

# Spectral cytopathology of cervical samples: detecting cellular abnormalities in cytologically normal cells

Jennifer M Schubert<sup>1</sup>, Benjamin Bird<sup>1</sup>, Kostas Papamarkakis<sup>1</sup>, Miloš Miljković<sup>1</sup>, Kristi Bedrossian<sup>2</sup>, Nora Laver<sup>2</sup> and Max Diem<sup>1</sup>

Spectral cytopathology (SCP) is a novel spectroscopic method for objective and unsupervised classification of individual exfoliated cells. The limitations of conventional cytopathology are well recognized within the pathology community. In SCP, cellular differentiation is made by observing molecular changes in the nucleus and the cytoplasm, which may or may not produce morphological changes detectable by conventional cytopathology. This *proof of concept study* shows SCP's potential as an enhancing tool for cytopathologists by aiding in the accurate and reproducible diagnosis of cells in all states of disease. Infrared spectra are collected from cervical cells deposited onto reflectively coated glass slides. Each cell has a corresponding infrared spectrum that describes its unique biochemical composition. Spectral data are processed and analyzed by an unsupervised chemometric algorithm, principal component analysis. In this blind study, cervical samples are classified by analyzing the spectra of morphologically normal looking squamous cells from normal samples and samples diagnosed by conventional cytopathology with low-grade squamous intraepithelial lesions. SCP discriminated cytopathological diagnoses amongst 12 different cervical samples with a high degree of specificity and sensitivity. SCP also correlated two samples with abnormal spectral changes: these samples had a normal cytopathological diagnosis but had a history of abnormal cervical cytology. The spectral changes observed in the morphologically normal looking cells are most likely because of an infection with human papillomavirus (HPV). HPV DNA testing was conducted on five additional samples, and SCP accurately differentiated these samples by their HPV status. SCP tracks biochemical variations in cells that are consistent with the onset of disease. HPV has been implicated as the cause of these changes detected spectroscopically. SCP does not depend on identifying the sparse number of morphologically abnormal cells within a large sample to make an accurate classification, as does conventional cytopathology. These findings suggest that the detection of cellular biochemical variations by SCP can serve as a new enhancing screening method that can identify earlier stages of disease.

*Laboratory Investigation* (2010) 90, 1068–1077; doi:10.1038/labinvest.2010.72; published online 5 April 2010

**KEYWORDS:** cervical cancer; cytopathology; human papillomavirus; infrared micro-spectroscopy; spectral cytopathology

In the past 10 years, infrared spectroscopic methods have emerged for the diagnosis of disease in cells and tissues. The spectroscopic approach is based on infrared micro-spectroscopy (IR-MSP), coupled to unsupervised multivariate methods of analysis, such as principal component analysis (PCA, see below). This combination of IR-MSP and PCA is referred henceforth as spectral cytopathology (SCP). The standard to which SCP results are placed in juxtaposition is conventional cytopathology, which relies on the well-trained eyes of a cytopathologist to differentiate between normal and abnormal cell types. This methodology, first introduced by

Papanicolaou in the early 1920s, uses exfoliated cells from the anatomical region of interest that are subsequently stained and examined under a microscope for morphological changes.<sup>1,2</sup> The cytopathologist's task is to positively identify and diagnose a small number of cells within a sample containing approximately 10 000 cells that may show subtle morphological changes in the early stages of dysplasia to direct patient care in the clinic.

The prevalence of cervical cancer decreased 75% between 1955 and 1992 as a direct result of the cervical cancer screening methods introduced by Papanicolaou. However, the

<sup>1</sup>Laboratory for Spectral Diagnosis, Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA, USA and <sup>2</sup>Cytopathology, Department of Pathology, Tufts Medical Center, Boston, MA, USA

Correspondence: J Schubert, Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Avenue, 102 Hurtig Hall, Boston, MA 02115, USA.

E-mail: jmschubert@gmail.com

Received 30 June 2009; revised 15 January 2010; accepted 8 February 2010

American Cancer Society estimates that 11 270 women will be diagnosed with invasive cervical cancer in 2009.<sup>3</sup> A review by the US Department of Health and Human Service's Agency for Healthcare Research and Quality that evaluated cervical cytology, found the specificity of the Pap smear screening to be 0.98, 95% confidence interval (CI) 0.97–0.99, and the sensitivity to be 0.51, 95% CI 0.37–0.66.<sup>4</sup> The low accuracy because of false-negative results is widely recognized and acknowledged within the pathology community. False-negative results are primarily because of a low number of abnormal cells in a sample, but may also be the result of poor sample collection; a gradual transition of cells from a normal morphology to an abnormal morphology; biological contaminants such as blood and infectious organisms; sample preparation; and the inherent subjectivity of the diagnosis. Consequences of false-negative test results are compounded by low rates of interobserver reproducibility in accurately grading dysplasia of the uterine cervix.<sup>5,6</sup> These results affect patient care, and frequently require follow-up testing in terms of another Pap test and, in the case of atypical squamous cells of undetermined significance (ASCUS) a human papillomavirus (HPV) DNA test. Current research has shown that combined Pap screening and HPV DNA testing may offer an improvement in sensitivity to over 99%. Although HPV DNA testing is the state of the art in terms of treating and tracking the course of the disease in infected women, co-testing would not be cost-effective.<sup>7</sup>

These diagnostic challenges are not limited to cervical smears alone; cytological diagnoses from other body sites have equal or worse sensitivity. In a study of the accuracy of urinary cytology in daily practice, it was found that the diagnostic sensitivity of high-grade transitional cell neoplasms was roughly 79% whereas the sensitivity in low-grade transitional cell neoplasms was confirmed at approximately 26%.<sup>8</sup> The accuracy of clinical diagnosis was also studied in interstitial lung diseases and for patients who underwent surgical lung biopsy, in these cases the diagnoses were 62% accurate.<sup>9</sup> The accuracy rate is dependent on the presence of diagnostic morphological features and on adequate tissue samples.

To overcome some of the inherent obstacles of conventional cytopathology, a significant amount of research has been focused on applying SCP to procure a more objective diagnosis of cells. The aim of these studies is to develop a synergistic methodology using SCP that would assist cytopathologists in improving the overall accuracy of cytological diagnoses through collaborative work between bio-spectroscopists and cytopathologists.

The study by Wong<sup>10–12</sup> in the early 1990s claimed that IR-spectroscopy could differentiate between normal, dysplastic, and cancerous cervical cells in pellets based on decreased glycogen peaks and increased symmetric and anti-symmetric phosphate stretching intensities. However, follow-up studies undertaken by other groups indicated that the spectral changes observed by Wong were not related to the

molecular composition of dysplastic cells, but to confounding contributions made by different cell types present within a smear. Benign variations such as inflammation, metaplasia, the ratio of non-dividing to dividing cells, and the overall divisional activity of the cells will also dramatically change the IR spectrum collected.<sup>13–19</sup> As these problems were recognized, it also became apparent that other contaminants may affect the spectra, including blood, mucus, micro-organisms, and semen.<sup>20</sup>

Today, advances in instrumentation technology permit the collection of a spectrum from an individual cell, opposed to cells in pellet form. In 2006, initial studies explored the spectral variance among individual cells from homogeneous cell samples.<sup>21,22</sup> A model system using individual cells from canines was used to examine effects of cellular maturation. Homogeneous squamous cells were exfoliated from the cervix of both estrus and non-estrus dogs. Unsupervised statistical analysis using PCA showed distinct separation between these two states of maturation. These observed differences are due to hormones that initiate cells to mature when canines are in estrus.<sup>21</sup>

The aim of this research paper is to establish a primary proof of concept for SCP as a potential diagnostic tool for cytopathologists evaluating gynecological samples. Although hundreds of samples have been analyzed by SCP to date, only a small number of samples with sufficient patient background information were incorporated into this preliminary study. However, over 3000 cells were included from 17 patient samples, which is statistically significant. The method of analysis, PCA, is not a diagnostic algorithm, but an unsupervised classification algorithm that does not require training or validation data: PCA simply identifies quantifiable spectral differences and classifies the data accurately without user input. Therefore, in terms of this proof of concept study with a limited number of patient samples, PCA is a suitable classification algorithm.

SCP is unique in that it takes a snapshot of a cell's biochemical composition. This snapshot encompasses all the biochemical processes in the cell at the time of exfoliation. This includes, but is not limited to, any defining hormonal influences or inherent viral infections, which may not be reflected in the cell's morphology. The cervical samples used in the SCP analysis were correlated with cytopathological interpretations. Non-diagnostic, or normal looking, cells were selected from samples with a cytological diagnosis of low-grade dysplasia. We show that SCP has the sensitivity required to differentiate these cervical cells, from normal cells of healthy patients based on the infrared analysis of their biochemical composition. DNA testing was performed on five additional samples, diagnosed with normal cytopathology and low-grade squamous intraepithelial lesion (LSIL) to confirm the presence of any high-risk HPV (hrHPV) strains. Morphologically normal cells of hrHPV + samples (one LSIL sample, one normal cytology sample) were differentiated from cells of cytologically normal hrHPV– samples

by PCA. This study parallels oral cytology studies in our laboratory that distinguish cells from a normal sample to those in a sample diagnosed as reactive or cancerous.<sup>23</sup>

## MATERIALS AND METHODS

### Sample Preparation

All cervical samples were obtained in collaboration with the Cytopathology Division of the Pathology Department at Tufts Medical Center (Boston, MA, USA) after routine testing and follow-up had been performed. Samples on cytological brushes were preserved in SurePath® solution (Burlington, NC, USA), and adequate cellularity remained for SCP analysis. Subsequently, cells were vortexed off the brushes, filtered to remove debris, and deposited onto 'low-e' microscope slides (Kevley Technologies, Chesterland, OH, USA) using cytocentrifugation (CytoSpin, Thermo, Waltham, MA, USA). Ethical approval for this study was provided by a local institutional review board and supported by the National Institute of Health (NIH).

### Data Collection

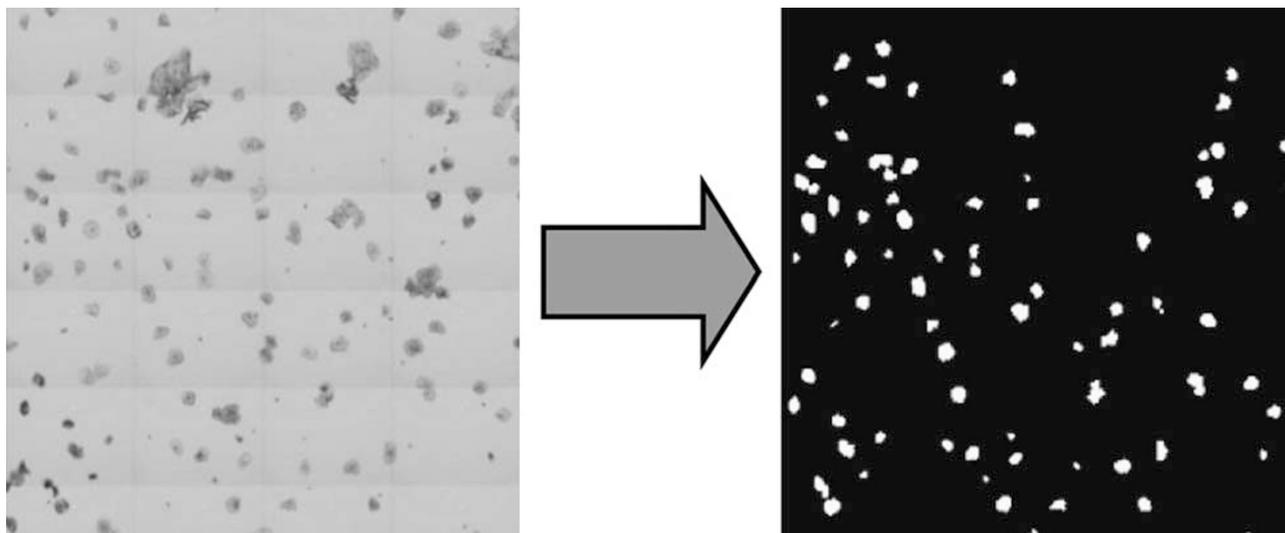
Infrared spectral data were collected from a 4 mm × 4 mm area of the sample deposited on a low-e slide in imaging mode using one of two Perkin Elmer Spectrum One/Spotlight 400 imaging IR micro-spectrometers (Sheldon, CT, USA) in the Laboratory for Spectral Diagnosis at Northeastern University. The instrument optical bench, the infrared microscope, and an external microscope enclosure box were purged with a continuous stream of dry air (−40 °C dew point). Inside the purge chamber, the relative humidity is below the limits of detectability using standard commercial hygrometers (<5% relative humidity). The following data acquisition parameters were used: 4 cm<sup>−1</sup> spectral resolution,

6.25 μm × 6.25 μm pixel size, Norton–Beer apodization, one level of zero-filling, no atmospheric background correction. Two co-added interferograms for each pixel were Fourier transformed to yield spectral vectors, each covering the 4000–700 cm<sup>−1</sup> range at 2 cm<sup>−1</sup> intervals. Background spectra for all 16 detector elements were collected using 128 co-added interferograms. Data were collected in reflectance mode. Raw data sets consist of 409 600 spectra, occupy *ca.* 2.54 GByte each and are stored in native instrument data format (.fsm).

### Image Processing

Raw data sets from the infrared micro-spectrometers were imported into a program developed in the investigators' laboratory and referred to as PapMap.<sup>24</sup> This program is written in 64-bit MATLAB (The Mathworks, Natick, MA, USA) to accommodate the large data matrices. PapMap reconstructs the spectra of individual cells collected in mapping mode from between 9 and 100 individual pixel spectra for each cell area (corresponding to cells with a diameter between *ca.* 19 and 63 μm). To this end, PapMap first establishes which pixel spectra belong to a given cell. This is accomplished by constructing a binary mask in which contiguous regions belonging to individual cells are identified. Such a binary mask is shown in Figure 1. The larger white areas, corresponding to squamous cells, may consist of up to 100 individual pixels, whereas the smallest white areas may contain as few as 9 pixels. Cell clumps are eliminated because they occupy areas larger than 100 pixels. This mask is established by defining a threshold for the amide I intensity (1650 cm<sup>−1</sup>), which is a specific signature of protein abundance (see below).

For each contiguous area occupied by a cell (i.e. the white pixels), the cellular spectrum is calculated, starting from the



**Figure 1** (Left) Visual image of a 2 mm × 2 mm area on a sample spot on a low-e slide collected on the PE Spotlight microscope. (Right) 'Binary mask' image, constructed through PapMap, of the image shown on the left. The white spots constitute the area occupied by a cell. The cellular diameter ranges from 3 to 10 pixels, or 19 μm to 63 μm (1 pixel equals 6.25 μm).

spectrum with the largest amide I intensity. This spectrum is presumably from the nucleus of the cell, which always shows the strongest protein intensity. Subsequently, all spectra identified by the binary mask to be associated with a cell are co-added and subject to several constraints, to prevent very weak spectra with poor signal-to-noise contaminating the cell spectrum and to prevent spectra from the edges of a cell, which may be contaminated by scattering<sup>25–27</sup> to be co-added. The co-added cellular spectra, as well as the coordinates of each cell, are exported for further data analysis.

After infrared data collection, the cells on a slide are manually stained using standard cytological stain combinations: Protocol OG6 (Fisher Scientific, Kalamazoo, MI, USA); EA-50 (Surgipath Medical Industries, Richmond, IL, USA); Hemotoxylin 1, Clarifier 1, and Bluing (Richard-Allan Scientific, Kalamazoo, MI, USA). Tap water and solutions of ethanol are used in the washing steps. Finally, to avoid degradation, slides are dipped in xylene and cover-slipped for cytological analysis. Next, visual images at  $40\times$  magnification of each stained cell are collected at the coordinates indicated by the PapMap algorithm, using an Olympus BX40 microscope fitted with a computer-controlled stage and a QImaging GO3 3MB digital color camera. The images and cellular spectra are linked and stored in a database for easy identification. The cell images are diagnosed by a cytopathologist, and the resulting medical diagnosis is correlated to spectral and cytologic data.

### Data Analysis

Data were pre-processed as follows: the spectral range for analysis of cervical cells was restricted to 3100–2800 and 1700–1200  $\text{cm}^{-1}$ , because the 2800 to 1700  $\text{cm}^{-1}$  is devoid of spectral information. Subsequently, second derivatives of the spectral intensities vs wavenumber were calculated, using a nine-point sliding window<sup>28</sup> and vector normalized.

The spectral range from 1700 to 1200  $\text{cm}^{-1}$  was chosen for subsequent analysis because it eliminated confounding effects caused by glycogen contribution. Glycogen shows a triad of specific peaks between 1250 and 1000  $\text{cm}^{-1}$ , and could be used to follow cyclic changes.<sup>29,30</sup> However, variations in glycogen abundance in cells can be caused by a number of conditions, and it is difficult to correlate these causes in a diagnostic application. The 1700–1480  $\text{cm}^{-1}$  spectral region contains two dominant protein peaks, known as the amide I (1650  $\text{cm}^{-1}$ , C=O stretching) and amide II ( $\sim 1550$   $\text{cm}^{-1}$ , C–N stretching and N–H deformation) vibrations in the primary protein structure. Changes in these peaks, with regards to band shape, position, and appearance of shoulders, are due to overall changes in the abundance of specific proteins. The region between 1480–1200  $\text{cm}^{-1}$  contains spectral bands related to other biochemical components, including DNA and RNA, phosphates and phospholipids. An increase in intensity of the DNA band may be indicative of a viral infection because of its high replication rate. The

3100–2800  $\text{cm}^{-1}$  region contains the C–H stretching vibrations of mainly lipids and phospholipids.

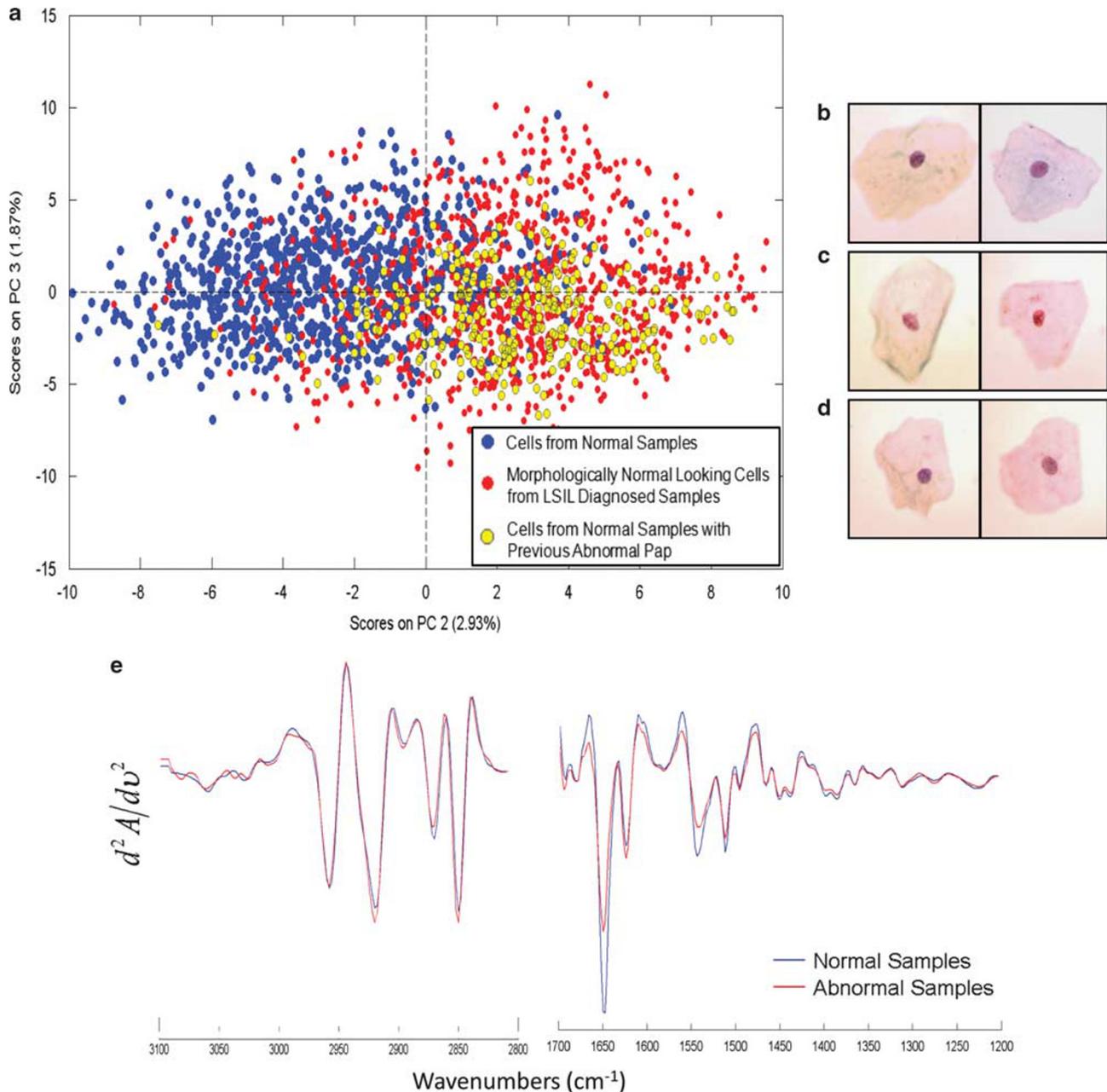
At this point, PCA, an unsupervised method of multivariate analysis, was performed. The term ‘unsupervised analysis’ means that training data and validation data are not needed to accurately classify data. For a detailed explanation of PCA, the reader is referred to Adams.<sup>31</sup> PCA was carried out using the PLS Toolbox 402 (Eigen Vector Research, Wenatchee, WA, USA) in MATLAB. The results from PCA are presented in the form of ‘scores plots’, in which the spectrum of each cell is represented by a dot in a coordinate system that indicates the contributions of principal component (PC) 2 and 3 to the spectrum of the cell. The PCs are obtained from the eigenvectors of the correlation matrix of the data set, and represent a totally unbiased decomposition of cellular spectra. Although this method is not suitable for diagnostic purposes, we use it at this stage to determine whether or not there are systematic changes in cellular spectra and to classify these spectral changes.

### Sample Selection

The cyclic fluctuations of estrogen and progesterone, the hormones influencing the proliferation and differentiation of the cervical epithelium, cause inevitable biochemical effects that make diagnosis by SCP more difficult because the phase of the menstrual cycle has to be considered. Therefore, all cervical samples used in this study were from patients who were taking hormonal contraceptives. Hormonal contraceptives prevent pregnancy by maintaining the patient’s hormone levels to that of the ovulation–luteal phase of a typical cycle. For this study, it is assumed that all patients have a similar estrogen and progesterone profile.

Superficial squamous cells from five samples diagnosed by a cytopathologist as ‘negative for intraepithelial lesion or malignancy’ (referred to as normal for the remainder of this paper) and from five samples diagnosed by a cytopathologist with ‘epithelial cell abnormality: LSIL, encompassing HPV’ were scrutinized by SCP. Two additional samples that had a normal cytopathological diagnosis at the time of sample collection, but had a recent history of cervical disease, were also included. As data were acquired for unstained cells, both morphologically normal and abnormal cells are included in the data sets from LSIL samples. However, PCA was executed in two steps: first, normal cells from normal samples were compared with morphologically normal looking cells from LSIL samples (Figure 2); second, cells with morphological abnormalities were compared with morphologically normal looking cells from LSIL samples (Figure 3).

An additional study evaluated the presence of hrHPV strains (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68) by performing a DNA test using the Digene Hybrid Capture II technique (Qiagen, Valencia, CA, USA). Five samples were included in this study (all on hormonal contraceptives): four had a normal cytopathological diagnosis, and one was

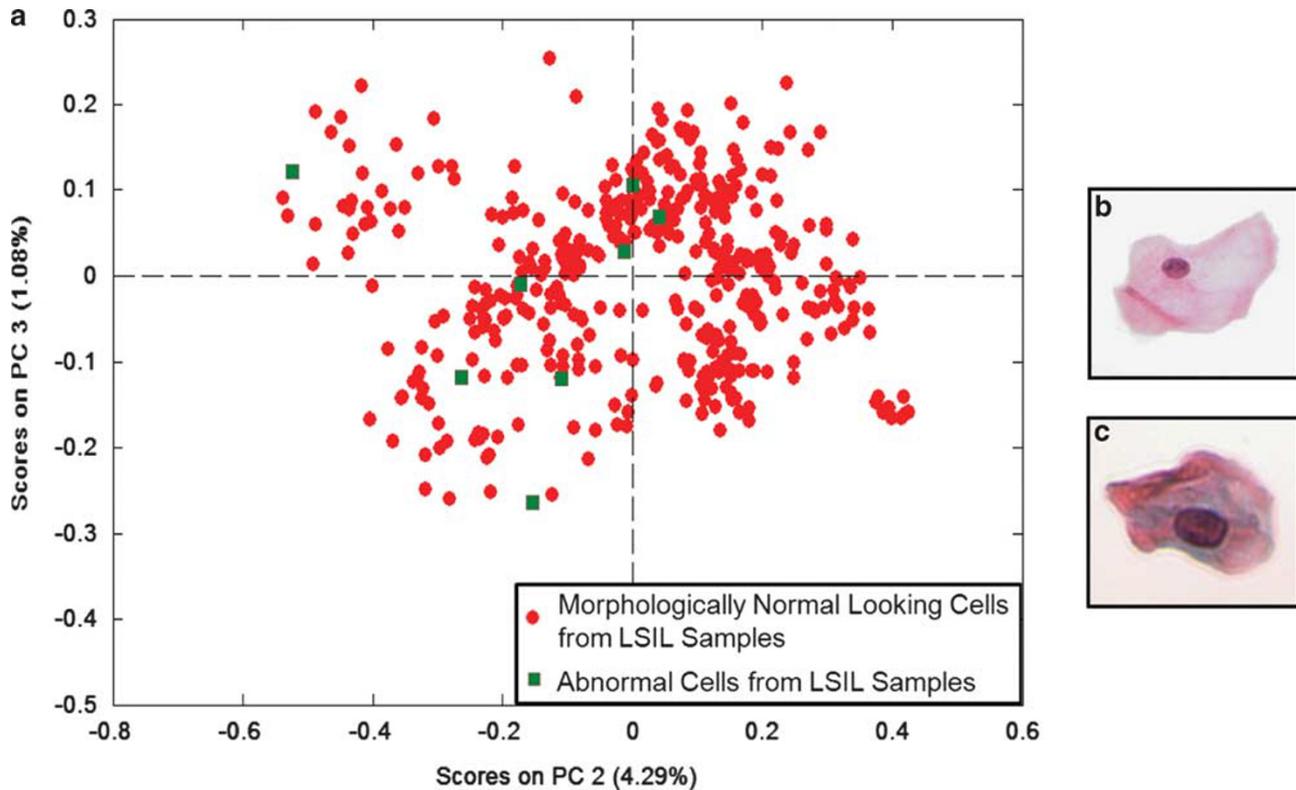


**Figure 2** (a) PCA scores plot of cervical cells from five normal samples (blue), five abnormal samples (red), and two samples from patients with a history of an abnormal Pap test (yellow). The number of cells included from each sample is listed in Table 1. Each symbol represents the spectrum of an individual cell. There is a separation of the blue and red classes along PC2. The cellular spectra from samples of patients with a history of abnormal cervical cytology (yellow) co-cluster with the abnormal samples (red). All cells were morphologically normal as shown in (b–d), representing high-resolution  $40\times$  images of cells from each classification. (e) Mean second derivative vector normalized spectra of the normal samples and abnormal samples for the spectral range  $3100\text{--}2800$  and  $1700\text{--}1200\text{ cm}^{-1}$ .

diagnosed with low-grade dysplasia. Three of the normal samples were hrHPV negative. The dysplastic sample and one of the normal samples were hrHPV positive. Only morphologically normal looking cells were analyzed. Figure 4 shows the PCA results.

The investigators acknowledge the low number of samples in this proof of concept study. This paper reports partial

results from a much larger study involving well over 100 patients; however, only those patients for whom the hormonal status was known accurately were included. Thus, the 3000 cellular spectra from 17 patient samples represent a data set, which is devoid of confounding hormonal influences. In the first part of the study (Figures 2 and 3), we show the spectral differentiation of truly normal cells from patients



**Figure 3** (a) PCA scores plot of eight cells that show morphological abnormalities (green) and the morphologically normal looking cells (red) from the same two samples. (b, c) Representative high-resolution  $40\times$  images of a morphologically normal looking cell and a morphologically abnormal cell, respectively.

diagnosed as normal and from morphologically normal looking cells from patients diagnosed as LSIL or history of LSIL. In the second part of the study (Figure 4), we provide evidence that the spectral differences between healthy samples and diseased samples is due to the presence of hrHPV. The same trends were observed in a study of oral cells submitted previously.<sup>2,3</sup>

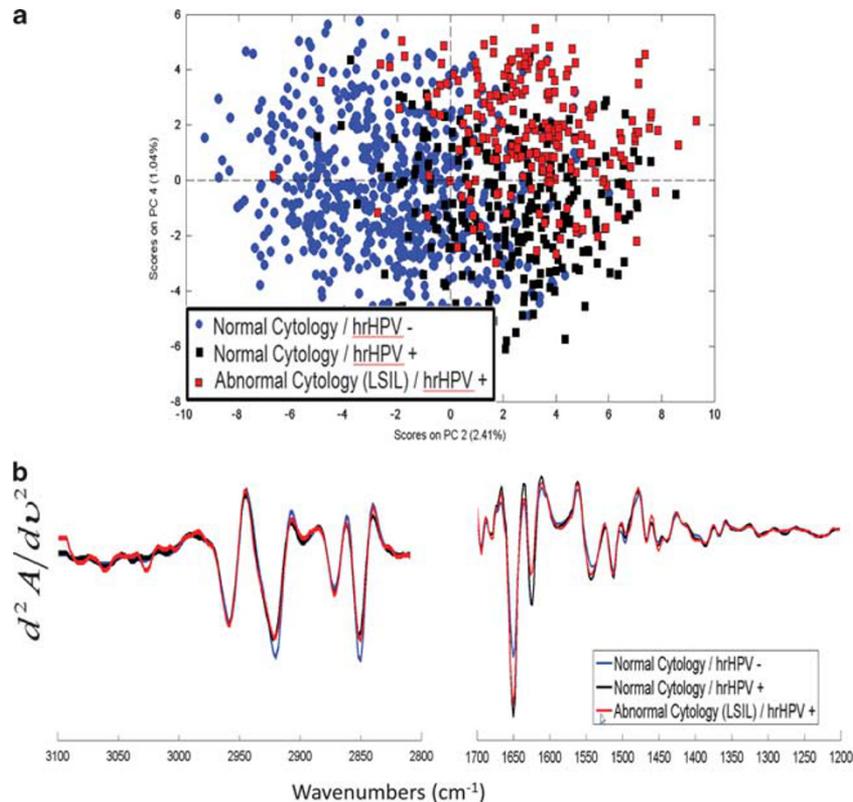
## RESULTS

Figure 2a depicts a PCA-based scores plot that blindly differentiates the majority of cells into two classes: cells from normal samples (Table 1, samples 1–5, blue) and morphologically normal looking cells from abnormal samples (Table 1, samples 6–10, red). In addition, two samples, shown in yellow (Table 1, samples 11–12), were diagnosed by the cytopathologist as normal, but had a recent history of abnormal cervical cytology. SCP classified both of these samples with the abnormal samples. Table 1 outlines the original cytological diagnosis as well as the number of individual cells used from each sample in the SCP analysis. Representative cells from each classification: normal, morphologically normal looking cells from abnormal samples, and samples with a history of abnormal Pap test results, are depicted in Figures 2b–d, respectively.

Figure 2e depicts the mean cellular spectral changes of the second derivative spectra between the normal samples and

the abnormal samples. The most notable spectral changes are in the amide I and amide II manifolds. The term ‘manifold’ is used here because these bands are superpositions of the protein spectral features of hundreds of cellular proteins, and therefore, do not represent the vibrations of one, but multiple proteins. Small variations in the peak height ratio for the amide I and amide II bands are apparent. More significantly, there is a discernable change in the ratio of the amide I and its low-frequency component at  $1624\text{ cm}^{-1}$ . These differences may be interpreted in terms of a change in the proteome of the cells. Changes in band shape and peak position are also apparent in the rest of the spectral region analyzed. This figure is the first report of biochemical compositional changes between morphologically normal looking cells from normal and diseased cervical samples. Excellent spectral quality, and the averaging of hundreds of individual cells, allow for the characterization of these spectral differences, which are directly related to the discrimination of classes shown in Figure 2a.

Conventional cytopathology requires the identification of abnormal cells, which may be few, within a sample of thousands of non-diagnostic cells, to make a diagnosis. SCP does not rely on the chance of picking out diagnostic sparse cells. In fact, the infrared spectral signatures for cells showing morphological abnormalities and normal looking cells from the same abnormal samples are nearly indistinguishable,



**Figure 4** (a) PCA scores plot of five samples that were tested for hrHPV DNA: three samples with normal cytology were hrHPV negative (blue circles), one sample with normal cytology was hrHPV positive (black squares), and one sample was diagnosed by cytology with low-grade dysplasia and was hrHPV positive (red squares). The samples differentiate along PC2 based on their HPV status. (b) Mean second derivative vector normalized spectra for each class for the spectral range 3100–2800 and 1700–1200  $\text{cm}^{-1}$ .

but quite different from those of normal cells from healthy patients. Figure 3a depicts a PCA-based scores plot in which eight morphologically abnormal cells from two samples diagnosed with LSIL were compared with the morphologically normal looking cells from the same samples. These eight cells, with reactive and atypical morphological features, co-cluster with the morphologically normal looking cells. High-resolution images of representative cells from each classification are shown in Figures 3b and c. Although these cells are undoubtedly very different, the spectral patterns of the cells showing abnormal morphologies and morphologically normal looking cells are nearly equivalent.

The PCA scores plot in Figure 4 shows for the first time that there are quantifiable differences between hrHPV-infected and hrHPV-negative cells, even when analyzing only normal looking cells. The three hrHPV– samples with corresponding normal cytology (Table 1, samples 13–15, blue circles) cluster away from the two hrHPV+ samples (Table 1, samples 16–17). One of these hrHPV+ samples (sample 16, black squares) had a normal cytological diagnosis with no known history of cervical disease. The most notable spectral changes shown in Figure 4b are in the amide I and amide II regions; however, there are some distinct changes in the DNA, RNA, and phosphate bands as well as in the C–H stretching region.

## DISCUSSION

In the first part of this preliminary study, SCP was used to successfully differentiate cytologically indistinguishable squamous cells from normal samples and LSIL samples, and show how cells showing abnormal patterns are not necessary for SCP an accurate diagnosis. In the second part of this study, SCP distinguished spectral differences between cells based on hrHPV status.

Studies have shown that HPV DNA can be found in 99.7% of all cervical cancers,<sup>32</sup> and it has been proven that infection of hrHPV is a necessary prerequisite for the development of cervical cancer.<sup>33</sup> Cervical dysplasia is commonly an early manifestation of HPV infection, typically appearing 4–24 months after exposure. Invasive cervical cancer is a rare and late manifestation of HPV infection, with an average detection period of 10–20, even 30 years.<sup>34</sup>

The PCA scores plot in Figure 2a undeniably illustrates the power of SCP, because there is a strong separation of the red and blue classes along PC2. PC2 is directly related to the most significant spectral differences between normal healthy cells and morphologically normal looking cells from abnormal samples. These changes are described in the mean second derivative spectra in Figure 2e: a dramatic increase of the amide I low-frequency component and the decrease in the amide I to amide II peak height ratio. The source of these

**Table 1 Information for samples included in SCP analysis**

Sample no.	Cytopathological diagnosis of sample	No. individual cells for SCP
1	Normal	105
2	Normal	174
3	Normal	200
4	Normal	191
5	Normal	197
6	LSIL, encompassing HPV	201
7	LSIL, encompassing HPV	155
8	LSIL, encompassing HPV	291
9	LSIL, encompassing HPV	221
10	LSIL, encompassing HPV	192
11	Normal, history of abnormality	132
12	Normal, history of abnormality	369
13	Normal, hrHPV negative	136
14	Normal, hrHPV negative	263
15	Normal, hrHPV negative	159
16	Normal, hrHPV positive	243
17	LSIL, hrHPV positive	216

This table lists the original cytological diagnosis for each sample included in this study, as well as the number of individual cells from each respective sample, which were included in the SCP analysis. All cells listed are morphologically normal looking, in spite of the sample's original cytological diagnosis. On review of the slides from samples 11 to 12, no morphological abnormalities were identified. However, these patients had a history of abnormal cervical cytology. Samples 1–12 were used in part 1 of the study, Figures 2 and 3. Samples 13–17 were tested for hrHPV strains via DNA analysis. These samples were used in part 2 of the study, Figure 4.

changes in the protein composition may be the result of a viral infection, HPV, or other, disease-specific changes in protein content. These samples were not tested for HPV DNA, but there is strong evidence that cervical abnormalities are due to viral infection. Similar results have been observed in oral SCP studies of patients infected with the herpes simplex virus: although the majority of the cells are morphologically normal, a progression of disease can be observed ranging from healthy normal cells, to morphologically normal infected cells, and finally, cells infected with the herpes simplex virus.<sup>23</sup> Although the cells from the abnormal samples used in this analysis were morphologically normal, they showed abnormal spectral changes in the majority of the cells using SCP. Some morphologically normal looking cells did co-cluster with the normal cells in Figure 2a. This is expected because all cells may not show the same biochemical changes.

The biochemistry of the cells in the samples with a history of cellular abnormality (yellow) in Figure 2a deserves further investigation. These samples were initially diagnosed as normal by the cytopathologist, but classified with the

abnormals by SCP analysis. With a clinical history of previously abnormal Pap tests, it is possible that these samples are still infected with HPV. Studies have shown that some HPV infections are cleared by the body's immune system within 2 years.<sup>3</sup> In cases wherein a lesion was detected during morphological analysis, an HPV DNA test was positive at the previous cytologically normal Pap screening test. The HPV DNA remained positive in one or more follow-up Pap screening tests while the cytological evaluation was normal.<sup>34–39</sup> Therefore, the SCP classification is most likely not a false-positive in these two instances, but rather a confirmation for the sensitivity of the technique in detecting a latent infection.

In general, it is not the actual infection of HPV, but rather a persistent infection by the virus that is risky. HPV infects the basal layer cells of immature metaplastic epithelium in the transformation zone. HPV can remain in these cells as low copy episomal viral genomes (50–100 per cell), which are replicated only once per cycle. Therefore, these infected cells can provide a reservoir of virus in morphologically normal looking cells. Normally, the virus is cleared in a year or two, but can persist in this stable episomal state for extended periods of time. This is a non-productive or latent infection.<sup>40</sup> In the two cases with a history of abnormal Pap test results, SCP has shown its potential in monitoring such latent infections, which possibly could lead to cervical cancer if they are not shed from the cervical epithelium.

Figure 3 illustrates that the cells, which show morphological patterns of abnormality show the same spectral patterns as the cells with normal morphology from abnormal samples. SCP reaches the same diagnosis as the cytopathologist without the presence of any diagnostic cells and using only unstained cells. Only eight cells from two samples diagnosed with LSIL were identified as reactive or atypical—there were no characteristic low-grade dysplastic cells identified from any of the samples, which would have been present in the original sample screened by the cytopathologist. In Figure 3a, the eight morphologically abnormal cells were analyzed in conjunction with the morphologically normal looking cells from the same two samples. The scores plot shows co-clustering of the two groups; therefore, there are no statistically significant spectral differences between these two groups. The drastic differences in morphology are shown in Figures 3b and c. This analysis shows strong evidence that SCP can accurately classify samples without cells showing morphological changes as conventional cytopathology requires. SCP would not only improve the sensitivity of the Pap test, but potentially also improve patient care.

In another blind test, PCA accurately differentiated cells based on hrHPV status, Figure 4a. By performing additional DNA testing on five other samples, we were able to confidently attribute the spectral changes to the presence of hrHPV strains. Four of the samples precisely correlated with their cytological diagnosis; however, one sample with a normal cytological diagnosis, with no known history of

cervical disease, was hrHPV positive. This sample clustered with the other hrHPV positive sample diagnosed with low-grade dysplasia. This result is not only important to show that SCP can distinguish HPV infection, but also shows that SCP has the potential to provide more diagnostic information than standard cytology alone.

The most notable spectral changes observed in Figure 4b are in the protein amide I and amide II bands (1670–1610 and 1550–1500  $\text{cm}^{-1}$ , respectively), which most likely can be attributed to a degradation of cellular proteins and increased production of viral proteins. Changes in the DNA and RNA regions are due to the high replication rate of the viral genome.

This proof of concept study shows that SCP can be used successfully to detect abnormalities using morphologically un-differentiable cervical cells. Samples were used from patients who were taking hormonal contraceptives to eliminate variance because of the menstrual cycle. SCP reveals intrinsic spectral changes in the biochemical composition that differentiates cells from normal samples and morphologically normal looking cells from LSIL samples. SCP does not require the identification of a few abnormal cells to classify cells precisely, but rather detects biochemical compositional changes in cells that still show normal morphology. Although HPV DNA testing is currently the state of the art for diagnostic cytopathology, SCP differentiated cells based on their individual hrHPV status. In conclusion, there is substantial evidence that SCP has the potential to provide an objective, unsupervised and unbiased method of detecting and classifying exfoliated cells as a novel tool in diagnostic cytopathology. Studies including a wider patient population are currently underway for patients not taking hormonal contraceptives and from samples diagnosed with more progressive diseases, mainly high-grade dysplasia and carcinoma *in situ*, as well as ASCUS.

#### ACKNOWLEDGEMENT

We gratefully acknowledge the National Institute of Health (CA090346) for supporting this research.

#### DISCLOSURE/CONFLICT OF INTEREST

The authors declare no conflict of interest.

- Papanicolaou GN, Traut H. The diagnostic value of vaginal smears in carcinoma of the uterus. *Am J Obstet Gynecol* 1941;42:193–206.
- Papanicolaou GN. New cancer diagnosis. *Proc Third Race Betterment Conference*, January 2–6, 1928. 1928; 528–534.
- Cervical Cancer. June 2009. Available from: [www.cancer.org/docroot/CRI/content/CRI\\_2\\_4\\_IX\\_what\\_are\\_the\\_key\\_statistics\\_for\\_cervical\\_cancer\\_8.asp?rnav=cri](http://www.cancer.org/docroot/CRI/content/CRI_2_4_IX_what_are_the_key_statistics_for_cervical_cancer_8.asp?rnav=cri).
- Evaluation of Cervical Cytology. Summary, Evidence Report/Technology Assessment June 2009. Available from: <http://www.ahrq.gov/clinic/epcsums/cervsumm.htm>.
- Solomon D, Davey D, Kurman R, *et al*. The 2001 Bethesda system terminology for reporting results of cervical cytology. *JAMA* 2002;287:2114–2119.
- Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations. *JAMA* 2001;285:1500–1505.
- Brink AATP, Zielinski GD, Steenbergen RDM, *et al*. Clinical relevance of human papillomavirus testing in cytopathology. *Cytopathology* 2005;16:7–12.
- Bastacky S, Ibrahim S, Wilczynski SP, *et al*. The accuracy of urinary cytology in daily practice. *Cancer* 1999;87:118–128.
- Raghu G, Mageto YN, Lockhart D, *et al*. The accuracy of the clinical diagnosis of new-onset idiopathic pulmonary fibrosis and other interstitial lung disease. *Chest* 1999;116:1168–1174.
- Wong PTT, Papavassiliou ED, Rigas B. Phosphodiester stretching bands in the infrared spectra of human tissues and cultured cells. *Appl Spectrosc* 1991;45:1563–1567.
- Wong PTT, Wong RK, Caputo TA, *et al*. Infrared spectroscopy of exfoliated human cervical cells: evidence of extensive structural changes during carcinogenesis. *Proc Nat Acad Sci USA* 1991;88:10988–10992.
- Wong PTT, Lacelle S, Fung MFK, *et al*. Characterization of exfoliated cells and tissues from human endocervix and ectocervix by FTIR and ATR/FTIR spectroscopy. *Biospectroscopy* 1995;1:357–364.
- Cohenford M, Godwin T, Cahn F, *et al*. Infrared spectroscopy of normal and abnormal cervical smears: evaluation by principal component analysis. *Gynecol Oncol* 1997;66:59–65.
- Cohenford M, Rigas B. Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: implications for tumor biology. *Proc Nat Acad Sci USA* 1998;95:15327–15332.
- Wood BR, Quinn MA, Burden FR, *et al*. An investigation into FTIR spectroscopy as a biodiagnostic tool for cervical cancer. *Biospectroscopy* 1996;2:143–153.
- Wood BR, Quinn MA, Tait B, *et al*. A FTIR microspectroscopic study of cell types and potential confounding variables in screening for cervical malignancies. *Biospectroscopy* 1998;4:75–91.
- Boydston-White S, Gopen T, Houser S, *et al*. Infrared spectroscopy of human tissue V. Infrared spectroscopic studies of myeloid leukemia (ML-1) cells at different phases of the cell cycle. *Biospectroscopy* 1999;5:219–227.
- Chiriboga L, Xie P, Yee H, *et al*. Infrared spectroscopy of human tissue. I. Differentiation and maturation of epithelial cells in the human cervix. *Biospectroscopy* 1998;4:47–53.
- Chiriboga L, Xie P, Vigorita V, *et al*. Infrared spectroscopy of human tissue. II. A comparative study of spectra of biopsies of cervical squamous epithelium and of exfoliated cervical cells. *Biospectroscopy* 1998;4:55–59.
- Chiriboga L, Xie P, Yee H, *et al*. Infrared spectroscopy of human tissue IV. Detection of dysplastic and neoplastic changes of human cervical tissue via infrared microspectroscopy. *Cell Mol Biol* 1998;44: 219–229.
- Romeo MJ, Mohlenhoff B, Jennings M, *et al*. Infrared microspectroscopic studies of epithelial cells. *Biochim Biophys Acta* 2006;1758:915–922.
- Romeo MJ, Mohlenhoff B, Diem M. Infrared microspectroscopy of human cells: causes for the spectral variance of oral mucosa (buccal) cells. *Vib Spectrosc* 2006;42:9–14.
- Papamarkakis K, Bird B, Schubert JM, *et al*. Cytopathology by optical methods: spectral cytopathology of the oral mucosa. *Lab Invest* 2010; 90:589–598.
- Diem M, Miljkovic M, Romeo M, *et al*. Method of reconstituting cellular spectra from spectral mapping data. *USA* 2009.
- Bassan P, Byrne HJ, Bonnier F, *et al*. Resonant Mie scattering infrared spectroscopy of biological materials - understanding the 'dispersion artefact'. *Analyst* 2009;134:1586–1593.
- Bassan P, Byrne HJ, Lee J, *et al*. Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells. *Analyst* 2009;134:1171–1175.
- Bassan P, Kohler A, Martens H, *et al*. Resonant Mie scattering (RMieS) correction of infrared spectra from highly scattering biological samples. *Analyst* 2010;135:268–277.
- Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 1964;36:1627–1639.
- Tay SK, Singer A. The effects of oral contraceptive steroids, menopause and hormone replacement therapy on the cervical epithelium. In: Jordan JA, Singer A (eds). *The Cervix*, 2nd edn. Blackwell Publishing: Malden, 2006, pp 128–144.
- Romeo MJ, Wood BR, McNaughton D. Observing the cyclical changes in cervical epithelium using infrared microspectroscopy. *Vib Spectrosc* 2002;28:167–175.

31. Adams MJ. *Chemometrics in Analytical Spectroscopy*, 2nd edn, Royal Society of Chemistry: Cambridge, 2004.
32. Walboomers JMM, Jacobs MV, Manos MM, *et al*. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* 1999;189:12–19.
33. Bosch FX, Lorincz A, Munoz N, *et al*. The causal relation between human papillomavirus and cervical cancer. *J Clin Pathol* 2002;55: 244–265.
34. Woodman CBJ, Collins S, Winter H, *et al*. Natural history of cervical human papillomavirus infection in young women: a longitudinal cohort study. *Lancet* 2001; 1931–1936.
35. Winer RL, Kiviat NB, Hughes JP, *et al*. Development and duration of human papillomavirus lesions, after initial infection. *J Infect Dis* 2005;191:731–738.
36. Zielinski GD, Snijders PFJ, Rozendaal L, *et al*. HPV presence precedes abnormal cytology in women developing cervical cancer and signals false negative smears. *Br J Cancer* 2001;85:398–404.
37. Hildesheim A, Schiffman MH, Gravitt PE, *et al*. Persistence of type-specific human papillomavirus infection among cytologically normal women. *J Infect Dis* 1994;169:235–240.
38. Wallin K, Wiklund F, Angstrom T, *et al*. Type specific persistence of human papillomavirus DNA before the development of invasive cervical cancer. *New Engl J Med* 1999;341:1633–1638.
39. Bulk S, Rozendaal L, Zielinski GD, *et al*. High-risk human papillomavirus is present in cytologically false-negative smears: an analysis of 'normal' smears preceding CIN2/3. *J Clin Pathol* 2008;61:385–389.
40. DeMay RM. *The Pap Test*. ASCP Press: Chicago, 2005.