Biochemistry and Biophysics Reports 26 (2021) 100993

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

journal homepage: www.elsevier.com/locate/bbrep

PCR-RFLP and FTIR-based detection of high-risk human papilloma virus for cervical cancer screening and prevention

Igor Martins Alves Melo [a], Magda Rogéria Pereira Viana [a], Breno Pupin [a], Tanmoy Tapobrata Bhattacharjee [b], Renata de Azevedo Canevari [a,*]

[a] Laboratório de Biologia Molecular do Câncer, Universidade do Vale do Paraíba, UNIVAP, Instituto de Pesquisa e Desenvolvimento, Avenida Shishima Hifumi 2911, Urbanova, São José dos Campos, 12244-000, São Paulo, SP, Brazil

[b] Sir John Walsh Research Institute, University of Otago, 310 Great King St, North Dunedin, Dunedin, 9016, New Zealand

ARTICLE INFO

Keywords:
HPV screening
PCR-RFLP
FTIR
Cervical cancer prevention

ABSTRACT

Background: Approximately 70% of cervical carcinoma cases show the presence of high-risk Human Papilloma Virus (HPV), especially HPV-16 and HPV-18, and can be used to stratify high risk patients from low risk and healthy. Currently, molecular biology techniques such as polymerase chain reaction (PCR) are used to identify the presence of virus in patient samples. While the methodology is highly sensitive, it is labor intensive and time-consuming. Alternative techniques, such as vibrational spectroscopy, has been suggested as a possible rapid alternative. Therefore, in this study, we evaluate the efficiency of cervical fluid Fourier Transform Infrared spectroscopy (FTIR) in patient risk stratification informed by PCR.

Methods: Cervical fluid samples (n = 91) were obtained from patients who have undergone routine Papanicolaou (Pap) test. Viral genome was identified and classified as high/low-risk by PCR-Restriction Fragment Length Polymorphism (PCR-RFLP). FTIR spectra were acquired from samples identified by PCR-RFLP as No-HPV (n = 10), high-risk HPV (n = 7), and low-risk HPV (n = 7).

Results: Of the 91 samples, was detected the viral genome by PCR in 36 samples. Of these 36 samples, nine samples were identified to contain high-risk HPV (HR-HPV) and nine samples were found to have low-risk HPV (LR-HPV). The FTIR spectra acquired from No-HPV, LR-HPV, and HR-HPV showed differences in 1069, 1437, 1555, 1647, 2840, 2919, and 3287 cm^{-1} bands. Principal Component Analysis (PCA) showed distinct clusters for No-HPV and HR-HPV and No-HPV and LR-HPV, but there was significant overlap in the clusters of HR-HPV and LR-HPV. PCA-Linear Discriminant Analysis (PC-LDA) after Leave One Out Cross Validation (LOOCV) classified No-HPV from HR-HPV and No-HPV from LR-HPV with 100% efficiency in the 1400-1800 cm^{-1} spectral range. LOOCV classifications for LR-HPV and HR-HPV from each other were 71 and 75%, respectively, in the 2800-3400 cm^{-1} spectral range.

Conclusions: The results highlight the high sensitivity of PCR-RFLP in HPV identification and show that FTIR can classify samples identified as healthy, low, and high-risk samples by PCR-RFLP.

General significance: We show the possibility of using FTIR for initial cervical cancer risk stratification followed by detailed PCR-RFLP investigations for suspect cases.

1. Introduction

Cervical cancer is the fourth most common cancer worldwide, with estimated 570,000 new cases every year [1]. Early diagnosis of the cancer can greatly enhance therapeutic outcome and Papanicolaou (Pap) cytology test is the most widely used method for detecting pre-cancerous lesions for early diagnosis [2]. The technique, however, has low sensitivity and specificity and is highly subjective. Another approach that has been adopted for early cervical cancer diagnosis is detecting the presence of Human Papilloma Virus (HPV).

HPV is classified to the Papillomaviridae family. This virus is composed of a circular DNA double helix surrounded by an icosahedral protein capsid. To date, 207 different types of HPV have been identified and more than 170 are completely sequenced. The different types of

* Corresponding author.
E-mail address: rcanevari@univap.br (R. de Azevedo Canevari).

https://doi.org/10.1016/j.bbrep.2021.100993
Received 4 November 2020; Received in revised form 5 March 2021; Accepted 25 March 2021

HPV are due to variations in the sequence of the L1 gene present in the viral DNA, being classified as low or high carcinogenic risk according to their respective potentials for cervical cancer development. The highly oncogenic variants can cause dysplasia that, in most cases, can progress to a carcinoma, represented mainly by HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68. On the other hand, types 6, 11, 40, 42, 43, 44 and 54 are considered of low oncogenic risk because they only provide the formation of warts and other minimal lesions that will rarely evolve to malignancies [3]. Epidemiological studies have shown that HPV types 16 and 18 are associated with 70% cervical cancer cases worldwide [4].

Molecular biology techniques such as Polymerase Chain Reaction (PCR) combined with Restriction Fragment Length Polymorphism (RFLP) can be used to find the presence of specific HPV type from samples, and enable patient stratification as at no, low, or high risk of cancer [5]. The methodology has shown high sensitivity and specificity, even in patients who do not show clinical signs and symptoms, because the virus may be in latent state. This association has allowed the identification of single nucleotide polymorphisms based on the genomic sequence of the various HPV variants. The identification helps guide clinical management.

Despite its advantages, the PCR-RFLP procedure is labor intensive and time consuming [6]; and cannot be used for mass-screening. Other HPV detection techniques such as immunohistochemistry, electron microscopy, Western blot, southern blot, in situ hybridization, and ELISA techniques have low clinical sensitivity [6] and are time-consuming as well. One of the techniques investigated for rapid means of risk stratification, vibrational spectroscopy, has shown promising results [7]. The complementary vibrational spectroscopy techniques Raman and infrared spectroscopy provide complete chemical profiles of the study samples. Multivariate statistical analysis tools such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) can be used to decipher differences between spectral patterns of samples and identify them. For example, subjecting spectra of healthy, low risk HPV (LR-HPV) and high-risk HPV (HR-HPV) to PCA and LDA can lead to recognition of spectral signatures unique to each group, which can then be used to distinguish HR-HPV containing patient samples from LR-HPV and healthy samples.

PCA is an unsupervised technique wherein all spectra are used as input without providing group labels. The analysis finds a mean for all spectra, calculates difference in each spectrum from the mean, computes a series principal components (PCs) that represent these differences in descending order, and assigns scores of PCs to each spectrum. By plotting the scores of two PCs (say PC1 and PC2), a 2D scatter plot is generated. If there exist differences between groups, their spectra appear as separate clusters; if not, the spectra overlap. Thus, if spectra of HR-HPV are different from LR-HPV, they will form two distinct clusters; if not, they will overlap. However, overlap does not necessarily mean there is no difference between the spectra. A 3D plot of PCs (say PC 1, 2, and 3) may yield separate clusters. Or it is possible that other PC combinations (PC3 and PC 4 or PC 1, 2, and 5) yield distinct clustering.

Instead of plotting different combinations, a supervised analysis can be used that provides an overall result based on all PCs. LDA is such a method, where the spectra used as input are provided with group label (No-HPV/LR-HPV/HR-HPV). The method finds the best possible orientation in n-space that results in best classification and gives output in form of a confusion matrix. A confusion matrix shows how many spectra were correctly identified as a group (No-HPV/LR-HPV/HR-HPV). PC-LDA method uses PCs as input instead of raw spectra, enabling removal of noise before LDA. However, LDA tries to find the best classification and may present an erroneous result due to overfitting of data to get best result. To avoid this, the LDA outcome is subjected to cross validation. One such method is leave one out cross validation (LOOCV), wherein LDA model is built after removing one spectrum and then predicting the group of the spectrum; and the process is repeated till each spectrum have been left out once. The confusion matrix is built by placing each spectrum in the group it had been predicted by the above process.

In this study, we aimed to detect the presence and type of HPV in cervical fluids of patients using the PCR-RFLP molecular technique and studied the efficacy of Fourier Transform Infrared spectroscopy (FTIR) in differentiating cervical fluid sample containing HR-HPV from samples containing LR-HPV and healthy subject samples (No-HPV) by using PCA, PC-LDA and LOOCV.

2. Materials and methods

2.1. Subjects

This study was approved by the Research Ethics Committee of the University of Paraíba Valley (n° 1,542,726) and the Research Ethics Committee of the University Center of the Faculty of Health, Human and Technological Sciences of Piauí (n° 2,434,076), according to the established by resolution 466/2012, of the National Health Council.

The inclusion criteria of the patients consisted of age between 18 and 65 years and no evident gynecological changes, such as hyperemia, presence of fluid discharge and irritation of the cervix. The exclusion criteria were women in menstrual periods, with pre-existing malignant uterine neoplasia, infectious diseases, carriers of sexually transmitted infections and pregnant patients.

The procedure was explained to all the subjects and signed consent was obtained before sample collection. They also answered a questionnaire asking about their clinical and behavioral characteristics, including: age, weight, height, smoking habit, age at menarche, age at first sexual intercourse, sexual orientation, number of sexual partners, number of pregnancies and deliveries, occurrence of abortions and the presence of signs or symptoms of gynecological diseases. Statistical analysis between patient characteristics and HPV infection was performed using the chi-squared test.

2.2. Samples

In total, samples were collected from 91 patients (41 in Cascavel, Paraná and 50 in Teresina, Piauí). Four cervical fluid samples were obtained from each patient using Ayre spatula and endocervical brush. Two samples were used for Pap tests, one for PCR-RFLP and one for FTIR. All samples were stored at −20 °C till respective analyses were carried out.

2.3. PCR-RFLP

The extraction of deoxyribonucleic acid (DNA) from the cervical fluid collected for the detection of Human Papillomavirus occurred according to the nucleic acid purification protocol of the QIAamp® MinElute® Virus Spin kit (Qiagen). The quantification and concentration of the genetic material present in each sample was evaluated by the ultraviolet absorption spectroscopy of the NanoDrop equipment (ND-1000 Spectrophotometer v.3.0.1, Labtrade).

The measure of the purity of the extracted DNA was analyzed in the spectrophotometer for the reasons 260/280 and 260/230 which represent, respectively, contamination by proteins and reagents. Ideal values are between 1.8 and 2.0 for the first ratio presented and between 1.8 and 2.2 for the second. In addition, DNA integrity and quality were assessed using electrophoresis in 2.0% agarose gel. In samples with extracted DNA concentrations below 100 ng/uL, the Concentrator plus equipment (Eppendorf Concentrator plus™) was used to reduce the amount of water present in the tube and, after 8 min of vacuum centrifugation, to increase the DNA concentration in the sample.

All DNA samples extracted from the fluids collected from the cervix were subjected to the Polymerase Chain Reaction (PCR) for the amplification of the L1 gene, the most conserved region of the genome of the different types of HPV, using the MY09/11 (MY09: 5' - CGT CCM AAR

GGA WAC TGA TC – 3' and MY11: 5' – GCM CAG GGW CAT AAY AAT GG – 3') and GP5+/GP6+ (GP5+: 5' – TTT GTT ACT GTG GTA GAT ACT AC – 3' and GP6+: 5' – GAA AAA TAA ACT GTA AAT CAT ATT C – 3') primer sets. These primers amplify a region of approximately 450 and 150 base pairs (bp), respectively, and its specificities were analyzed using the BLAST platform (available at: https://blast.ncbi.nlm.nih.gov/Blast.cg), showing 100% correspondence with the virus genome.

Aiming at the quality of the reaction, each system for amplification contained a final volume of 25 μl, which consisted of 50 mM KCl, 10 mM Tris-HCl with pH 8.3, 4 mM MgCl2, 200 mM deoxynucleotide triphosphate (dNTP), 50 pmol of each primer and 2U of Easy TAQ DNA Polymerase. The amplification conditions, performed in the thermocycler (Veriti™ 96-Well Thermal Cycler) for primers MY09 and MY11, occurred in 35 cycles consisting of: denaturation in 45 s at 94 °C, annealing in 45 s at 58 °C and extension at 45 s at 72 °C, accompanied by 5 min at 72 °C; while the conditions used for amplification with the GP5+ and GP6+ primers consisted of an initial denaturation at 94 °C for 4 min, followed by 40 cycles in which the denaturation occurred at 94 °C for 1 min, annealing at 50 °C for 1 min and 15 s and extension in 1 min at 72 °C and, in addition, a final extension at 72 °C for 10 min.

Subsequently, the amplified products obtained by the reaction were analyzed using electrophoresis in a 1.3% agarose gel. As a molecular standard, a low molecular weight marker (100 bp) was used and the result was visualized through the interaction of the amplified product with the fluorescent dye Ethide Bromide (3 μg/ml), which is intercalated between nitrogenous bases that constitute the DNA. The incidence of ultraviolet (UV) light, which allows the visualization of bands on the gel, was controlled by the Transilluminator-D Pro MiniBIS equipment (DNR Bio-Imaging Systems Ltd), and the image capture was obtained using the GelCapture software for further analysis.

After detecting the presence of HPV as evidenced by PCR amplification and to perform viral typing, the amplified product was digested by the restriction enzyme RsaI, which recognizes the target sequence and breaks the DNA phosphodiester bonds at the GT↓AC site. The solution for enzymatic digestion was prepared using 20 μL of PCR product and 0.5 μL of restriction enzyme, the mixture being incubated at 37 °C for 2 h.

To identify the type of HPV present in each sample positive for infection with this virus, an analysis of the enzymatic activity was carried out by cleaving the amplified product by visualizing the agarose gel (2.0%) in ultraviolet light in the Transilluminator-D Pro MiniBIS equipment (DNR Bio-Imaging Systems Ltd). The bands visualized on the gel, which represent the amplified product digested from the virus genome, were compared with the polymorphisms related to the different types of HPV [8,9].

2.4. FTIR

The samples (stored at −20 °C) were thawed (10 No-HPV, 7 LR-HPV, and 7 HR-HPV) and processed. The processing was characterized by the initial dilution in 0.5 mL of saline and homogenization by centrifugation at 4000 rpm for 2 min. Subsequently, 1 mL of SLH buffer (red cell lysis solution) was added to the fluid, which was centrifuged at 5000 rpm for 15 min. After this procedure, an additional centrifugation was performed at 5000 rpm for 15 min and the supernatant was discarded. This procedure was repeated three times for each sample. The formed pellet was used to carry out spectral measurements in the FTIR.

After processing each sample, 2 μl was applied to the Calcium Fluoride (CaF_2) slide, four spectra per sample with 32 scans in the range of 750–4000 cm^{-1} at 4 cm^{-1} resolution were obtained FTIR spectrometer (Spotlight 400, PerkinElmer, USA) equipped with a microscope (Spotlight PerkinElmer 400, USA). All spectra were preprocessed by first derivatization, spectral range selection (1400-1800 cm^{-1} or 2800-3400 cm^{-1}) and area normalization; and subjected to multivariate analysis Principal Component Analysis (PCA), PCA-Linear Discriminant Analysis (PC-LDA) and Leave One Out Cross Validation (LOOCV) in MATLAB 2015.

3. Results

3.1. Questionnaire answers

The responses to the questionnaire revealed that 75.00% of patients infected with HPV (HPV +) were younger than 30 years old ($p < 0.0001$), 44.44% reported having their first sexual intercourse before reaching 20 years of age ($p < 0.0001$), 41.67% had only one sexual partner ($p = 0.1277$), 13.89% reported that they smoke daily ($p = 0.8993$), 11.11% ingest alcoholic beverages ($p = 0.5525$) and 33.33% have completed higher education ($p = 0.3022$). Besides that, 16.67% of HPV+ patients had completed elementary school, 25.00% completed high school and 33.33% completed higher education (Table 1).

Regarding patients negative for HPV infection (HPV-), 65.45% of them have only one sexual partner and 34.55% have two or more partners. 14.55% of HPV- patients were smokers, while 81.82% of these do not practice this habit. On the other hand, 32.73% of HPV- patients ingested alcohol. Moreover, 7.27% of HPV- patients revealed they had not completed elementary school, 12.73% had completed elementary school, 38.18% had completed high school and 25.45% have completed higher education.

3.2. PCR-RFLP

The presence of HPV genome was observed in 36 (39.56%) of the 91 samples. Viral type could be identified by PCR-RFLP in 18 samples, nine

Table 1
Personal characteristics and relation to HPV infection of 91 patients analyzed. *; ***: with statistical significance. ns: no statistical significance.

FEATURE	HPV +	%	HPV -	%	*p* value
Age of the patient					< 0.0001***
≤ 30 years old	27	75,00%	18	32,73%	
> 30 yearsold	9	25,00%	37	67,27%	
Age of first sexual intercourse					**< 0.0001*** **
≤ 20 years old	16	44,44%	11	20,00%	
> 20 years old	3	8,33%	35	63,64%	
No information	17	47,22%	9	16,36%	
Multiplicity of sexual partners					**0.1277^{ns}**
1	15	41,67%	36	65,45%	
2	6	16,67%	18	32,73%	
3 or more	3	8,33%	1	1,82%	
No information	12	33,33%	0	0,00%	
Smoking habits					**0.8993^{ns}**
Yes	5	13,89%	8	14,55%	
No	26	72,22%	45	81,82%	
No infornation	5	13,89%	2	3,64%	
Alcoholism habits					**0.5525^{ns}**
Yes	4	11,11%	18	32,73%	
No	3	8,33%	22	40,00%	
No infornation	29	80,56%	15	27,27%	
Degree of schooling					**0.3022^{ns}**
Elementary education incomplete	0	0,00%	4	7,27%	
Elementary education complete	6	16,67%	7	12,73%	
Incomplete Secondary Education	1	2,78%	5	9,09%	
Complete Secondary Education	9	25,00%	21	38,18%	
Higher Education	12	33,33%	14	25,45%	
Higher Education Incomplete	0	0,00%	1	1,82%	
No infornation	8	22,22%	3	5,45%	

of which showed HR-HPV types 16, 31, 39, 51, 58, and 68, while the rest showed presence of LR-HPV types 6b, 11, 44, and 55 (Fig. 1).

3.3. FTIR

Spectra (n = 4) were acquired from each sample identified as either No-HPV (n = 10), HR-HPV (n = 7) or LR-HPV (n = 7). Spectral differences were observed in 1069 (nucleic acid), 1437 (lipids), 1555 (protein), 1647 (protein Amide I), 2840 (lipids), 2919 (lipids), and 3287 (lipids) cm^{-1} bands (Fig. 2).

PCA shows clear separation of No-HPV from HR-HPV (Fig. 3a.1) and No-HPV from LR-HPV (Fig. 3b.1) in the range 1400–1800 cm^{-1}. There was no separation between low and high-risk virus in this range, but clustering with overlap was observed in 2800–3400 cm^{-1} (Fig. 3c.1). The loading factor (PCs) responsible for separating No-HPV from high-risk shows 1472, 1562, and 1657 cm^{-1} bands, suggested differences in lipids and proteins (Fig. 3a.2). Loading factor suggests differences in protein (1550 and 1655 cm^{-1}) between No-HPV and LR-HPV samples (Fig. 3b.2). A difference in lipid (2856 cm^{-1}) is observed in case of LR-HPV versus HR-HPV samples (Fig. 3c.2).

PCA-LDA LOOCV shows that high-risk (Table 2a) and LR-HPV (Table 2b) can be classified from No-HPV samples with 100% efficiency at 1400-1800 cm^{-1} range. LOOCV between LR-HPV and HR-HPV sample spectra (Table 2c) shows they can be classified with 71 and 75% efficiency, respectively, in the 2800-3400 cm^{-1} spectral range.

The misclassifications between LR-HPV and HR-HPV (Table 2c) were studied. Four spectra from cervical fluid sample shown by PCR-RFLP to contain HR-HPV type 39 were wrongly classified as low-risk. Similarly, two and one out of four spectra from sample containing HR-HPV type 31 and type 58, respectively, were incorrectly classified as low risk. All four spectra of sample showing presence of LR-HPV type 11 were classified as high-risk, while the same was observed for two spectra out of four for LR-HPV types 55 and 44, respectively. This may be because the biochemical characteristics of cervical fluid samples HPV types 39, 31, and 58 resemble that of those containing low-risk virus, which may be influenced by the extent and kind of changes the virus induces in the tissue from where the fluid was collected. The same can be an explanation of some low-risk virus containing samples misclassifying with

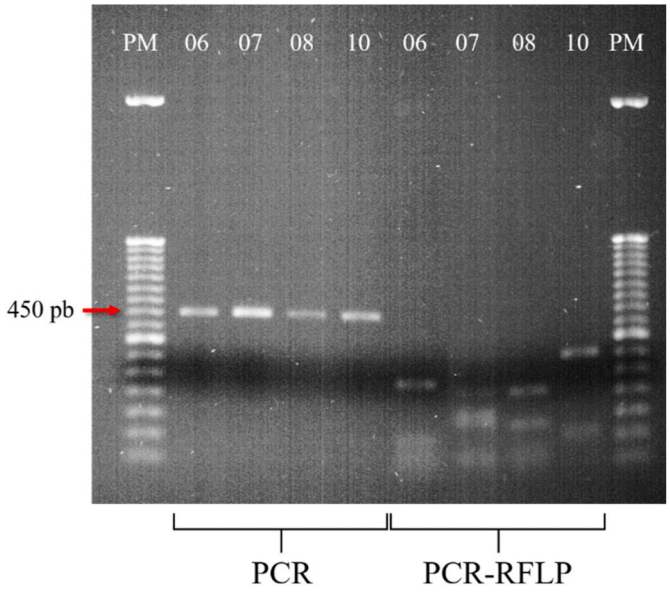

Fig. 1. Agarose gel electrophoresis (2.0%) of the PCR and PCR-RFLP product of 4 HPV + samples. Legend: PM: Molecular Weight Standard (50 bp); Sample 06: HPV 39 (bands: 260/123/72); Sample 07: HPV 44 (bands: 222/161/72); Sample 08: HPV 11 (bands: 216/135/72/26); Sample 10: HPV 58 (bands: 306/111/32).

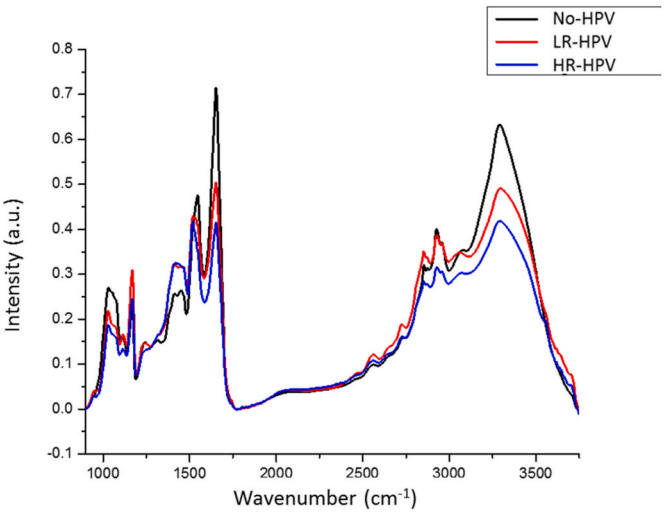

Fig. 2. Mean spectra of cervical sample containing No-HPV (normal/healthy), Low-risk HPV (LR-HPV), and High-risk HPV (HR-HPV).

those containing high risk.

4. Discussion

Early detection of cervical cancer is critical to efficacious treatment. Several techniques are in use for screening for cervical cancer, such as visual inspection, colposcopy, and Pap smears, but these techniques have low or moderate clinical sensitivity with respect to detection of precancer lesions [6]. In contrast, PCR and real time PCR based HPV genome detection has very high sensitivity. Recent studies have shown that use of complementary primers for PCR, MY09/11 and GP5+/GP6+, leads to 99.7% identification of samples, as opposed to use of only MY09/11 primer [10–12]. In this study, we have used the combination of primers, although this is the first study to report RFLP using only a single restriction enzyme Rsa I. We found HPV to be present in 36 out of 91 samples. Of these 36 samples detected to contain HPV, 83% did not show any cytological changes by PAP smear analysis (data not shown), highlighting the sensitivity of the PCR technique.

Despite this high sensitivity, the PCR-RFLP based detection suffers from several disadvantages. The technique is expensive, time consuming, and labor intensive. It also requires extensive training and expertise to be performed efficiently and consistently. These shortcomings prevent it from becoming a mass screening tool. For mass-screening, a simple methodology that is inexpensive, rapid, and requiring little expertise and training is necessary. Fourier Transform Infrared (FTIR) spectroscopy is an ideal tool in this respect. Therefore, in this study, we have investigated the ability of FTIR to distinguish samples that does not contain HPV (No-HPV) from samples containing high risk (HR-HPV) and samples with low-risk HPV (LR-HPV) based on PCR-RFLP results.

Using multivariate analysis, we found that 40 out of 40 No-HPV spectra (100%) were correctly classified as No-HPV spectra, while 28 out of 28 HR-HPV spectra (100%) were correctly classified as HR-HPV (Table 2a). This suggests that FTIR can differentiate No-HPV from HR-HPV with 100% efficacy. Similarly, FTIR could distinguish No-HPV from LR-HPV with 100% efficacy (Table 2b). Table 2c shows that 21 out of 28 HR-HPV spectra (75%) could be classified as HR-HPV, while 7 (25%) were wrongly classified; and that 20 out of 28 LR-HPV (71%) were correctly distinguished as LR-HPV, while 8 spectra (29%) were wrongly classified. Thus, when it comes to distinguishing HR-HPV from LR-HPV, the efficacy of identification using FTIR is 75% and 71%, respectively. Other studies have shown similar results. For example, Viana et al. [13] showed the possibility of distinguishing HPV infected

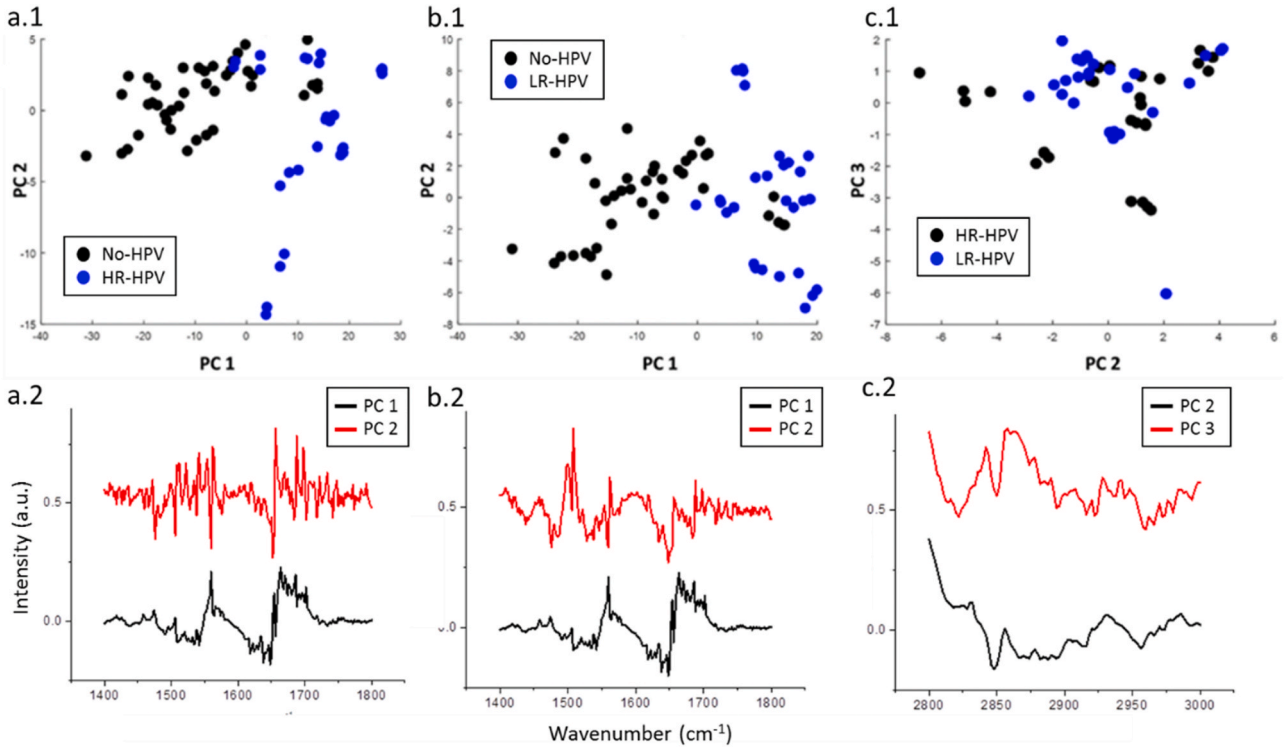

Fig. 3. PCA scatter plots and loading factors (PCs): scatter plot of scores of PC 1 and 2 and PCs for No-HPV vs HR-HPV (a.1) and plot of PC 1 and 2 (a.2), scatter plot of scores of PC 1 and 2 for No-HPV vs LR-HPV (b.1) and plot of PC 1 and 2 (b.2), and scatter plot of scores of PC 2 and 3 for HR-HPV vs LR-HPV (c.1) and plot of PC 2 and 3 (c.2). Legends: No-HPV (normal/healthy), Low-risk HPV (LR-HPV), and High-risk HPV (HR-HPV).

Table 2
PC-LDA LOOCV confusion matrix.

(a)	Normal	High-Risk HPV
Normal	40 (100%)	0
High-Risk HPV	0	28 (100%)
(b)	Normal	Low-Risk HPV
Normal	40 (100%)	0
Low-Risk HPV	0	28 (100%)
(c)	High-Risk HPV	Low-Risk HPV
High-Risk HPV	21 (75%)	7 (25%)
Low-Risk HPV	8 (29%)	20 (71%)

samples from normal samples with 76% efficiency. Interestingly, the study by Rymsza et al. [14] showed that FTIR could not distinguish between samples identified as low/high risk by cytology, but PCR showed that all samples contained HPV, proving that FTIR results were more accurate compared to cytology.

With respect to correlating spectral peaks with presence of HPV, it is difficult to attribute a biochemical change to HPV. The nature of multivariate analysis and its biggest plus point is that it considers multiple peaks for classification, providing more specificity than tests using single markers. In our study, visual analysis found differences in 1069, 1437, 1555, 1647, 2840, 2919, and 3287 cm^{-1} bands. Principal components suggest 1472, 1562, and 1657 cm^{-1} bands (lipids and proteins) to play a role in classifying No-HPV from HR-HPV, 1550 and 1655 cm^{-1} (proteins) to be responsible for classification between No-HPV and LR-HPV, and 2856 cm^{-1} (lipids) to distinguish High-risk from LR-HPV. It is important to note here that these changes probably not due to HPV itself, but changes in cells/tissues induced by HPV, because we are acquiring spectra from cells.

5. Conclusions

Overall, the study further demonstrates that PCR-RFLP is a highly sensitive technique for detecting the type of HPV in cervical samples and distinguishing High-risk HPV containing samples from Low-risk HPV samples; and that FTIR can rapidly identify No-HPV samples with 100% efficacy in relation to HR/LR-HPV and High/Low-risk HPV with 75% and 71%, respectively.

Author statement

Igor Martins Alves Melo: Conceptualization, methodology, software, formal analysis, investigation and writing.

Magda Rogéria Pereira Viana: Conceptualization and methodology.

Breno Pupin: Methodology and validation.

Tanmoy Tapobrata Bhattacharjee: Formal analysis and writing (review and editing),

Renata de Azevedo Canevari: Supervision, project administration, writing (review and editing) and funding acquisition

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported in part by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Project 2013/14464–9) and additional funding was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Project 800985/2018–4). We appreciate Taciana Rymsza for their collaboration in collecting the samples and providing the clinicopathological of patients.

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA. Cancer J. Clin. 68 (2018) 394–424, https://doi.org/10.3322/caac.21492.

[2] G.A. Mishra, S.A. Pimple, S.S. Shastri, An overview of prevention and early detection of cervical cancers, Indian J. Med. Paediatr. Oncol. 32 (2011) 125–132, https://doi.org/10.4103/0971-5851.92808.

[3] E.M. Burd, Human papillomavirus and cervical cancer, Clin. Microbiol. Rev. 16 (2003) 1–17, https://doi.org/10.1128/CMR.16.1.1.

[4] E.J. Crosbie, M.H. Einstein, S. Franceschi, H.C. Kitchener, Human papillomavirus and cervical cancer, Lancet 382 (2013) 889–899, https://doi.org/10.1016/S0140-6736(13)60022-7.

[5] J. Coser, T. da R. Boeira, A.S. Kazantzi Fonseca, N. Ikuta, V.R. Lunge, Human papillomavirus detection and typing using a nested-PCR-RFLP assay, Braz. J. Infect. Dis. 15 (2011) 467–472, https://doi.org/10.1016/S1413-8670(11)70229-X.

[6] L.L. Villa, L. Denny, CHAPTER 7 Methods for detection of HPV infection and its clinical utility, Int. J. Gynecol. Obstet. 94 (2006) S71–S80, https://doi.org/10.1016/S0020-7292(07)60013-7.

[7] B.R. Wood, M. Kiupel, D. McNaughton, Progress in fourier Transform infrared spectroscopic imaging applied to venereal cancer diagnosis, Vet. Pathol. 51 (2014) 224–237, https://doi.org/10.1177/0300985813501340.

[8] P.J. Snijders, A.J. van den Brule, H.F. Schrijnemakers, G. Snow, C.J. Meijer, J. M. Walboomers, The use of general primers in the polymerase chain reaction permits the detection of a broad spectrum of human papillomavirus genotypes, J. Gen. Virol. 71 (Pt 1) (1990) 173–181, https://doi.org/10.1099/0022-1317-71-1-173.

[9] H.U. Bernard, S.Y. Chan, M.M. Manos, C.K. Ong, L.L. Villa, H. Delius, C.L. Peyton, H.M. Bauer, C.M. Wheeler, Identification and assessment of known and novel human papillomaviruses by polymerase chain reaction amplification, restriction fragment length polymorphisms, nucleotide sequence, and phylogenetic algorithms, J. Infect. Dis. 170 (1994) 1077–1085, https://doi.org/10.1093/infdis/170.5.1077.

[10] M. Aref-Adib, T. Freeman-Wang, Cervical cancer prevention and screening: the role of human papillomavirus testing, Obstet. Gynaecol. 18 (2016) 251–263, https://doi.org/10.1111/tog.12279.

[11] C.M. de Oliveira, J.H.T.G. Fregnani, J.P. Carvalho, A. Longatto-Filho, J.E. Levi, Human papillomavirus genotypes distribution in 175 invasive cervical cancer cases from Brazil, BMC Canc. 13 (2013), https://doi.org/10.1186/1471-2407-13-357.

[12] M.P.D.N. Sartori, L.P. Agostini, T.T. Tovar, R.S. Reis, R.S. Dettogni, E. de V.W. dos Santos, A.C. Filho, I.D. Louro, Presence of infection and analysis of HPV subtypes in girls younger than 9 years old attended at a referral service in Espírito Santo, Brazil, J. Med. Virol. 90 (2018) 761–766, https://doi.org/10.1002/jmv.24992.

[13] M.R.P. Viana, I.M.A. Melo, B. Pupin, L.J. Raniero, R. de A. Canevari, Molecular detection of HPV and FT-IR spectroscopy analysis in women with normal cervical cytology, Photodiagnosis Photodyn. Ther. 29 (2020) 1–6, https://doi.org/10.1016/j.pdpdt.2019.101592.

[14] T. Rymsza, E.A. Ribeiro, L.F. das, C.e.S. de Carvalho, T. Bhattacharjee, R. de Azevedo Canevari, Human papillomavirus detection using PCR and ATR-FTIR for cervical cancer screening, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 196 (2018) 238–246, https://doi.org/10.1016/j.saa.2018.02.004.