Translational Oncology

www.transonc.com

Volume 5 Number 3 June 2012 pp. 172–179 172

Near-infrared Raman Microspectroscopy Detects High-risk Human Papillomaviruses[1]

Elizabeth Vargis*,[2], Yi-Wei Tang†,[3], Dineo Khabele‡ and Anita Mahadevan-Jansen*

*Department of Biomedical Engineering, Vanderbilt University, Nashville, TN; †Departments of Pathology and Medicine, Vanderbilt University Medical Center, Nashville, TN; ‡Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN

Abstract

BACKGROUND: Detecting human papillomaviruses (HPVs) infection in cervical cells is an exceedingly important part of the clinical management of cervical dysplasia. Current guidelines in women's health outline the need for both the Papanicolaou test as well as high-risk HPV testing. Testing for HPV is expensive, is time-consuming, and requires experienced technicians. *METHODS:* Two sets of near-infrared Raman microspectroscopy experiments were conducted using a Raman confocal microscope system. First, Raman spectra were acquired from four different cell culture lines, two positive for HPV (HeLa, SiHa), one negative for HPV, but malignant (C33A), and one normal, HPV-negative line (NHEK). The three malignant lines were all derived from cervical cells. Second, Raman spectra were acquired from deidentified patient samples that were previously tested for the presence of high-risk HPV. *RESULTS:* The spectra from the cell culture lines and the patient samples contained many statistically significant differences. Using sparse multinomial logistic regression to classify the data led to classification accuracies of 89% to 97% for the cell culture samples and 98.5% for the patient samples. *CONCLUSIONS:* Raman microspectroscopy can be used to detect HPV and differentiate among specific HPV strains. This technique may provide health providers with a new method for quickly testing cell samples for the presence of HPV.

Translational Oncology (2012) 5, 172–179

Introduction

Persistent infections with sexually transmitted human papillomaviruses (HPVs) are correlated with essentially all cases of cervical cancer [1]. Of more than 150 HPV strains, a limited number of high-risk strains (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, and 69) have been found to cause cervical cancer [2,3]. Of these, two high-risk HPV strains, 16 and 18, together cause 70% of cervical cancer cases worldwide. These high-risk strains predominantly infect skin and mucosal membranes to promote epithelial proliferation, resulting in uninhibited proliferation and malignant transformations. Three oncogenes, E5, E6, and E7, are found in high-risk HPV strains and cause cell damage and abnormal cell proliferation by cooperatively interfering with the functions of cellular tumor suppressor proteins, specifically p53 and pRb [4]. Because HPV infects basal layer cells and only replicates in fully differentiated cells, it can usually remain undetected by the immune system, avoiding a humoral or cell-mediated immune response for years [5].

The Papanicolaou (Pap) test, where cells are scraped from the cervix and examined for atypical cytologic features under a microscope, was introduced in the early 1920s. Since then, clinical management of cervical dysplasia has been achieved by administering regular Pap

Address all correspondence to: Anita Mahadevan-Jansen, PhD, VU Station B Box 351631, Nashville, TN 37235. E-mail: anita.mahadevan-jansen@vanderbilt.edu

[1]The authors acknowledge the financial support of the National Institutes of Health (grant no. R01-CA-095405), a predoctoral fellowship (grant no. T32-HL7751-15), and the Lai Sulin Scholarship for E.V. The authors have no commercial affiliations or financial interests that would be considered conflicts of interest.

[2]Current address: Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831.

[3]Current address: Department of Laboratory Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY 10065.

Received 30 January 2012; Revised 30 January 2012; Accepted 1 March 2012

1944-7124/12 DOI 10.1593/tlo.12106

tests. However, Pap tests alone do not provide a complete picture of the possible malignant changes that occur in the cervix due to HPV infection. Since 2009, the American Society of Colposcopy and Cervical Pathology has recommended including HPV testing for the clinical management of cervical dysplasia in women [6]. Therefore, in the past few years, testing for all high-risk HPV strains or only strains 16 and 18 has been incorporated into routine cervical cancer screening in women older than 30 years, leading to more accurate results and a reduction in the overall number of Pap tests as women who have both a negative Pap result and a negative HPV-DNA test result are only retested every 2 to 3 years.

Current testing methods rely on obtaining a cellular sample, fixing the cells on a slide, and then transporting the slide to a pathology laboratory that runs the HPV test. After 7 to 12 days, the test results, whether the cell sample is positive or negative for a high-risk strain of HPV, are reported. No tests currently used for routine screening report the specific high-risk strain. The HPV tests developed by Qiagen, Roche, Gen-Probe, and Hologic are the most commonly used [7,8]. Cervista only tests for HPV types 16 and 18 and is approved for use in conjunction with Pap tests in women older than 30 years [9]. In cases where the Pap smear shows malignant cells and the sample is positive for high-risk HPV, the patient then needs to return for a follow-up appointment for either another Pap test or a colposcopy-guided biopsy. A diagnostic tool that can rapidly identify specific high-risk HPV strains *in vivo* or *ex vivo* would help medical providers determine appropriate follow-up treatment to diagnose cervical dysplasia and prevent its progression.

Two HPV vaccines, Gardasil and Cervarix, prevent infection from high-risk HPV strains 16 and 18 [9,10]. These vaccines are available to women ages 9 to 26 and some men only on a voluntary basis. However, these vaccines may not protect against the other high-risk HPV types, which are common in countries other than the United States [11,12]. A tool for identifying cervical samples that are positive for high-risk strains of HPV remains necessary even with the increasing use of HPV vaccines.

Optical techniques, such as drug-mediated fluorescence [13], autofluorescence [14], Fourier transform infrared (FTIR) spectroscopy [15], and Raman spectroscopy (RS) [16–18] have been used to detect cervical dysplasia. As the clinical management of cervical cancer has shifted more recently, some of these techniques have been extended to the detection of infection with HPV. Such methods can be used noninvasively, directly on the cervix either to detect dysplasia or to detect the presence of HPV in small volumes of *ex vivo* samples. The results from these studies have ranged from detecting the difference between HPV-positive and -negative samples with sensitivities and specificities ranging from 90% to 94% [13,15] and from 92% to 96% [13,15,17], respectively, to understanding the important spectral signatures that are associated with different types of HPV infection without a diagnostic application. These optical techniques do have their drawbacks, however, because fluorescence signal increases with HPV infection, regardless of strain, and FTIR is significantly hindered by water content, a significant hurdle for potential *in vivo* applications.

In this study, we use a confocal Raman microspectroscopy system to study the biochemical characteristics of various HPV strains and malignant cervical cells. RS is based on the inelastic scatter effect, where incident light is scattered from a molecular or cellular sample. This scattered light exhibits an energy shift that reports the energy of specific molecular vibrations within the sample, effectively providing a detailed biochemical composition or fingerprint. A Raman spectrum, therefore, plots the energy shift away from the incident wavelength, usually measured in relative wavenumbers, *versus* scattering intensity. This optical technique can objectively characterize samples from multiple tissues based solely on their biochemical constituents, without relying on specific chemical markers or exogenous agents [19–21].

Previously, we have acquired Raman spectra from bulk cervical tissue *in vivo* using a fiber-optic tool to detect and differentiate types of cervical dysplasia within a diverse patient population [18,22]. Recently, we have found that accounting for patient variables, such as menopausal status or previous pregnancies, can lead to an increase in disease classification accuracy [16,22]. Infection with HPV is also an important variable that has yet to be included in these analyses because not every patient is tested for HPV, and the most widely used testing standards do not differentiate between various HPV strains.

The goal of this study was to evaluate the ability of RS to detect the presence of HPV and the differences between specific HPV strains. Therefore, in this study, two sets of experiments were conducted to determine whether RS is sensitive to HPV infection. First, Raman spectra were acquired using a Raman confocal microscope from four different cell lines: HPV-16–positive SiHa cells, HPV-18–positive HeLa cells, HPV-negative but malignant C33A cells, and benign NHEK cells. Next, Raman spectra were obtained from HPV-positive and -negative patient samples. Logistic regression algorithms were then used to classify spectra into either specific strain categories for the cell culture studies or HPV-positive or -negative for the patient samples. Other research groups have performed similar tests acquiring Raman spectra from HPV-expressing cell lines [23,24]. However, these studies omitted either a positive or a negative control. Also, many of these studies only reported on the oncogenic component of the virus (i.e., E7), instead of a complete HPV infection. Furthermore, to our knowledge, no study has incorporated Raman spectra acquired from HPV-positive and HPV-negative patient samples into their analysis of the sensitivity of RS toward HPV.

Materials and Methods

Two sets of experiments were conducted to evaluate the ability of RS to differentiate between HPV strains in cell culture lines and the presence of high-risk HPV strains in patient samples.

Cell Culture and Sample Preparation

In the first set of experiments, four cell types with different HPV strains or no HPV infection (Table 1) were used. First, the two HPV-expressing cells, SiHa and HeLa (ATCC, Manassas, VA), were grown in RPMI with L-glutamate medium with 10% fetal bovine serum and antibiotics (Gibco, Carlsbad, CA). SiHa cells express

Table 1. Description of Cell Culture and Patient Samples Used in This Study.

	HPV-Positive	Transformed (Malignant)
Cell culture sample		
SiHa cell culture	Yes – HPV-16, 1-2 copies	Yes
HeLa cell culture	Yes – HPV-18, 10-50 copies	Yes
C33A cell culture	No	Yes
NHEK cell culture	No	No
Patient samples		
HPV-positive ($n_{samples}$ = 25)	Yes — one or more high-risk strains	Potentially
HPV-negative ($n_{samples}$ = 25)	Negative for high-risk strains	No

HPV-16 and HeLa cells express HPV-18, two high-risk strains. A C33A cell line (ATCC), transformed but HPV negative, was grown in Dulbecco modified Eagle medium with 10% fetal bovine serum and antibiotics. NHEK cells (a human keratinocyte cell line) were grown in EpiLife Media (Gibco) and antibiotics.

Once the cells reached approximately 90% confluence in a T-75 flask (on average, 18.2×10^6 cells), the cells were washed with PBS and trypsinized (medium was used to stop the trypsinization process). The cell mixture was then centrifuged at 300*g* for 5 minutes. After the supernatant was removed, 10 ml of PBS was added, and the mixture was centrifuged. This process was repeated three times. Next, 10 ml of sterile water was added, and the cell-water mixture was centrifuged at 300*g* for 5 to 7 minutes, also repeated three times. Finally, the supernatant was removed, and the cellular pellets were transferred onto CaF_2 slides (Crystran, Dorset, United Kingdom). The slides were allowed to air dry in a sterile environment overnight. Raman spectra were acquired the next day. This protocol was developed to minimize the effects of the different types of medium used to grow each cell type, while maximizing cell growth.

Patient Samples and Preparation

After approval from the Vanderbilt University Institutional Review Board was obtained, HPV-positive ($n_{samples}$ = 25) and -negative ($n_{samples}$ = 25) deidentified patient samples were obtained from the Molecular Infectious Diseases Laboratory at Vanderbilt University (Table 1). The samples were initially acquired for standard HPV testing, which includes cells from the cervix also used in liquid-based Pap tests. Vanderbilt University uses the digene HC2 HPV DNA test developed by Qiagen (Valencia, CA). After 8 weeks, when samples are typically discarded, we were able to acquire 50 samples. Each sample contained 20 µl, with an average of 9×10^4 cells (fewer than the cell culture samples). Each cell sample was centrifuged at 300*g* for 5 minutes. After the supernatant was removed, 10 ml of sterile water was added, and the mixture was centrifuged again. This process was repeated three times. Finally, the supernatant was removed, and the cellular pellets were transferred onto CaF_2 slides (Crystran). The slides were allowed to air dry in a sterile environment overnight. Raman spectra were acquired the next day.

Raman Microspectroscopy Measurements

Confocal Raman microspectroscopy provides a platform for acquiring detailed Raman spectra from small volumes. The RenishawInVia Raman microscope (Gloucestershire, United Kingdom; Figure 1) consists of a temperature-stabilized diode near-infrared laser with a maximum power of 120 mW (Innovative Photonics, Monmouth Junction, NJ) that operates at 785 nm and delivers ~30 mW to the sample. Light was guided through a collimator onto a series of mirrors that focused the light through an open field 50× microscope lens. Dried cell pellets were illuminated with the laser beam using a microscope objective lens (Leica N Plan 50×/0.75, Wetzlar, Germany). Scattered light from a 2- to 3-µm illuminated spot was collected with the same objective lens via a 180-degree backscatter sampling geometry. Rayleigh scattering intensity was reduced by the use of angle-tuned dielectric edge filters allowing OD6 rejection of the laser line. The inelastic (Raman) scattered light was then focused through a slit (100-µm width) and dispersed by a diffraction grating (600 grooves per millimeter) onto an air-cooled, deep-depleted CCD detector (576 × 384 pixels; each pixel is 22 × 22 µm, –70°C). Raman spectra were then processed as described in the next paragraph. This system yields a resolution of ~6 cm^{-1}, compared with the resolution of ~8 cm^{-1} found with portable probe-based systems used by this research group. For the cell culture samples, spectra were acquired with an exposure time of 30 seconds. For the patient samples,

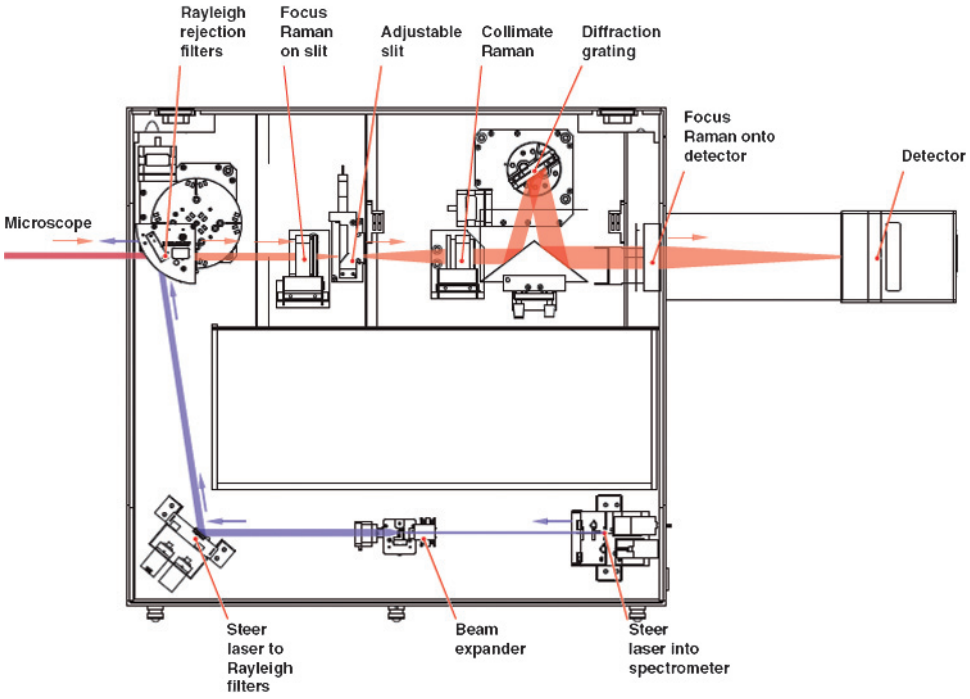

Figure 1. Schematic of Renishaw confocal Raman system. (© Renishaw)

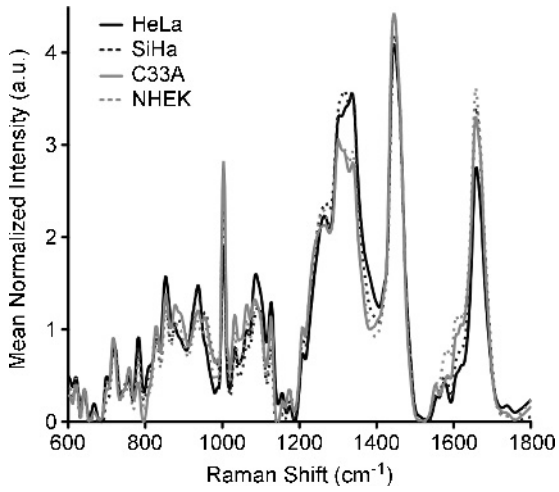

Figure 2. Spectra of HeLa, SiHa, C33A, and NHEK cell culture samples.

spectra were acquired with an exposure time of 45 to 60 seconds. Three accumulations were acquired at each acquisition point, with a binning of 3.

The spectra were processed for fluorescence subtraction and noise smoothing using the modified polynomial fit and Savitzky-Golay methods, described previously [25]. After data processing, each spectrum was normalized to its mean spectral intensity across all Raman bands to account for intensity variability.

Analysis and Classification of Raman Spectra

A Student's *t* test (unpaired two-sample, equal variance) was first conducted to find regions of significant differences among the four cell culture samples and between the two patient samples. Next, classification algorithms were used to tease out subtle differences among spectra acquired from different samples. For this study, a logistic regression method called sparse multinomial logistic regression (SMLR) was used [26]. In brief, SMLR is a Bayesian machine learning framework that computes the posterior probability of a spectrum belonging to each pathology class based on a labeled training set. For these analyses, a composite spectrum averaging Raman measurements from each cell culture or patient sample was used. A range of input parameters to SMLR was tested. The settings that provided the most accurate classification while also maximizing sparsity among the samples were a Laplacian prior, a direct kernel, a lambda value of 0.01, and no additional bias term. After every analysis, a confusion matrix that displays how each spectrum classified was produced, which can be presented as either the total number of spectra or a percentage. Also, each SMLR analysis provides a training algorithm that can be used for validation.

For the cell culture samples, three sets of SMLR analyses were used. First, HPV-positive samples (HeLa, SiHa) were compared with HPV-negative samples (C33A, NHEK). Next, malignant samples (HeLa, SiHa, C33A) were compared with the benign sample (NHEK). Finally, SMLR was used to classify each cell line. For the patient samples, HPV-positive and HPV-negative samples were

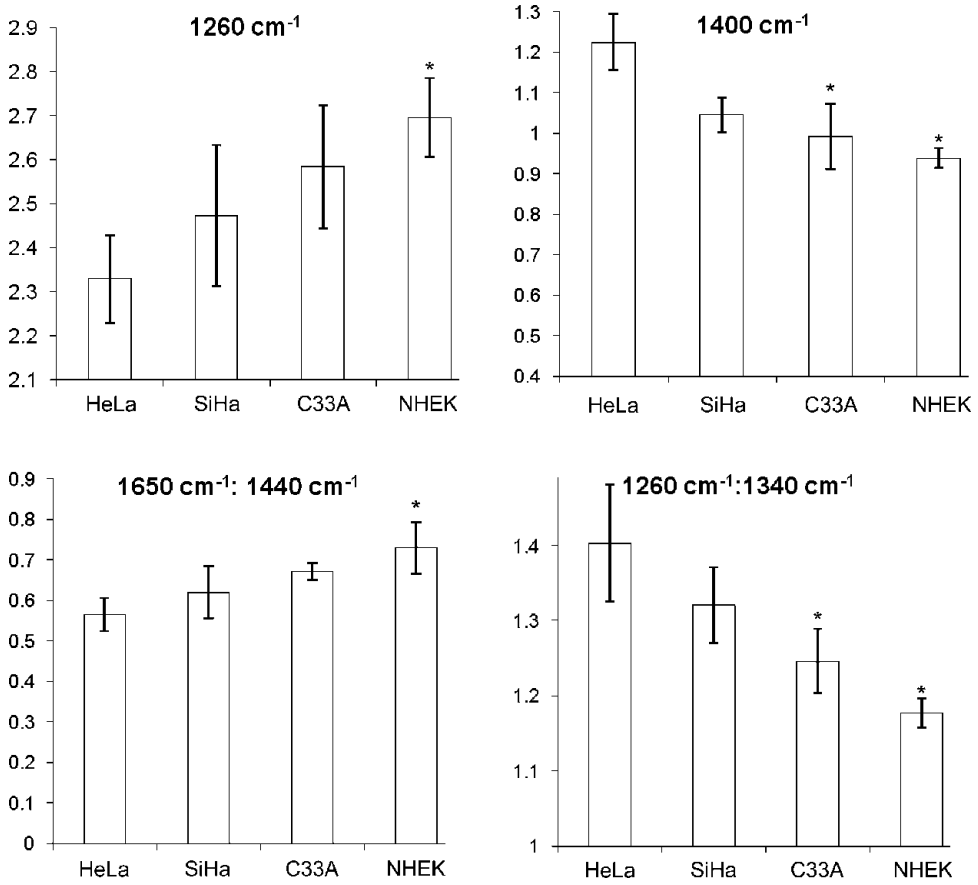

Figure 3. Specific wavenumbers and ratios of wavenumbers from spectra of cell culture samples with ±standard error. *$P < .01$ when compared to HeLa cells.

Table 2. Classification Table for HPV-Positive *Versus* HPV-Negative Cell Culture Samples.

Classification Accuracy: 97%		Raman Classification, Output of SMLR	
		HPV-Positive (HeLa, SiHa)	HPV-Negative (C33A, NHEK)
Cell culture sample type	HPV-positive (HeLa, SiHa, $n_{spectra}$ = 1365)	**98%** (1338)	2% (27)
	HPV-negative (C33A, NHEK, $n_{spectra}$ = 1390)	4% (56)	**96%** (1334)

The number of Raman spectra from each samples type is in parentheses. **Bolded** numbers along the diagonal represent the spectra that classified correctly.

Table 3. Classification Table for Malignant and Normal Cell Culture Samples.

Classification Accuracy: 92%		Raman Classification, Output of SMLR	
		Malignant (HeLa, SiHa, C33A)	Normal (NHEK)
Cell culture sample type	Malignant (HeLa, SiHa, C33A, $n_{spectra}$ = 2075)	**90%** (1868)	10% (207)
	Normal (NHEK, $n_{spectra}$ = 680)	6% (41)	**94%** (639)

The number of Raman spectra from each samples type is in parentheses. **Bolded** numbers along the diagonal represent the spectra that classified correctly.

classified using SMLR. The algorithm developed from the patient samples was then applied to classify the cell samples.

Results

The goal of this study was to determine whether the differences between HPV-positive and -negative samples, malignant and benign samples, and specific HPV strains from cell culture and patient samples can be detected using RS (Table 1). Approximately 30 Raman spectra were obtained from 15 independent samples, consisting of at least 1×10^4 cells per sample. The power of the cell culture experiments is greater than 90% (two-sided statistical test, $\alpha = 0.05$) with an effect size of 1; the power of the human sample study is greater than 80% under the same criteria. Figure 2 displays the Raman spectra acquired with the Raman microscopy system from four cell culture samples: HeLa, SiHa, C33A, and NHEK. Spectra acquired from the fixed HeLa cells seem to be the most different visually across the range of the spectrum, specifically around 990, 1080 to 1160, around 1400, and around 1670 cm^{-1}. These areas have been shown in previous work to correspond to C–C stretching in lipids and proteins, DNA content, and CH_2 deformation [27].

In Figure 3, peaks and peak ratios of the spectra in Figure 2 are displayed to provide an in-depth view of the changes that occur in the spectra from different cell lines. Figure 3 (*A* and *B*) demonstrates that concentrations of molecules vary depending on cell type, HPV strain, and the number of HPV copies. Specifically, wavenumber 1260 has been tentatively assigned to the deformation of CH_3 bonds and amide III [28]. Wavenumber 1400 cm^{-1} has been shown to correspond to DNA and RNA components, such as uracil and adenine [27]. Looking at the changes among ratios of wavenumbers is another method commonly used to interpret Raman spectra. Figure 3*C* shows the ratio of 1650 to 1440 cm^{-1}, which is one way of describing the intensity of the CH_2 bending and the intensity of the C=C stretch [29]. In these samples, this ratio has a negative linear dependence on HPV infection and malignancy, leading to a decrease in the ratio. In Figure 3*D*, the ratio of 1260 cm^{-1} to 1340 cm^{-1} or the ratio of Amide III to amino acids is displayed [30]. In this case, as the number of copies of HPV increases in each sample (1-2 in SiHa cells, 10-50 in HeLa cells), this ratio increases, resulting in a linear correlation between HPV infection and this ratio [31]. A Student's *t* test was also performed to compare the differences in the peak intensities to HeLa cells. $P < .01$ was defined as significant.

The first step of this analysis to determine the clinical significance of this technique was to use SMLR to classify the cell culture samples as HPV-positive or -negative (Table 2). Table 2 shows that the presence of HPV is correctly identified using RS with a classification accuracy of 97%. Next, the analyses classified the spectra as malignant or benign (Table 3). Combining the malignant cell lines for comparison to the NHEK benign cell line resulted in a lower classification accuracy of 92%, with 10% of the Raman spectra obtained from malignant cells classifying as normal. Finally, SMLR was used to classify the spectra from the four different cell culture types with an accuracy of 89% (Table 4). The HeLa samples had the highest classification accuracy (98%), whereas the HPV-negative, malignant cells (C33A) had the lowest (84%) with 8% of its spectra classifying as normal.

Figure 4 shows the spectra acquired from patient samples that were tested for the presence of high-risk strands for HPV. Spectra from these two samples vary significantly across the 600- to 1800-cm^{-1} range, in many of the same regions listed above, but also in areas corresponding to both Amide III and Amide I (1200-1300 and 1660 cm^{-1}, respectively) [30]. *P* values less than .001, resulting from a Student's *t* test, are also shown on this graph, demonstrating that many regions of the spectra obtained from HPV-positive and -negative samples are significantly different.

The classification algorithm SMLR was also used to classify patient samples as HPV positive or negative, and the results were compared with the HPV test result. These spectra from patient samples were classified with an accuracy of 98.5% (Table 5). The training algorithm that resulted from using SMLR on the patient samples was then applied to both the malignant and the benign classifications (data not shown) and the classification of the four different cell culture lines. When the classification algorithm derived from the patient data was used to classify malignant and normal samples, the classification accuracy increased from 92% to 93%. For classifying the four cell culture lines, using the patient sample algorithm instead of the cell culture algorithm resulted in an increase in classification accuracy from 89% to 93% (Table 6). The spectra acquired from HeLa and SiHa cells maintained their previous classification accuracies, but the C33A and NHEK spectra increased their classification accuracies by 5% and 1%, respectively.

Discussion

The results of this study demonstrate that RS is able to differentiate between various cell culture samples as well as patient samples, based on the presence of HPV alone and the specific HPV strain. These results may positively contribute to screening for cervical disease by determining if cervical samples are positive for high-risk strains of HPV.

In Figure 2, there are certain regions that are more variable, including those corresponding to C=C stretching in lipids and

Table 4. Classification Table for All Four Cell Culture Samples (HeLa, SiHa, C33A, and NHEK).

Classification Accuracy: 89%		Raman Classification, Output of SMLR			
		HeLa	SiHa	C33A	NHEK
Cell culture sample type	HeLa ($n_{spectra}$ = 675)	**98%** (661)	1% (7)	1% (7)	0% (0)
	SiHa ($n_{spectra}$ = 690)	2% (14)	**96%** (662)	1% (7)	1% (7)
	C33A ($n_{spectra}$ = 710)	3% (21)	5% (36)	**84%** (596)	8% (57)
	NHEK ($n_{spectra}$ = 680)	1% (7)	2% (14)	9% (61)	**88%** (598)

The number of Raman spectra from each samples type is in parentheses. **Bolded** numbers along the diagonal represent the spectra that classified correctly.

proteins, DNA content, and CH_2 deformation. Also, the SiHa, C33A, and NHEK spectra seem to resemble each other more than the HeLa spectra. HeLa cells have the most copies of HPV, which may result in a more malignant cell line that has a higher concentration of DNA than the other cells types. The deformation of both proteins and lipids, found within cells and the phospholipid membrane, is observed in the Raman spectra acquired from more advanced malignant cell lines and those with a higher number of copies of high-risk strains of HPV [31]. These changes may be associated with the increased disorganization caused by HPV infection and more advanced malignancy.

When smaller regions of the spectra are considered, differences can be observed among the cell culture samples. These differences may be related to HPV infection, the number of copies of HPV, and the transformation of the cell line. The two ratios that have been graphed (Figure 3, *C* and *D*) demonstrate some of the biochemical changes that occur as cells with and without HPV infection advance malignantly. The 1650-to-1440-cm^{-1} ratio demonstrates that, as the cell develops increasingly malignant characteristics and as the number of HPV copies increases, the ratio of CH_3 bending to C=C stretching decreases. This result suggests an increase in disorganization occurring at a cellular level. The 1260-to-1340-cm^{-1} ratio corresponds to Amide III and the amount of amino acids (Figure 3*D*), suggesting that there are slight increases in the amount of protein as the cells become malignant and infected with HPV. It is also interesting to note that, although the C33A cell line is a transformed, malignant cell line, it mostly resembles the spectra obtained from the normal

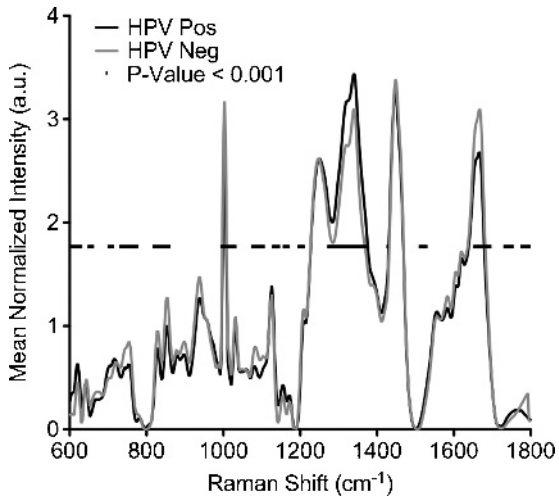

Figure 4. Spectra of HPV-positive *versus* HPV-negative patient samples. Black line represents regions of significant difference ($P < .001$) when the two samples are compared.

Table 5. Classification Table for HPV-Positive and -Negative Patient Samples.

Classification Accuracy: 98.5%		Raman Classification, Output of SMLR	
		HPV-Positive	HPV-Negative
Pathologic diagnosis	HPV-positive ($n_{spectra}$ = 840)	**98% (823)**	2% (17)
	HPV-negative ($n_{spectra}$ = 875)	1% (9)	**99% (866)**

The number of Raman spectra from each samples type is in parentheses. **Bolded** numbers along the diagonal represent the spectra that classified correctly.

NHEK cell line. This result suggests that infection with HPV is the more pertinent feature affecting Raman spectra.

The spectra obtained from patient samples seem to be dramatically different based only on infection with high-risk strains of HPV (Figure 4), verifying the hypothesis that RS is sensitive to HPV infection. The differences between the spectra are seen in multiple regions, such as those corresponding to lipid, amino acid, and DNA content, as well as CH stretching and bending regions that have been assigned to proteins such as albumin and collagen [31]. Many of these differences correlate with changes that occur in patient samples that are positive for at least one high-risk HPV strain. Specifically, an increased DNA content and density is found with an increased amount of phosphate and an increasing amount of disorganization, seen throughout the spectra as the deformation and breakdown of CH_x bonds [31]. Furthermore, the concentration of lipids (tentatively assigned to 1450 cm^{-1}) seems to decrease as the number of copies of HPV increases.

Developing SMLR algorithms based on the spectra from the cell culture samples led to classification accuracy rates from 89% to 97%. Spectra of HPV-positive and HPV-negative cells were classified with an accuracy of 97% (Table 2). As discussed, HPV infection leads to modifications within the cellular environment by circumventing normal cell growth pathways, specifically by increasing DNA synthesis. This phenomenon can be observed in the spectra because the density or organization of DNA is lower in HPV-positive cells because the DNA is transcriptionally active, as opposed to densely packed and quiescent [32]. Next, SMLR differentiated between three malignant cell lines and one benign cell line with a classification accuracy of 92% (Table 3). This result is lower than the result from discriminating the HPV-positive and -negative cell lines, most likely due to the addition of the HPV-negative cell line (C33A) into the malignant category. C33A is a transformed, malignant epithelial cervical cell line; however, its spectra is closer to the spectra obtained from NHEK cells than the HPV-positive cell lines. Therefore, it is not surprising that 10% of the spectra obtained from C33A samples were misclassified. Clinically, a patient having malignant cervical cells without the presence of HPV is a very rare occurrence, accounting for approximately 0.01% of all cases of cervical dysplasia [33].

Table 6. Classification Table for Cell Culture Samples (HeLa, SiHa, C33A, and NHEK) Using the Algorithm Derived from Patient Samples (used in Table 5).

Classification Accuracy: 93%		Raman Classification, Output of SMLR			
		HeLa	SiHa	C33A	NHEK
Cell culture sample type	HeLa ($n_{spectra}$ = 675)	**98% (661)**	1% (7)	1% (7)	0% (0)
	SiHa ($n_{spectra}$ = 690)	2% (14)	**96% (662)**	1% (7)	1% (7)
	C33A ($n_{spectra}$ = 710)	2% (14)	4% (28)	**89% (632)**	5% (36)
	NHEK ($n_{spectra}$ = 680)	1% (7)	2% (14)	8% (54)	**89% (605)**

The number of Raman spectra from each samples type is in parentheses. **Bolded** numbers along the diagonal represent the spectra that classified correctly.

Management of such patients, therefore, would depend solely on the results of their Pap tests.

The lowest classification accuracy from the cell culture study was achieved when all four cell types were classified independently, resulting in an accuracy of 89% (Table 4). The normal NHEK cells classified correctly only 89% of the time, perhaps because of the similarities between the NHEK and C33A spectra from the lack of HPV infection, as discussed above. It is also important to note that SMLR was able to correctly classify greater than 95% of the spectra obtained from HeLa and SiHa cells. The difference between these cells is likely due to the type of HPV (18 *vs* 16) infection as well as the number of copies of HPV in each line (estimated 10-50 copies of HPV 18 in HeLa cells *vs* 1-2 copies of HPV 16 in SiHa cells).

HPV-positive and -negative patient samples were also classified using SMLR with an accuracy of 98.5% (Table 5). This result corresponds with the differences observed between the two spectra, quantified by the low *P* values across the spectral range. Similar to the cell culture studies, classifying patient samples based on the presence of HPV, as opposed to HPV type or copy number, leads to extremely promising results. The algorithm generated by SMLR based on the patient data set was then applied to previous data obtained from the cell culture samples to determine if the differences between the patient samples were more representative of the differences between HPV-positive and -negative cell lines. This algorithm was applied to the sets of the malignant and benign samples (data not shown) and the four different cell lines (Table 6). Both sets of data were classified with a higher accuracy when the patient-data algorithm was used. More specifically, using the patient-data algorithm led to an increase in the classification of both the C33A and NHEK cells by decreasing their incorrect assignments to each other (i.e., when C33A spectra classified as NHEK spectra), implying that the algorithm developed from the patient data contains more accurate information about HPV-negative cells. It is also interesting to note that only a few cells in a patient sample needed to be infected with high-risk HPV strands to be considered HPV positive. On the other hand, it is assumed that all the cells from HPV-negative samples were HPV negative. Future work will determine the detection limit of our RS system for detecting patient samples that are positive and negative for HPV in comparison to current HPV tests.

A few studies by other groups have shown that the presence of the biomarker protein $p16^{INK4A}$ has a great influence on cells infected with HPV [34–36]. $p16^{INK4A}$ is a minichromosome maintenance protein found normally in cervical cells [33]. This protein is one of many that regulate the level of active cyclin D/CDK, part of the feedback loop involved in maintaining levels of minichromosome maintenance proteins, proliferating cell nuclear antigens, and cyclin E. $p16^{INK4A}$ is one of the biomarkers that correspond to elevated E7 expression and, therefore, to high-risk HPV infection. Overexpression of $p16^{INK4A}$ has been detected in all grades of cervical lesions [34,36]. Results from FTIR and confocal fluorescence microscopy studies on $p16^{INK4A}$ has shown that overexpression of $p16^{INK4A}$ and HPV infection result in an increase in nucleic acid levels, a decrease in lipid levels, and a moderate to low change in protein levels [35]. The Raman spectra correspond to these results, showing a biochemical increase in nucleic acid levels and a decrease in lipid levels (Figure 3). Current studies are focused on correlating the presence of biomarkers with HPV-positive or -negative status within the Raman data.

Recent studies have shown that combining HPV testing with the Pap test is more effective than the Pap test alone at detecting clinically relevant high-grade dysplasia early, and results led to increased prevention against more aggressive forms of cervical dysplasia in women older than 30 years [37]. In fact, dual screening detected 25% more potentially cancerous lesions than a Pap test alone and resulted in fewer cancer cases 5 years later compared with patients who were only screened with a Pap test. Whereas previous work shows that RS is capable of detecting dysplastic areas of the cervix [16–18], this research demonstrates that RS is capable of detecting HPV-infected cells as well. Therefore, RS could potentially be used as an alternative method to Pap tests and HPV screening to detect abnormal areas of the cervix and the presence of HPV *in vivo* and in real time. Although *in vivo* samples were not used in this study, because HPV testing is now recommended for women older than 30 years, *in vivo* Raman measurements can be taken concurrently with HPV DNA testing. This work is being pursued currently.

Results from this article demonstrate that RS is sensitive to changes occurring in cells due to HPV infection, HPV type, and the number of copies of HPV. This technique can be combined with current methods used to screen for cervical dysplasia to provide a tool that can detect the presence of HPV immediately without the need for extensive sample preparation, quickly leading to accurate results. Current studies within our laboratory are focused on using the same technique to study the differences between low-risk and high-risk HPV strains. HPV testing is being introduced in conjunction with our ongoing *in vivo* Raman study for cervical dysplasia detection to obtain *in vivo* measurements with the presence of HPV.

Acknowledgments

Special thanks go to Isaac Pence and Cat Majors for their help with the upkeep of the cell culture, to Criziel Quinn for preparing the patient samples, to Amy Rudin for proofreading this article, and to Ninell Mortensen for help with plotting the figures.

References

[1] Coleman DV, Wickenden C, and Malcolm ADB (1986). Association of human papillomavirus with squamous carcinoma of the uterine cervix. *Ciba Found Symp* **120**, 175–189.

[2] Wright T, Kurman RJ, and Ferenczy A (1994). *Blaustien's Pathology of the Female Genital Tract*. Springer-Verlag, New York, NY.

[3] Rowson KE and Mahy BW (1967). Human papova (wart) virus. *Bacteriol Rev* **31**, 110–131.

[4] Brokaw JL, Yee CL, and Munger K (1994). A mutational analysis of the amino terminal domain of the human papillomavirus type 16 E7 oncoprotein. *Virology* **205**, 603–607.

[5] Walboomers JMM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJF, Peto J, Meijer CJLM, and Munoz N (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* **189**, 12–19.

[6] Cox JT (2003). The clinician's view: role of human papillomavirus testing in the American Society for Colposcopy and Cervical Pathology Guidelines for the management of abnormal cervical cytology and cervical cancer precursors. *Arch Pathol Lab Med* **127**, 950–958.

[7] Hesselink AT, van Ham MA, Heideman DA, Groothuismink ZM, Rozendaal L, Berkhof J, van Kemenade FJ, Massuger LA, Melchers WJ, Meijer CJ, et al. (2008). Comparison of $GP5^+/6^+$-PCR and SPF10-line blot assays for detection of high-risk human papillomavirus in samples from women with normal cytology results who develop grade 3 cervical intraepithelial neoplasia. *J Clin Microbiol* **46**, 3215–3221.

[8] Huh W, Einstein MH, Herzog TJ, and Franco EL (2010). What is the role of HPV typing in the United States now and in the next five years in a vaccinated population? *Gynecol Oncol* **117**, 481–485.

[9] Khan MJ, Castle PE, Lorincz AT, Wacholder S, Sherman M, Scott DR, Rush BB, Glass AG, and Schiffman M (2005). The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or

18 and the possible utility of type-specific HPV testing in clinical practice. *J Natl Cancer Inst* **97**, 1072–1079.

[10] Siddiqui MAA and Perry CM (2006). Human papillomavirus quadrivalent (types 6, 11, 16, 18) recombinant vaccine (Gardasil). *Drugs* **66**, 1263–1271.

[11] American Cancer Society (2007). *Cervical Cancer Resource Center*. American Cancer Society, Atlanta, GA.

[12] Au WW, Abdou-Salama S, Sierra-Torres CH, and Al-Hendy A (2007). Environmental risk factors for prevention and molecular intervention of cervical cancer. *Int J Hyg Environ Health* **210**, 671–678.

[13] Schneede P, Munch P, Wagner S, Meyer T, Stockfleth E, and Hofstetter A (2001). Fluorescence urethroscopy following instillation of 5-aminolevulinic acid: a new procedure for detecting clinical and subclinical HPV lesions of the urethra. *J Eur Acad Dermatol* **15**, 121–125.

[14] Ramanujam N, Mitchell MF, Mahadevan A, Warren S, Thomsen S, Silva E, and Richards-Kortum R (1994). *In vivo* diagnosis of cervical intraepithelial neoplasia using 337-nm-excited laser-induced fluorescence. *Proc Natl Acad Sci USA* **91**, 10193.

[15] Cohenford MA and Rigas B (1998). Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: implications for tumor biology. *Proc Natl Acad Sci USA* **95**, 15327–15332.

[16] Kanter EM, Majumder S, Vargis E, Robichaux-Viehoever A, Kanter GJ, Shappell H, Jones HW III, and Mahadevan-Jansen A (2009). Multiclass discrimination of cervical precancers using Raman spectroscopy. *J Raman Spectrosc* **40**, 205–211.

[17] Kanter EM, Vargis E, Majumder S, Keller MD, Woeste E, Rao GG, and Mahadevan-Jansen A (2009). Application of Raman spectroscopy for cervical dysplasia diagnosis. *J Biophotonics* **2**, 81–90.

[18] Vargis E, Kanter EM, Majumder SK, Keller MD, Beaven RB, Rao GG, and Mahadevan-Jansen A (2011). Effect of normal variations on disease classification of Raman spectra from cervical tissue. *Analyst* **136**, 2981–2987.

[19] Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, and Feld MS (2005). Diagnosing breast cancer by using Raman spectroscopy. *Proc Natl Acad Sci USA* **102**, 12371–12376.

[20] Chrit L, Bastien P, Sockalingum GD, Batisse D, Leroy F, Manfait M, and Hadjur C (2006). An *in vivo* randomized study of human skin moisturization by a new confocal Raman fiber-optic microprobe: assessment of a glycerol-based hydration cream. *Skin Pharmacol Physiol* **19**, 207–215.

[21] Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, Yeoh KG, So JB, and Huang Z (2011). Characterizing variability in *in vivo* Raman spectra of different anatomical locations in the upper gastrointestinal tract toward cancer detection. *J Biomed Opt* **16**, 037003.

[22] Vargis E, Byrd T, Logan Q, Khabele D, and Mahadevan-Jansen A (2011). Sensitivity of Raman spectroscopy to normal patient variability. *J Biomed Opt* **16**, 117004.

[23] Jess PR, Smith DD, Mazilu M, Dholakia K, Riches AC, and Herrington CS (2007). Early detection of cervical neoplasia by Raman spectroscopy. *Int J Cancer* **121**, 2723–2728.

[24] Ostrowska KM, Malkin A, Meade A, O'Leary J, Martin C, Spillane C, Byrne HJ, and Lyng FM (2010). Investigation of the influence of high-risk human papillomavirus on the biochemical composition of cervical cancer cells using vibrational spectroscopy. *Analyst* **135**, 3087–3093.

[25] Lieber CA and Mahadevan-Jansen A (2003). Automated method for subtraction of fluorescence from biological Raman spectra. *Appl Spectrosc* **57**, 1363–1367.

[26] Krishnapuram B, Carin L, Figueiredo MA, and Hartemink AJ (2005). Sparse multinomial logistic regression: fast algorithms and generalization bounds. *IEEE Trans Pattern Anal Mach Intell* **27**, 957–968.

[27] Utzinger U, Heintzelman DL, Mahadevan-Jansen A, Malpica A, Follen M, and Richards-Kortum R (2001). Near-infrared Raman spectroscopy for *in vivo* detection of cervical precancers. *Appl Spectrosc* **55**, 955–959.

[28] Miller FA, Mayo DW, and Hannah RW (2004). *Course Notes on the Interpretation of Infrared and Raman Spectra*. Wiley-Interscience, Hoboken, NJ.

[29] Stone N, Kendall C, Shepherd N, Crow P, and Barr H (2002). Near-infrared Raman spectroscopy for the classification of epithelial precancers and cancers. *J Raman Spectrosc* **33**, 564–573.

[30] Williams RW (1986). Protein secondary structure analysis using Raman Amide I and Amide III spectra. *Method Enzymol* **130**, 311–331.

[31] Meissner JD (1999). Nucleotide sequences and further characterization of human papillomavirus DNA present in the CaSki, SiHa and HeLa cervical carcinoma cell lines. *J Gen Virol* **80**, 1725–1733.

[32] Arends MJ, Buckley CH, and Wells M (1998). Aetiology, pathogenesis, and pathology of cervical neoplasia. *J Clin Pathol* **51**, 96–103.

[33] Martin CM, Kehoe L, Spillane CO, and O'Leary JJ (2007). Gene discovery in cervical cancer: towards diagnostic and therapeutic biomarkers. *Mol Diagn Ther* **11**, 277–290.

[34] Kalof AN, Evans MF, Simmons-Arnold L, Beatty BG, and Cooper K (2005). $p16^{INK4A}$ immunoexpression and HPV *in situ* hybridization signal patterns: potential markers of high-grade cervical intraepithelial neoplasia. *Am J Surg Pathol* **29**, 674–679.

[35] Ostrowska KM, Garcia A, Meade AD, Malkin A, Okewumi I, O'Leary JJ, Martin C, Byrne HJ, and Lyng FM (2011). Correlation of p16(INK4A) expression and HPV copy number with cellular FTIR spectroscopic signatures of cervical cancer cells. *Analyst* **136**, 1365–1373.

[36] Dray M, Russell P, Dalrymple C, Wallman N, Angus G, Leong A, Carter J, and Cheerala B (2005). p16(INK4a) as a complementary marker of high-grade intraepithelial lesions of the uterine cervix. I: Experience with squamous lesions in 189 consecutive cervical biopsies. *Pathology* **37**, 112–124.

[37] Rijkaart DC, Berkhof J, Rozendaal L, van Kemenade FJ, Bulkmans NW, Heideman DA, Kenter GG, Cuzick J, Snijders PJ, and Meijer CJ (2012). Human papillomavirus testing for the detection of high-grade cervical intraepithelial neoplasia and cancer: final results of the POBASCAM randomised controlled trial. *Lancet Oncol* **13**, 78–88.