

2021

## Raman Spectroscopy of Liquid-Based Cervical Smear Samples as a Triage to Stratify Women Who Are HPV-Positive on Screening

Damian Gordon

*Technological University Dublin, damien.traynor@mydit.ie*

Cara Martin

*Trinity College Dublin*

Christine White

*Trinity College Dublin*

*See next page for additional authors*

Follow this and additional works at: <https://arrow.tudublin.ie/aegart>

 Part of the [Medicine and Health Sciences Commons](#)

### Recommended Citation

Traynor, D., Martin, C.M., & White, C. (2021). Raman Spectroscopy of Liquid-Based Cervical Smear Samples as a Triage to Stratify Women Who Are HPV-Positive on Screening. *Cancers*, vol. 13, no. 9. doi:10.3390/cancers13092008

This Article is brought to you for free and open access by the Applied Electrochemistry Group at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [aisling.coyne@tudublin.ie](mailto:aisling.coyne@tudublin.ie), [gerard.connolly@tudublin.ie](mailto:gerard.connolly@tudublin.ie).



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](#)

---

**Authors**

Damian Gordon, Cara Martin, Christine White, Stephen Reynolds, Tom D'Arcy, John J. O'Leary, and Fiona Lyng

## Article

# Raman Spectroscopy of Liquid-Based Cervical Smear Samples as a Triage to Stratify Women Who Are HPV-Positive on Screening

Damien Traynor<sup>1,2</sup>, Cara M. Martin<sup>3,4,5</sup>, Christine White<sup>3,4</sup>, Stephen Reynolds<sup>3,4</sup>, Tom D'Arcy<sup>6</sup>, John J. O'Leary<sup>3,4,5</sup> and Fiona M. Lyng<sup>1,2,\*</sup> 

- <sup>1</sup> Centre for Radiation and Environmental Science, FOCAS Research Institute, Technological University Dublin, Kevin St, D08 NF82 Dublin, Ireland; damien.traynor@tudublin.ie
- <sup>2</sup> School of Physics & Clinical & Optometric Sciences, City Campus, Technological University Dublin, Kevin Street, D08 NF82 Dublin, Ireland
- <sup>3</sup> Discipline of Histopathology, University of Dublin Trinity College, D08 NHY1 Dublin, Ireland; cara.martin@tcd.ie (C.M.M.); chwhite@tcd.ie (C.W.); stephen.reynolds@tcd.ie (S.R.); olearyjj@tcd.ie (J.J.O.)
- <sup>4</sup> CERVIVA Molecular Pathology Research Laboratory, The Coombe Women and Infants University Hospital, D08 XW7X Dublin, Ireland
- <sup>5</sup> The Trinity St. James's Cancer Institute, D08 NHY1 Dublin, Ireland
- <sup>6</sup> Department of Obstetrics and Gynaecology, Coombe Women and Infants University Hospital, D08 XW7X Dublin, Ireland; tomдарcy@coombe.ie
- \* Correspondence: fiona.lyng@tudublin.ie



**Citation:** Traynor, D.; Martin, C.M.; White, C.; Reynolds, S.; D'Arcy, T.; O'Leary, J.J.; Lyng, F.M. Raman Spectroscopy of Liquid-Based Cervical Smear Samples as a Triage to Stratify Women Who Are HPV-Positive on Screening. *Cancers* **2021**, *13*, 2008. <https://doi.org/10.3390/cancers13092008>

Academic Editor: Ruud L. M. Bekkers

Received: 11 February 2021

Accepted: 19 April 2021

Published: 22 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** Persistent high-risk human papillomavirus (HPV) infection can lead to cervical precancer and cancer. Recently, HPV testing has been introduced for primary cervical screening, but the HPV DNA test cannot distinguish between transient and persistent HPV infection. Thus, there is an unmet clinical need to develop a new test to identify women with a high-risk persistent HPV infection. Raman spectra were recorded from cervical smear samples ( $n = 60$ ) and, on the basis of HPV DNA and HPV mRNA test results, a classifier was developed to identify persistent HPV infection. A further blinded independent test set ( $n = 14$ ) was used to validate the model, and sensitivity of 90% and specificity of 100% were achieved. Improved triage would allow women with a high-risk persistent HPV infection to be referred for immediate treatment, while women with a low-risk transient infection could avoid overtreatment.

**Abstract:** The role of persistent high-risk human papillomavirus (HPV) infection in the development of cervical precancer and cancer is now well accepted, and HPV testing has recently been introduced for primary cervical screening. However, the low specificity of HPV DNA testing can result in large numbers of women with an HPV-positive result, and additional triage approaches are needed to avoid over-referral to colposcopy and overtreatment. The aim of this study was to assess Raman spectroscopy as a potential triage test to discriminate between transient and persistent HPV infection. HPV DNA status and mRNA status were confirmed in ThinPrep<sup>®</sup> cervical samples ( $n = 60$ ) using the Cobas 4800 and APTIMA HPV test, respectively. Raman spectra were recorded from single-cell nuclei and subjected to partial least squares discriminant analysis (PLSDA). In addition, the PLSDA classification model was validated using a blinded independent test set ( $n = 14$ ). Sensitivity of 85% and specificity of 92% were achieved for the classification of transient and persistent HPV infection, and this increased to 90% sensitivity and 100% specificity when mean sample spectra were used instead of individual cellular spectra. This study showed that Raman spectroscopy has potential as a triage test for HPV-positive women to identify persistent HPV infection.

**Keywords:** Raman spectroscopy; cytology; HPV; biomarkers; ThinPrep; exfoliated cells; cervical cancer; cervical precancer

## 1. Introduction

Cervical cancer ranks fourth for both incidence and mortality with an estimated 570,000 cases and 311,000 deaths worldwide in 2018 [1]. In order to achieve the World Health Organization's call to action to eliminate cervical cancer, improved screening and early detection are required [2,3].

Persistent infection with high-risk human papillomavirus (HPV) is accepted as the major cause of the development of cervical precancer and cancer [4]. Over 100 different types of HPV have been identified, and 14 are considered as high-risk HPV types (hrHPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68) [5]. HPV is a common sexually transmitted infection, and most HPV infections are transient and are resolved in 1–2 years [6]. Persistent hrHPV infection, however, followed by integration of the HPV genome into the host chromosomes, results in production of E6/E7 messenger RNA (mRNA) transcripts and E6/E7 oncoproteins causing cell-cycle deregulation, which can lead to precancer (cervical intraepithelial neoplasia (CIN)) [7]. This persistent form of HPV infection can be termed a transcriptionally active HPV infection.

HPV testing has recently replaced cytology as the primary cervical screening method in many countries, including Ireland [8]. HPV testing has a higher sensitivity than cytology for detection of high-grade cervical precancer (CIN2/3). However, HPV DNA testing is limited by its lower specificity to cytology; thus, additional approaches are required to triage HPV DNA-positive women to avoid over-detection of clinically insignificant, transient HPV infections as these can result in over-referral to colposcopy and overtreatment [9].

Cytology-based triage, with and without HPV16/18 genotyping, is generally used to identify women with low-grade squamous intraepithelial lesions (LSIL) and high-grade squamous intraepithelial lesions (HSIL). However, cytology alone has a sensitivity of only 59–70% for detection of HSIL due to sampling, technical, and/or inter-observer errors mainly associated with the subjectivity of the morphological assessment [10].

Thus, there is an unmet need for additional triage tests that can identify disease-specific biomarkers. A number of molecular biomarkers have recently been investigated to detect transforming HPV infections including E6/E7 mRNA, methylation markers, miRNAs, and protein biomarkers, such as p16INK4a, Ki67, minichromosome maintenance protein 2 (MCM), and topoisomerases type IIa (TOP2A) [11].

As an alternative approach to single molecular biomarkers, Raman spectroscopy can provide a biochemical fingerprint of a cell or tissue [12,13] and, thus, offers an opportunity to capture molecular changes in one test. Many studies have shown the potential of Raman spectroscopy for noninvasive, objective detection of cervical cancer and precancer in cells and tissues [14–30], but there are limited studies on detection of HPV infection.

Jess et al. [31] showed that Raman microspectroscopy could distinguish between primary human keratinocytes (PHK) and PHK cells expressing the E7 gene of HPV16 and between PHK cells and cervical cancer cells expressing HPV16 (CaSki). Both infrared absorption and Raman spectroscopy were shown to distinguish between an HPV-negative cervical cancer cell line (C33a) and cervical cancer cell lines with different HPV copy number (SiHa, HeLa, and CaSki) as a function of the spectral profile of the cells [32]. Vargis et al. [16] showed discrimination between different HPV-negative and HPV-positive cell lines and cytology samples using Raman spectroscopy. More recently, biochemical changes due to HPV infection in HSIL cells were shown to be more marked than biochemical changes due to the menstrual cycle or the use of hormonal contraceptives [29].

To our knowledge, no studies to date have investigated the potential of Raman spectroscopy for detection of transcriptionally active HPV infection. Thus, the main aim of this study was to assess Raman spectroscopy as a potential triage test for HPV-positive cervical samples. Raman spectra were recorded from liquid-based cytology (LBC) samples prepared as ThinPrep slides ( $n = 60$ ); then, on the basis of HPV DNA and HPV mRNA test results, a classification model was developed to identify transcriptionally active HPV infections. A further blinded independent test set ( $n = 14$ ), which was completely separate

from the training set, was utilized to validate the model, and the results were compared to cytology, histopathology, and HPV DNA and mRNA test results.

## 2. Materials and Methods

### 2.1. Sample Collection

Residual material from cervical smear samples collected in PreservCyt solution was obtained from women attending the colposcopy clinic at the Coombe Women and Infants University Hospital (CWIUH), Dublin, Ireland, for a routine colposcopy visit. Women gave written informed consent for their sample to be used for this research study. Ethical approval was obtained from the CWIUH Research Ethics Committee (Study No. 28-2014). A subset of samples was obtained from the residual cervical smear samples from women enrolled in the CERVIVA HPV Primary Screening Trial. Ethical approval for this was obtained from the ICGP Research Ethics Committee. This study is part of a larger study underway within CERVIVA and, thus, samples were specifically selected that were HPV DNA/mRNA-positive or HPV DNA-positive only. A total of 60 cervical LBC samples were used for this study to develop the classification model (Table 1). HPV mRNA-positive samples represented the transcriptionally active form of infection ( $n = 30$ ) and HPV DNA-positive mRNA-negative samples represented the non-transcriptionally active form ( $n = 30$ ).

**Table 1.** Summary of training set of samples, including cytology, HPV DNA, and HPV mRNA test results.

Scheme	Cytology	HPV DNA	HPV mRNA	Transcriptionally Active HPV Infection
1–30	HSIL	positive	negative	no
31–60	HSIL	positive	positive	yes

HSIL: high-grade squamous intraepithelial lesion.

HPV DNA status was confirmed using the HPV Cobas 4800 test which detects the presence of hrHPV DNA. This test detects HPV16 and HPV18 separately, as well as a pool of 12 other high-risk types (HPV 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68).

HPV mRNA status was confirmed using the Aptima HPV Assay. The Aptima HPV assay is a target amplification nucleic acid probe test for the in vitro qualitative detection of E6/E7 viral messenger RNA (mRNA) from 14 high-risk types of HPV (16/18/31/33/35/39/45/51/52/56/58/59/66/68). The Aptima HPV assay does not discriminate between the 14 high-risk types.

A further set of blinded samples ( $n = 14$ ), selected as above but completely separate from the training set, was used as a test set to validate the classification model and consisted of a mixture of non-transcriptionally active and transcriptionally active HPV infected samples.

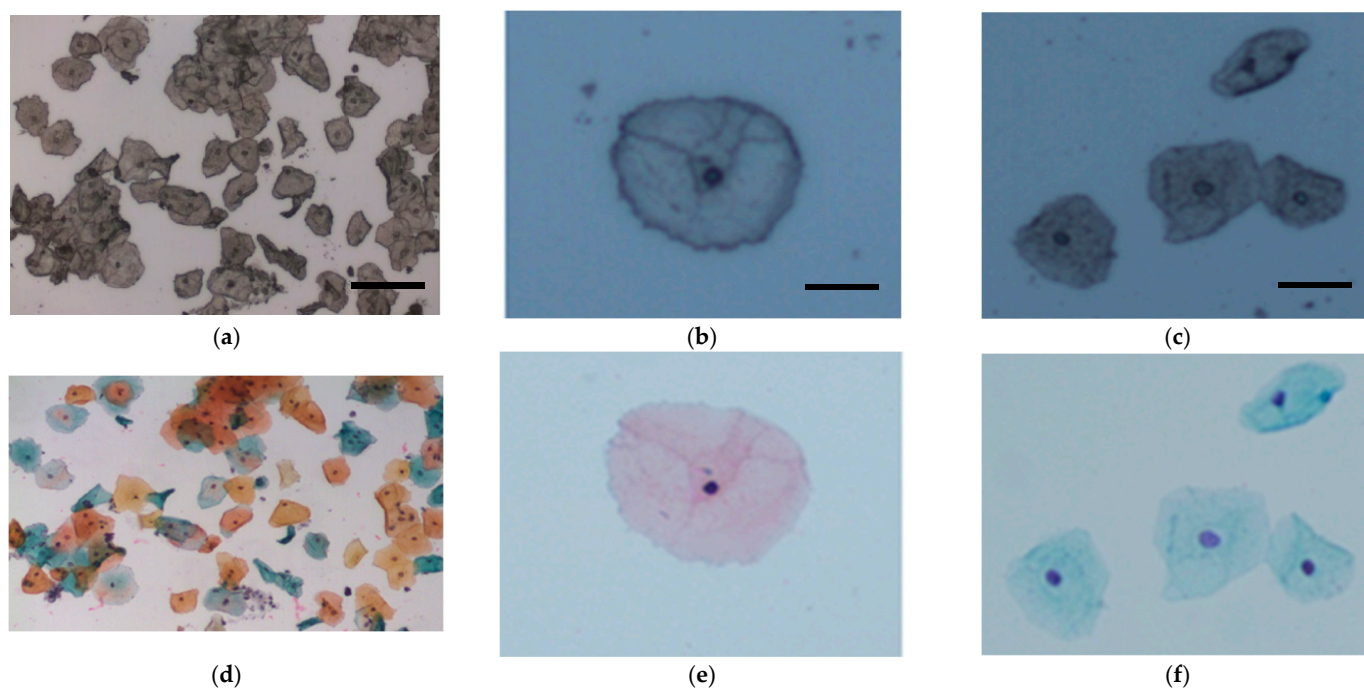
### 2.2. ThinPrep Slide Preparation

Samples were prepared for Raman spectroscopy using the ThinPrep 2000 processor (Hologic Inc., Marlborough, MA, USA). The ThinPrep processor homogenizes the sample by spinning the filter, creating shear forces that break up any clumped material (blood, mucin and non-diagnostic material). The cells are then transferred onto the TransCyt filter and transferred onto a glass slide to produce a circular monolayer of cells approximately 20 mm in diameter. The slide is then ejected into a fixative bath of 95% ethanol.

The slides then underwent a pretreatment step to remove any molecular contamination by hemoglobin, which obscures several features of the cellular spectrum as described previously [28]. Briefly, slides were treated with a 30% solution of H<sub>2</sub>O<sub>2</sub> at room temperature for 3 min, followed by a 70% solution of industrial methylated spirits (IMS) for 3 min, followed by multiple dips into 100% IMS to remove any remaining cellular debris and H<sub>2</sub>O<sub>2</sub>, before being air-dried.

### 2.3. Raman Spectroscopy

A HORIBA Jobin Yvon XploRA system (Villeneuve d'Ascq, France), with a 532 nm diode laser source, which incorporates an Olympus microscope BX41 equipped with a X100 objective (MPlan, Olympus, NA = 0.9) was utilized to record Raman spectra. Laser power was set to 100% resulting in 16 mW at the objective, and the confocal hole, coupled to a slit aperture of 100  $\mu\text{m}$ , was set at 100  $\mu\text{m}$ . Raman signals were detected using a spectrograph with a 1200 g/mm grating coupled to a charge-coupled device (Andor, 1024  $\times$  256 pixels) and the spectrometer was controlled by Labspec V6.0 software. The spectrometer was calibrated daily using the 520  $\text{cm}^{-1}$  line of silicon before sample acquisition. Calibration spectra of 1,4-bis(2-methylstyryl) benzene (Sigma Aldrich, Arklow, Ireland) were also recorded at the time of each sample acquisition. All spectra were subsequently wavenumber calibrated using in-house developed procedures in Matlab v.9.3 (Mathworks Inc., Natick, MA, USA). For sample acquisition, a Raman spectrum was acquired for each cell from the nucleus in the fingerprint region, 400 to 1800  $\text{cm}^{-1}$ , with an integration time of 30 s averaged over two accumulations. As spectra from the cell nucleus have been found to be more reproducible and consistent than spectra from the cell cytoplasm [25], only nuclear spectra were recorded. Where possible, spectra were recorded from at least 20–30 randomly selected morphologically normal superficial and intermediate cells from each unstained Pap smear, as shown in Figure 1. This resulted in a total of 1500 spectra for the training set of samples (750 spectra for samples 1–30 and 750 spectra for samples 31–60) and a total of 350 spectra for the blinded test set. The time taken to record spectra from 20–30 cells per sample was approximately 30 min.



**Figure 1.** Representative cervical cells from (a) unstained and (d) stained Pap smears (bar = 50  $\mu\text{m}$ ). Superficial cells (b) before and (e) after staining (bar = 20  $\mu\text{m}$ ) and intermediate cells (c) before and (f) after staining (bar = 20  $\mu\text{m}$ ).

### 2.4. Data Preprocessing and Analysis

Data preprocessing was performed using Matlab software (Mathworks Inc., Natick, MA, USA) and in-house scripts, including smoothing (Savitzky-Golay  $K = 5$ ,  $K = 13$ ), baseline correction (rubberband), and vector normalization. A non-negative least-squares (NNLS) method was used for glass correction as described previously [27,29]. The data were mean-centered and partial least squares discriminant analysis (PLS-DA) using the PLS toolbox (Eigenvector Research, Washington, DC, USA) in the Matlab (Mathworks Inc.,

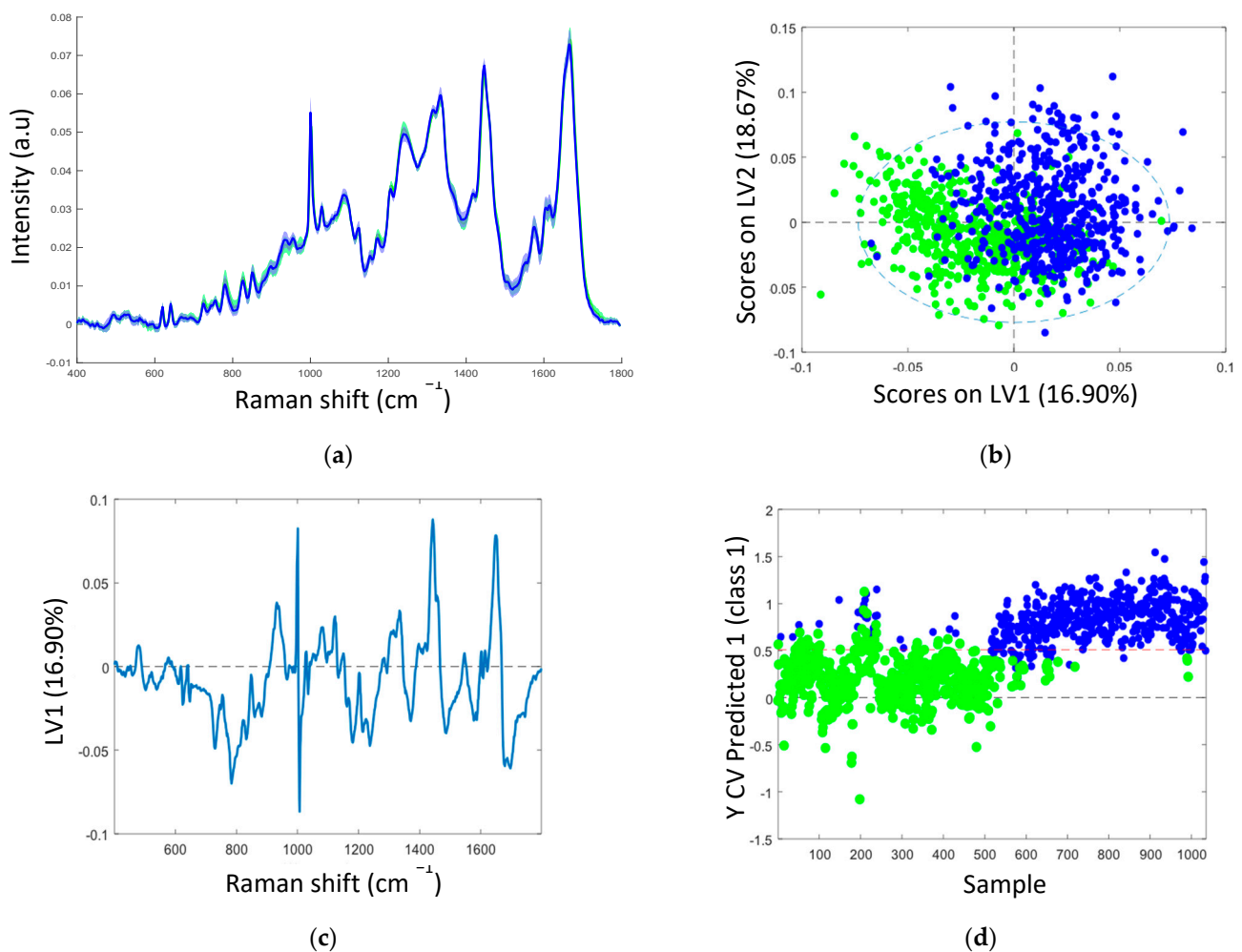
Natick, MA, USA) environment was used to build classification models. Leave-one-patient-out cross-validation (LOPOCV) was used, which involved removing all data from one patient sample from the model and repeating this process until all patient samples were left out once.

For further validation of the model, a blinded independent test set was employed ( $n = 14$ ). Firstly, all spectra were tested individually and, secondly, the mean spectrum of each patient sample was tested. The positive class comprised samples with a transcriptionally active HPV infection because the aim of screening is to identify patients most at risk who will require further treatment. The negative class comprised samples with a non-transcriptionally active HPV infection because they are deemed to be at a lower risk of developing cervical cancer in the next 5 years and can return to routine screening.

### 3. Results

The mean Raman spectra of samples with non-transcriptionally active and transcriptionally active HPV infection showed similar spectral profiles, as can be seen in Figure 2a. Table 2 illustrates the tentative Raman band assignments used in the present study. Differences were noted at Raman peak positions  $782\text{ cm}^{-1}$  (nucleic acids),  $1238\text{ cm}^{-1}$  (amide III), and  $1670\text{ cm}^{-1}$  (amide I). A classification model based on the Raman spectral data was developed using PLS-DA with LOPOCV. The scatter plot shows that latent variable LV1 mainly contributed to the differentiation of samples with non-transcriptionally active and transcriptionally active HPV infection (Figure 2b). The loadings from LV1 are shown in Figure 2c and show that discrimination is based around Raman peaks at 482 (glycogen), 727 (nucleic acids), 782 (nucleic acids), 826 (nucleic acids), 852 (glycogen), 937 (glycogen), 1082 (glycogen), 1123 (glycogen), 1152 (proteins), 1334 (glycogen), 1380 (glycogen), 1238 (amide III), 1450 (proteins/lipids), 1485 (nucleic acids), 1580 (nucleic acids), 1642 (proteins), and  $1670\text{ cm}^{-1}$  (amide I). Although a shift around  $1002\text{ cm}^{-1}$  (phenylalanine) can indicate calibration issues, a rigorous wavenumber calibration procedure was carried out; therefore, it is likely that this is a protein-related difference between the two groups. A sensitivity of 88% and a specificity of 88% were achieved by PLS-DA classification with LOPOCV (Figure 2d), indicating that Raman spectroscopy can distinguish non-transcriptionally active HPV infection from transcriptionally active HPV infection.

The next step was to validate the results with a blinded test set. The results are shown in Table 3 and show the result of (1) testing all sample spectra individually and (2) testing the mean spectrum of each patient sample. Nine out of 10 transcriptionally active samples and four out of four non-transcriptionally active samples were classified correctly using both types of testing regime. A sensitivity of 85% and a specificity of 92% were achieved for testing all sample spectra individually, and this increased to 90% and 100% for sensitivity and specificity, respectively, when the mean spectrum of each sample was tested. In both cases, sample 010 was incorrectly classified. When each cellular spectrum was tested individually, 10 out of 16 of the individual cellular spectra were classified incorrectly as non-transcriptionally active when they were from a transcriptionally active sample. When the mean spectrum of sample 010 as a whole was tested, it was again classified incorrectly as a non-transcriptionally active sample.



**Figure 2.** (a) Mean spectra of samples with non-transcriptionally active HPV infection (green,  $n = 750$  spectra) and transcriptionally active HPV infection (blue,  $n = 750$  spectra); shading indicates the standard deviation; (b) latent variable (LV) score scatter plot of LV1 and LV2 of samples with non-transcriptionally active HPV infection (green) and transcriptionally active HPV infection (blue); (c) LV1 loadings; (d) PLSDA plot.

The Raman classification results, together with the cytology, histology, HPV DNA, and HPV mRNA results, are shown in Table 4 for the blinded test set samples. Samples 001–009, which were classified correctly by Raman spectroscopy as transcriptionally active samples, were HSIL on cytology and, apart from sample 005, were confirmed as CIN2/3 on histology. In addition, each of these samples (001–009) was both HPV DNA-positive and mRNA-positive. Similarly, samples 011–014, which were classified correctly as non-transcriptionally active samples, were either negative or LSIL on cytology and negative on histology. Although these samples 011–014 were HPV DNA-positive, each one was HPV mRNA-negative. Sample 010, which was classified incorrectly by Raman spectroscopy as a non-transcriptionally active sample, was LSIL on cytology, CIN1 on histology, and both HPV DNA-positive and mRNA-positive.



**Table 2.** Tentative peak assignments [33].

Raman Peak Position ( $\text{cm}^{-1}$ )	Proteins	Lipids	Carbohydrates	Nucleic Acids
482			Glycogen	
577			Glycogen	
622	C–C twist Phe			
643	C–C twist Tyr			
727	CH <sub>2</sub> def	C–C head		A
752	Sym br. Trp			
781				U, C, T ring br
826	Out of Plane ring br. Tyr			PO2 a.str
851	Ring br. Tyr, C–C str. Pro			
852			Glycogen	
937			Glycogen	
985		C–C head		
1002	Sym. Ring br. Phe			
1033	C–H in plane Phe, C–C str			
1060	C–N str			
1082	C–N str	Chain C–C str	C–O str, Glycogen	
1096		Chain C–C str	C–C str	
1123	C–N str	Chain C–C str	C–O str, Glycogen	
1152	C–N str			
1207	C–C <sub>6</sub> H <sub>5</sub> str. Phe, Trp			
1238	C–N str, Amide III			
1334			Glycogen	
1338	Trp			G
1366		Sym. str. CH <sub>3</sub>		
1381			Glycogen	
1450	CH <sub>2</sub> def	CH <sub>2</sub> def		
1485	CH <sub>2</sub> def			G, A
1560	Tyr, Trp			
1580				A, G ring br
1584	C=C str, C=C bend. Trp, Phe			
1605	C=C Phe, Tyr			
1642	C=O str, C=C sym. str.			
1669	C=O str. Amide I			

**Table 3.** Classification of blinded test set samples ( $n = 14$ ) as transcriptionally active or non-transcriptionally active by (1) testing each cellular spectrum individually and (2) testing the mean spectrum of each sample.

Sample	Correct Classification	Individual Cellular Spectra		Mean Sample Spectra	
		Non-Transcriptionally Active	Transcriptionally Active	Non-Transcriptionally Active	Transcriptionally Active
001	Transcriptionally active	0	16	0	1
002	Transcriptionally active	15	16	0	1
003	Transcriptionally active	1	12	0	1
004	Transcriptionally active	4	33	0	1
005	Transcriptionally active	7	25	0	1
006	Transcriptionally active	0	19	0	1
007	Transcriptionally active	2	17	0	1
008	Transcriptionally active	3	16	0	1
009	Transcriptionally active	4	44	0	1
010	Transcriptionally active	10	6	1	0
011	Non-transcriptionally active	29	0	1	0
012	Non-transcriptionally active	26	3	1	0
013	Non-transcriptionally active	21	3	1	0
014	Non-transcriptionally active	22	2	1	0

**Table 4.** Raman classification, cytology results, histology results, and HPV testing results for the blinded test set ( $n = 14$ ).

Sample Number	Raman Classification	Cytology	Histology	HPV DNA	HPV mRNA
001	Transcriptionally active	HSIL	CIN2	Positive	Positive
002	Transcriptionally active	HSIL	CIN3	Positive	Positive
003	Transcriptionally active	HSIL	CIN3	Positive	Positive
004	Transcriptionally active	HSIL	CIN2	Positive	Positive
005	Transcriptionally active	HSIL	CIN1	Positive	Positive
006	Transcriptionally active	HSIL	CIN2	Positive	Positive
007	Transcriptionally active	HSIL	CIN2	Positive	Positive
008	Transcriptionally active	HSIL	CIN3	Positive	Positive
009	Transcriptionally active	HSIL	CIN2	Positive	Positive
010	Non-transcriptionally active	LSIL	CIN1	Positive	Positive
011	Non-transcriptionally active	LSIL	Negative	Positive	Negative
012	Non-transcriptionally active	Negative	Negative	Positive	Negative
013	Non-transcriptionally active	LSIL	Negative	Positive	Negative
014	Non-transcriptionally active	Negative	Negative	Positive	Negative

#### 4. Discussion

In this study, the aim was to investigate the use of Raman spectroscopy as a triage test for HPV-positive cases. Initially, Raman spectra were recorded from a training set consisting of cervical LBC samples ( $n = 60$ ). Samples were specifically selected for the training set as HPV-positive and high-grade HSIL cytology, and half of the samples ( $n = 30$ ) tested positive for HPV mRNA. The mean Raman spectra and LV1 loadings showed that the discrimination was mostly based around increased nucleic acids (727, 781, 826, 1485, and 1580  $\text{cm}^{-1}$ ), decreased glycogen (482, 852, 937, 1082, 1123, 1334, and 1380  $\text{cm}^{-1}$ ), and changes in protein features (1152, 1240, 1450, 1640, and 1670  $\text{cm}^{-1}$ ), indicating increased proliferation and altered protein expression due to the overexpression of E6/E7 viral proteins in the samples with a transcriptionally active HPV infection. These discriminating spectral features are consistent with previous studies on exfoliated cells showing discrimination between samples with negative cytology and HSIL cytology [25,27,28,30,34] and between samples with negative cytology and HPV DNA-positive HSIL samples [29]. PLSDA classification with LOPOCV achieved a sensitivity of 88% and a specificity of 88% in distinguishing non-transcriptionally active HPV infection from transcriptionally active HPV infection.

This classification model was further tested using a blinded test set ( $n = 14$ ). Raman spectra were recorded from this test set, and both individual cellular spectra and mean sample spectra were used to test the model. Both methods provided successful validation of the classification model with nine out of 10 samples with a transcriptionally active HPV infection and four out of four samples with a non-transcriptionally active HPV infection classified correctly. Sensitivity of 85% and specificity of 92% were achieved using individual cellular spectra, and this increased to 90% sensitivity and 100% specificity using mean sample spectra. The variability within the dataset was reduced by utilizing the mean spectra, resulting in an increase in sensitivity and specificity.

The need to record multiple cellular spectra from the same sample is highlighted in the data for sample 002, as, although this sample was classified correctly using individual cellular spectra, there was only one cellular spectrum in the difference. This suggests that not every cell within an HPV-infected sample will exhibit the associated biochemical signature of a persistent HPV infection. Factors such as sampling technique, the size of the sampled area, or the presence of HPV mRNA in every cell tested could account for this. Mean spectral data can reduce the variability in each sample, and it is expected that correct classification should be achieved if the mean sample spectrum shows the biochemical changes associated with a transcriptionally active HPV infection.

All the transcriptionally active samples tested in this study were reported as HSIL; thus, currently, it is not possible to determine if the cytology grade (LSIL or HSIL) affects the number of cells classified into each category. LSIL samples that are HPV DNA/mRNA-

positive need to be included in future studies to see if this affects the correct classification of individual cellular spectra.

Notably, very good agreement was observed among the cytology, histology, HPV tests, and the Raman classification result. The majority of samples that were classified as having a transcriptionally active HPV infection were both HPV DNA-positive and mRNA-positive and were reported as HSIL on cytology. The cytology result was confirmed as CIN2/3 by histology in all but one sample. Four out of five samples that were classified as having a non-transcriptionally active HPV infection were HPV DNA-positive but mRNA-negative and were reported as either negative or LSIL on cytology. All four samples were negative on histology. Interestingly, the sample that was classified incorrectly as having a non-transcriptionally active HPV infection was HPV DNA-positive and mRNA-positive but was reported as LSIL on cytology and confirmed as CIN1 on histology. A limitation of our study is that the training set consisted only of HSIL cases; hence, future work would need to include LSIL cases to fully train the classification model.

An additional limitation of this study is the relatively low number of samples with a non-transcriptionally active HPV infection in the blinded test set. As the main source of sample recruitment was a colposcopy clinic, where most patients were referred for further assessment of a HSIL cytology result, the majority of patients had a transcriptionally active HPV infection. This proof-of-principle study provides sufficient evidence to warrant a larger study in a screening population where HPV is the primary screening test and the clinical value of Raman spectroscopy for HPV triage can be realized. This would provide a larger sample base of samples with non-transcriptionally active HPV infection for testing compared to colposcopy-based recruitment.

## 5. Conclusions

In conclusion, this study showed that Raman spectroscopy has potential as a triage test for HPV-positive women to identify transcriptionally active HPV infections, although further work is necessary in a HPV primary screening setting. Unnecessary referral of women with a non-transcriptionally active HPV infection for cytological evaluation could be prevented by an improved triage test, while women with an increased risk of disease are referred to colposcopy. Women most at risk can be identified using a Raman spectroscopy-based test without the subjectivity of morphological assessment. The method can be performed on the same initial Pap test, and, if required, cytology and/or immunocytochemistry for molecular biomarkers can be carried out following the Raman spectral analysis.

**Author Contributions:** Conceptualization, F.M.L., D.T., C.M.M., and J.J.O.; methodology, D.T., S.R., and C.W.; software, D.T.; validation, D.T.; formal analysis, D.T.; investigation, D.T., S.R., and C.W.; resources, F.M.L., C.M.M., J.J.O., and T.D.; data curation, D.T. and F.M.L.; writing—original draft preparation, D.T. and F.M.L.; writing—review and editing, F.M.L., D.T., C.M.M., and J.J.O.; visualization, D.T.; supervision, F.M.L.; project administration, F.M.L., C.M.M., and J.J.O.; funding acquisition, F.M.L., C.M.M., and J.J.O. All authors read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Health Research Board Collaborative Applied Research Grant, CARG2012/29, and Enterprise Ireland co-funded by the European Regional Development Fund (ERDF) and Ireland's EU Structural Funds Program 2007–2013, CF2011 1045.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Research Ethics Committee of the Coombe Women and Infants University Hospital (Study No. 28-2014, date of approval: 31 October 2014).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** This research was undertaken as part of CERVIVA, the Irish Cervical Screening Research Consortium. We thank the women who agreed to participate in the study. We thank the Cytology staff and Colposcopy Staff at the Coombe Women and Infants University Hospital, Dublin, for facilitating the study and technical support staff at the FOCAS Research Institute, Technological University Dublin.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [[CrossRef](#)] [[PubMed](#)]
2. Simms, K.T.; Steinberg, J.; Caruana, M.; Smith, M.A.; Lew, J.B.; Soerjomataram, I.; Castle, P.E.; Bray, F.; Canfell, K. Impact of Scaled up Human Papillomavirus Vaccination and Cervical Screening and the Potential for Global Elimination of Cervical Cancer in 181 Countries, 2020–2099: A Modelling Study. *Lancet Oncol.* **2019**, *20*, 394–407. [[CrossRef](#)]
3. Canfell, K. Towards the Global Elimination of Cervical Cancer. *Papillomavirus Res.* **2019**, *8*, 100170. [[CrossRef](#)] [[PubMed](#)]
4. Walboomers, J.M.M.; Jacobs, M.V.; Manos, M.M.; Bosch, F.X.; Kummer, J.A.; Shah, K.V.; Snijders, P.J.F.; Peto, J.; Meijer, C.J.L.M.; Muñoz, N. Human Papillomavirus Is a Necessary Cause of Invasive Cervical Cancer Worldwide. *J. Pathol.* **1999**, *189*, 12–19. [[CrossRef](#)]
5. Paavonen, J.; Naud, P.; Salmerón, J.; Wheeler, C.M.; Chow, S.N.; Apter, D.; Kitchener, H.; Castellsague, X.; Teixeira, J.C.; Skinner, S.R.; et al. Efficacy of Human Papillomavirus (HPV)-16/18 AS04-Adjuvanted Vaccine against Cervical Infection and Precancer Caused by Oncogenic HPV Types (PATRICIA): Final Analysis of a Double-Blind, Randomised Study in Young Women. *Lancet* **2009**, *374*, 301–314. [[CrossRef](#)]
6. De Sanjosé, S.; Brotons, M.; Pavón, M.A. The Natural History of Human Papillomavirus Infection. *Best Pract. Res. Clin. Obstet. Gynaecol.* **2018**, *47*, 2–13. [[CrossRef](#)]
7. Tropé, A.; Sjøborg, K.; Eskild, A.; Cuschieri, K.; Eriksen, T.; Thoresen, S.; Steinbakk, M.; Laurak, V.; Jonassen, C.M.; Westerhagen, U.; et al. Performance of Human Papillomavirus DNA and MRNA Testing Strategies for Women with and without Cervical Neoplasia. *J. Clin. Microbiol.* **2009**, *47*, 2458–2464. [[CrossRef](#)]
8. Tota, J.E.; Bentley, J.; Blake, J.; Coutlée, F.; Duggan, M.A.; Ferenczy, A.; Franco, E.L.; Fung-Kee-Fung, M.; Gotlieb, W.; Mayrand, M.H.; et al. Introduction of Molecular HPV Testing as the Primary Technology in Cervical Cancer Screening: Acting on Evidence to Change the Current Paradigm. *Prev. Med.* **2017**, *98*, 5–14. [[CrossRef](#)]
9. Bonde, J.; Floore, A.; Ejegod, D.; Vink, F.J.; Hesselink, A.; Ven, P.M.; Valenčak, A.O.; Pedersen, H.; Doorn, S.; Quint, W.G.; et al. Methylation Markers *FAM19A4* and *miR124-2* as Triage Strategy for Primary Human Papillomavirus Screen Positive Women: A Large European Multicenter Study. *Int. J. Cancer* **2020**, *148*, 369–405. [[CrossRef](#)]
10. Koliopoulos, G.; Nyaga, V.N.; Santesso, N.; Bryant, A.; Martin-Hirsch, P.P.L.; Mustafa, R.A.; Schünemann, H.; Paraskevaidis, E.; Arbyn, M. Cytology versus HPV Testing for Cervical Cancer Screening in the General Population. *Cochrane Database Syst. Rev.* **2017**, *2017*, CD008587. [[CrossRef](#)]
11. Shiraz, A.; Crawford, R.; Egawa, N.; Griffin, H.; Doorbar, J. The Early Detection of Cervical Cancer. The Current and Changing Landscape of Cervical Disease Detection. *Cytopathology* **2020**, *31*, 258–270. [[CrossRef](#)] [[PubMed](#)]
12. Santos, I.P.; Barroso, E.M.; Bakker Schut, T.C.; Caspers, P.J.; Van Lanschot, C.G.F.; Choi, D.H.; Van Der Kamp, M.F.; Smits, R.W.H.; Van Doorn, R.; Verdijk, R.M.; et al. Raman Spectroscopy for Cancer Detection and Cancer Surgery Guidance: Translation to the Clinics. *Analyst* **2017**, *142*, 3025–3047. [[CrossRef](#)] [[PubMed](#)]
13. Upchurch, E.; Isabelle, M.; Lloyd, G.R.; Kendall, C.; Barr, H. An Update on the Use of Raman Spectroscopy in Molecular Cancer Diagnostics: Current Challenges and Further Prospects. *Expert Rev. Mol. Diagn.* **2018**, *18*, 245–258. [[CrossRef](#)]
14. Kanter, E.M.; Vargis, E.; Majumder, S.; Keller, M.D.; Woeste, E.; Rao, G.G.; Mahadevan-Jansen, A. Application of Raman Spectroscopy for Cervical Dysplasia Diagnosis. *J. Biophotonics* **2009**, *2*, 81–90. [[CrossRef](#)]
15. Vargis, E.; Kanter, E.M.; Majumder, S.K.; Keller, M.D.; Beaven, R.B.; Rao, G.G.; Mahadevan-Jansen, A. Effect of Normal Variations on Disease Classification of Raman Spectra from Cervical Tissue. *Analyst* **2011**, *136*, 2981–2987. [[CrossRef](#)]
16. Vargis, E.; Tang, Y.-W.; Khabele, D.; Mahadevan-Jansen, A. Near-Infrared Raman Microspectroscopy Detects High-Risk Human Papillomaviruses. *Transl. Oncol.* **2012**, *5*, 172–179. [[CrossRef](#)]
17. Kamemoto, L.E.; Misra, A.K.; Sharma, S.K.; Goodman, M.T.; Hugh, L.U.K.; Dykes, A.C.; Acosta, T. Near-Infrared Micro-Raman Spectroscopy for in Vitro Detection of Cervical Cancer. *Appl. Spectrosc.* **2010**, *64*, 255–261. [[CrossRef](#)]
18. Tan, K.M.; Herrington, C.S.; Brown, C.T.A. Discrimination of Normal from Pre-Malignant Cervical Tissue by Raman Mapping of de-Paraffinized Histological Tissue Sections. *J. Biophotonics* **2011**, *4*, 40–48. [[CrossRef](#)]
19. Rashid, N.; Nawaz, H.; Poon, K.W.C.; Bonnier, F.; Bakhiet, S.; Martin, C.; O’Leary, J.J.; Byrne, H.J.; Lyng, F.M. Raman Microspectroscopy for the Early Detection of Pre-Malignant Changes in Cervical Tissue. *Exp. Mol. Pathol.* **2014**, *97*. [[CrossRef](#)] [[PubMed](#)]

20. Lyng, F.M.; Traynor, D.; Ramos, I.R.M.; Bonnier, F.; Byrne, H.J. Raman Spectroscopy for Screening and Diagnosis of Cervical Cancer. *Anal. Bioanal. Chem.* **2015**, *407*, 8279–8289. [[CrossRef](#)] [[PubMed](#)]
21. Ramos, I.; Meade, A.D.; Ibrahim, O.; Byrne, H.; McMenamin, M.; McKenna, M.; Malkin, A.; Lyng, F. Raman Spectroscopy for Cytopathology of Exfoliated Cervical Cells. *Faraday Discuss.* **2015**, *187*, 187–198. [[CrossRef](#)]
22. Duraipandian, S.; Zheng, W.; Ng, J.; Low, J.J.H.; Ilancheran, A.; Huang, Z. Near-Infrared-Excited Confocal Raman Spectroscopy Advances in Vivo Diagnosis of Cervical Precancer. *J. Biomed. Opt.* **2013**, *18*, 067007. [[CrossRef](#)]
23. Duraipandian, S.; Zheng, W.; Ng, J.; Low, J.J.H.; Ilancheran, A.; Huang, Z. Simultaneous Fingerprint and High-Wavenumber Confocal Raman Spectroscopy Enhances Early Detection of Cervical Precancer in Vivo. *Anal. Chem.* **2012**, *84*, 5913–5919. [[CrossRef](#)]
24. Traynor, D.; Duraipandian, S.; Martin, C.M.; O’Leary, J.J.; Lyng, F.M. Improved Removal of Blood Contamination from ThinPrep Cervical Cytology Samples for Raman Spectroscopic Analysis. *J. Biomed. Opt.* **2018**, *23*, 1. [[CrossRef](#)] [[PubMed](#)]
25. Kearney, P.; Traynor, D.; Bonnier, F.; Lyng, F.M.; O’Leary, J.J.; Martin, C.M. Raman Spectral Signatures of Cervical Exfoliated Cells from Liquid-Based Cytology Samples. *J. Biomed. Opt.* **2017**, *22*, 1. [[CrossRef](#)] [[PubMed](#)]
26. Daniel, A.; Prakasarao, A.; Ganesan, S. Near-Infrared Raman Spectroscopy for Estimating Biochemical Changes Associated with Different Pathological Conditions of Cervix. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2018**, *190*, 409–416. [[CrossRef](#)]
27. Traynor, D.; Duraipandian, S.; Bhatia, R.; Cuschieri, K.; Martin, C.M.; O’Leary, J.J.; Lyng, F.M. The Potential of Biobanked Liquid Based Cytology Samples for Cervical Cancer Screening Using Raman Spectroscopy. *J. Biophotonics* **2019**, *12*, e201800377. [[CrossRef](#)] [[PubMed](#)]
28. Bonnier, F.; Traynor, D.; Kearney, P.; Clarke, C.; Knief, P.; Martin, C.; O’Leary, J.J.; Byrne, H.J.; Lyng, F. Processing ThinPrep Cervical Cytological Samples for Raman Spectroscopic Analysis. *Anal. Methods* **2014**, *6*, 7831–7841. [[CrossRef](#)]
29. Traynor, D.; Kearney, P.; Ramos, I.; Martin, C.M.; O’Leary, J.J.; Lyng, F.M. A Study of Hormonal Effects in Cervical Smear Samples Using Raman Spectroscopy. *J. Biophotonics* **2018**, *11*, e201700240. [[CrossRef](#)]
30. Rubina, S.; Amita, M.; Kedar, K.D.; Bharat, R.; Krishna, C.M. Raman Spectroscopic Study on Classification of Cervical Cell Specimens. *Vib. Spectrosc.* **2013**, *68*, 115–121. [[CrossRef](#)]
31. Jess, P.R.T.; Smith, D.D.W.; Mazilu, M.; Dholakia, K.; Riches, A.C.; Herrington, C.S. Early Detection of Cervical Neoplasia by Raman Spectroscopy. *Int. J. Cancer* **2007**, *121*, 2723–2728. [[CrossRef](#)]
32. Ostrowska, K.M.; Malkin, A.; Meade, A.; O’Leary, J.; Martin, C.; Spillane, C.; Byrne, H.J.; Lyng, F.M. Investigation of the Influence of High-Risk Human Papillomavirus on the Biochemical Composition of Cervical Cancer Cells Using Vibrational Spectroscopy. *Analyst* **2010**, *135*, 3087–3093. [[CrossRef](#)] [[PubMed](#)]
33. Movasaghi, Z.; Rehman, S.; Rehman, I.U. Raman Spectroscopy of Biological Tissues. *Appl. Spectrosc. Rev.* **2007**, *42*, 493–541. [[CrossRef](#)]
34. Duraipandian, S.; Traynor, D.; Kearney, P.; Martin, C.; O’Leary, J.J.; Lyng, F.M. Raman Spectroscopic Detection of High-Grade Cervical Cytology: Using Morphologically Normal Appearing Cells. *Sci. Rep.* **2018**, *8*, 15048. [[CrossRef](#)] [[PubMed](#)]