presented in Fig. 3A–D. Representative unprocessed FTIR maps, generated by the Spectrum Image software, showing spectrally averaged absorbances, recorded from the regions of interest within the investigated samples are shown in Fig. 3E–H, for C33A, SiHa, HeLa and CaSki, respectively. Nuclear and cytoplasmic regions are well defined, due to their varying density and thickness, in the FTIR maps. Spectra representing nuclear and cytoplasmic regions of the cells were extracted from the FTIR maps using the Perkin Elmer Spectrum IMAGE software and delivered mean spectra of cellular structures characteristic of the cell lines (Fig. 4). Spectra extracted from the maps were chosen from the middle of the nucleus or cytoplasm. As a result, from 60–80 spectra recorded for one cell, only a few of them were used in the analysis. This procedure ensured that only nuclear or cytoplasmic signals were compared rather than a mix of the cellular signatures. Each spectrum from Fig. 4 was calculated as an

Table 1 Quantitative analysis of fluorescence intensity related to $p16^{INK4A}$ expression level and negative control samples with no secondary antibody for cervical cancer cell lines obtained from flow cytometry studies

		$p16^{INK4A}$ expression		Negative controls (NC1)	
Cell line	Reported number of HPV copies per cell	Fluorescence intensity (a.u.)	Standard deviation for fluorescence intensity (a.u.)	Fluorescence intensity (a.u.)	Standard deviation for fluorescence intensity (a.u.)
C33A	0	61.92	1.42	1.32	0.06
SiHa	1–2	68.14	1.23	1.18	0.08
HeLa	20–50	80.10	1.92	1.17	0.01
CaSki	60–600	94.31	1.77	1.41	0.11

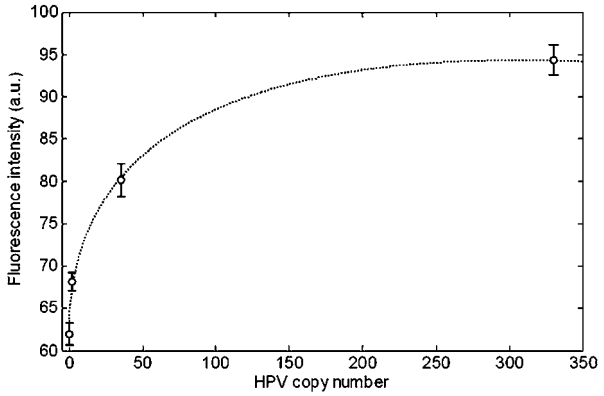

Fig. 2 Fluorescence intensity related to the $p16^{INK4A}$ expression level in cervical cancer cell lines plotted against the average HPV copy number present in a cell with the fitted function. The dotted line is a fit of the data with the equation $y = a + b\sqrt{x} + cx$, where $a = 62.4$, $b = 3.5$ and $c = -0.1$.

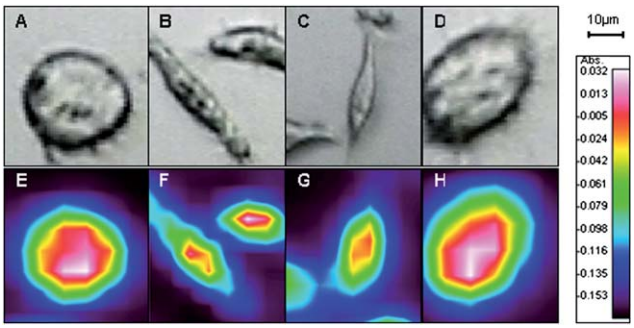

Fig. 3 Visual images (A–D) and average absorbance FTIR maps (E–H) representing cells present in the investigated samples. (A and E) C33A cell, (B and F) SiHa cells, (C and G) HeLa cells, (D and H) CaSki cells. Red colour represents region of a nucleus, green and light blue—region of a cytoplasm, violet—background.

average of 30–40 signals extracted from the FTIR maps characteristic of the defined cellular structure and was subjected to pre-processing using the RMieS-EMSC algorithm. Assignments of the main FTIR bands (numbered in Fig. 4) are presented in Table 2. Three main biochemical cellular components feature strongly: proteins, lipids and nucleic acids. For all cell lines, the nuclear and cytoplasmic regions exhibited similar FTIR bands

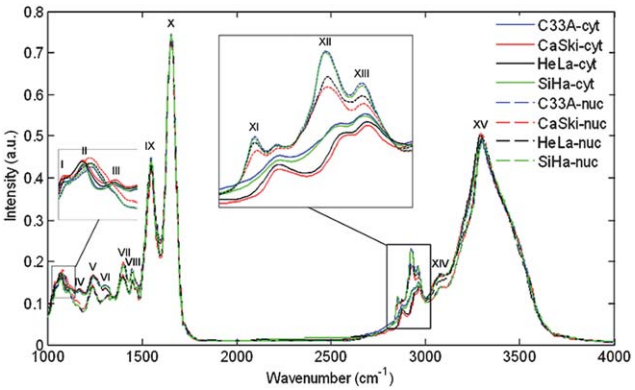

Fig. 4 Mean FTIR spectra of cervical cancer cell lines representing regions of cells (nucleus and cytoplasm).

related to the vibrations of these constituent components. There were, however, differences in the band intensities. A distinctive reduction of lipids was noticed in the IR signal of the cytoplasm compared to the signal of the nucleus, manifest in the CH_2 and CH_3 groups stretching bands in the range of 2850–2970 cm^{-1}, due to the double lipid bilayer nuclear envelope, compared to the phospholipid bilayer membrane of the cytoplasm. Furthermore, the peak at 1120 cm^{-1} is present only in the cytoplasm. This feature is related to vibrations of RNA components, which are mainly present in the cytoplasmic organelles. Three peaks related to cytoplasmic constituent vibrations were also noticed to be shifted to higher wavenumbers (1082 cm^{-1}, 1316 cm^{-1}, 3300 cm^{-1}) compared to nuclear bands related to the same types of vibration (1070 cm^{-1}, 1310 cm^{-1}, 3292 cm^{-1}). A detailed examination of the relative intensities of peaks related to vibrations of the cellular components (nucleus and cytoplasm) of the C33A, SiHa, HeLa and CaSki infrared spectra revealed the following tendencies:

- Increase in nucleic acid levels with increasing number of HR-HPV copies in the cell or increasing $p16^{INK4A}$ expression (Fig. 5A and D).
- Decrease in lipid levels with increasing number of HR-HPV copies in the cell or increasing $p16^{INK4A}$ expression (Fig. 5B and E).
- No tendency for changes in protein levels related to HR-HPV copy number or increasing $p16^{INK4A}$ expression (Fig. 5C and F).

As the RMieS-EMSC correction uses the protein based Matrigel as a reference and effectively normalises the protein levels, the third observation is not surprising, and variations in nucleic acids and lipid related features should be considered relative to the protein levels.

The observed increase in nucleic acid levels may be related to an increased number of chromosomes present in HPV infected nuclei. It was reported by Mehés *et al.* that HPV presence facilitates polyploidisation (increase in chromosome number in a cell nucleus) in cervical squamous cells.[59] Additionally, it is known that binding of HPV DNA to host DNA disrupts the normal function of the cellular proteins and as a consequence, the host cell accumulates more and more damaged DNA that cannot be repaired.[60]

CaSki cells are known to be the most malignant and in the spectra of the cytoplasm and nucleus representing this cell line the highest level of nucleic acids was observed. A similar increase of the nucleic acid related peaks in cancerous cervical cells was noticed in previous studies[33,42] and is confirmed by these observations. Changes in lipid levels are possibly associated with the disruption of the membrane functionality caused by the virus which influences lipid rafts.[61] A similar behaviour was noted and reported in our previous study on cervical cancer cell lines, wherein both Raman and FTIR (point mode) techniques were utilised.[62] Moreover, it was reported previously by Wong *et al.* that the degree of disorder of methylene chains in membrane lipids increases in cervical neoplastic cells.[63] The absence of any clear relationship between protein levels and HPV copy number is expected, as the presence of HR-HPV in the cell results in functional over- and under-expression of only select biomarker proteins,[5] while the level of the other proteins would not be affected. The variation of the spectroscopic features associated

Table 2 Assignments[55] of FTIR bands of cellular spectra recorded from cervical cancer cells. Wavenumbers in brackets are attributed to peaks characteristic of cytoplasmic spectra that were noticed to be shifted with respect to peaks corresponding to nuclear constituents

Peak no.	Wavenumber (cm^{-1})	Assignment	Proteins	Lipids	Nucleic acids
I	1036	C–C skeletal stretching	+		
II	1070 (1082)	Symmetric PO_2 str. mode			DNA, RNA
III	1120	C–O stretching			RNA
IV	1170	CO–OC asymmetric stretching	+		
V	1240	Amide III, asymmetric PO_2 str. mode	+		DNA, RNA
VI	1310 (1316)	CH_2 stretching		+	
VII	1400	CH_3 symmetric deformations	+		
VIII	1450	CH_3 asymmetric deformations	+	+	
IX	1550	C–N stretching and CHN bending—amide II	+		
X	1650	C═O stretching—amide I	+		
XI	2854	CH_2 symmetric stretching		+	
XII	2875	CH_3 symmetric stretching		+	
XIII	2926	CH_2 asymmetric stretching		+	
XIV	2962	CH_3 asymmetric stretching		+	
XV	3292 (3300)	N–H stretching (amide A)	+		DNA, RNA

with the nuclear lipids, proteins and nucleic acids as a function of the HPV copy number was found to be sublinear and to be well fitted with a similar function to that employed to describe the variation of the $p16^{INK4A}$ levels with HPV copy number in Fig. 2. Notably, the dependences of the cytoplasmic lipids, proteins and nucleic acids on the $p16^{INK4A}$ levels, as identified by fluorescence were found to be linear.

Principal component analysis

Fig. 6 shows the results of Principal Component Analysis (PCA) of the spectral data for the four cell lines. The analysis clearly differentiates between nuclear and cytoplasmic FTIR signals for each cell line. The percentage variance explained by the first three principal components was 88.34%, 4.44% and 2.23% for PC1, PC2 and PC3, respectively. The PCA scatterplot (Fig. 6A) reveals that the cytoplasm spectra of C33A and SiHa are close to each other, suggesting similarities in cytoplasmic composition of these cell lines. However, in terms of the nuclear data, the cell lines are very well separated. C33A are HPV negative, while SiHa contain a very low number of integrated HPV copies per cell (1–2). HPV DNA and host cell DNA integration occurs within the nucleus of a cell. Thus, the differences between C33A (HPV negative—no integration) and SiHa (HPV positive—integration) at the nuclear level are understandable. It is possible that in the cytoplasm, biochemical changes in the cell affected by a very low number of HPV copies are not as prominent as in the nucleus. In the case of cells infected by 20–50 HPV copies (HeLa) and 60–600 copies (CaSki), biochemical changes within the cytoplasm become more prominent resulting in, for example, a stronger under- and overexpression of cytoplasmic proteins. Thus the cytoplasmic signals of C33A and SiHa are distributed close to each other, while HeLa and CaSki are very well separated. In general, the clusters representing the spectra of cytoplasm are more dispersed than those corresponding to cell nuclei. This could be explained by the fact that the cytoplasm is much more variable in its composition compared to the nucleus, therefore exhibiting more variation in the FTIR signals. Although, the Resonant Mie scattering correction algorithm cannot be considered perfect as evidenced by the incomplete baseline correction in the loadings plots of the PCs, (Fig. 6B), it is considered the most effective method currently available allowing for removing the Mie scattering influence from the FTIR signals. The separation between the nuclear and the cytoplasmic clusters is clearly due to the first principal component (PC1). Based on the PC1 loading (Fig. 6B), it can be concluded that the biochemical variability between these two cellular components primarily arises from the lipid levels (region of 2800–3000 cm^{-1}). In PC1, protein contributions are also notable (1400–1700 cm^{-1}, 3000–3500 cm^{-1}). PC1 also shows nucleic acid variations (1050–1200 cm^{-1}), the symmetric phosphodiester group stretching band in RNA (peak at 1120 cm^{-1}) being strongly accented in cytoplasm spectra. This observation is consistent with the mean spectra analysis, whereby the peak at 1120 cm^{-1} was seen to be prominent in the spectra of the cytoplasm of all cell lines and absent in the signals from nuclei (Fig. 4—peak III in enlarged nucleic acid region). PC2 indicates a high variability of the proteins (amide I, amide II, amide III, amide A). This separation of the cell lines confirms the study of Kelly *et al.*, where amide bands were also determined to be a source of differentiation of the cervical cancer cells.[27] In PC2 there is also a lipid contribution present originating from symmetric and asymmetric vibrations of the CH_2 and CH_3 groups. A two dimensional representation of the scatterplot PC1 *vs.* PC2 (Fig. 6A) shows that the internal variability of the cytoplasm signals is extended along PC2. Therefore, it can be concluded that it originates mainly from variations in the protein levels within the cell. Groups of points in the PCA scatterplot representing the spectroscopic signatures of C33A, SiHa, HeLa and CaSki cell lines recorded from nuclei are separated along the third principal component. In PC3 contributions of all components are present; nucleic acids, lipids and proteins. This illustrates that the HR-HPV infection influences the entire biochemistry of the cell through a chain of reactions.[64]

Partial least squares analysis

In order to further elucidate the multivariate spectral signatures which are specifically related to HPV infection, Partial Least Squares (PLS) analysis was performed. Fig. 5 demonstrates that $p16^{INK4A}$ levels as measured by fluorescence intensity are

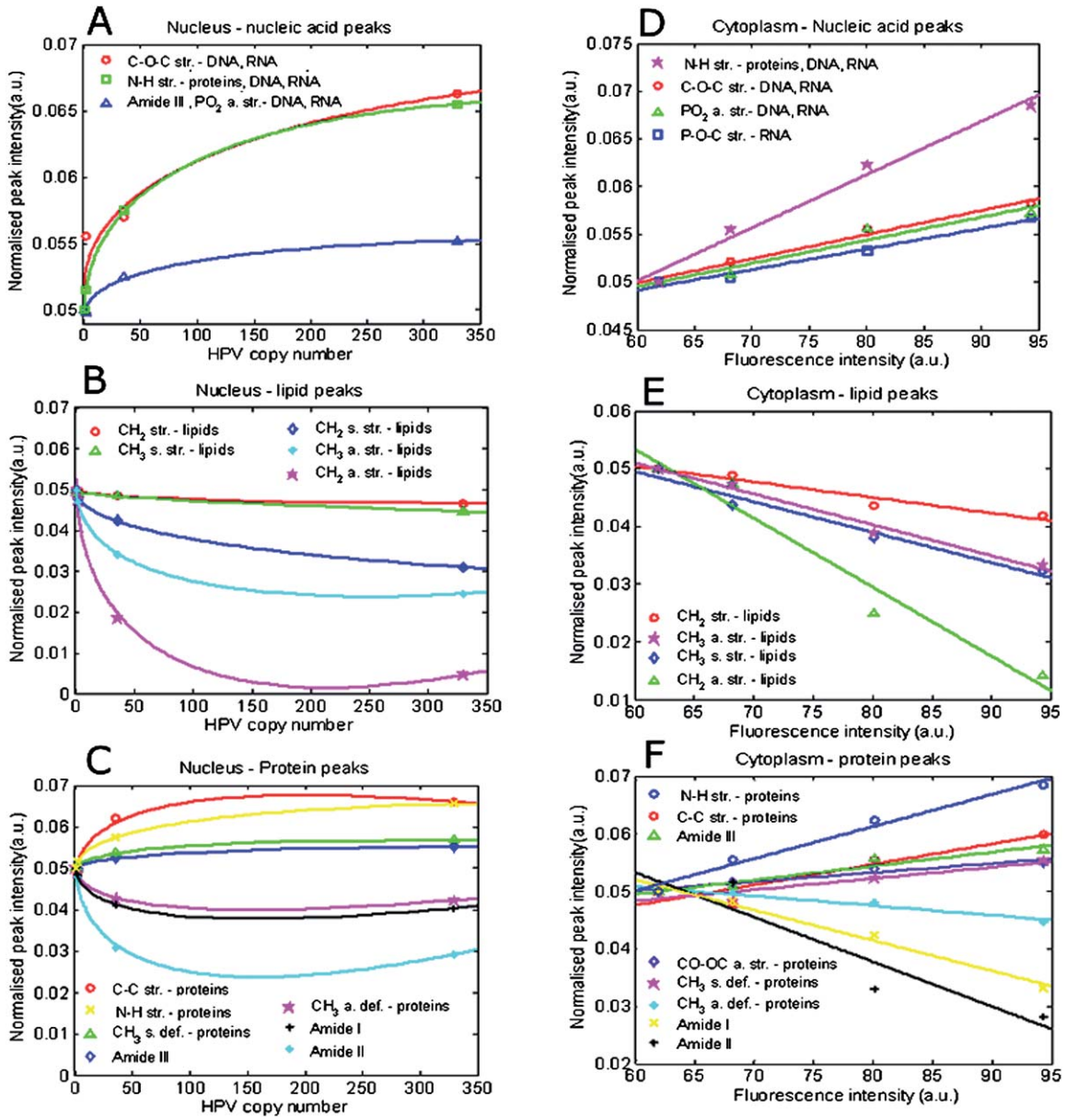

Fig. 5 Peak intensity analysis for FTIR spectra of nuclear and cytoplasmic regions of cervical cancer cells. Dependence of peak intensities *vs.* HPV copy number was fit with $y = a + b\sqrt{x} + cx$ function, while peak intensities *vs.* fluorescence intensities ($p16^{INK4A}$ expression level) was fit with a linear function, $y = a + bx$.

approximately linearly correlated with the univariate spectral features and thus, as PLS is a linear model, these values are employed as targets for the PLS analysis. The model can therefore be applied to the FTIR data to elucidate multivariate signatures which are correlated to $p16^{INK4A}$ expression level, and therefore, in accordance with Fig. 2, to HPV infection levels. Once established, these variation patterns can then be applied to unknown samples to screen for the biomarker levels. As $p16^{INK4A}$ was found to be predominantly expressed in cytoplasmic regions of the cells, signals recorded from the cytoplasm were utilised in the analysis. Calibration (based on 40% of the data) and test set performance (based on 60% of the data) are presented in Fig. 7A and demonstrate a very good fit to the model. As shown in Fig. 7B, the PLS loading exhibits variations originating from proteins, primarily as sub-bands of amide A (3000–3700 cm^{-1}) and amide III (1230–1250 cm^{-1}), and are associated with the chain of biochemical disruption in protein regulation caused by HPV presence.[64] Another prominent spectral feature differentiating the cell lines, and present in PC loadings and the PLS loading, is the lipid contribution. Again, the presence of HR-HPV is seen to significantly influence the lipid balance within the cell,[61] and thus it can be expected that the vibrations of lipid component groups differentiate cells with various HPV copy number. This differentiation was distinctly manifest in the mean spectra of the cytoplasm and nucleus (Fig. 4—enlarged lipid region of 2800–3050 cm^{-1}). Furthermore, the DNA and RNA related peaks are exhibited in the range of 1050–1250 cm^{-1} (sym. and asym. PO_2 str. CO str.) and 3200–3400 cm^{-1}. Such activity in the region of the nucleic acid features is consistent with the profile of PC1 and PC3 (Fig. 6B), which differentiates the cell line

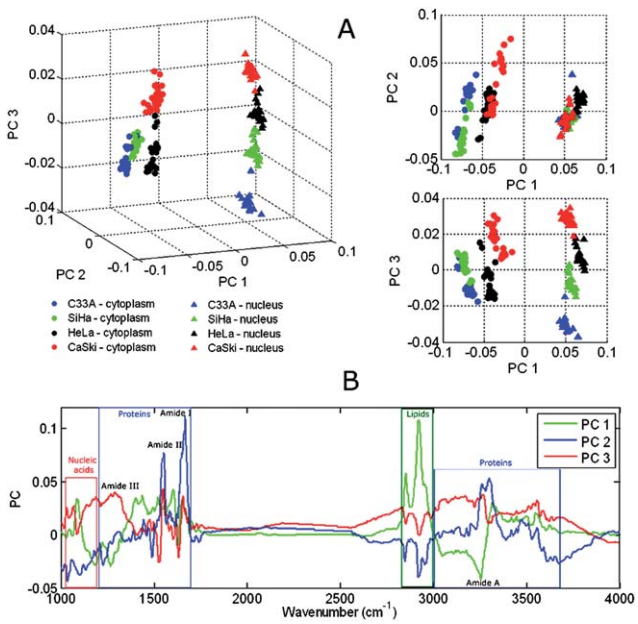

Fig. 6 Principal component analysis results. (A) Scatterplots presented in three and two dimensional form, (B) principal component loadings with marked regions of the main biocomponent contribution.

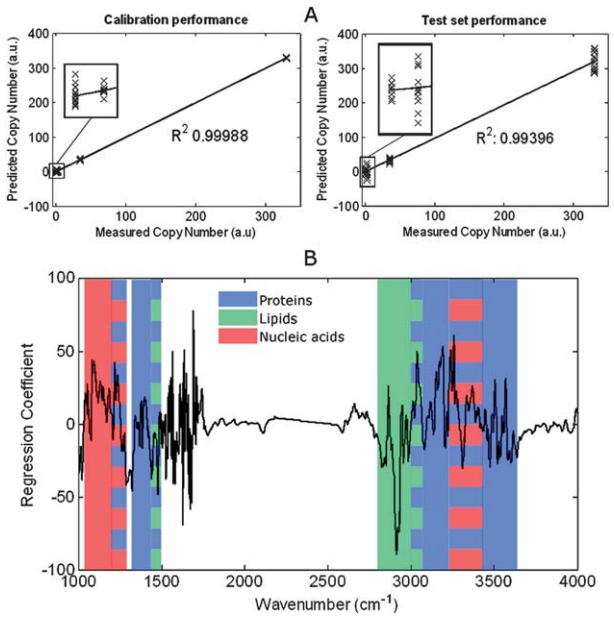

Fig. 7 Partial Least Squares analysis results targeting the p16^{INK4A} expression level prediction based on the FTIR spectral features.

signals, and as a consequence distinguishes cells as a function of HPV copy number.

Conclusions

Confocal fluorescence microscopy and flow cytometry revealed that p16^{INK4A} expression levels were correlated with HPV infection levels in cervical cancer cell lines. This confirms p16^{INK4A} as a potential diagnostic marker of cervical cancer, supporting the results of many other studies.[47,48,65] Integration of HR-HPV sequences into the cell genome is considered to be an important event in the progression of cervical neoplasia. In this study, a correlation between the integrated HPV copy number and intensity of p16^{INK4A} was observed. FTIR imaging allowed for identification of the specific spectral features of nuclear and cytoplasmic regions of the cervical cancer cells, indicating the differences between the two, in terms of lipid and RNA levels. PCA analysis revealed that the spectral analysis clearly separates the nuclear and cytoplasmic characteristic signals, and additionally distinguishes the cell lines. At low copy number, HPV predominantly influences the nucleus of the cell, whereas increased copy number impacts on the cytoplasmic signatures. A model of p16^{INK4A} expression level prediction based on FTIR spectroscopy was constructed utilising PLS analysis. Modelling was predominantly based on spectral features arising from nucleic acid, lipid and protein contributing features, which are influenced by the HPV interaction with the host cell. Notably, this work proves that the wavenumbers above 2700 cm^{-1} may exhibit crucial information. This region of the FTIR spectrum has been neglected in many studies conducted on biological samples.

A potential application of vibrational spectroscopy to cervical cancer screening and diagnosis requires a full understanding of the spectral information and its correlation with existing screening and diagnostic methods such as Pap testing or colposcopy. However, the field of the cervical cancer recognition and detection is still developing with new biomarkers[5] being identified that may be considered as adjuncts to existing cervical cytology and pathology methods. Biomarkers such as p16^{INK4A} have been found to be useful for low-grade lesions[66–69] making them particularly attractive. Thus, a correlation and explanation of the spectral information with these new biomarkers is a new challenge for biospectroscopists working on future cancer diagnostic systems.

Acknowledgements

This work was funded by the Department of Education and Science Technological Sector Research Strand III programme and enabled through the National Biophotonics and Imaging Platform, Ireland, and the Integrated NanoScience Platform for Ireland, both funded by the Irish Government's Programme for Research in Third Level Institutions, Cycle 4, National Development Plan 2007-2013.

References

1 D. Parkin, F. Bray, J. Ferlay and P. Pisani, *CA Cancer J. Clin.*, 2002, **55**, 74–108.
2 D. Parkin and F. Bray, *Vaccine*, 2006, **24**(Suppl. 3), S3/11–25.
3 J. Walboomers, M. Jacobs, M. Manos, F. Bosch, J. Kummer, K. Shah, P. Snijders, J. Peto, C. Meijer and N. Muñoz, *J. Pathol.*, 1999, **189**, 12–19.
4 N. Munoz, F. X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K. V. Shah, P. J. Snijders and C. J. Meijer, *N. Engl. J. Med.*, 2003, **348**, 518–527.
5 C. M. Martin, L. Kehoe, C. O. Spillane and J. J. O'Leary, *Mol. Diagn. Ther.*, 2007, **11**, 277–290.
6 A. Kalof, M. Evans, L. Simmons-Arnold, B. Beatty and K. Cooper, *Am. J. Surg. Pathol.*, 2005, **29**, 674–679.
7 M. Dray, P. Russell, C. Dalrymple, N. Wallman, G. Angus, A. Leong, J. Carter and B. Cheerala, *Pathology*, 2005, **37**, 112–124.
8 T. Yu, M. J. Ferber, T. H. Cheung, T. K. Chung, Y. F. Wong and D. I. Smith, *Cancer Genet. Cytogenet.*, 2005, **158**, 27–34.

9 N. Ozgul, A. Cil, G. Bozdayi, A. Usubutun, D. Bulbul, S. Rota, M. Kose, A. Biri and A. Haberal, *J. Obstet. Gynaecol. Res.*, 2008, **34**, 865–871.
10 N. Murphy, M. Ring, C. Heffron, B. King, A. Killalea, C. Hughes, C. Martin, E. McGuinness, O. Sheils and J. O'Leary, *J. Clin. Pathol.*, 2005, **58**, 525–534.
11 N. Missaoui, S. Hmissa, L. Frappart, A. Trabelsi, A. Ben Abdelkader, C. Traore, M. Mokni, M. T. Yaacoubi and S. Korbi, *Virchows Arch.*, 2006, **448**, 597–603.
12 N. Murphy, M. Ring, A. G. Killalea, V. Uhlmann, M. O'Donovan, F. Mulcahy, M. Turner, E. McGuinness, M. Griffin, C. Martin, O. Sheils and J. J. O'Leary, *J. Clin. Pathol.*, 2003, **56**, 56–63.
13 S. Candefjord, K. Ramser and O. A. Lindahl, *J. Med. Eng. Technol.*, 2009, **33**, 585–603.
14 C. Kendall, M. Isabelle, F. Bazant-Hegemark, J. Hutchings, L. Orr, J. Babrah, R. Baker and N. Stone, *Analyst*, 2009, **134**, 1029–1045.
15 C. Krafft, G. Steiner, C. Beleites and R. Salzer, *J. Biophotonics*, 2009, **2**, 13–28.
16 F. Bazant-Hegemark, K. Edey, G. R. Swingler, M. D. Read and N. Stone, *Technol. Cancer Res. Treat.*, 2008, **7**, 483–496.
17 M. J. Walsh, M. J. German, M. Singh, H. M. Pollock, A. Hammiche, M. Kyrgiou, H. F. Stringfellow, E. Paraskevaidis, P. L. Martin-Hirsch and F. L. Martin, *Cancer Lett.*, 2007, **246**, 1–11.
18 N. S. Eikje, K. Aizawa and Y. Ozaki, *Biotechnol. Annu. Rev.*, 2005, **11**, 191–225.
19 R. J. Meier, *Chem. Soc. Rev.*, 2005, **34**, 743–752.
20 M. Diem, J. M. Chalmers and P. R. Griffiths, *Vibrational Spectroscopy for Medical Diagnosis*, John Wiley & Sons, Chichester, England, Hoboken, NJ, 2008.
21 J. I. Chang, Y. B. Huang, P. C. Wu, C. C. Chen, S. C. Huang and Y. H. Tsai, *Gynecol. Oncol.*, 2003, **91**, 577–583.
22 L. Chiriboga, P. Xie, H. Yee, D. Zarou, D. Zakim and M. Diem, *Cell Mol. Biol. (Noisy-le-grand)*, 1998, **44**, 219–229.
23 L. Chiriboga, H. Yee, M. Diem and B. Wood, *Gynecol. Oncol.*, 2003, **91**, 275–276, author reply 276–277.
24 M. A. Cohenford, T. A. Godwin, F. Cahn, P. Bhandare, T. A. Caputo and B. Rigas, *Gynecol. Oncol.*, 1997, **66**, 59–65.
25 S. G. El-Tawil, R. Adnan, Z. N. Muhamed and N. H. Othman, *Pathology*, 2008, **40**, 600–603.
26 M. Fung Kee Fung, M. Senterman, P. Eid, W. Faught, N. Z. Mikhael and P. T. Wong, *Gynecol. Oncol.*, 1997, **66**, 10–15.
27 J. G. Kelly, K. T. Cheung, C. Martin, J. J. O'Leary, W. Prendiville, P. L. Martin-Hirsch and F. L. Martin, *Clin. Chim. Acta*, 2010, **411**, 1027–1033.
28 C. M. Krishna, G. D. Sockalingum, B. M. Vadhiraja, K. Maheedhar, A. C. Rao, L. Rao, L. Venteo, M. Pluot, D. J. Fernandes, M. S. Vidyasagar, V. B. Kartha and M. Manfait, *Biopolymers*, 2007, **85**, 214–221.
29 S. R. Lowry, *Cell Mol. Biol. (Noisy-le-grand)*, 1998, **44**, 169–177.
30 S. Mark, R. K. Sahu, K. Kantarovich, A. Podshyvalov, H. Guterman, J. Goldstein, R. Jagannathan, S. Argov and S. Mordechai, *J. Biomed. Opt.*, 2004, **9**, 558–567.
31 S. Mordechai, R. K. Sahu, Z. Hammody, S. Mark, K. Kantarovich, H. Guterman, A. Podshyvalov, J. Goldstein and S. Argov, *J. Microsc.*, 2004, **215**, 86–91.
32 B. J. Morris, C. Lee, B. N. Nightingale, E. Molodysky, L. J. Morris, R. Appio, S. Sternhell, M. Cardona, D. Mackerras and L. M. Irwig, *Gynecol. Oncol.*, 1995, **56**, 245–249.
33 S. Neviliappan, L. Fang Kan, T. Tiang Lee Walter, S. Arulkumaran and P. T. Wong, *Gynecol. Oncol.*, 2002, **85**, 170–174.
34 C. H. Petter, N. Heigl, M. Rainer, R. Bakry, J. Pallua, G. K. Bonn and C. W. Huck, *Curr. Med. Chem.*, 2009, **16**, 318–326.
35 A. Podshyvalov, R. K. Sahu, S. Mark, K. Kantarovich, H. Guterman, J. Goldstein, R. Jagannathan, S. Argov and S. Mordechai, *Appl. Opt.*, 2005, **44**, 3725–3734.
36 B. Rigas, K. LaGuardia, L. Qiao, P. S. Bhandare, T. Caputo and M. A. Cohenford, *J. Lab. Clin. Med.*, 2000, **135**, 26–31.
37 R. Sindhuphak, S. Issaravanich, V. Udomprasertgul, P. Srisookho, S. Warakamin, S. Sindhuphak, R. Boonbundarlchai and N. Dusitsin, *Gynecol. Oncol.*, 2003, **90**, 10–14.
38 W. Steller, J. Einenkel, L. C. Horn, U. D. Braumann, H. Binder, R. Salzer and C. Krafft, *Anal. Bioanal. Chem.*, 2006, **384**, 145–154.
39 M. J. Walsh, M. N. Singh, H. M. Pollock, L. J. Cooper, M. J. German, H. F. Stringfellow, N. J. Fullwood, E. Paraskevaidis, P. L. Martin-Hirsch and F. L. Martin, *Biochem. Biophys. Res. Commun.*, 2007, **352**, 213–219.
40 B. R. Wood, L. Chiriboga, H. Yee, M. A. Quinn, D. McNaughton and M. Diem, *Gynecol. Oncol.*, 2004, **93**, 59–68.
41 B. R. Wood, M. A. Quinn, B. Tait, M. Ashdown, T. Hislop, M. Romeo and D. McNaughton, *Biospectroscopy*, 1998, **4**, 75–91.
42 H. M. Yazdi, M. A. Bertrand and P. T. Wong, *Acta Cytol.*, 1996, **40**, 664–668.
43 J. M. Schubert, B. Bird, K. Papamarkakis, M. Miljkovic, K. Bedrossian, N. Laver and M. Diem, *Lab. Invest.*, 2010, **90**, 1068–1077.
44 K. M. Ostrowska, A. Malkin, A. Meade, J. O'Leary, C. Martin, C. Spillane, H. J. Byrne and F. M. Lyng, *Analyst*, 2010, **135**, 3087–3093.
45 A. D. Meade, C. Clarke, F. Draux, G. D. Sockalingum, M. Manfait, F. M. Lyng and H. J. Byrne, *Anal. Bioanal. Chem.*, 2010, **396**, 1781–1791.
46 P. Bassan, A. Kohler, H. Martens, J. Lee, H. J. Byrne, P. Dumas, E. Gazi, M. Brown, N. Clarke and P. Gardner, *Analyst*, 2010, **135**, 268–277.
47 T. Sano, T. Oyama, K. Kashiwabara, T. Fukuda and T. Nakajima, *Pathol. Int.*, 1998, **48**, 580–585.
48 R. Klaes, T. Friedrich, D. Spitkovsky, R. Ridder, W. Rudy, U. Petry, G. Dallenbach-Hellweg, D. Schmidt and M. von Knebel Doeberitz, *Int. J. Cancer*, 2001, **92**, 276–284.
49 N. Murphy, M. Ring, A. Killalea, V. Uhlmann, M. O'Donovan, F. Mulcahy, M. Turner, E. McGuinness, M. Griffin, C. Martin, O. Sheils and J. O'Leary, *J. Clin. Pathol.*, 2003, **56**, 56–63.
50 K. Milde-Langosch, S. Riethdorf, A. Kraus-Poppinghaus, L. Riethdorf and T. Loning, *Virchows Arch.*, 2001, **439**, 55–61.
51 S. N. Agoff, P. Lin, J. Morihara, C. Mao, N. B. Kiviat and L. A. Koutsky, *Mod. Pathol.*, 2003, **16**, 665–673.
52 N. Murphy, C. C. Heffron, B. King, U. G. Ganuguapati, M. Ring, E. McGuinness, O. Sheils and J. J. O'Leary, *Virchows Arch.*, 2004, **445**, 610–615.
53 N. Murphy, M. Ring, C. C. Heffron, B. King, A. G. Killalea, C. Hughes, C. M. Martin, E. McGuinness, O. Sheils and J. J. O'Leary, *J. Clin. Pathol.*, 2005, **58**, 525–534.
54 S. S. Wang, M. Trunk, M. Schiffman, R. Herrero, M. E. Sherman, R. D. Burk, A. Hildesheim, M. C. Bratti, T. Wright, A. C. Rodriguez, S. Chen, A. Reichert, C. von Knebel Doeberitz, R. Ridder and M. von Knebel Doeberitz, *Cancer Epidemiol. Biomarkers Prev.*, 2004, **13**, 1355–1360.
55 Z. Movasaghi, S. Rehman and I. U. Rehman, *Appl. Spectrosc. Rev.*, 2007, **42**, 493–541.
56 J. W. Black and P. Leff, *Proc. R. Soc. London, Ser. B*, 1983, **220**, 141–162.
57 G. Lizard, M. C. Chignol, Y. Chardonnet, C. Souchier, M. Bordes, D. Schmitt and J. P. Revillard, *J. Immunol. Methods*, 1993, **157**, 31–38.
58 K. Adler, T. Erickson and M. Bobrow, *Histochem. Cell Biol.*, 1997, **108**, 321–324.
59 G. Mehés, N. Speich, M. Bollmann and R. Bollmann, *Pathol. Oncol. Res.*, 2004, **10**, 142–148.
60 E. M. Burd, *Clin. Microbiol. Rev.*, 2003, **16**, 1–17.
61 F. A. Suprynowicz, G. L. Disbrow, E. Krawczyk, V. Simic, K. Lantzky and R. Schlegel, *Oncogene*, 2008, **27**, 1071–1078.
62 K. M. Ostrowska, A. Malkin, A. D. Meade, J. J. O'Leary, C. Martin, C. O. Spillane, H. J. Byrne and F. M. Lyng, *Analyst*, 2010, **135**, 3087–3093.
63 P. T. Wong, R. K. Wong, T. A. Caputo, T. A. Godwin and B. Rigas, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 10988–10992.
64 J. Doorbar, *Clin. Sci.*, 2006, **110**, 525–541.
65 N. Murphy, C. C. Heffron, B. King, U. G. Ganuguapati, M. Ring, E. McGuinness, O. Sheils and J. J. O'Leary, *Virchows Arch.*, 2004, **445**, 610–615, epub 2004 Sep 2018.
66 M. Benevolo, M. Mottolese, F. Marandino, G. Vocaturo, R. Sindico, G. Piperno, L. Mariani, I. Sperduti, P. Canalini, R. P. Donnorso and A. Vocaturo, *Mod. Pathol.*, 2006, **19**, 384–391.
67 B. Passamonti, D. Gustinucci, P. Recchia, S. Bulletti, A. Carlani, E. Cesarini, M. R. D'Amico, V. D'Angelo, E. Di Dato, N. Martinelli, M. Malaspina and N. Spita, *Pathologica*, 2010, **102**, 6–11.
68 G. Negri, G. Bellisano, G. F. Zannoni, F. Rivasi, A. Kasal, F. Vittadello, S. Antoniazzi, G. Faa, R. Ambu and E. Egarter-Vigl, *Am. J. Surg. Pathol.*, 2008, **32**, 1715–1720.
69 G. R. Focchi, I. D. Silva, N. C. Nogueira-de-Souza, C. Dobo, C. T. Oshima and J. N. Stavale, *J. Low. Genit. Tract Dis.*, 2007, **11**, 98–104.