Human Papillomavirus DNA extraction and genotype analysis by multiplex real time polymerase chain reaction from formalin fixed paraffin wax-embedded cervical carcinoma specimens

Brendon Price

Student number: PRCBRE004

A dissertation submitted in partial fulfilment of the requirements for the degree of Master of Medicine (Anatomical Pathology)

Division of Anatomical Pathology

Department of Pathology

Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

Date of submission: 19 July 2019

Supervisor: Professor D Govender

Co-supervisor: Associate Professor R Naidoo

Division of Anatomical Pathology

DECLARATION

I, Dr Brendon Price, hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither with whole work nor any part of it has been, is been, or is to be submitted for another degree is this or any other university. I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 15/07/2019

ACKNOWLEDGEMENTS

Prof Dhirendra Govender. Thank you for allowing me freedom to choose this MMed topic and to plan and direct experiments.

Dr Amsha Ramburan. Thank you for your assistance in the lab, advice regarding various aspects of molecular biology and results interpretation. Thank you also for kindly helping go through the drafts of this thesis and being so generous with your time.

Dr Jason Sutherland from Celtic Diagnostics for his advice and support regarding the RT-PCR experiments.

Subash Govender. Thank you very much for specimen sectioning, performing part of the DNA extractions and initial housekeeping gene end point PCR.

Associate Professor Richard Naidoo. Thank you for your support and advice and permitting access to the LightCycler™ 480 II real time PCR machine.

Raymond Kriel. Thank you for assistance in lab functioning and procurement of lab consumables.

The NHLS Research Trust for help in funding this study.

ABSTRACT

Introduction:

Cervical squamous cell carcinoma is most commonly caused by persistent infection by high risk human papillomavirus (hrHPV) genotypes. The exact type of hrHPV varies geographically and is the basis for HPV–based vaccination for cervical squamous cell carcinoma prevention. Little is known regarding local hrHPV genotypes within the Western Cape population of South Africa.

Aims and objectives:

This was a pilot study aiming to extract of high quality genomic DNA from archival FFPE cervical squamous cell carcinoma cases and identify hrHPV genotypes by multiplex real time PCR (RT-PCR).

Materials and methods:

A retrospective search identified a total of 57 cases of cervical squamous cell carcinoma for the period 2004-2014. This was reduced to a final number of 23 that exhibited sufficient tumour burden for DNA extraction. The most common age group was 40-49 years. HIV status was as follows: two HIV-positive, 14 HIV-negative and 7 HIV unknown. DNA was extracted from archival FFPE cervical squamous cell carcinoma samples using QIAGEN QIAamp® DNA FFPE Tissue kit. Housekeeping genes were detected by endpoint PCR using standard primers for *glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*) sequences to determine the quality and integrity of extracted DNA for downstream PCR amplification experiments. HrHPV DNA amplification was optimised using a touchdown PCR technique with L1 consensus gene GP5+/GP6+ primers. HrHPV genotypes were detected using a four colour multiplex hrHPV genotyping kit. Samples showing positive results in overlapping probe filter

detection spectra were subjected to DNA Sanger sequencing for final confirmation of specific hrHPV genotype.

Results:

Standard xylene DNA extraction methods using QIAamp® system yielded adequate amounts of DNA with average final concentration of 463.2 ng/μl and A_{260}/A_{280} ratio of 1.86. Housekeeping genes were successfully detected in all samples, confirming that no significant DNA degradation of target sequences occurred within the archival time range of 2004-2014. HPV L1 detection via GP5+/GP6+ primers with endpoint PCR was not achieved via standard cycling conditions and required the use of a touchdown technique with gradually decreasing annealing temperatures. This method successfully identified HPV L1 sequences in 22 out of 23 cases.

Multiplex RT-PCR with four colour hydrolysis probes identified hrHPV genotypes in 22 of 23 cases with relative frequencies of HPV genotypes: 16>>18=39=45>33. Most cases showed infection with a single hrHPV genotype (HPV 16 and one case with HPV 33) with four cases demonstrating two genotypes (two with HPV 16&18, one with 16&33 and one with 39&45) and one case with three genotypes (HPV 16, 39, 45). Interestingly, none of the HIV-positive cases showed multiple hrHPV genotype infection. Four hrHPV cases with overlapping spectra for HPV 18/31 and 45/59 were subjected to Sanger sequencing for confirmation of genotype. Three of four cases showed 100% match for genotypes 18 and 45 with the final case demonstrating only co-infective HPV 16.

Conclusion:

Commercial DNA extraction kits yield adequate amounts of intact, amplifiable DNA in archival FFPE cervical carcinoma specimens. Touchdown PCR is necessary for HPV detection in extracted FFPE DNA cases using GP5+/GP6+ L1 primers. RT-PCR using multicolour hydrolysis probes is a rapid, sensitive technique for hrHPV genotype screening of cervical squamous cell carcinoma specimens. A three colour detection system rather than four colour kit is recommended for future studies in order to avoid extra cost in DNA sequencing cases with overlapping spectra. This pilot study demonstrates hrHPV genotype prevalence similar to that in other populations and suggests that vaccination with currently available formulations would provide a sufficiently wide coverage of HPV genotypes. Future studies will include application of the FFPE DNA extraction, endpoint PCR and RT-PCR techniques to the remainder of the cases in the original cohort.

TABLE OF CONTENTS

DECLARATION ii
ACKNOWLEDGEMENTS iii
ABSTRACT iv
TABLE OF CONTENTS vii
LIST OF TABLES ix
LIST OF FIGURES x
LIST OF ABBREVIATIONS: xi
CHAPTER ONE 1
INTRODUCTION 1
1.1 HPV and cervical carcinoma 1
1.2 The HPV genome, life cycle and classification 4
1.3 Local HPV genotype knowledge and the influence of HIV 8
1.4 Other HPV-associated carcinomas 10
1.5 Methods employed for HPV DNA extraction and identification of hrHPV in clinical samples 11
1.5.1 DNA extraction techniques 11
1.5.2 Methods for detection of HPV and associated genotypes from extracted DNA 12
1.5.3 HPV genotyping at Anatomical Pathology, GSH 21
1.6 Aims and objectives of the current study 22
CHAPTER TWO 23
MATERIALS AND METHODS 23
2.1 Ethics approval 23
2.2 Study design 23
2.3 Case selection 23
2.4 Sectioning of tissue for DNA extraction 24
2.5 DNA extraction and quantitation 24
2.6 DNA spectrophotometry 26
2.7 Reference gene PCR 26
2.8 HPV L1 primer synthesis 27
2.9 Optimisation of HPV L1 consensus sequence endpoint PCR using GP5+/GP6+ primers . 28
2.10 Reamplification endpoint PCR of HPV L1 gene using GP5+/GP6+ primers 29
2.11 Touchdown PCR optimisation using GP5+/GP6+ primers 29
2.12 Application of optimised GP5+/GP6+ HPV L1 touchdown PCR conditions to all extracted samples 30
2.13 DNA agarose electrophoresis 30
2.14 Detection of hrHPV genotypes by multiplex real time PCR (RT-PCR) 31

2.15 Confirmation of hrHPV genotypes by DNA sequencing and BLAST alignment 33
CHAPTER THREE 35
RESULTS 35
3.1 Sample group 35
3.1.1 Patient demographics 35
3.1.2 HIV distribution within cases 35
3.2 DNA extraction from FFPE archival tissue 36
3.3 Endpoint PCR 36
3.3.1 Housekeeping gene PCR using reference gene primers 36
3.3.2 HPV L1 end point PCR using GP5+/GP6+ primers 37
3.3.3 Reamplification HPV L1 end point PCR using GP5+/GP6+ primers 37
3.3.4 Touchdown PCR optimisation of HPV L1 using GP5+/GP6+ primers 38
3.3.5 Touchdown PCR of cervical carcinoma samples 39
3.4 hrHPV genotypes detected by multiplex RT-PCR on LightCycler™ 480 II 40
3.4.1 RT-PCR curves of supplied positive control DNA samples for hrHPV and human internal control, *β-globin* 41
3.4.2 Detection of FFPE internal control (housekeeping) genes by RT-PCR 43
3.4.3 Detection of hrHPV genotypes by multiplex RT-PCR 45
3.5 Confirmation of hrHPV L1 sequences via DNA Sanger sequencing and BLAST alignment 53
CHAPTER FOUR 54
DISCUSSION 54
4.1 Case demographics 54
4.2 DNA extraction and housekeeping gene endpoint PCR 55
4.3 HPV L1 amplification by endpoint PCR 56
4.4 Optimisation of touchdown PCR parameters for GP5+/GP6+ primers 57
4.5 Real time PCR (RT-PCR) detection of hrHPV genotypes 58
4.6 HIV status and hrHPV distributions 62
4.7 Limitations of the current study 63
4.8 Application to other HPV related carcinomas in different anatomical sites 64
CHAPTER FIVE 65
CONCLUSION 65
REFERENCES 66
APPENDIX 1 72
APPENDIX 2 74
APPENDIX 3 76

LIST OF TABLES

Table 1. 1 Various modalities available for hrHPV genotype identification.....20

Table 2. 1 PCR Master mix recipe for housekeeping gene.....27
Table 2. 2 End point PCR for housekeeping gene amplification.....27
Table 2. 3 Primer sets used in the amplification of housekeeping gene and HPV L1 consensus gene.....28
Table 2. 4 End point PCR optimization for HPV L1 amplification by GP5+/GP6+ primers 29
Table 2. 5 Touchdown PCR setup for HPV L1 amplification by GP5+/GP6+ primers...30
Table 2. 6 Absoprtion and emission spectra of dyes with available filter set.....31
Table 2. 7 Recipe for Sacace RT-PCR kit 32
Table 2. 8 RT-PCR conditions for multiplex identification of hrHPV genotypes.....33

Table 3. 1 hrHPV genotypes identified by RT-PCR.....51
Table 3. 2 BLAST alignment of DNA sequences from selected cases.....53

Table 4. 1 HIV prevalence in South African women with CIN and cervical carcinoma..55
Table 4. 2 Most common hrHPV genotypes in cervical carcinoma in sub-Saharan Africa 62

LIST OF FIGURES

Figure 1. 1 Age standardised rates (ASR) of cervical cancer incidence in sub-Saharan Africa. 2
Figure 1. 2 Age standardised rates (ASR) of cervical cancer mortality in sub-Saharan Africa. 2
Figure 1. 3 Schematic diagram of the HPV genome and associated genes (Riemer *et al*., 2010) 6
Figure 1. 4 Diagram showing the development and progression of cervical intraepithelial neoplasia to invasive carcinoma (taken from Cohen *et al*., 2019)....8
Figure 1. 5 Diagram of HPV L1 consensus gene with various primers and products available for detection and amplification (taken from Molijn *et al*., 2005)....12
Figure 1. 6 Diagram illustrating the principles of line probe assay (LiPA) chromogenic detection of HPV....14
Figure 1. 7 Diagram outlining the principles of real time PCR (RT-PCR)......15
Figure 1. 8 Fluorescence curve for the ideal RT-PCR reaction. 16
Figure 1. 9 Detection of multiple hrHPV genotypes simultaneously using melting curve analysis 17
Figure 1. 10 Sanger dideoxynucleotide chain termination method of DNA sequencing.....18

Figure 3. 1 Number of cases of cervical squamous cell carcinoma per age group 35
Figure 3. 2 DNA electrophoresis of selected cases amplified by end-point PCR using *GAPDH* housekeeping primers.....37
Figure 3. 3 DNA electrophoresis of selected cases amplified by reamplification end-point PCR using GP5+/GP6+ HPV L1 primers.....38
Figure 3. 4 DNA electrophoresis of optimised touchdown end-point PCR using GP5+/GP6+ HPV L1 primers.....39
Figure 3. 5 DNA electrophoresis of all cases amplified by touchdown end-point PCR using GP5+/GP6+ HPV L1 primers. 40
Figure 3. 6 RT-PCR curves of supplied positive controls for HPV 16, 33, 39 and 58.. 41
Figure 3. 7 RT-PCR curves of supplied positive controls for HPV 18, 31, 45, 59, 35, 68, 52 and 66. 42
Figure 3. 8 RT-PCR curves of supplied positive controls for HPV 51, 56 and *β-globin*. 42
Figure 3. 9 RT-PCR curves of *β-globin* housekeeping gene.. 44
Figure 3. 10 RT-PCR curves of HPV 16.....46
Figure 3. 11 RT-PCR curves of HPV 18.....47
Figure 3. 12 RT-PCR curve of HPV 33. 48
Figure 3. 13 RT-PCR curves of HPV 39.....49
Figure 3. 14 RT-PCR curves of HPV 45.....50
Figure 3. 15 Frequency distribution of hrHPV genotypes identified by RT-PCR....52
Figure 3. 16 Number of hrHPV genotypes per case identified by RT-PCR 52

LIST OF ABBREVIATIONS:

A_{260}/A_{280}	Ratio of absorbance at 260 nanometers to 280 nanometers wavelength
ASR	Age standardised rates
BLAST	Basic local alignment search tool
bp	Base pairs
CIN	Cervical intraepithelial neoplasia
Cy5	Cyanine5 dye
Ct	Threshold cycle
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
ddNTP	Dideoxynucleotide triphosphate
EDTA	Ethylenediaminetetra-acetic acid
FAM	Fluorescein amidite dye
FDA	Food and Drug Administration
FFPE	Formalin-fixed paraffin-wax embedded
FRET	Fluorescence resonance energy transfer
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
gDNA	Genomic (template) DNA
GSH	Groote Schuur Hospital
GP5+/GP6+	General primer with elongation (+) at 3' ends
H&E	Haematoxylin and eosin
HIV	Human immunodeficiency virus

HPV	Human papillomavirus
HREC	Human Research Ethics Committee
hrHPV	high risk HPV genotype
lrHPV	low risk HPV genotype
HSIL	High grade squamous intraepithelial lesion
JOE	6-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein
L1	Late gene (major capsid protein)
LCR	Long control region
LiPA	Line probe assay
LSIL	Low grade squamous intraepithelial lesion
µl	Microliter
Mol	Mole
mM	Millimolar
µM	Micromolar
NHLS	National Health Laboratory Service
nm	Nanometers
OD	Optical density
ORF	Open reading frame(s)
PCR	Polymerase chain reaction
ROX	6-carboxyl-X-rhodamine
RT	Room temperature
RT-PCR	Real time PCR
SEER	Surveillance, epidemiology and end results program
SCC	Squamous cell carcinoma
SPF_{10}	Short PCR fragment

Taq	*Thermus aquaticus*
TBE	Tris-borate-EDTA buffer
Tm	Melting temperature
UCT	University of Cape Town
UV	Ultraviolet
WHO	World Health Organisation

CHAPTER ONE

INTRODUCTION

1.1 HPV and cervical carcinoma

Human papillomavirus (HPV) is a double stranded deoxyribonucleic acid (DNA) virus that belongs to the genus *Alphapapillomavirus* (Kocjan *et al*., 2016). The other four HPV genera include beta, gamma, mu and nu. HPV demonstrates tropism towards epithelial cells, infecting genital and oral skin and mucosa (Schiffman *et al*., 2007). The HPV family of viruses includes over 200 types but only a minority of these are implicated in oncogenic transformation (Choi and Park, 2015). The identification of HPV as the aetiological cause of cervical cancer was made in the 1970's by German virologist Dr Harald zur Hausen, for which he won the 2008 Nobel Prize in Physiology or Medicine. Each year approximately 266 000 women worldwide die of cervical cancer, 90% of whom live in middle to low income countries. It is the leading cause of cancer deaths in Eastern and Central Africa (World Health Organisation, 2014).

South Africa has both a relatively high incidence and mortality rate for cervical carcinoma, similar to other sub-Saharan African countries (Figures 1.1 and 1.2). In South Africa, cervix carcinoma is the second most common female cancer and has the highest age-standardised death rate of all the top female cancers from 2000-2012 (World Health Organisation, 2014). This is thought to be due to a host of contributing factors including poor screening rate, lack of awareness, poor access to health care facilities, late presentation of disease (De Vuyst *et al*., 2013, Denny *et al*., 2014). The high rate of HIV infection is also thought to contribute to the high incidence rate (by accelerating the progression from cervical dysplasia to carcinoma) and mortality rate (Denny *et al*., 2014).

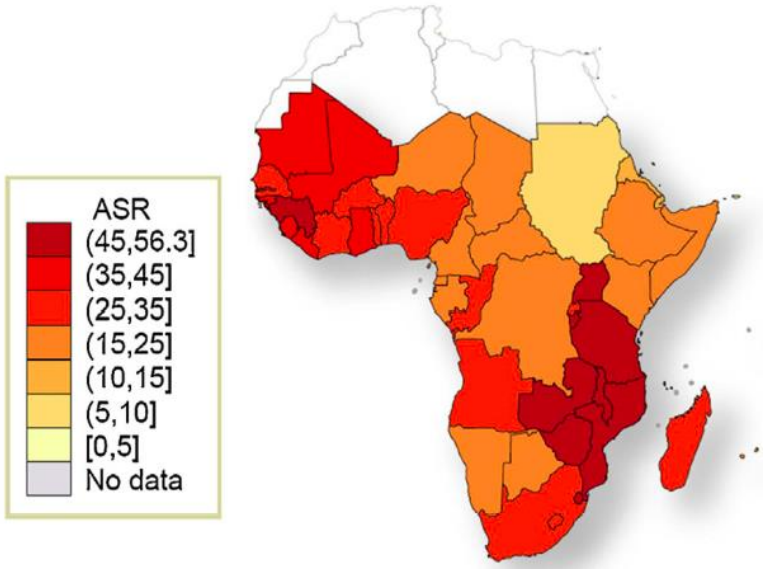


Figure 1. 1 Age standardised rates (ASR) of cervical cancer incidence in sub-Saharan Africa. Numbers are per 100,000 women (from De Vuyst *et al*., 2013).

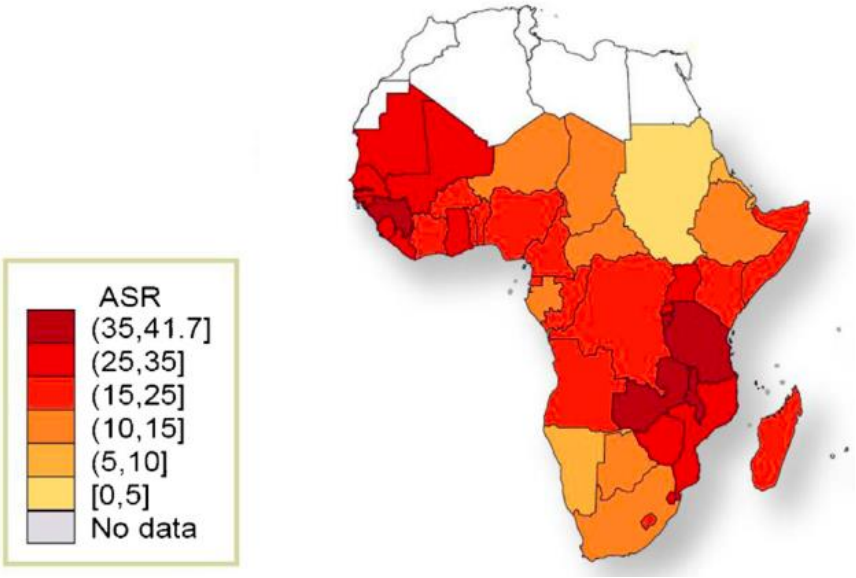


Figure 1. 2 Age standardised rates (ASR) of cervical cancer mortality in sub-Saharan Africa. Numbers are per 100,000 women (from De Vuyst *et al*., 2013)**.**

Cervical carcinoma encompasses squamous cell carcinoma and adenocarcinoma with squamous cell carcinoma accounting for at least three quarters of all carcinomas (Vinh-Hung *et al*., 2007). HrHPV infection is associated with the majority of cervical squamous cell carcinomas (reaching almost 100%) and adenocarcinomas (62-100%) (Pirog, 2017) however a small portion of either carcinoma arises by non-HPV oncogenic pathways.

Non-HPV associated adenocarcinomas of the cervix include gastric, endometrial, villoglandular, serous and mesonephric adenocarcinoma (Rodriguez-Carunchio *et al*., 2015). Non-HPV related squamous cell carcinoma of the cervix is associated with a poorer prognosis compared to conventional hrHPV-related cervical squamous cell carcinoma. This poorer prognosis is also noted in non-HPV related squamous cell carcinomas of the head and neck region arising from alcohol and tobacco use (Mirghani and Blanchard, 2018).

Surveillance, Epidemiology and End Results (SEER) data from the United States of America of 30 989 samples for the period 1973-2002 show that the numbers of new cases of squamous cell carcinoma have steadily declined (as a result of increased Pap screening) while some histological subtypes have increased, such as adenocarcinoma, mucinous and adenosquamous (Vinh-Hung *et al*., 2007). This trend of decreasing squamous cell carcinoma is not reflecting in developing countries where screening facilities are limited and HPV vaccination unavailable.

Cervical carcinoma variants are associated with different prognoses. For example, patients with small cell carcinoma of the cervix have worse 10-year overall survival data (31.6%) than mucinous (43%) or adenosquamous carcinoma (50.7%) who, in turn, have worse overall survival compared to squamous cell carcinoma (52.3%) (Vinh-Hung *et al*., 2007).

The WHO currently recognises superficially invasive squamous cell carcinoma (depth of invasion less than 3 mm) and invasive squamous cell carcinoma. Variants of invasive squamous cell carcinoma include keratinsing, non-keratinising, basaloid, warty, papillary, verrucous, lymphoepithelial and squamotransitional (Vinh-Hung *et al*., 2007).

1.2 The HPV genome, life cycle and classification

The structure of HPV includes an icosahedral capsid housing viral DNA. The HPV genome is circular and divided into three regions with eight overlapping open reading frames (ORF) (Riemer *et al*., 2010). These include an early region (E), late region (L) and a long control region (LCR) (Figure 1.3).

1. The early region contains genes *E1, E2, E4, E5, E6* and *E7*. E1 and E2 are involved in transcription and replication. E6 and E7 bind to and degrade tumour suppressor gene products p53 and retinoblastoma (Rb) respectively and play critical roles in oncogenesis. E7 also interacts with other protein targets and is involved in allowing the host cell to evade apoptosis and immune surveillance (Choi and Park, 2015). The roles of E4 and E5 are less well understood.

2. The late region contains genes *L1* and *L2*. These encode major and minor capsid proteins, respectively. These genes are expressed prior to viral escape from host cells. L2 is also essential for permitting viral entry into host cells.

3. Long control region (LCR). This is a non-coding regulatory region that is positioned upstream between genes E6 and L1. This region contains a core promoter sequence which is involved in regulating DNA replication and gene transcription. The DNA sequence shows variation between HPV genotypes (Choi and Park, 2015).

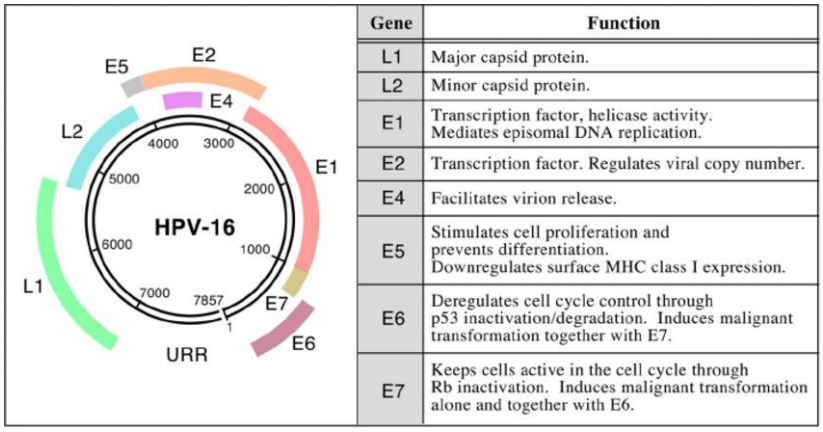
HPV genotypes are classified on the basis of differing L1 gene sequences. Those with >10% differences are considered different genotypes whereas those with >1% and <1% are considered intratypic and sub-lineage variants, respectively (Choi and Park, 2015). HPV genotypes are subdivided into low risk and high risk.

1. High risk HPV genotypes (hrHPV) include 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68. These genotypes are implicated in the development of cervical intra-epithelial neoplasia (CIN) and cervical cancer (squamous cell carcinoma and adenocarcinoma).

2. Low risk HPV genotypes (lrHPV) include 6, 11, 34, 40, 42, 43, 44, 53, 54, 70 and 74. The genotypes are implicated in the development of genital warts, which usually require local surgical treatment but are not implicated in the development of CIN or cancer (Bernard *et al*., 2010).

Most sexually active women become infected with HPV during their lifetime but only approximately 10% will become persistently infected. HrHPV 16 and 18 are responsible for ~70% of cervical cancer (Bateman *et al*., 2015).

Recent years have seen the development and U.S. Food and Drug Administration (FDA) approval of three different vaccines targeting HPV genotypes implicated in genital warts (lrHPV 6 and 11) and cervical cancer (hrHPV 16 and 18) (Schiffman *et al*., 2007; Denny *et al*., 2014). Vaccination of female children prior to sexual debut is envisaged to greatly reduce the burden of cervical cancer. These include bivalent vaccine Gardasil (Merck, 2006), quadrivalent vaccine Cervarix (Glaxo Smith-Kline, 2007) and nonavalent vaccine Gardasil 9 (Merck, 2014) (Denny *et al*., 2014). Vaccines contain E6/E7 oncoprotein epitopes that elicit a strong cellular immune response and can clear hrHPV lesions. Studies have suggested cross-protection against other less common hrHPV genotypes (Denny *et al*., 2014; Salazar *et al*., 2015).



Gene	Function
L1	Major capsid protein.
L2	Minor capsid protein.
E1	Transcription factor, helicase activity. Mediates episomal DNA replication.
E2	Transcription factor. Regulates viral copy number.
E4	Facilitates virion release.
E5	Stimulates cell proliferation and prevents differentiation. Downregulates surface MHC class I expression.
E6	Deregulates cell cycle control through p53 inactivation/degradation. Induces malignant transformation together with E7.
E7	Keeps cells active in the cell cycle through Rb inactivation. Induces malignant transformation alone and together with E6.

Figure 1. 3 Schematic diagram of the HPV genome and associated genes (Riemer *et al.*, 2010)

HrHPV genotypes responsible for the development of cervical, penile, anal and head and neck cancers are sexually transmitted. The life cycle is most well understood in the context of cervical cancer where the squamocolumnar junction of the transformation zone has been identified as the point of HPV infection (Figure 1.4) (Schiffman *et al.*, 2007; Cohen *et al.*, 2019). HPV infects the basal, replicating cells of the epithelium. Initially HPV replication occurs in tandem with host cell replication. E1 and E2 HPV gene products are required for the maintenance of viral infection within the host cell and recruit host DNA polymerase to initiate viral genome replication. E6 and E7 allow for bypassing of normal host apoptotic activation with viral infection and they promote increased cellular proliferation and viral production. As the cervical epithelial cells mature as they progress towards the surface, activation of the late genes L1 and L2 allow for viral protein capsid formation so that intact virions may be assembled and shed when the uppermost layer of host cells are shed. E4 interacts with host cytoskeletal proteins and is responsible for the formation of koilocytes within the uppermost regions of the epithelium (Schiffman *et al.*, 2007; Choi and Park, 2015). The actions of E6 and E7 oncoproteins may initiate additional genetic injuries to the host cell as a result of p53 and Rb destruction and the resultant dysplastic cells undergo a switch from "productive" infection, where HPV DNA is incorporated into capsids to produce infectious virions, to "non-

productive" infection where late gene expression is down regulated and virions are not assembled (Depuydt *et al*., 2016).

Infection of cervical epithelial cells within the transformation zone evolves through a stepwise progression of cytomorphological changes within host cells, termed CIN I-III (or low grade squamous intra-epithelial lesion/LSIL to high grade squamous intra-epithelial lesion/HSIL). HPV may exist episomally (as seen in non-malignant and pre-malignant tissue) or integrated within the host genome (largely seen in malignancies) (Cohen *et al*., 2019).

CIN I/LSIL is characterised by the appearance of koilocytic changes predominantly within the upper epithelial layers together with increased nuclear:cytoplasmic ratios and vertical orientation of basal epithelial cells that occupies less than one third the thickness of the epithelium (Cohen *et al*., 2019). Koilocytic atypia refers to the appearance of the host nucleus and includes nuclear enlargement with nuclear membrane irregularity and perinuclear halos. CIN II and III (HSIL) describe the progression of cellular atypia with dysplastic cells occupying less than two-thirds and full thickness of the epithelium, respectively. Koilocytosis is not a feature in CIN III, reflecting the down regulation of HPV viral capsid proteins involved in disrupting host cellular cytoskeletal proteins. Progressive accumulation of further genetic mutations is associated with development of invasive squamous cell carcinoma from CIN III.

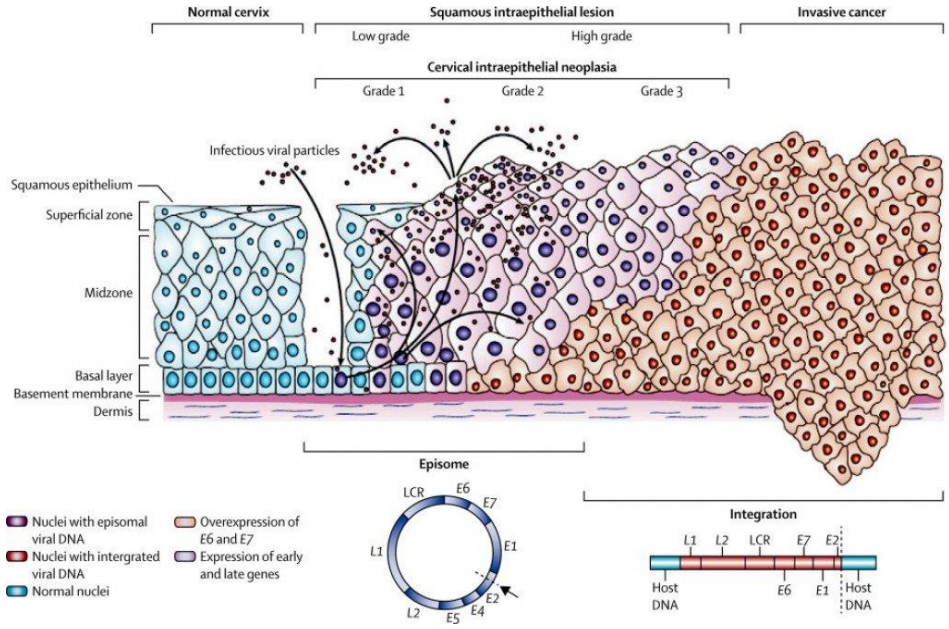


Figure 1. 4 Diagram showing the development and progression of cervical intraepithelial neoplasia to invasive carcinoma (taken from Cohen *et al*., 2019)

1.3 Local HPV genotype knowledge and the influence of HIV

The development of HPV vaccines to provide immunity against lrHPV and hrHPV genotypes are based upon the most common genotypes associated with the development of cervical cancer in the developed world. There is some concern that the hrHPV strains found most commonly in developed nations may not reflect the range seen in developing countries where cervical cancer burdens are greater. In addition, the co-infection with HIV in women may result in different hrHPV genotype prevalence in high grade CIN or cervical carcinoma. In many African countries, the local prevalence and genotypic spread of HPV is unknown (Bateman *et al*., 2015).

Recent studies have shown that HIV-positive women carry a higher risk of infection with non-HPV 16, non-HPV 18 high risk genotypes which may not be covered by currently available commercial vaccines (McDonald *et al*., 2012). In addition, HIV-positive women carry a higher risk of infection with multiple HPV genotypes and infections are more likely to be persistent (Denny *et al*., 2014).

Studies of HPV prevalence in HIV negative and positive South African men and women (Mbulawa *et al*., 2015) have shown:

1. HPV prevalence of 36.7% in HIV negative women compared to 86.4% in HIV positive women.

2. HPV prevalence decreased with increasing age in HIV negative women but not with HIV positive women.

3. HIV positive women had a greater prevalence of one or more HPV genotypes compared to HIV negative women.

HPV genotype analyses have been described for invasive cervical carcinoma in Zambian and Ugandan women (Odida *et al*., 2008; Bateman *et al*., 2015). The Zambian study showed HPV 16 and/or 18 genotypes in 69.9% of women. The Ugandan cohort showed that the majority of women had a single infection (96.3%) and HPV 16 and 18 genotypes accounted for 80% of cases with single infection. In general, women in Zambia and Uganda have similar hrHPV genotype profiles in cervical carcinoma and will likely benefit from the commercially available vaccines.

A multicentre study that included women from South Africa, Ghana and Nigeria investigated the HPV prevalence and type distribution in invasive cervical cancer (Denny *et al*., 2014). The HPV prevalence rate in invasive cervical carcinoma was 90.4% with the vast majority of women exhibiting a single infection (86.8%). The prevalence of HPV infection and the presence of multiple HPV infections was higher in HIV-positive women, however the high rate of unknown HIV status in women from Ghana and Nigeria (96.9% and 82.1%, respectively) limited thorough analysis. In contrast, only 11% of South African women had unknown HIV status. The most common HPV genotypes, in descending order were: 16, 18, 35, 45, 33 and

52. With the high burden of HIV in sub-Saharan Africa and reports showing differing HPV genotype associations with cervix cancer, future vaccine design may be required to reflect these differences.

1.4 Other HPV-associated carcinomas

As anal, vulval, penile and head and neck cancers are also related to oncogenic HPV infection (Molijn *et al*., 2005; De Vuyst *et al*., 2013), vaccination has the potential benefit to also prevent these cancers. Penile carcinomas, particularly the warty and basaloid subtypes are associated with HPV 16 and 18 genotypes. Oropharyngeal squamous cell carcinoma (predominantly tonsillar and base of tongue) is associated with HPV 16 infection, together with smoking and alcohol (Mirghani and Blanchard, 2018). Recent years have shown a sharp increase in the incidence of HPV-associated oropharyngeal carcinoma (Van Dyne *et al*., 2018). HPV-associated oropharyngeal carcinoma has been associated with improved disease-free survival post chemoradiation and surgery compared to HPV-negative oropharyngeal squamous cell carcinoma (Mirghani and Blanchard, 2018). The implication of hrHPV 16 and 18 in non-cervical carcinomas raises the possibility that vaccination may also provide benefit for the prevention of these HPV-related cancers. This would mean widening the vaccination population to also include boys. Data regarding the hrHPV genotype range in HIV-positive patients with penile, anal and oropharyngeal carcinoma is limited and is therefore an important area to investigate in populations with high HIV incidence, such as sub-Saharan Africa.

1.5 Methods employed for HPV DNA extraction and identification of hrHPV in clinical samples

1.5.1 DNA extraction techniques

Extraction of high quality DNA from fixed archival tissues poses a number of challenges. The process of formalin fixation results in a number of crosslinks and DNA strand breaks (Kocjan *et al*., 2016). This may interfere or completely inhibit subsequent amplification attempts by polymerase chain reaction (PCR). In addition, unbuffered formalin is commonly used in some centres, increasing the risk of damaging the structural integrity of DNA. Common steps for DNA extraction involve deparaffinisation with xylene, Proteinase K lysis followed by precipitation (or via another matrix) and reconstitution (Zhong *et al*., 2013). Improvements in DNA extraction have been previously investigated and may be used if standard commercial extraction kits do not yield adequate amounts of DNA. These additional methods include, but are not limited to:

1.5.1.1 Heat treatment

This is thought to partially reverse formalin DNA crosslinks. This method has recently been successfully applied in the extraction of FFPE HPV DNA in a Zambian cross-sectional study where mean extracted nucleic acid concentration using heat methods (99°C, 30 min) prior to xylene extraction was significantly higher than standard room temperature xylene extraction (117.4 ng/μl versus 38.2 ng/μl, $p<0.001$) (Bateman *et al*., 2015).

1.5.1.2 Removal of the xylene deparaffinisation step

Xylene is thought to inhibit PCR and removal of this step improves the yield of DNA extracted (Kocjan *et al*., 2016). The xylene-free, prolonged proteinase K method of DNA extraction has been shown to allow amplification of a 650 bp fragment of human β-actin DNA of in 100% of FFPE samples (Cannavo *et al*., 2012). This amplicon size is much larger than those amplified by commercially available HPV SPF_{10} primers (65 bp) and GP5+/6+ (150 bp) primers.

1.5.1.3 Mineral oil

A recent paper described efficient extraction of high quality DNA from archival FFPE samples using mineral oil for deparaffinisation (Heikal *et al*., 2014). Paraffin dissolved in mineral oil is not thought to interfere with PCR (Lin *et al*., 2009). Mineral oil also improved the time required for DNA extraction (20 min versus 75 min) and significantly reduced the number of centrifugation steps involved.

1.5.2 Methods for detection of HPV and associated genotypes from extracted DNA

1.5.2.1 PCR amplification of HPV DNA

PCR amplification is conventionally performed using primers directed at HPV L1 gene sequences. Commonly used primers include MY09/11, GP5+/6+ and SPF_{10} (Kleter *et al*., 1998; Molijn *et al*., 2005; Micalessi *et al*., 2013; Kocjan *et al*., 2016) (Figure 1.5) which produce amplicons of varying lengths. As the efficiency of PCR decreases with increasing amplicon base pair length, shorter sequences would theoretically have a higher success rate.

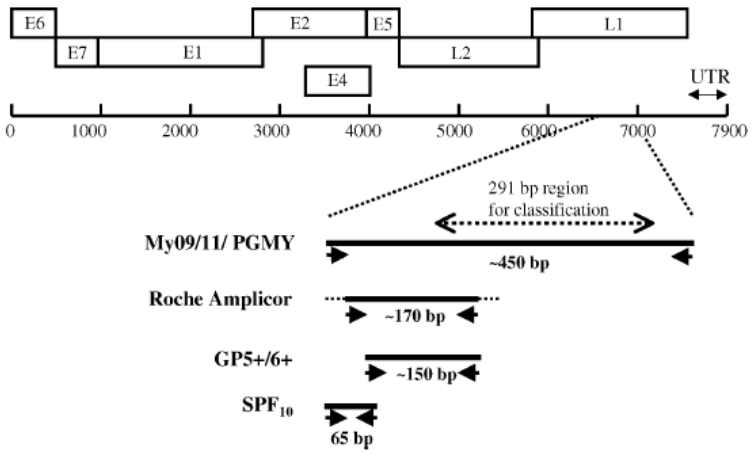


Figure 1. 5 Diagram of HPV L1 consensus gene with various primers and products available for detection and amplification (taken from Molijn *et al*., 2005)

Detection of PCR product/amplicon is performed by size separation with electrophoresis and comparison of the size of the amplicon to standard DNA base pair ladder and positive control. The larger amplicons (150 and 450 bp) are separated by agarose gel electrophoresis while the 50 bp amplicon would require polyacrylamide electrophoresis. Visualisation of the double stranded amplicons depends on the electrophoresis method employed. Intercalating fluorescent dyes (e.g., ethidium bromide) can be used for agarose and polyacrylamide gels with alternative silver precipitation methods available for polyacrylamide gels.

1.5.2.2 Detection of specific HPV genotypes from PCR amplified consensus L1 ORF DNA

For the detection of the specific HPV genotype, a number of commercial kits and sequencing-based methods are available, which are described below:

1.5.2.2.1 Line probe assay (LiPA)

LiPA assays require the least amount of specialised equipment but have the advantage of low cost and simplicity (Figure 1.6). HPV DNA is amplified by end point PCR using consensus sequence primers with inclusion of biotinylated dNTPs into the PCR reaction mix. The biotinylated PCR product is then allowed to anneal to a solid matrix to which genotype-specific complimentary DNA sequences have been attached. Complimentary binding of biotinylated DNA to specific hrHPV genotype sequences is detected using streptavidin-enzyme conjugate and chromogenic substrate (van den Brule *et al*., 2002; Molijn *et al*., 2005). The identity of the specific hrHPV band is determined by comparing the position of the band with a standard chart.

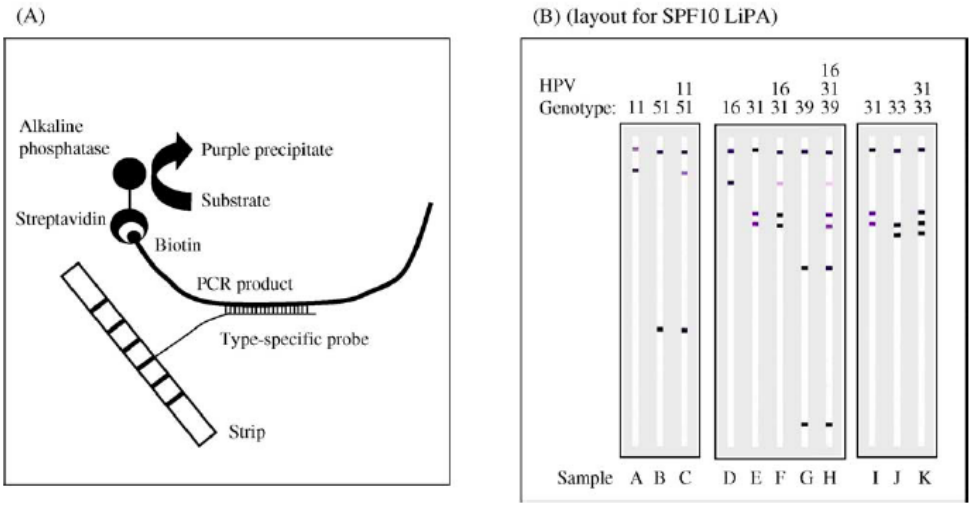


Figure 1. 6 Diagram illustrating the principles of line probe assay (LiPA) chromogenic detection of HPV. Biotinylated PCR product is annealed to immobilised specific HPV genotype sequences and detected using streptavidin-alkaline phosphatase system (A). Identification of specific HPV genotypes is accomplished by comparing position of enzyme product band with standard positions (B) (taken from Molijn *et al*., 2005).

1.5.2.2.2 Multiplex PCR

Multiplex PCR systems make use of real time PCR (RT-PCR) and hydrolysis probe technology for detection of DNA target sequences (Kocjan *et al*., 2016). Hydrolysis probes are oligonucleotide probes specific to the region of interest to be amplified but have two dyes covalently bonded to either end of the probe sequence (Figure 1.7). The fluorophore (reporter dye) is covalently liked to the 5' end of the primer with a quencher dye attached to the 3' end. The close proximity of the reporter and quencher dye to each other ensures that no reporter fluorescence is emitted. This transfer of energy between the reporter and quencher dyes is known as fluorescence resonance energy transfer (FRET). During PCR, the 5'-3' exonuclease component of *Taq* polymerase removes the reporter dye from the annealed primer sequence and thereby separates it from the quencher dye (and therefore, FRET). This results in the reporter dye emitting fluorescence when excited by light of the appropriate wavelength. In this manner, a number of different probes, each with their unique reporter-quencher dye

combination can be added to a single DNA template and different sequences detected in the same sample by using different wavelength channels for detection.

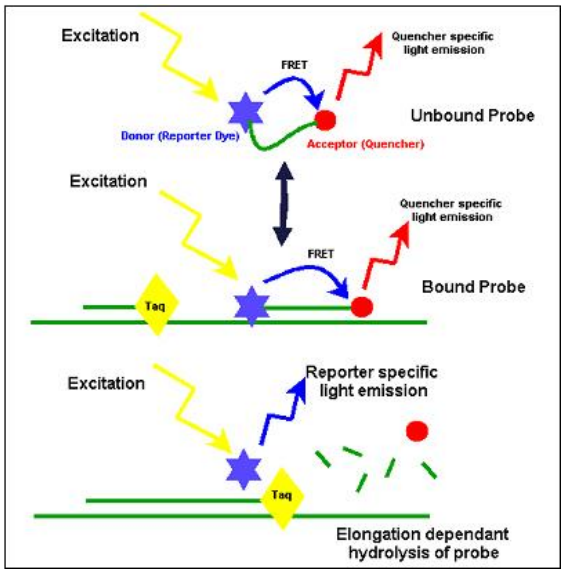


Figure 1. 7 Diagram outlining the principles of real time PCR (RT-PCR). Before PCR amplification occurs, the probe reporter dye fluorescence is absorbed by a nearby quencher by the process for fluorescence resonance energy transfer (FRET), thereby producing no detectable fluorescence in the emission spectrum of the reporter dye. Once PCR begins (at the 3' end of the separate primer) and *Taq* polymerase 5'-3' exonuclease activity releases the reporter dye from the 5' end bound probe, FRET cannot persist over wider distances and the report dye fluorescence can be detected (taken from http://ipc.nxgenomics.org/teaching/qpcr/qpcr.html).

In a RT-PCR protocol, fluorescence is detected during the extension phase of each cycle throughout the PCR (i.e., in real time). The fluorescence trace is roughly sigmoid in shape. The threshold cycle (C_t) refers to the cycle number at which the trace begins to rapidly increase in intensity (exponential phase) and crosses the threshold (i.e. increases in intensity above that of background) (Figure 1.8). C_t is a relative measure of the concentration of the target DNA within the sample.

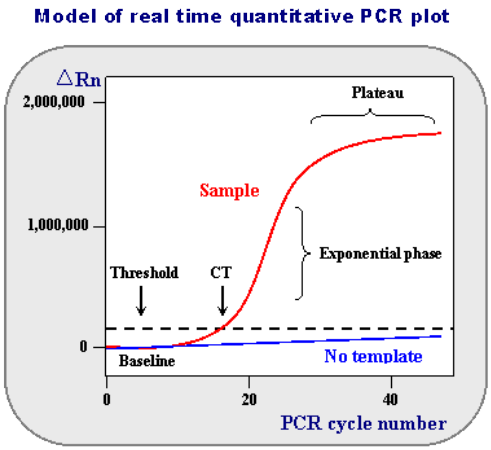


Figure 1. 8 Fluorescence curve for the ideal RT-PCR reaction. Diagram showing the curve obtained from an ideal RT-PCR reaction. The threshold cycle (C_t) is the cycle number at which the reporter dye fluorescence is greater than background. After this C_t has been reached, the increase in fluorescence enters an exponential phase until it plateaus. ΔRn = increment in fluorescence signal at each time point (Image taken from http://www.ncbi.nlm.nih.gov/probe/docs/techqpcr/).

1.5.2.2.3 Melting curve analysis

Melting curve analysis allows for the detection of multiple PCR products using various combinations of limited fluorescent dyes and take advantage of changes in fluorescence that occurs when double stranded DNA is denatured at high temperatures and allowed to reanneal in an environment of slowly cooling temperature. Either fluorescent dyes attached to probes or an intercalating dye (e.g., SYBR® Green) can be used, the latter showing fluorescence only when intercalated between annealed, double stranded DNA and not single stranded DNA. An example of melting curve analysis product for hrHPV identification is PapillomaFinder SMART20 (Ossel *et al.*, 2014) (Figure 1.9). HPV DNA is enriched in the sample by PCR amplification using universal forward and reverse primers (Step 1). Probe pairs (5'/green and 3'/black) are annealed to the amplicons and ligated using *Taq* ligase. The red region is a stuffer sequence complimentary to the genotype specific probe. The blue and brown regions are target specific regions of the probe (Steps 2 and 3). The ligated probes are amplified using universal primers with the 5' primer containing FAM dye (Step 4). ROX and Cy5 labelled detection probes are added and produce fluorescence via FRET if hybridised to their complimentary

stuffer sequence in close proximity to FAM. Owing to the difference in nucleotide sequences between genotypes, the annealed probe-stuffer sequence duplex will melt/denature at different temperatures, resulting in loss of FRET and therefore fluorescence. The negative derivative of the change of fluorescence is plotted against temperature change. Combination of dyes and differing melting temperatures results in detection of multiple genotypes in a single tube.

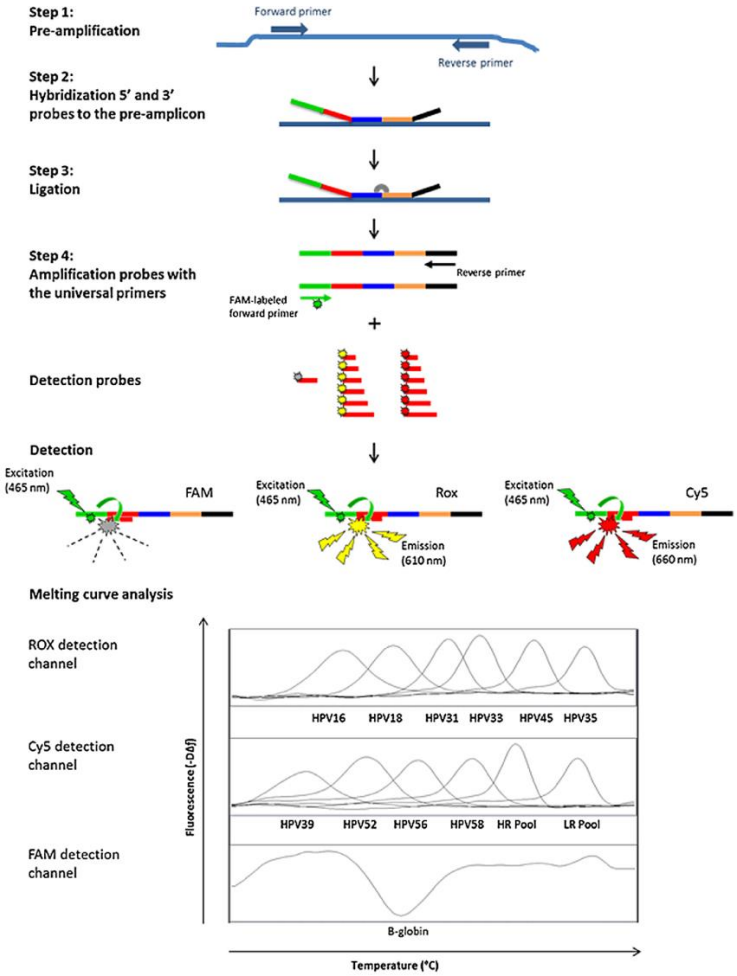


Figure 1. 9 Detection of multiple hrHPV genotypes simultaneously using melting curve analysis

1.5.2.2.4 Genotype specific primers and Sanger dideoxynucleotide sequencing

It is possible to design genotype specific primers by comparing L1 ORF hrHPV DNA sequences by alignment methods (Gharizadeh *et al*., 2005). Detection of non-overlapping regions have to however be aligned against nucleotide databases to ensure non-overlap with other viral or human gene targets. Sanger dideoxynucleotide (ddNTP) sequencing is based on

the principle of DNA chain termination during PCR using ddNTP analogues of all dNTPs, each also containing a fluorescent dye. ddNTPs are incorporated into the DNA chain synthesised by *Taq* polymerase and once incorporated, prevent further elongation. The variably sized PCR products, each with one of four ddNTP fluorophores are separated by size electrophoresis and the dyes detected by laser. The sequence of the PCR product can then be read out from 5'-3', one nucleotide at a time (Figure 1.10). This method was used in this study.

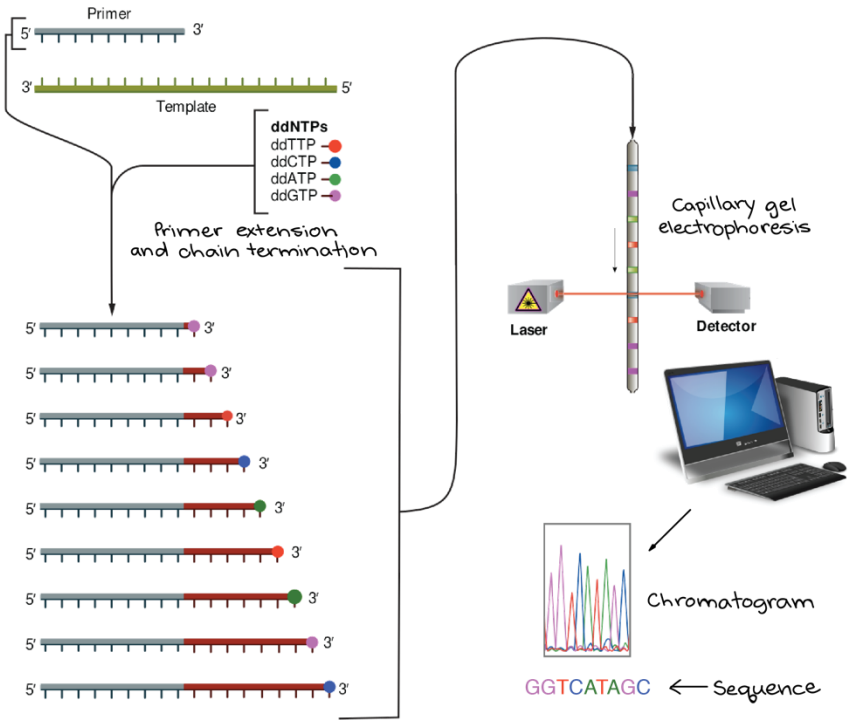


Figure 1. 10 Sanger dideoxynucleotide chain termination method of DNA sequencing. A target sequence is amplified from a forward primer with a mixture of dye-labelled dideoxynucleotides (ddNTPs) mixed with normal unlabeled deoxynucleotides (dNTPs). Competition of ddNTPs and dNTPs occur during the PCR extension step. Once a complimentary ddNTP is incorporated, the sequence is terminated as a phosphodiester bond cannot be created with the next dNTP. Varying lengths of terminated chains with dye labels are produced and separated by size using capillary electrophoresis. As the dye labelled strands pass a laser, the fluorescence peak generated corresponds to the specific ddNTP. Each strand differs in size by a single nucleotide corresponding to the target DNA sequence (5' to 3') (Image taken from http://www.orbitbiotech.com).

1.5.2.2.5 Next generation sequencing

This is also known as high throughput sequencing or massive parallel sequencing. It encompasses a number of different methodologies with the common feature of simultaneous sequencing of multiple DNA sequences. These methodologies include sequencing by synthesis (NextSeq [Illumina®]), semiconductor sequencing (Ion Torrent™ [Thermo Fisher Scientific]) and pyrosequencing (454 [Roche]) (Radford *et al*., 2012). An article exploring the use of next generation sequencing in the context of HPV in a local (Cape Town) population showed that hrHPV genotypes were identifiable via this technology which were not detected by commercial kits (Meiring *et al*., 2012). At present, access to this technology is limited and prohibitively expensive for routine diagnostic use.

The associated advantages and disadvantages of each hrHPV detection technique are tabulated below (Table 1.1).

Table 1. 1 Various modalities available for hrHPV genotype identification

HPV genotype detection system	Advantages	Disadvantages
Line probe assay (LiPA)	Cheap, multiple genotypes detection simultaneously. No need for additional equipment apart from endpoint PCR machine (i.e., RT-PCR setup, DNA sequencer).	Chromogenic substrate system, time consuming. Each sample requires separate strip for development. Bands fade over time.
Multiplex RT-PCR	Detection of multiple hrHPV genotypes within a single tube using different fluorophore spectral features.	Requires RT-PCR equipment with multiple fluorophore filters. Expensive.
Melting curve analysis	Potential for simultaneous detection of multiple hrHPV genotypes within a single reaction tube	As for multiplex RT-PCR. Multiple steps involved for probe ligation/PCR. Require appropriate software for interpretation of melt curve profile.
Sanger dideoxynucleotide DNA sequencing	Quick, relatively inexpensive. No additional expertise required for analysis of DNA sequence results.	Multiple single reactions required. Require non-overlapping genotype specific DNA sequences for primer design.
Next generation sequencing	Identification of hrHPV genotypes otherwise not identified by other techniques.	High cost, requires specialised equipment and technical skills for data analysis.

1.5.3 HPV genotyping at Anatomical Pathology, GSH

Currently, no HPV genotyping platform is present in the Division of Anatomical Pathology, Groote Schuur Hospital for FFPE tissue specimens or cytology specimens (Pap smears).

At present, HPV-related cancers are identified using the surrogate marker, p16. P16 (also known as $p16^{INK4a}$) is a tumour suppressor gene that plays a role in control of the cell cycle by inhibiting cyclin dependant kinase and preventing progression of G1 to S phase. This protein is overexpressed as a response to destruction of p53 and Rb *via* HPV proteins E6 and E7. Increased p16 expression therefore is a surrogate for hrHPV-related carcinogenesis. P16, however is not hrHPV-specific and can be upregulated in a number of high grade malignancies (e.g., high grade serous carcinoma of the endometrium). P16 expression also does not provide information regarding the specific hrHPV genotype.

Alternative methods for hrHPV detection is fluorescence *in situ* hybridisation (FISH) which currently only tests for HPV 16 and 18 (Kobayashi *et al*., 2018). This technology, though sensitive, is time consuming and is not well suited for screening multiple samples simultaneously.

Knowledge of HPV status of tumours is becoming increasingly important as HPV+ tumours are associated with better prognosis especially in the context of head and neck cancers (Kobayashi *et al*., 2018). Therefore, a need exists for extraction of high quality DNA from FFPE tissue and accurate, sensitive detection of hrHPV genotypes in cancer samples. This project aims to use commercially available kits for DNA extraction and detection of hrHPV genotypes using RT-PCR technology as this allows for the simultaneous, real time detection of multiple hrHPV genotypes in multiple samples. This has a potential advantage over traditional colorimetric assays such as LiPA in terms of shortened time required and digital output of results for immediate analysis *via* downstream software programs e.g., MSExcel.

As the incidence of HPV-related malignancies (particularly oropharyngeal carcinomas) is increasing rapidly in developed countries, the need for hrHPV detection is likely to increase.

In the context of cervical carcinoma, hrHPV genotyping is of particular importance especially in precursor lesions (SIL) and in routine cytology (Pap smears). hrHPV detection in Pap smears has the potential to reduce pathologist and cytologist workloads, increase productivity and rapidly streamline patients to colposcopic clinics. hrHPV detection has the further advantage of reducing the number of cases diagnosed as ASC-H (atypical squamous cells, cannot exclude high grade SIL), thereby reducing potentially unnecessary colposcopy visits (+/- biopsy) or repeated Pap smears.

This study is intended as a pilot study for the investigation of the feasibility of using FFPE tissue for the detection of hrHPV in cervical carcinoma specimens. For this purpose, only cases with high tumour burdens were chosen. Based on the success of the high burden cases, screening of the entire cohort would be performed together with testing other HPV related tumours. Future application to cervical Pap smears could be considered if the RT-PCR platform is successful.

1.6 Aims and objectives of the current study

1. To identify HPV DNA, by polymerase chain reaction (PCR) using HPV L1 consensus primers, in a cohort of cervical cancer patients diagnosed at GSH.
2. To determine the specific HPV genotype/s in each sample by using a commercially available real-time PCR kit.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Ethics approval

The protocol was approved by the Human Research Ethics Committee (HREC) of the Faculty of Health Sciences, UCT (HREC reference number: 247/2016). The ethics approval was renewed annually for the duration of the study.

2.2 Study design

This was a retrospective study in the Division of Anatomical Pathology. Funding was received from the National Health Laboratory Service (NHLS) Research Trust.

2.3 Case selection

The sample population chosen for this study is the same as that of Dr G Skead (Division of Anatomical Pathology, Groote Schuur Hospital/NHLS) (Skead, 2016). Dr Skead's study required samples demonstrating both SIL and invasive cervical squamous cell carcinoma on paraffin wax blocks in sufficient quantity for construction of tissue microarrays.

These cases were identified by a search of the NHLS DISA database in the division of Anatomical Pathology, Groote Schuur Hospital (Laboratory Systems Technologies, SA). The following criteria were used to select cases: excision specimens (including LLETZ, cone biopsy and hysterectomy) of primary cervical squamous cell carcinoma with normal epithelium, HSIL and invasive carcinoma in the same case. Exclusion criteria included superficially invasive carcinoma, non-squamous carcinomas, metastatic tumours, biopsies with insufficient material and any case where tissue extraction would result in loss of diagnostic material. HIV status was obtained from DISA results or from the clinical history provided on the histology request form. Information of each case (date, age, number of blocks) was entered

into a Microsoft Excel spreadsheet. The cases were arranged in temporal sequence and each SCA number given a new numerical designation. Names were removed to ensure patient confidentiality.

Slides and paraffin blocks from 2004-2014 were retrieved from archives. The diagnoses were reviewed with haematoxylin and eosin (H&E)-stained slides. From the original number of records of squamous cell carcinoma (284 cases), a total of 57 cases were chosen for Dr Skead's study. Many of the original 284 cases were duplicated cases (biopsies and excision) and biopsies were unsuitable.

In this pilot study, of the 57 cases, 23 contained sufficient tumour burden (>75% surface area) for optimal DNA extraction and further. The remaining 34 cases (demonstrating smaller tumour burden) were planned to be extracted in the future (not part of this MMed thesis) to reach the total of 57.

2.4 Sectioning of tissue for DNA extraction

For each case, four sections of 4 μm thickness were cut on a rotary microtome. Since all the tissue blocks represented tumour only, the sections were immediately placed into a 1.5 ml microcentrifuge tube. Between blocks, various means were employed to reduce cross-contamination of sample DNA. These included using a new microtome blade used for each block, wiping down the microtome with xylene and ethanol, sectioning paraffin-only blocks (negative DNA control) at end of each sectioning session. Further, filtered micropipette tips were used for all pipetting steps during DNA extraction and PCR.

2.5 DNA extraction and quantitation

DNA was extracted from FFPE tissue using the QIAamp DNA FFPE Tissue kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's guidelines.

Briefly, to each of the labelled tubes containing the tissue, tissue sectioning controls and DNA extraction negative control, 600µl of xylene was added. The tubes were vortexed vigorously and then incubated for 5 minutes at room temperature (RT). Thereafter, 300µl of ethanol (96-100%) was added to each tube and vortexed. The tubes were centrifuged at full speed (13 000rpm) for 2 minutes at RT. The supernatant was removed without disturbing the pellet. Nine hundred microlitres of ethanol (96-100%) was added to each tube and vortexed. The samples were centrifuged at full speed for 2 minutes at RT. The supernatant was removed without disturbing the pellet.

The open microcentrifuge tube was incubated at 37ºC on a Dri-block DB-3 (Bio-Techne, Minneapolis, MN, USA) until the ethanol had evaporated. One hundred and eighty microlitres of buffer ATL and 20µl Proteinase K were added to each tube and vortexed. The samples were incubated at 37ºC overnight. The next morning, the samples were removed from the heating block and the heating block was set to 90°C. Parafilm was wrapped over the microcentrifuge caps. The tubes were incubated for 1 hour at 90°C in order to partially reverse formaldehyde modification of nucleic acids and deactivate Proteinase K activity. The tubes were briefly centrifuged to remove drops from the inside of the lids. Two hundred microlitres of buffer AL and 200 µl ethanol (100%) were added to the sample and mixed by pipetting to yield a homogenous solution. This mixture was added to labelled assembled QIAamp® MinElute columns and 2 ml collection tubes without wetting the rim of the columns. The tubes were capped and centrifuged at 8000 rpm for 1 minute. The filtrate and collection tubes were discarded and the column placed in new 2 ml collection tubes. Five hundred microlitres of buffer AW1 was added to the columns without wetting the rim, capped and centrifuged at 8000 rpm for 1 minute. The filtrate and collection tubes were discarded and the column placed in new 2 ml collection tubes. Five hundred microlitres of buffer AW2 was added to the columns without wetting the rim, capped and centrifuged at 8000 rpm for 1 minute. The filtrate and

collection tubes were discarded and the column placed in new 2 ml collection tubes. Thereafter, the columns were centrifuged at full speed (13000 rpm) for 3 minutes to eliminate buffer carry over. The columns were placed in a clean labelled 1.5 ml microcentrifuge tube. Twenty seven microlitres of sterile water was added carefully to the columns, incubated at RT for 5 minutes and centrifuged at 13 000 rpm for 1 minute. This sterile water step was repeated once. The eluted DNA was either used immediately in the next step or stored at -20^{o}C.

2.6 DNA spectrophotometry

The quality and quantity of the extracted DNA was assessed using the NanoDrop™ One^{C} system (Thermo Fisher Scientific Inc., Waltham, MA, USA) (Appendix 1).

2.7 Reference gene PCR

Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene was amplified to assess the quality of extracted DNA. The QIAGEN HotStarTaq Master Mix kit (Qiagen Inc., Valencia, CA, USA) was used according to manufacturer instructions. Briefly, all solutions were thawed gently vortexed and kept on ice until required. The master mix was made up as per Table 2.1. The volumes for one reaction was multiplied to the appropriate number of reactions, including controls. Twenty three microlitres of the master mix was aliquoted into labelled 0.2 ml PCR tubes. Thereafter, 2 µl of sample DNA (100 ng) was added to each tube. A previously tested sample was used as a positive control. Two negative controls were used included to monitor cross-contamination/spill during PCR setup. This included a tissue sectioning control which comprised a paraffin wax block sectioned at the end of the samples to exclude DNA cross contamination and a DNA extraction control.

PCR was performed on a GeneAmp® PCR system 9700 (Applied Biosystems, Foster City, CA, USA). The conditions for this PCR are listed in Table 2.2 and the primer sequences are listed in Table 2.3.

Table 2. 1 PCR Master mix recipe for housekeeping gene

PCR component	**Volume (µl)**
HotStarTaq® Master mix (2x)	12.5
Forward primer (150 ng/µl)	0.5
Reverse primer (150 ng/µl)	0.5
Sterile water	9.5
gDNA template (100 ng)	2
Total	**25**

Table 2. 2 End point PCR for housekeeping gene amplification

PCR step	**Condition**
Activation	95°C/15 min
Denaturation (39 cycles)	94°C/20 sec
Anneal (39 cycles)	60°C/20 sec
Extension (39 cycles)	72°C/20 sec
Final extension	72°C/5 min
Hold	4°C/∝

2.8 HPV L1 primer synthesis

GP5+/GP6+ primers were synthesised on an Applied Biosystems 394 DNA synthesiser at the Department of Molecular & Cellular Biology, University of Cape Town (UCT) (Applied Biosystems, Foster City, CA, USA) (Table 2.3). The primers were diluted in water and added

to the master mix to yield final concentrations of 0.5-1 µM. Primer stock and dilutions were stored at -20°C.

Table 2. 3 Primer sets used in the amplification of housekeeping gene and HPV L1 consensus gene

Primer set	DNA sequence (5'-3')	Product size (bp)
GAPDH (forward)	AGGTCATCCATGACAACTTTGGTATC	222
GAPDH (reverse)	TGAGGCCCTGCAGCGTACTC	
GP5+ (forward)	TTTGTTACTGTGGTAGATACTAC	150
GP6+ (reverse)	GAAAAATAAACTGTAAATCATATTC	

2.9 Optimisation of HPV L1 consensus sequence endpoint PCR using GP5+/GP6+ primers

The HotStartTaq master mix was made up as per Table 2.1. The volumes for one reaction was multiplied to the appropriate number of reactions, including controls. Twenty three microlitres of the master mix was aliquoted into labelled 0.2 ml PCR tubes. Thereafter, 2 µl of sample DNA (100 ng) was added to each tube. PCR was performed on a GeneAmp® PCR system 9700 (Applied Biosystems, Foster City, CA, USA).

The following parameters were subject to optimisation for this PCR (Table 2.4):

1. Amplification conditions

   The initial experiment used standard PCR cycling parameters. Run 2 reduced the annealing temperature from 50°C to 40°C as the melting temperature (T_m) of the GP5+/GP6+ primers are low (45°C and 41°C, respectively) (Evans *et al.*, 2008). Run 3 kept the temperatures constant but extended the duration of annealing and extension from 30 seconds to 2 minutes (Table 2.4).

2. Magnesium titration

In Run 3, the $MgCl_2$ concentration was increased from 1.5 mM to 3.0 mM by addition of stock 25 mM $MgCl_2$ (Table 2.4). The volume of sterile water was adjusted accordingly to keep the final volume of the PCR mix at 25 µl (Table 2.1).

Table 2. 4 End point PCR optimization for HPV L1 amplification by GP5+/GP6+ primers

PCR step	Initial run	Run 2	Run 3
Activation	95°C/15 min	95°C/15 min	95°C/15 min
Denaturation (39 cycles)	94°C/30 sec	94°C/30 sec	94°C/30 sec
Anneal (39 cycles)	50°C/30 sec	40°C/30 sec	40°C/2 min
Extension (39 cycles)	72°C/30 sec	72°C/30 sec	72°C/2 min
Final extension	72°C/5 min	72°C/5 min	72°C/5 min
Hold	4°C/∝	4°C/∝	4°C/∝

2.10 Reamplification endpoint PCR of HPV L1 gene using GP5+/GP6+ primers

To increase the sensitivity of this PCR, reamplification was performed. This is akin to a "nested PCR" except the same primer set is used in a second round of PCR. The PCR master mix was the same as for Run 3 except that two microlitres of the first PCR reaction was used as the new template.

2.11 Touchdown PCR optimisation using GP5+/GP6+ primers

A published method for improved detection of HPV L1 gene using GP5+/GP6+ primers using a touchdown technique was used (Evans *et al*., 2005). This modified PCR protocol involved reducing the annealing temperatures in 0.5°C decrements per cycle and increasing total cycles to 50 (Table 2.5). The PCR conditions were further optimised by titration of genomic DNA, primer concentrations and $MgCl_2$ concentrations as follows:

Genomic DNA template: 50, 100, 250 and 500 ng

Primer concentration: 0.5 μM and 1.0 μM

$MgCl_2$ concentration: 1.5 mM and 4.0 mM

Table 2. 5 Touchdown PCR setup for HPV L1 amplification by GP5+/GP6+ primers

Touchdown PCR step	**Conditions**
Activation	95°C/15 min
Denaturation (21 cycles)	94°C/2 min
Anneal (21 cycles)	50°C/2 min to 40°C/2 min, decreasing 0.5°C/cycle
Extension (21 cycles)	72°C/1.5 min
Denaturation (29 cycles)	94°C/2 min
Anneal (29 cycles)	40°C/2 min
Extension (29 cycles)	72°C/1.5 min
Final extension	72°C/4 min
Hold	4°C/∝

2.12 Application of optimised GP5+/GP6+ HPV L1 touchdown PCR conditions to all extracted samples

Extracted DNA from all the cases were amplified using the optimised touchdown protocol with 100 ng template gDNA, 0.5 μM forward and reverse primers and 4 mM $MgCl_2$.

2.13 DNA agarose electrophoresis

Post PCR DNA samples were size separated on a 1.5 % agarose gel (Appendix 2). The running buffer used was 1xTBE (Appendix 2). PCR products were prepared by mixing 10 μl of each with bromophenol blue loading dye (1 μl; Appendix 2) and dispensed into separate wells of

the gel. Five microliters of a ready-to use molecular weight marker, HyperLadder™ 100 bp (Bioline, London, UK) was also included. This ladder produces bands ranging from 100-1300 bp. The gel was run at 100V (constant) for 60-90 min. The gel was visualised by UV transillumination and images were captured digitally via GeneSnap software (Syngene, Cambridge, UK) in the Division of Forensic Medicine, UCT. The images were saved in either tiff or png formats.

2.14 Detection of hrHPV genotypes by multiplex real time PCR (RT-PCR)

Detection of hrHPV genotypes was performed using Sacace HPV Genotypes 14 Real-TM Quant Kit (Como, Italy) which uses hydrolysis probe technology to simultaneously detect up to 14 hrHPV genotypes (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) with additional probes included as positive controls (*β-globin*). The HPV target sequences for this kit are within the E6 and E7 genes and not L1 as used in the end point PCR experiments. This kit is optimised for cervical swabs, fluid cytology and mucosal biopsy material. It has also been successfully used on DNA extracted from FFPE tissue (Zohoncon *et al*., 2016, Snietura *et al*., 2017).

The hydrolysis probes included FAM (HPV 16, 33, 39 and 58), JOE (HPV 31, 35, 45 and 52), ROX (HPV 18, 59, 66 and 68) and Cy5 (internal control [*β-globin*], HPV 51 and 56).

Table 2. 6 Absoprtion and emission spectra of dyes with available filter set

Fluorophore	**Excitation wavelength**	**Emission wavelength**	**LightCycler™ 480 II filter set available**
FAM	495 nm	520 nm	498/580 nm
JOE	529 nm	555 nm	533/610 nm
ROX	575 nm	602 nm	533/610 nm
Cy5	617nm	667 nm	618/660 nm

A stock solution of *Taq* polymerase ("Hot Start DNA polymerase") and buffer was made by adding 600 µl PCR buffer mix to 60 µl Hot Start DNA Polymerase. This mixture, called "PCR-buffer-FRT", was briefly vortexed and kept on ice. The total volume of each of the 4 DNA probe mixes required was calculated according to the number of samples using the formula 10xN+2 (where N=number of samples for analysis) while the total volume of master mix DNA polymerase was 5xN+2.

Thawed extracted gDNA from each case was diluted with sterile water to a final concentration of 20 ng/µl and kept on ice until required. The reaction consisted of 10 µl (containing 200 ng) gDNA with 15 µl PCR-buffer-FRT. The total reaction volume was 25 µl (Table 2.6).

The samples were dispensed into LightCycler™ 480 96 multi-well white plates (Roche, Basel, Switzerland). The plates were sealed with clear adhesive film to reduce evaporation. Negative controls included no template DNA. Positive controls (provided in the kit) corresponding to 1000 genome equivalents/ml included all hrHPV genotypes and human *β-globin*. RT-PCR was performed on a LightCycler™ 480 II (Roche, Basel, Switzerland) instrument housed in the Division of Immunology, UCT Medical School.

Table 2. 7 Recipe for Sacace RT-PCR kit

PCR component	**Volume (µl)**
PCR-buffer-FRT mix	15
gDNA template (200 ng)	10
Total	**25**

LightCycler™ 480 1.5.1.62 SP3 software was used with the following parameters:

- Detection format: 4 color hydrolysis probe.
- Workflow: Abs quant.
- Filter combinations: 498-580 nm (FAM), 533-610 nm (Red 610), 618-660 nm (Cy5).

The RT-PCR parameters used were those provided with the Sacace kit:

Table 2. 8 RT-PCR conditions for multiplex identification of hrHPV genotypes

PCR step	Condition	Ramp rate
Activation	95°C/15 min	4.4°C/sec
Denaturation (5 cycles)	95°C/5 sec	4.4°C/sec
Annealing (5 cycles)	60°C/20 sec	2.2°C/sec
Extension (5 cycles)	72°C/15 sec	4.4°C/sec
Denaturation (40 cycles)	95°C/5 sec	4.4°C/sec
Annealing (40 cycles)	60°C/20 sec	2.2°C/sec
Extension (40 cycles with fluorescence acquisition)	72°C/15 sec	4.4°C/sec
Hold	40°C/30 sec	4.4°C/sec

The data was analysed by selecting "Abs quant/2nd derivative max" for all samples. The raw fluorescence data and calculated C_t values were exported (in .txt format) and analysed with Microsoft Excel 2016.

2.15 Confirmation of hrHPV genotypes by DNA sequencing and BLAST alignment

DNA sequencing was performed to both confirm the L1 DNA sequence of the GP5+/GP6+ PCR products and to verify the hrHPV genotypes of samples that showed positive results in the overlapping filter spectra only.

The GP5+/GP6+ PCR products for cases 20 and 23 were randomly chosen for confirmation of hrHPV L1 DNA sequence as no separate positive control was available. The GP5+ forward primer was used for this purpose. DNA sequencing (via Sanger method) was performed by Inqaba Biotechnical Industries (Pretoria, South Africa) on an ABI3500XL system (Applied Biosystems, Foster City, CA, USA).

Genotype-specific forward sequencing primers were synthesised for cases 33, 34, 42 and 55 using sequences previously described by Gharizadeh *et al*., (2005) as these cases showed potential overlap with types 18/31 and 45/59. The primer sequences were: HPV 18 (5'-GCTTCTACACAGTCTCCTGT-3'), HPV 31 (5'-GTGCTGCAATTGCAAACAGT-3'), HPV 45 (5'-TATGTGCCTCTACACAAAAT-3'). A genotype specific primer for HPV 59 was not described in this paper. It was reasoned that if the HPV 45 sequencing produced no positive results, a separate HPV 59 specific probe would be designed and tested. The forward sequencing primers were synthesised and used for Sanger sequencing at Inqaba Biotechnical Industries (Pretoria, South Africa).

The resultant nucleotide sequences were generated in abi file format converted into FAST-all (FASTA) format using BioEdit, a freeware biological sequence alignment edit programme (Carlsbad, CA, USA). The FASTA-formatted DNA sequences were subjected to Basic Local Alignment Search Tool (BLAST) analysis using the National Center for Biotechnology Information (NCBI) nucleotide BLAST service (http://blast.ncbi.nlm.nih.gov). The "other - nucleotide" database was selected and the search was optimised for highly similar sequences (megablast).

CHAPTER THREE

RESULTS

3.1 Sample group

3.1.1 Patient demographics

The cohort comprised 23 cases. The most common age range was 40-49 years. Only 1 case of cervical squamous cell carcinoma was present in each of the extreme age ranges 20-29 and 70-79 (Figure 3.1).

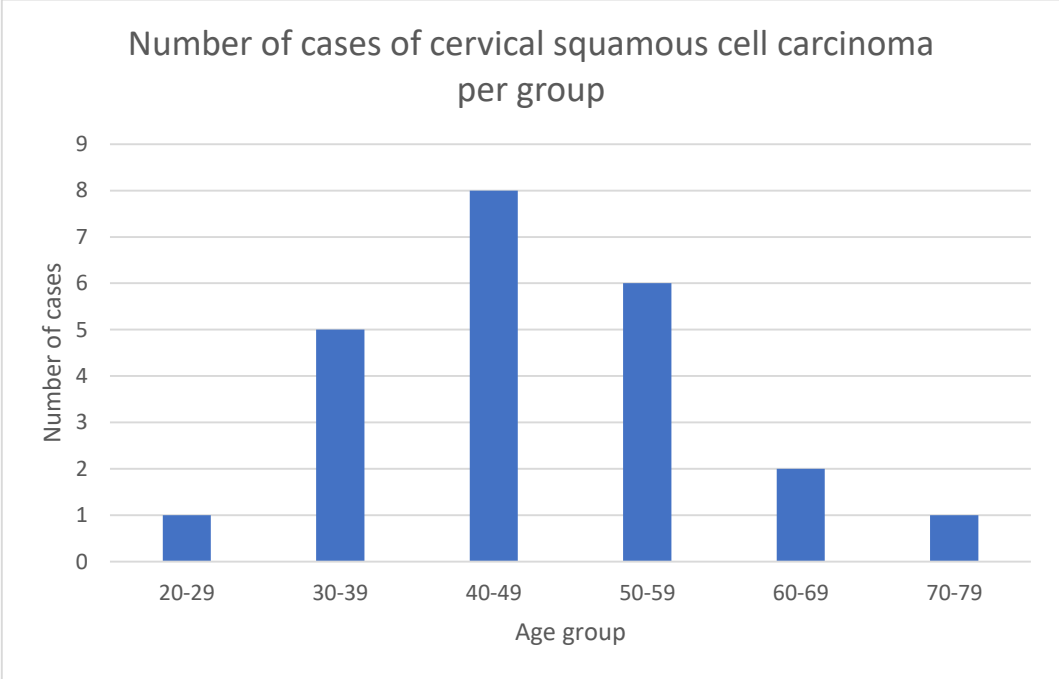


Figure 3. 1 Number of cases of cervical squamous cell carcinoma per age group

3.1.2 HIV distribution within cases

Two cases were HIV-positive, 14 HIV-negative and 7 HIV unknown (i.e., the status could not be determined from DISA lab investigations or from the clinical history provided on the specimen submission form).

3.2 DNA extraction from FFPE archival tissue

The average concentration of extracted gDNA was 463.2 ng/µl (range: 91.3-791.4) (see Appendix 2. The average A_{260}/A_{280} ratio was 1.86 (range: 1.67-1.94) with a median of 1.88. A calculated A_{260}/A_{280} ratio ≥ 1.80 is generally considered adequate purity (Wilfinger *et al*., 1997). The majority of cases (20/23, 87%) were of optimal purity, with cases 15, 26 and 51 achieving A_{260}/A_{280} ratios of 1.67, 1.78 and 1.67, respectively. A ratio lower than 1.8 may indicate the presence of proteins or other contaminants but these marginally suboptimal values did not appear to negatively affect downstream PCR reactions.

3.3 Endpoint PCR

3.3.1 Housekeeping gene PCR using reference gene primers

All samples showed amplification for *GAPDH* (results not shown). This indicated that the sample DNA was free of inhibitors and was of sufficiently good quality to proceed with the GP5+/6+ PCR and RT-PCR. The negative controls showed no amplification indicating that there were no cross-contamination events. The positive control amplified a band of the correct size (either 150 bp or 222 bp) suggesting that the PCR conditions were optimal. Some reduction in intensity of PCR amplicons was noted in the oldest specimen (case 3 from 2004, Figure 3.2 lanes 2 and 7) compared to the most recent (case 56 from 2014, Figure 3.2 lanes 4 and 9). Cases from approximately midpoint of the time range showed similar intensity to that obtained from 2014 (case 19 from 2008, Figure 3.2 lanes 3 and 8). This difference was likely due to increased DNA degradation over time however subsequent endpoint PCR experiments showed no such reduction in HPV L1 amplification (Figure 3.5). Therefore, it may be suggested that DNA fragmentation was more significant in housekeeping gene regions compared to HPV L1 sequences.

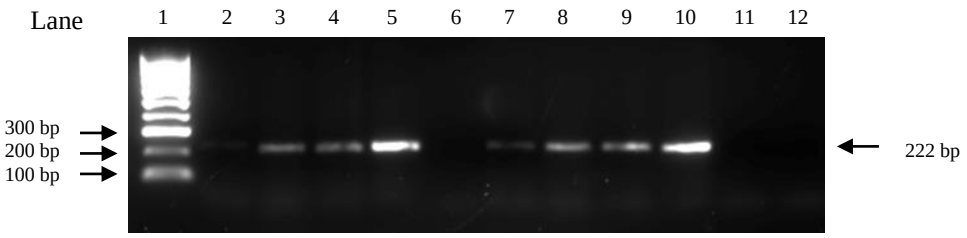


Figure 3. 2 DNA electrophoresis of selected cases amplified by end-point PCR using *GAPDH* housekeeping primers. Positive PCR amplicon = 222 bp (right arrow). Lane 1: 100 bp molecular weight marker ladder, lane 2: Case 3 (year 2004) (50 ng gDNA), lane 3: Case 19 (year 2008) (50 ng gDNA), lane 4: Case 56 (year 2014) (50 ng gDNA), lane 5: *GAPDH* positive control, lane 6: negative control, lane 7: Case 3 (100 ng gDNA), lane 8: Case 19 (100 ng gDNA), lane 9: Case 56 (100 ng gDNA), lane 10: *GAPDH* positive control, lanes 11 and 12: negative controls.

3.3.2 HPV L1 end point PCR using GP5+/GP6+ primers

Standard and modified PCR experiments failed to produce a 150 bp L1 amplicon in cases 3, 19 and 56 (results not shown). Reasons for this included suboptimal PCR conditions as no positive control was available and that the amount of HPV DNA in the samples may be below the detection limit. To increase the sensitivity of this PCR a nested PCR (reamplification PCR) was performed.

3.3.3 Reamplification HPV L1 end point PCR using GP5+/GP6+ primers

Reamplification PCR was performed on the same selected cases representing either time extreme plus one midpoint as shown in Figure 3.2. Interestingly, the oldest specimen (case 3, year 2004) produced the brightest amplicon (Figure 3.3, lane 2). This may reflect a greater HPV burden than case 19 (year 2008) which showed a fainter band (Figure 3.3, lane 3) . The most recent case, 56, (year 2014) showed no amplification (Figure 3.3, lane 4). The bands visible below 100 bp in lanes 3-5 are primer dimer bands and do not represent HPV L1 amplicons.

Reamplification PCR proved that HPV L1 DNA sequences could be detected in samples with sufficient HPV burden. Reamplification PCR is, however, time consuming by requiring two PCR runs and more costly with enzymes and reagents as each case effectively requires twice

the amount. Therefore, a published touchdown PCR protocol was optimised to permit L1 detection in one run.

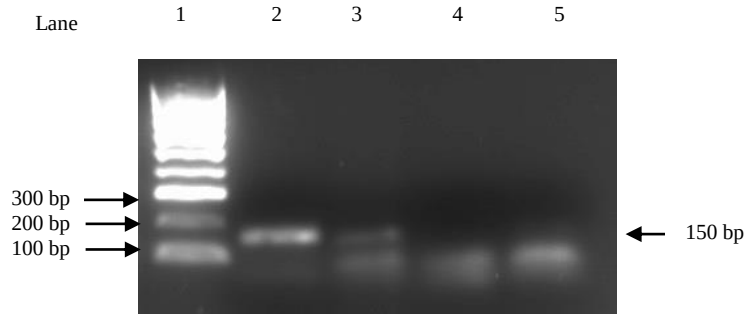


Figure 3. 3 DNA electrophoresis of selected cases amplified by reamplification end-point PCR using GP5+/GP6+ HPV L1 primers. Positive PCR amplicon = 150 bp (right arrow). Lane 1: 100 bp molecular weight marker ladder, lane 2: Case 3 (100 ng initial gDNA), lane 3: Case 19 (100 ng initial gDNA), lane 4: Case 56 (initial 100 ng gDNA), lane 5: negative control.

3.3.4 Touchdown PCR optimisation of HPV L1 using GP5+/GP6+ primers

Optimisation of the published touchdown protocol was performed in order to maximise the detection of HPV within cervical squamous cell carcinoma specimens. No modifications to the PCR parameters were made however a dilution series of genomic DNA was performed in order to determine the detection limits of the method and to further conserve gDNA for downstream hrHPV RT-PCR investigations. Primer concentrations and $MgCl_2$ concentrations were modified to maximise the efficiency and yield of the PCR reaction. Case 3 was chosen for this optimisation as it demonstrated adequate amount of amplifiable HPV L1 DNA in the reamplification experiment and theoretically contained the gDNA with the greatest degree of degradation (as noted in reference gene experiments, Figure 3.2).

Touchdown PCR demonstrated HPV L1 gene product could be detected with as low as 50 ng gDNA template (Figure 3.4, lanes 19-22). Primer concentration did not affect PCR product yield however 4.0 mM $MgCl_2$ resulted in improved PCR product yield versus 1.5 mM, especially at lower gDNA concentrations of 100 and 50 ng (Figure 3.4 lanes 18 and 20, respectively).

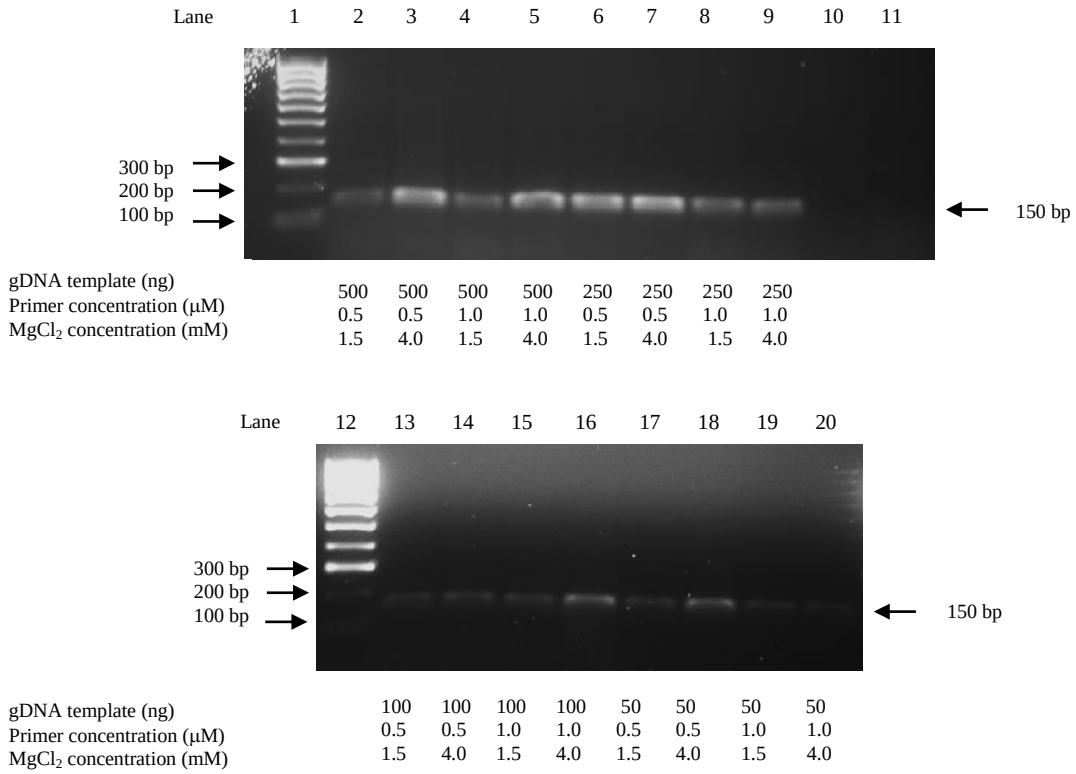


Figure 3. 4 DNA electrophoresis of optimised touchdown end-point PCR using GP5+/GP6+ HPV L1 primers. All samples are from case 3. Positive PCR amplicon = 150 bp (right arrow). Lanes 1 and 12: 100 bp molecular weight marker ladder, lanes 10 and 11: negative controls. Lanes 2-9 and 13-20 as described above.

3.3.5 Touchdown PCR of cervical carcinoma samples

Application of optimised touchdown PCR reaction conditions to all samples yielded positive 150 bp amplicons in 22/23 cases (case 56 did not amplify). The band intensities of the amplicons vary in line with differential HPV burden. Cases 19-23 (Figure 3.5 lanes 7-9) showed the greatest amount of amplicon while cases 39, 42, 51-56, 57 (Figure 3.5 lanes 20, 21, 23-25, 27) showed fainter bands. The faintest band was obtained from case 33 (Figure 3.5 lane 12). Negative controls showed no amplification and therefore no cross contamination (Figure 3.5 lanes 13 and 14). All positive sample bands were deemed to show adequate amounts of HPV DNA for hrHPV RT-PCR analysis. Though case 56 did not show a positive amplicon in this experiment (Figure 3.5, lane 26), it was included in the RT-PCR experiments.

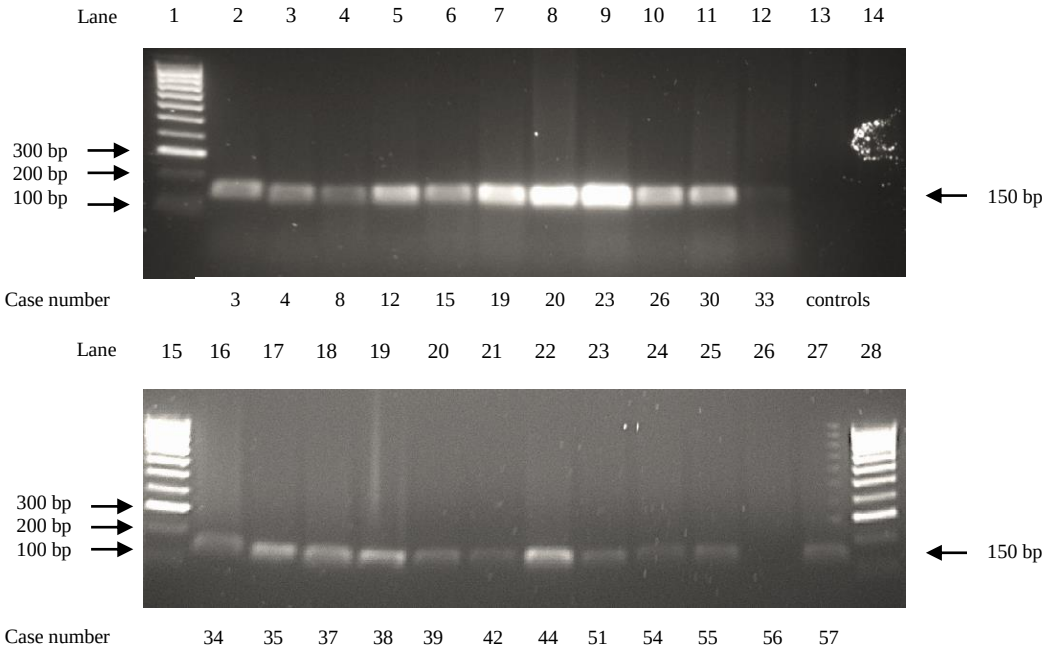


Figure 3. 5 DNA electrophoresis of all cases amplified by touchdown end-point PCR using GP5+/GP6+ HPV L1 primers. Positive PCR amplicon = 150 bp (right arrow). Standard conditions: 100 ng genomic DNA, 0.5 µM primers, 4 mM $MgCl_2$. Lanes 1, 15 and 28: 100 bp molecular weight marker ladder, lanes 13 and 14; negative controls. Case numbers as described above.

3.4 hrHPV genotypes detected by multiplex RT-PCR on LightCycler™ 480 II

Raw fluorescence data from the RT-PCR run was exported into Microsoft Excel. Background fluorescence was subtracted from the traces to yield net fluorescence. Graphs were plotted showing cycle number versus fluorescence for all samples in all detection channels. The cycle number refers to the fluorescence detected in the extension phase of PCR of a particular cycle. The threshold cycle (C_t), as previously described in Chapter 1.5.2.2.2, refers to the PCR cycle number at which the PCR amplicon fluorescence crosses the background/threshold fluorescence and begins exponential phase of fluorescence (Figure 1.9). The C_t value is calculated by the LightCycler™ software.

After the runs were complete, the positive controls were first analysed to ensure that all dye labelled probes worked correctly. Thereafter, negative controls and case samples were analysed and corresponding graphs plotted.

3.4.1 RT-PCR curves of supplied positive control DNA samples for hrHPV and human internal control, *β-globin*

RT-PCR of all of the positive controls was successful. All showed exponential phase fluorescence below cycle 30 and therefore individual C_t values are not displayed (Figures 3.6-3.8). Dyes in the FAM and Cy5 spectrum were individually resolved however an issue arose with the JOE/ROX dyes and the available detection filter. The Red 610 filter (533-610 nm) cannot distinguish the emission spectra of ROX and JOE dyes. Therefore, HPV types in the JOE probe set (types 31, 35, 45 and 52) cannot be distinguished from those in the ROX probe set (18, 59, 66 and 68) (Figure 3.7). It was therefore decided to establish exact hrHPV genotype in any sample that showed positivity in the Red 610 spectrum by DNA sequencing.

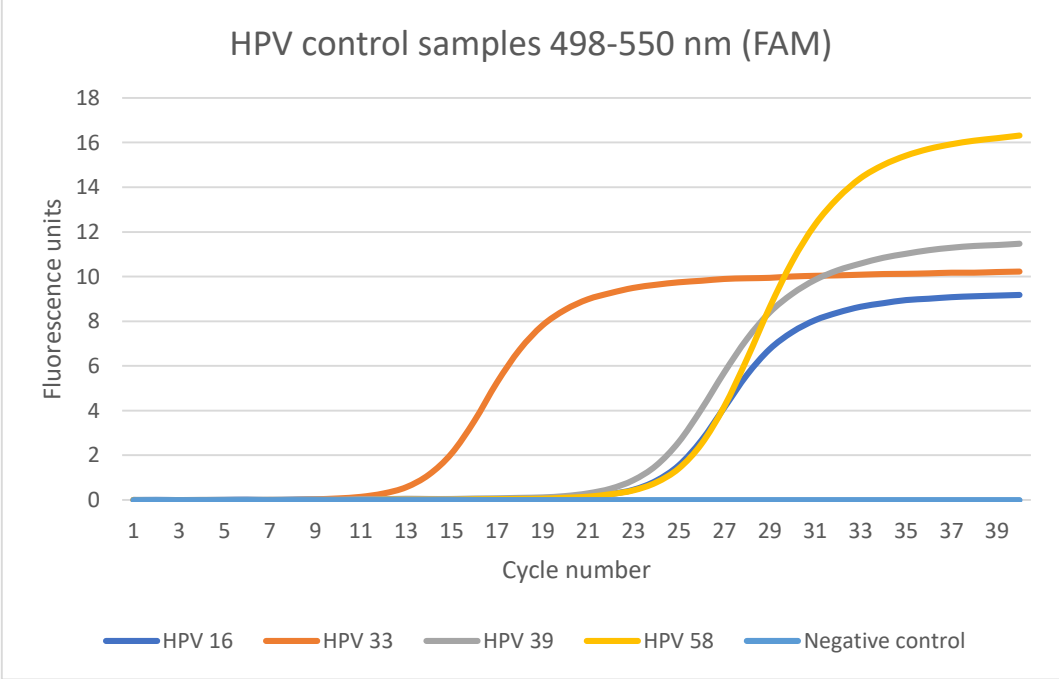


Figure 3. 6 RT-PCR curves of supplied positive controls for HPV 16, 33, 39 and 58. The genotype-specific hydrolysis probes were detected with FAM filter (498-550 nm).

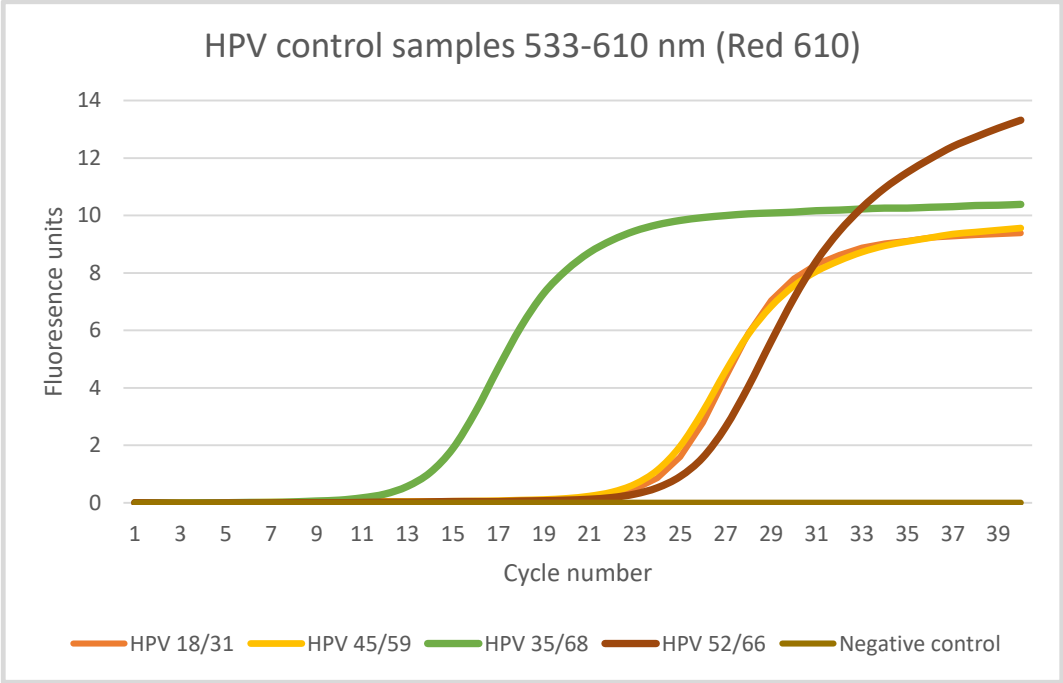


Figure 3. 7 RT-PCR curves of supplied positive controls for HPV 18, 31, 45, 59, 35, 68, 52 and 66. Due to overlapping spectra of JOE and ROX probes, the curves are grouped as 18/31, 45/59, 35/68 and 52/66. The genotype-specific hydrolysis probes were detected with Red 610 filter (533-610 nm).

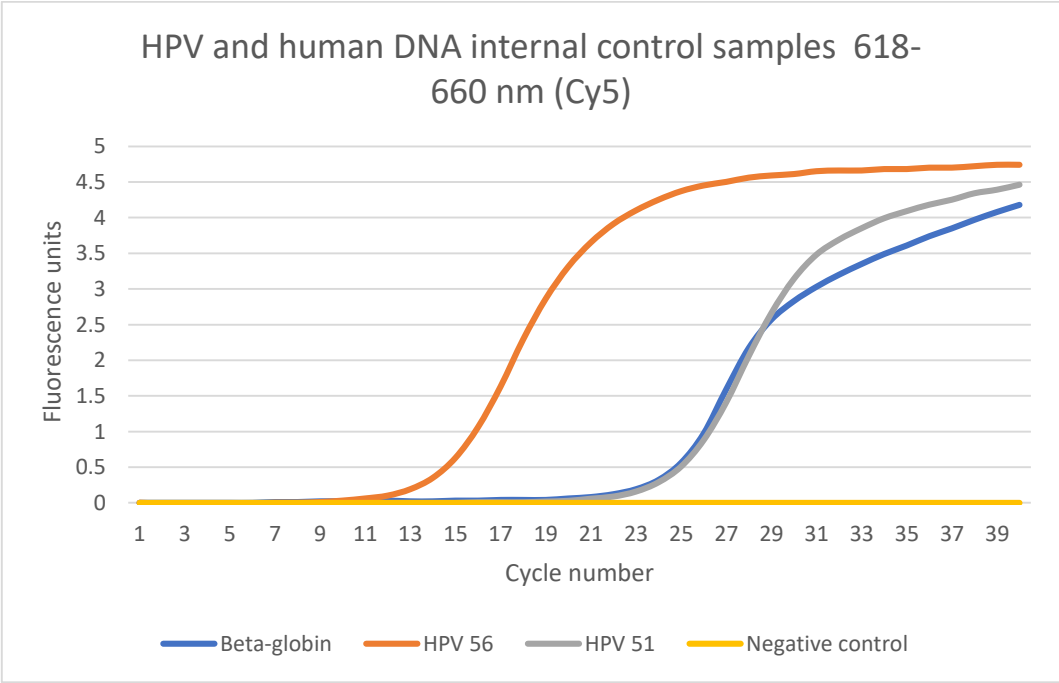


Figure 3. 8 RT-PCR curves of supplied positive controls for HPV 51, 56 and *β-globin*. The genotype-specific hydrolysis probes were detected with Cy5 filter (618-660 nm).

3.4.2 Detection of FFPE internal control (housekeeping) genes by RT-PCR

The Sacace kit provided two Cy5 labelled hydrolysis probes for the detection of internal control genes (housekeeping genes). Only detection of one of the probes is required for the specimen to be considered positive. A positive internal control with C_t>33 for hrHPV indicates that HPV concentration is low in the sample but that the sample DNA integrity is adequate. A negative internal control with negative hrHPV (i.e. no amplification signal at all for hrHPV) would imply that the specimen DNA integrity was not optimal. A negative internal control but positive hrHPV may occur and indicate that the DNA integrity of the target sequence(s) of the hrHPV genes may be better preserved than that of the internal control. Figure 3.9 shows the calculated C_t for internal control of all samples. Three cases (8, 26 and 39) showed no signal for internal controls however good amplicon detection was achieved with touchdown endpoint PCR (see Figure 3.5).

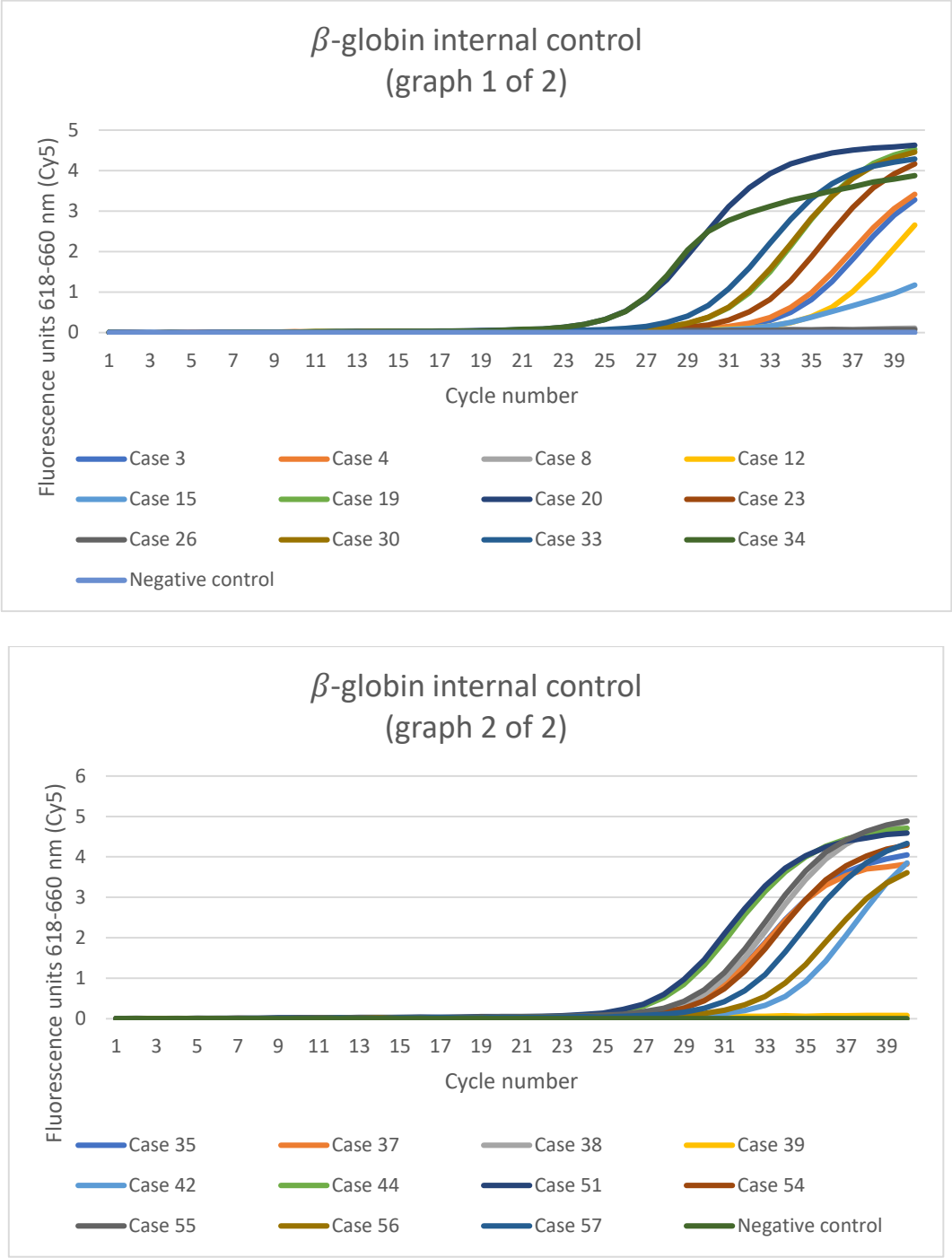


Figure 3. 9 RT-PCR curves of *β-globin* housekeeping gene. Cases 3 (C_t=34), 4 (C_t=33), 12 (C_t=35), 15 (C_t=33), 19 (C_t=29), 20 (C_t=26), 23 (C_t=32), 30 (C_t=29), 33 (C_t=28) , 34 (C_t=25), 35 (C_t=30), 37 (C_t=29), 38 (C_t=30), 42 (C_t=32), 44 (C_t=28), 51 (C_t=25), 54 (C_t=30), 55 (C_t=30), 56 (C_t=32) and 57 (C_t=31). Genotype-specific hydrolysis probe was detected with Cy5 filter (618-660 nm). Cases 8, 26 and 39 do not show amplification.

3.4.3 Detection of hrHPV genotypes by multiplex RT-PCR

Threshold cycles were calculated by the software program with the highest assigned threshold number being 35. The last 5 cycles have higher uncertainty and this is termed a "late call".

Twenty two out of twenty three cases showed the presence of hrHPV with case 54 representing the only case with no detection of hrHPV. The total number of hrHPV-positive cases is therefore 22. Case 54 showed a positive internal control signal (C_t=30) and demonstrated weak but positive GP5+/GP6+ amplicon with touchdown endpoint PCR (see Figure 3.5). Only the cases showing positive amplification curves for specific hrHPV genotypes are presented below (Figures 3.10-3.14). A summary of the cases and their hrHPV profiles is shown in Table 3.1

HPV 16

The most common hrHPV genotype was 16, which was identified in 21/22 hrHPV cases (95%). Cases with relatively late threshold cycles (C_t>33) (cases 12, 33, 37, 38 and 56) showed internal control (*β-globin*) curves with C_t<33 except for case 12 (Ct=35) (Figure 3.10). These cases are interpreted as representing low copy numbers of HPV 16.

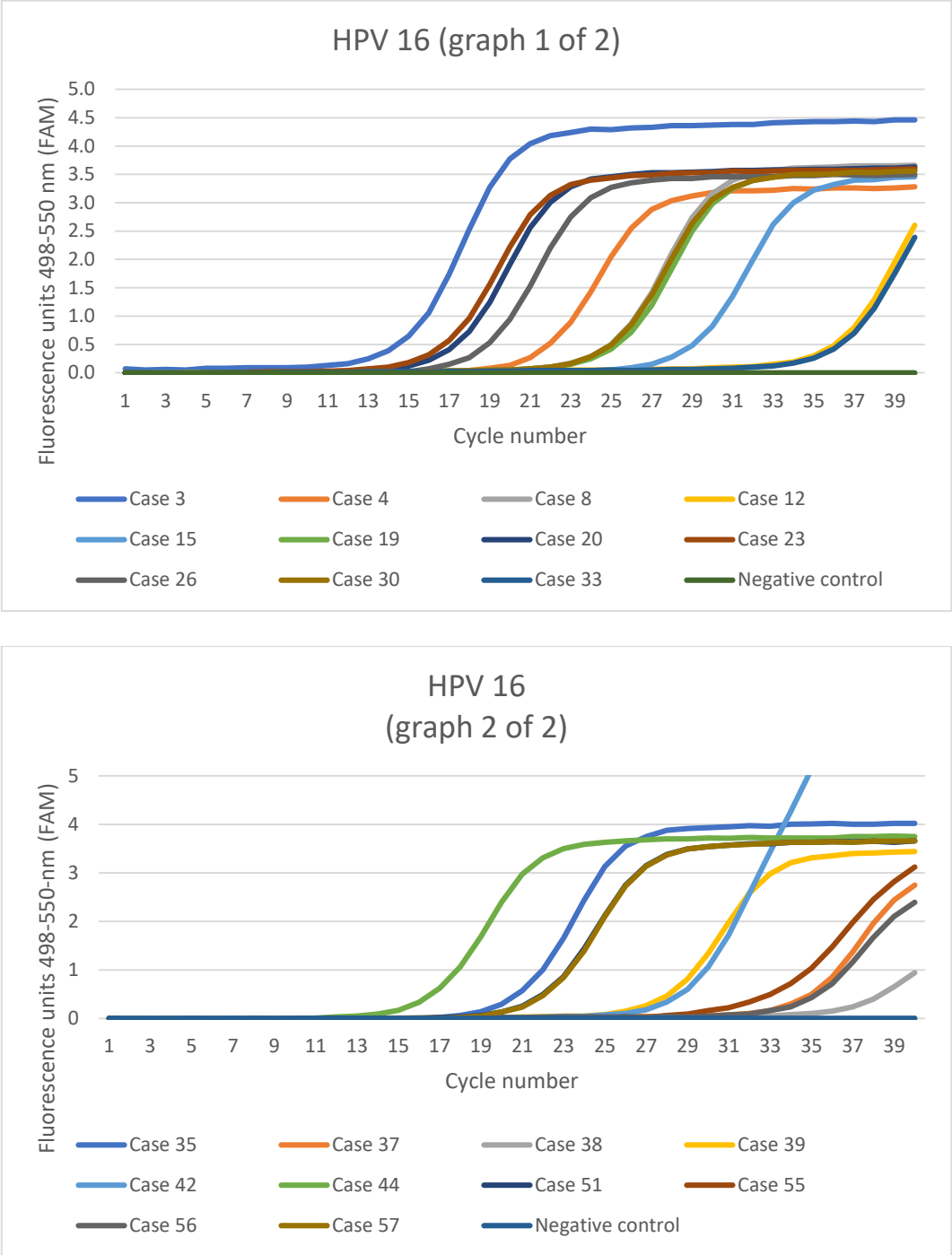


Figure 3. 10 RT-PCR curves of HPV 16. Cases 3 (C_t=15), 4 (C_t=21), 8 (C_t=25), 12 (C_t=35), 15 (C_t=29), 19 (C_t=25), 20 (C_t=17), 23 (C_t=16), 26 (C_t=18), 30 (C_t=25), 33 (C_t=35), 35 (C_t=20), 37 (C_t=35), 38 (C_t=35), 39 (C_t=28), 42 (C_t=29), 44 (C_t=16), 51 (C_t=21), 55 (C_t=33), 56 (C_t=35) and 57 (C_t=21). Case 34 does not contain HPV 16. Genotype-specific hydrolysis probe was detected with FAM filter (498-550 nm).

HPV 18

HPV 18 was identified in 2/22 hrHPV cases (9%). In both cases, co-infection with HPV 16 was present.

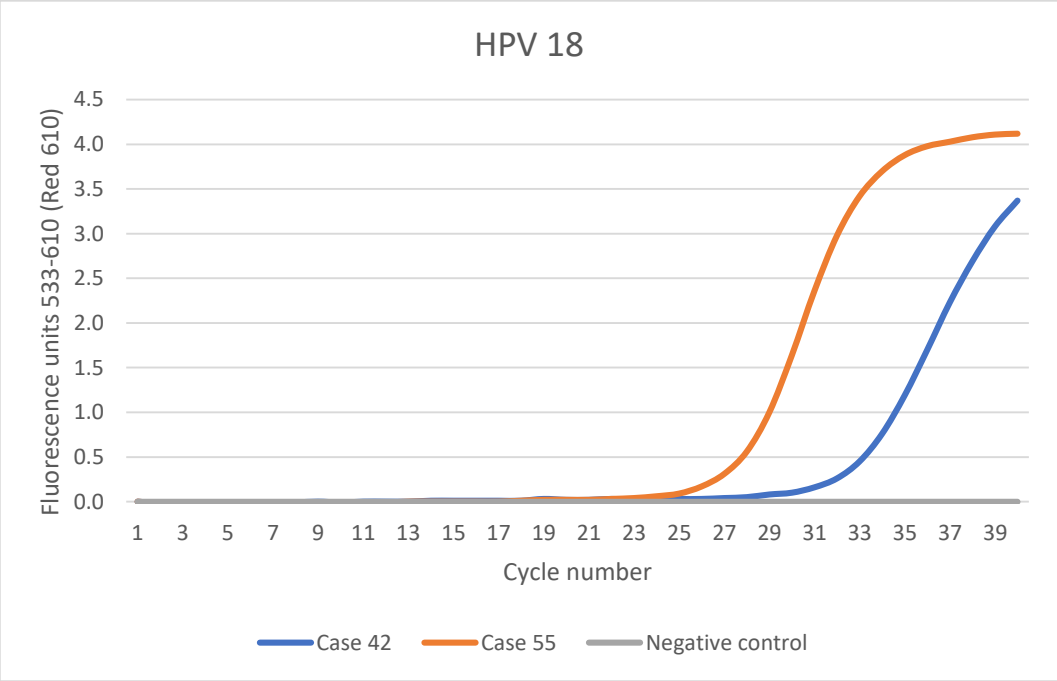


Figure 3. 11 RT-PCR curves of HPV 18. Cases 42 (C_t=33) and 55 (C_t=27). Genotype-specific hydrolysis probe detected with Red 610 filter (533-610 nm).

HPV 33

HPV 33 was identified in 1/22 (4.5%) hrHPV cases with HPV 16 co-infection (case 37).

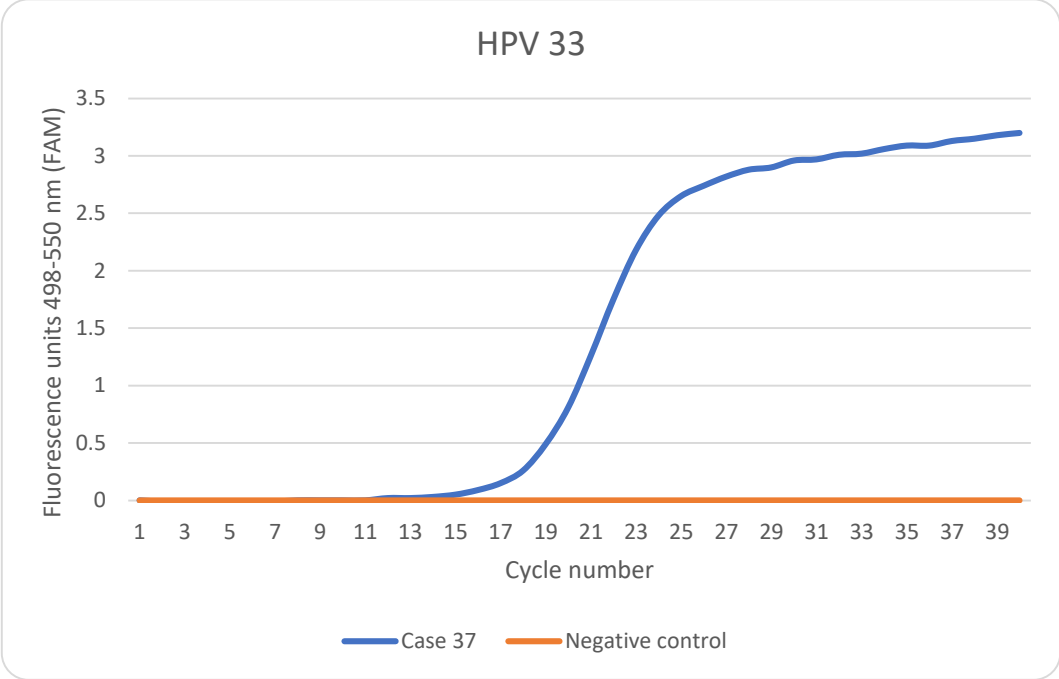


Figure 3. 12 RT-PCR curve of HPV 33. Case 37 (C_t=18). Genotype-specific hydrolysis probe detected with FAM filter (498-550 nm).

HPV 39 and 45

HPV 39 was identified in 2/22 cases (9%) with HPV 45 co-infection in one (case 34) and HPV 45 and 16 in the other (case 33).

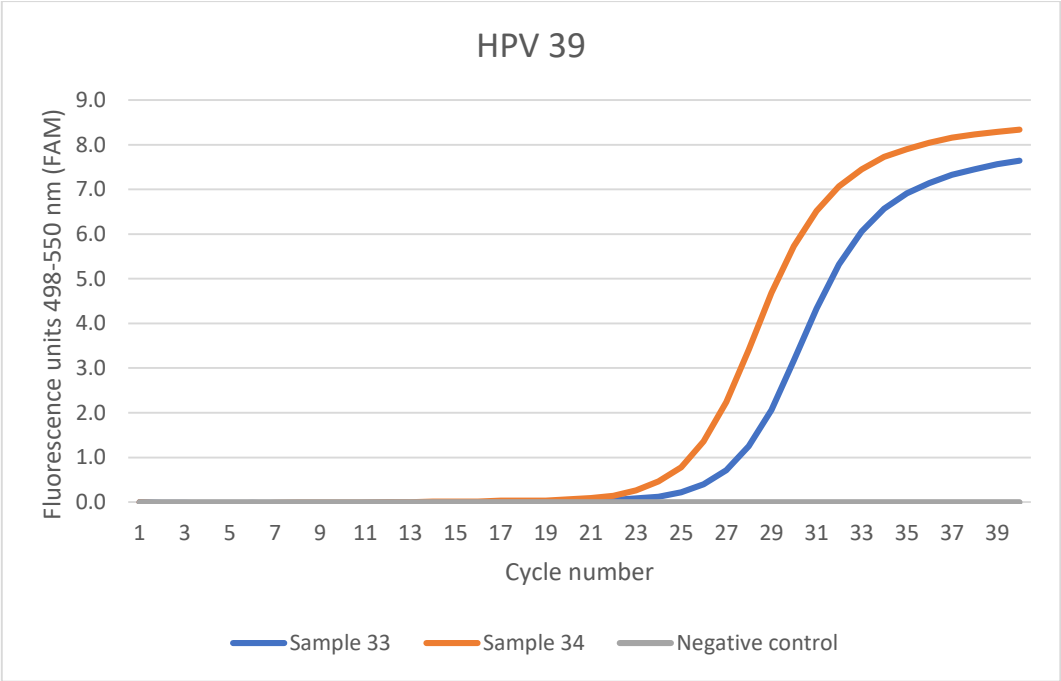


Figure 3. 13 RT-PCR curves of HPV 39. Cases 33 (C_t=27) and 34 (C_t=25). Genotype-specific hydrolysis probe detected with FAM filter (498-550 nm).

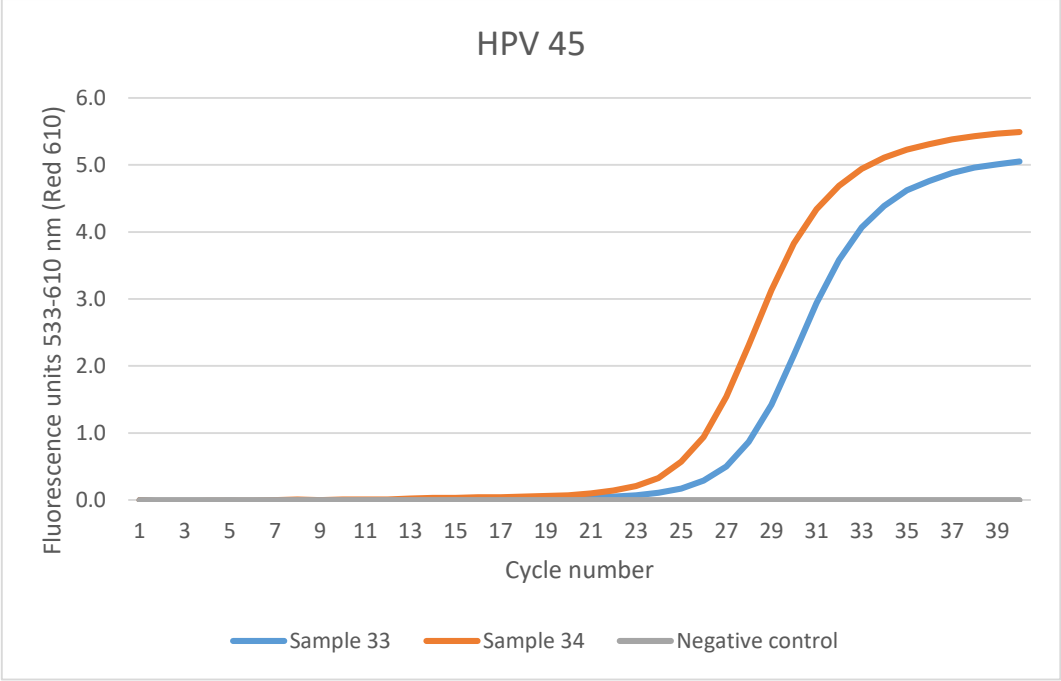


Figure 3. 14 RT-PCR curves of HPV 45. Cases 33 (C_t=27) and 34 (C_t=25). Genotype-specific hydrolysis probe detected with Red 610 filter (533-610 nm).

Table 3. 1 hrHPV genotypes identified by RT-PCR

Case	HPV 16	HPV 18	HPV 31	HPV 33	HPV 35	HPV 39	HPV 45	HPV 51	HPV 52	HPV 56	HPV 58	HPV 59	HPV 66	HPV 68
3														
4														
8														
12														
15														
19														
20														
23														
26														
30														
33														
34														
35														
37														
38														
39														
42														
44														
51														
55														
56														
57														

Green box indicates presence of hrHPV genotype and red marks absence

Overall, 17 cases showed infection with a single hrHPV genotype, four cases with 2 genotypes and one with three (Figure 3.16).

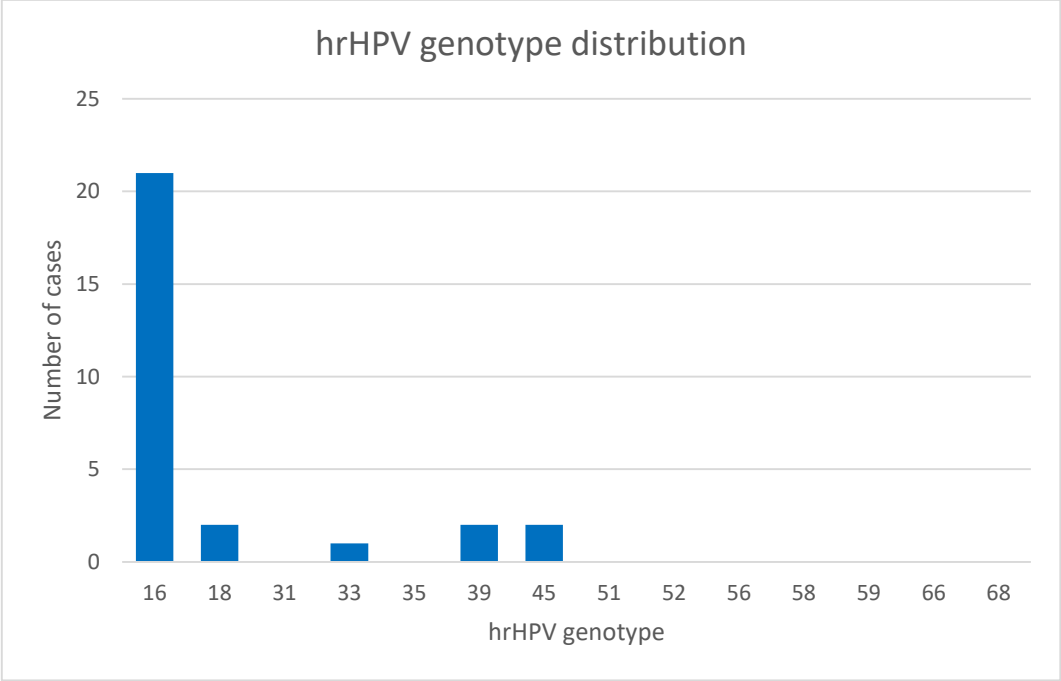


Figure 3. 15 Frequency distribution of hrHPV genotypes identified by RT-PCR

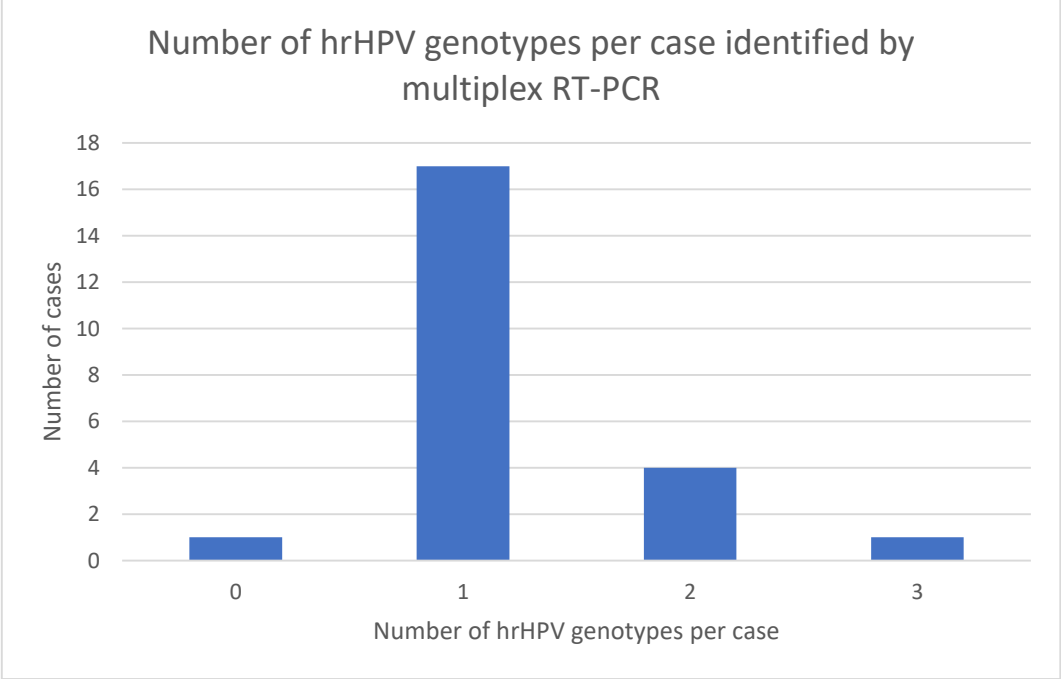


Figure 3. 16 Number of hrHPV genotypes per case identified by RT-PCR

3.5 Confirmation of hrHPV L1 sequences via DNA Sanger sequencing and BLAST alignment

As no specific positive control was used in the PCR experiments, DNA sequencing was chosen to prove that the amplicons generated by GP5+/GP6+ primers were truly HPV L1 in origin. Cases 20 and 23 were randomly chosen for sequencing for this purpose and sequenced using the GP5+ forward primer. BLAST analysis of the obtained sequences show 100% match for L1 gene of HPV 16 (Table 3.2).

The genotype specific primers used for cases 33, 34 and 55 show 100% matches for genotypes 45 (cases 33 and 34) and 18 (case 55). Case 42 did not show a match for HPV 18 or 31 (Table 3.2). The lack of sequencing trace in case 42 may reflect the low copy number of these genotypes (HPV 18/31) that may be below the threshold of the sequencing platform but within detection range of the RT-PCR kit. The individual Sanger sequence traces, FASTA format DNA sequence and BLAST alignments are displayed in Appendix 3.

Table 3. 2 BLAST alignment of DNA sequences from selected cases

Case number	Percent identity	E-value*	Aligned target sequence with accession number
20	100%	9E-24	HPV 16 L1 gene (LC456606.1)
23	100%	8E-36	HPV 16 L1 gene (LC456606.1)
33	100%	1E-15	HPV 45 L1 gene (KU049757.1)
34	100%	1E-23	HPV 45 L1 gene (KU049757.1)
55	100%	2E-08	HPV 18 L1 gene (MH028425.1)

*The E-value is the probability of obtaining the sequence alignment by chance.

CHAPTER FOUR

DISCUSSION

4.1 Case demographics

The initial sample size of 57 was reduced to 23. Exclusion of cases with sub minimal tumour burden reduced the initial cohort numbers by more than 50% however this was necessary in order to increase the probability of extracting sufficient high quality hrHPV DNA. At the onset of the study, the available literature regarding HPV DNA extraction from archival tissue suggested that the yield would be poor and that alternative methods for enrichment of HPV DNA would be necessary. As a result, we initially focused on cases with high tumour burden in order to maximise the amount of HPV DNA extracted for hrHPV genotypic analysis.

Cervical squamous cell carcinoma is usually diagnosed by biopsy or loop excision following abnormal/HSIL Pap and owing to late presentation, patients are at a clinical stage that precludes total abdominal hysterectomy and will only receive chemotherapy and radiotherapy. As such, tumour samples may consist of small biopsies or loop excisions where the malignant, invasive tumoural cells may form a small portion of the total area of the specimen. As this study also excluded areas of HSIL, the numbers were further reduced. By proving that housekeeping genes and hrHPV DNA were successfully amplified from archival tissue, future investigations will include the remainder of the outstanding cases and thereby reach the total number of 57. Thereafter, it is envisioned that the hrHPV genotyping RT-PCR will be applied to all cases for squamous cell carcinoma in that time period.

Cervical squamous cell carcinoma was lowest in extremes of ages as represented by 20-29 and 70-79 with the highest burden present in 40-49 followed by 30-39, 50-59, 60-69 (Figure 3.1). This age peak is similar to the mean age of 50.7 years described by van Aardt *et al*. (2015) in 299 cases of cervical squamous cell carcinoma in women living in the Gauteng province of

South Africa. Interestingly, the median age for HIV-negative women in that cohort was 55.8 years and HIV-positive women 41.4 years ($p<0.0001$). Previous studies investigating cervical dysplasia in HIV-positive and negative women in Cape Town showed that the age groups with highest prevalence of HSIL (CIN II and III) are 25-29 and 30-34 years (McDonald *et al*., 2014) which may account for the bulk of invasive disease occurring in a slightly older population in this study.

The percentage of confirmed HIV-positive patients in this pilot study was less than estimated in the general population (~30% amongst pregnant women, Mbulawa *et al*., 2015) and that previously published in investigations of cervical dysplasia and carcinoma in South African populations (Table 4.1). More HIV-positive cases may be present in the "HIV unknown" group however this cannot be confirmed.

Table 4. 1 HIV prevalence in South African women with CIN and cervical carcinoma

Publication	**HIV-positive**	**HIV-negative**	**HIV unknown**
Van Aardt *et al*, 2015	25.8%	51.5%	22.7%
Denny *et al*., 2014	24.6%	64.5%	11%
McDonald *et al*., 2014	14.6%	85.4%	-

4.2 DNA extraction and housekeeping gene endpoint PCR

As previously mentioned, initial review of the available literature regarding DNA extraction from FFPE implied potential limiting factors in achieving adequate yield of high quality DNA for downstream molecular analysis (Steinau *et al*., 2011, Cannavo *et al*., 2012, Zhong *et al*., 2013, Heikal *et al*., 2014,). The gDNA extracted from the pilot study, however, showed adequate concentrations and purity (see Appendix 1). This eliminated the need for further investigations into alternative extraction methodologies. The simple modification in this study

of overnight proteinase K digestion at 37ºC (versus 56ºC for 1 hour) yielded adequate results and was a more convenient workflow.

Housekeeping gene endpoint PCR was successful in initial experiments of cases 3 (year 2004), 19 (year 2008) and 56 (year 2014) using *GAPDH* gene primers (Figure 3.2). PCR product yield was highest in 2014 and lowest in 2004, in keeping with the widely held understanding that increased age of FFPE specimens resulted in DNA deterioration and less efficient amplification. This reduction in PCR product yield can theoretically be remedied by increasing gDNA template concentrations (assuming no PCR inhibitors are present) however subsequent reamplification PCR studies and the touchdown protocol showed no obvious difference in the ability of HPV L1 sequences to be successfully amplified over the entire time period studied (see Figures 3.5-3.8). The housekeeping genes chosen for this study amplify products of 222bp and 159bp which are longer than the PCR product of HPV L1 target sequence (150 bp) with GP6+/GP6+ primers. Successful amplification of housekeeping target sequences of these longer lengths suggested that amplification of the GP5+/GP6+ 150 bp product would be possible. Therefore, alternative shorter L1 targets (e.g., via MY09/11 and SPF_{10} primers) were not investigated further. In addition, a recent review discussing HPV detection from FFPE recommended the use of GP5+/GP6+ primers (Kocjan *et al.*, 2016). Ultimately, the amplification of the housekeeping genes are only to prove that DNA of sufficiently good quality was extracted so that HPV genotyping could proceed.

4.3 HPV L1 amplification by endpoint PCR

Targeting the L1 gene by PCR is a useful screening tool as it identifies all HPV genotypes. Optimisation of HPV L1 endpoint PCR was necessary to show that there was adequate amounts of HPV DNA in the samples to proceed with RT-PCR analysis owing to the high cost of commercial HPV genotyping kits (R280 per sample as per 2018 pricing for the Sacace kit).

Initial L1 amplification experiments using GP5+/GP6+ primers were unsuccessful following standard published PCR conditions. Attempts to modify the standard PCR conditions by adjusting various parameters (cycle numbers, annealing temperatures and times) were also unsuccessful (results not shown). A reamplification PCR experiment was also performed to assess whether lack of PCR product was due to low HPV target sequences within extracted gDNA. A PCR product of correct size was observed (Figure 3.3) but reamplification PCR was not a viable option for this project as it required twice the amount of reagents (*Taq* polymerase, buffer, primers) and two PCR experiments to identify HPV. This was considered both expensive in terms of biochemical consumables and time. Also, there would be an increased potential for cross contamination.

Investigation into alternative methods for L1 detection via GP5+/GP6+ PCR revealed a comprehensive paper outlining a touchdown PCR technique that increased the detection limit of L1 sequences by sequentially lowering the annealing temperature by 0.5°C/cycle in order to improve the yield of amplicons and increasing the total number of cycles to 50 (Evans *et al*., 2005). The GP5+/GP6+ primers have an unusually low annealing temperature and may therefore not anneal under standard temperatures used in PCR reactions. Touchdown protocol 3 (TDP3) was chosen from the paper (Evans *et al*., 2005, see Table 2.5) as it showed optimal amplification down to a template DNA amount of 1 ng.

4.4 Optimisation of touchdown PCR parameters for GP5+/GP6+ primers

A pilot experiment for touchdown PCR was performed on case 3 as this was the oldest and theoretically ought to have the least amount of intact, amplifiable HPV L1 DNA. If amplification was successful in this case, it was reasoned that later cases would be successfully amplified. In addition to applying the published touchdown PCR conditions, further optimisation of gDNA concentration, primer concentration and $MgCl_2$ concentration were

undertaken to maximise yield. Mg^{2+} is required for annealing of template DNA and primer and to increase efficiency of PCR amplification. Excessive Mg^{2+} can however decrease specificity leading to non-specific band smearing on DNA gels. Too little, however, may result in no PCR amplification. Decreasing gDNA also dilutes potential PCR inhibitors that may be present in the sample and also conserves gDNA for further experiments. An initial concern that the increased numbers of cycles in the touchdown protocol may result in depletion of adequate subminimum dNTP concentrations and *Taq* polymerase activity were not justified. Touchdown PCR was successful in case 3 (Figure 3.4) and the optimised PCR conditions were applied to the remainder of the cases.

Touchdown PCR using this method allowed for detection of HPV L1 sequences in 22/23 cases (Figure 3.5). PCR product yield was not reduced in older cases compared to newer cases with GP5+/GP6+ touchdown PCR, unlike the trend noted in the *GAPDH* experiment (Figure 3.2). Touchdown PCR was useful in screening samples to ensure sufficient concentration of HPV DNA and of adequate integrity for subsequent genotype PCR amplification was present.

In case 56 that did not show L1 amplification (Figure 3.5 lane 26), detection of housekeeping genes was successful (Figure 3.2). This lack of amplification may therefore reflect a lower concentration of HPV DNA in this sample (compared to host housekeeping gene), beyond the threshold detection limit of DNA agarose electrophoresis rather than other factors, *viz*., poor DNA quality, PCR inhibitors. Also possible is that the area for primer binding may have been fragmented and not available for annealing.

4.5 Real time PCR (RT-PCR) detection of hrHPV genotypes

The preliminary screening of samples for DNA integrity (housekeeping gene PCR) and HPV burden/dose (GP5+/GP6+ PCR) are necessary prior to RT-PCR due to cost concerns as mentioned above. The advantage of multiplex RT-PCR however is the potential identification

of up to 14 hrHPV genotypes (plus internal housekeeping gene control) per sample in 4 reaction tubes versus 14 separate reactions. RT-PCR also allowed for rapid determination of the presence/absence of target sequences in comparison to endpoint PCR that required an entire PCR run, gel electrophoresis, UV transillumination and digital image capture. The case that showed housekeeping gene amplification but not L1 amplification (case 56) was included in the RT-PCR experiments as the different amplification modality, coupled with different target HPV DNA sequences (E6/E7 versus L1) was expected to yield positive results.

One of the major limitations of the RT-PCR experiments involved the number of available fluorescence detection filters. The Sacace kit required four filter set to distinguish the four separate dye-tagged hydrolysis probes (FAM, JOE, ROX, Cy5). The available LightCycler™ 480 II only had three filter set combinations. The Red 610 filter spectra overlapped with the JOE and ROX dyes on the kit hydrolysis probes. In this kit, hrHPV genotypes fall in these spectra (HPV 18, 31, 35, 45, 52, 59, 66 and 68) therefore only positive samples in the Red 610 channel were confirmed by DNA sequencing and sequence alignment.

Though the positive DNA samples (included with the kit) are at equal gene equivalents, there is a wide range of RT-PCR amplification curves and ultimately C_t values (Figures 3.8-3.10). This reflected differences in hydrolysis probe length and annealing chemistry at a single annealing temperature. All positive controls however showed $C_t<33$.

Detection of *β-globin* internal control DNA sequences was achieved in 20/23 samples with cases 8, 26 and 39 not being amplified (Figure 3.9). These cases however showed housekeeping gene and L1 amplification in the endpoint PCR experiments (Fig 3.7) and subsequent RT-PCR with genotype specific hydrolysis probes (Fig 3.12). A possible reason for the discrepancy may be the fragmentation of the *β-globin* DNA compared to either *GAPDH* or β-actin. Of note, the internal control (*β-globin*) had a relatively late C_t of 25 and implies

that this target may potentially show false negative results if total gDNA was present in lower concentrations. Repeat experiments with increased gDNA concentrations may have potentially increased the detection of *β-globin* (not done in current study).

Overall, 22/23 (96%) of cases showed the presence of hrHPV. Case 54 did not show RT-PCR amplification despite exhibiting amplification with *β-globin* housekeeping gene RT-PCR (Figure 3.9) and GP5+/GP6+ primers (Figure 3.5). This may be due to the possible presence of an hrHPV gene not covered in the Sacace kit, similar to the findings by Meiring *et al*., 2012 where next generation sequencing identified hrHPV genotypes present that were not covered in the commercial kit.

HPV 16 was the most common genotype, accounting for 21/22 hrHPV cases (Figure 3.10). As all cases were standardised with regards to gDNA template concentration, comparison of Ct values for the various cases was possible. As such, cases 3, 23 and 44 showed the lowest Ct values (15, 16 and 16, respectively) indicating the highest concentrations of HPV 16 while cases 12, 33 and 38 showed the lowest with all having late calls at C_t=35 (Figure 3.10). The most common single hrHPV infection was also HPV 16 in previous studies (Odida *et al*., 2008; Denny *et al*., 2014; McDonald *et al*., 2014; van Aardt *et al*., 2015) and global trends (Schiffman *et al*., 2007).

HPV 18 was detected in 2/22 hrHPV cases (cases 42 and 55) with C_t values of 33 and 27, respectively (Figure 3.11). HPV 18 probes are labelled with ROX dye and cannot be distinguished from HPV 31 (JOE dye) via the LightCycler™ system. These cases were Sanger sequencing and subsequent BLAST analysis confirmed the genotype. Case 55 showed a 100% match for HPV 18 (Table 3.2). Case 42 however did not yield a sequence that matched to HPV 18 despite repeating the sequencing. The BLAST algorithm showed 100% match for HPV 16 which is a co-infecting genotype in this case. The lack of obtaining HPV 18 sequence in this

case is likely due to suboptimal target DNA concentration for Sanger sequencing, which may also be implied by noting that the RT-PCR curve C_t value for this case was fairly late at 33 (Figure 3.11).

HPV 33 was identified in only one case (case 37) (Figure 3.12). This is unusual as HPV 33 is regarded as one of the most common hrHPV genotypes in various studies in African populations (see Table 4.3). However, HPV prevalence varies from country to country and regions within countries and reflects difference in sexual behaviour and co- morbidities such as HIV infection and may account for this difference.

HPV 39 was identified in cases 33 and 34 (Figure 3.13). Identification of HPV 39 in these two cases is unusual as this genotype is rare in South African population (approximately 0.6%) while in Nigerian populations, it is 1.4% which is equivalent to HPV 33 (compared to 6.6% in South African populations) (Denny *et al*., 2014).

HPV 45 was identified in cases 33 and 34 as well (Figure 3.14). As this genotype probe was tagged with JOE probe and showed possible overlap with HPV 59, these cases were also Sanger sequenced with subsequent BLAST analysis. These additional investigations confirmed 100% match with HPV 45 (Table 3.2).

The relative frequency of hrHPV in this study is therefore HPV 16>>18=39=45>33.

The absence of HPV 35 is unusual as it is ranked third in prevalence in Ghanian, Nigerian and South Africans (Denny *et al*., 2014), Although tempting, care should be exercised when comparing the results of this limited pilot study to published results until the entire cohort has been analysed.

Table 4. 2 Most common hrHPV genotypes in cervical carcinoma in sub-Saharan Africa

Study	**Population location and period**	**hrHPV in descending order of prevalence**	**Method of hrHPV identification**
Van Aardt *et al*., 2015	South Africa (2003-2004, 2008-2011)	16>18>35>45>33 (HIV neg) 16>18>45>33>58 (HIV-positive)	LiPA with GP5+/GP6+ primers
Denny *et al*., 2014	Ghana, Nigeria, South Africa (2007-2010)	16>18>35>45>33 (South Africa)	LiPA with SPF_{10} primers
Odida *et al*., 2008	Uganda 1968-1992	16>18>45>31>35>39	LiPA with SPF_{10} primers
Ogembo *et al*., 2015	23 African countries	16>18>45>35>33	Various. Study was meta-analysis and systematic review

4.6 HIV status and hrHPV distributions

Out of the original cohort of 7 HIV-positive cases, only 2 HIV-positive cases contained sufficient tumour burden to be included in the final cohort. Both of these cases, numbers 39 and 56, showed infection by a single hrHPV genotype (HPV 16). The cases with 2 hrHPV genotypes (cases 34, 42 and 55) were all from HIV negative patients. The only case with 3 hrHPV genotypes (case 33) was also from an HIV negative patient. None of the HIV unknown cases showed more than one hrHPV genotype. A large proportion of cases did not have documented HIV test results either in the clinical history of the pathology specimen report or

on the DISA system. Nonetheless, the pilot study shows that multiple hrHPV genotypes are present in HIV negative cases.

HIV positive status would be expected to be higher in the local community and this may have been demonstrated if all cases had been tested. Van Aardt *et al* (2014) also showed that HIV-positive women had more multiple hrHPV than HIV negative women (27% versus 8%, p=0.001). Therefore, it is likely that future investigations into the remainder of the original cohort will include more HIV-positive women and may also reflect this trend.

4.7 Limitations of the current study

Limitations of the pilot study include low numbers of cases. Large proportion of cases were excluded on basis of low tumour burden. These could however be included in future with manual or laser microdissection techniques to preferentially extract small foci or tumour from surrounding involved stroma, thereby reducing the risk of diluting the HPV DNA with excess stromal tissue DNA.

An additional limitation pertained to the available detection filters in the LightCycler™ system. An alternative option would be to access another RT-PCR machine equipped with 4 filter detection capability (examples include Rotor-Gene™, ABI®7500), use of another dye that is standard with LightCycler™ 480 II in a non-overlapping spectrum (e.g., Cyan 500, 440/480 nm) or use of alternative commercially available kits using only 3 dyes (e.g., AmoyDx® HPV genotyping detection kit). With increased numbers of cases investigated, the numbers of detected genotypes in the JOE/ROX overlapping spectrum would increase and it would not be feasible to send all these for confirmation by sequencing.

4.8 Application to other HPV related carcinomas in different anatomical sites

The touch down GP5+/GP6+ PCR described here together with multiplex RT-PCR can be applied to other FFPE archival specimens of HPV-related ano-genital and head and neck carcinomas. As HPV associated carcinomas show better prognosis than HPV-negative in these anatomic sites, detection of hrHPV is essential. At present, guidelines recommend the use of p16 as a surrogate marker for hrHPV infection but future investigations may require definitive proof of hrHPV DNA and/or hrHPV subtyping to direct treatment and predict prognosis. Indeed, it may be of interest to compare the clinical progression and treatment response of hrHPV positive tumours based on their particular genotype and assess if multiple hrHPV genotypes adversely affect prognosis compared to single genotype infections.

The current study shows that the hrHPV genotype distribution overall mirrors that found in sub-Saharan and industrialised countries and vaccination of our local population with the commercially available HPV genotypes would be protective against cervical carcinoma. Investigation of hrHPV anogenital and oropharyngeal sites would also be of benefit to predict suitability for prevention of these cancers, the rates of which are increasing.

CHAPTER FIVE

CONCLUSION

This pilot study aimed to extract high quality DNA from FFPE cervical carcinoma cases with the aim to identify the hrHPV genotypes associated with each case.

Conventional kit based DNA extract methods were successful in obtaining DNA of adequate concentration and purity. *GAPDH* housekeeping gene and HPV L1 consensus sequences were successfully amplified by conventional and touchdown PCR techniques, respectively.

The multiplex RT-PCR kit from Sacace allowed for rapid detection of multiple hrHPV genotypes in the samples submitted with 22 showing the presence of hrHPV. Limitations of hrHPV detection in genotypes showing overlapping emission spectra in the RT-PCR setup were overcome by Sanger sequencing.

We conclude that adequate amounts of DNA can be extracted from archival FFPE tissue using standard techniques and that hrHPV DNA is present in adequate amounts for RT-PCR amplification and genotype identification. The positive results obtained in this study are encouraging and the remainder of the cohort of low tumour burden cases will be analysed in future experiments.

Touchdown end point PCR and RT-PCR are quick and efficient methods for the simultaneous detection of multiple hrHPV genotypes in FFPE case samples and are well suited to upscaling and potential application to other HPV associated carcinoma specimens.

REFERENCES

Bateman AC, Katundu K, Polepole P, Shibemba A, Mwanahamuntu M, Dittmer DP, et al. Identification of human papillomaviruses from formalin-fixed, paraffin-embedded pre-cancer and invasive cervical cancer specimens in Zambia: a cross-sectional study. Virol J. 2015;12:2.

Bernard HU, Burk RD, Chen Z, van Doorslaer K, zur Hausen H, de Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology. 2010;401(1):70-9.

Bishop JA, Andreasen S, Hang JF, Bullock MJ, Chen TY, Franchi A, et al. HPV-related multiphenotypic sinonasal carcinoma: An expanded series of 49 cases of the tumor formerly known as HPV-related carcinoma with adenoid cystic carcinoma-like features. Am J Surg Pathol. 2017;41(12):1690-701.

Cannavo I, Loubatier C, Chevallier A, Giordanengo V. Improvement of DNA extraction for human papillomavirus genotyping from formalin-fixed paraffin-embedded tissues. Biores Open Access. 2012;1(6):333-7.

Choi YJ, Park JS. Clinical significance of human papillomavirus genotyping. J Gynecol Oncol. 2016;27(2):e21.

Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet. 2019;393(10167):169-82.

Comprehensive cervical cancer control: A guide to essential practice. WHO guidelines approved by the guidelines review committee. Geneva 2014.

De Vuyst H, Alemany L, Lacey C, Chibwesha CJ, Sahasrabuddhe V, Banura C, et al. The burden of human papillomavirus infections and related diseases in sub-saharan Africa. Vaccine. 2013;31 Suppl 5:F32-46.

Denny L, Adewole I, Anorlu R, Dreyer G, Moodley M, Smith T, et al. Human papillomavirus prevalence and type distribution in invasive cervical cancer in sub-Saharan Africa. Int J Cancer. 2014;134(6):1389-98.

Depuydt CE, Beert J, Bosmans E, Salembier G. Human Papillomavirus (HPV) virion induced cancer and subfertility, two sides of the same coin. Facts Views Vis Obgyn. 2016;8(4):211-22.

Evans MF, Adamson CS, Simmons-Arnold L, Cooper K. Touchdown general primer (GP5+/GP6+) PCR and optimized sample DNA concentration support the sensitive detection of human papillomavirus. BMC Clin Pathol. 2005;5:10.

Gharizadeh B, Oggionni M, Zheng B, Akom E, Pourmand N, Ahmadian A, et al. Type-specific multiple sequencing primers: a novel strategy for reliable and rapid genotyping of human papillomaviruses by pyrosequencing technology. J Mol Diagn. 2005;7(2):198-205.

Heikal N, Nussenzveig RH, Agarwal AM. Deparaffinization with mineral oil: a simple procedure for extraction of high-quality DNA from archival formalin-fixed paraffin-embedded samples. Appl Immunohistochem Mol Morphol. 2014;22(8):623-6.

Kleter B, van Doorn LJ, ter Schegget J, Schrauwen L, van Krimpen K, Burger M, et al. Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. Am J Pathol. 1998;153(6):1731-9.

Kobayashi K, Hisamatsu K, Suzui N, Hara A, Tomita H, Miyazaki T. A review of HPV-related head and neck cancer. J Clin Med. 2018;7(9).

Kocjan BJ, Hosnjak L, Poljak M. Detection of alpha human papillomaviruses in archival formalin-fixed, paraffin-embedded (FFPE) tissue specimens. J Clin Virol. 2016;76 Suppl 1:S88-S97.

Lin J, Kennedy SH, Svarovsky T, Rogers J, Kemnitz JW, Xu A, et al. High-quality genomic DNA extraction from formalin-fixed and paraffin-embedded samples deparaffinized using mineral oil. Anal Biochem. 2009;395(2):265-7.

Mbulawa ZZ, Coetzee D, Williamson AL. Human papillomavirus prevalence in South African women and men according to age and human immunodeficiency virus status. BMC Infect Dis. 2015;15:459.

McDonald AC, Denny L, Wang C, Tsai WY, Wright TC, Jr., Kuhn L. Distribution of high-risk human papillomavirus genotypes among HIV-negative women with and without cervical intraepithelial neoplasia in South Africa. PLoS One. 2012;7(9):e44332.

McDonald AC, Tergas AI, Kuhn L, Denny L, Wright TC, Jr. Distribution of human papillomavirus genotypes among HIV-positive and HIV-negative women in Cape Town, South Africa. Front Oncol. 2014;4:48.

Meiring TL, Salimo AT, Coetzee B, Maree HJ, Moodley J, Hitzeroth, II, et al. Next-generation sequencing of cervical DNA detects human papillomavirus types not detected by commercial kits. Virol J. 2012;9:164.

Micalessi MI, Boulet GA, Pillet S, Jacquet J, Pozzetto B, Bogers JJ, et al. Comparison of SPF10 real-time PCR and conventional PCR in combination with the INNO-LiPA HPV Genotyping Extra assay for the detection and typing of human papillomavirus in cervical samples. J Virol Methods. 2013;194(1-2):113-7.

Mirghani H, Blanchard P. Treatment de-escalation for HPV-driven oropharyngeal cancer: Where do we stand? Clin Transl Radiat Oncol. 2018;8:4-11.

Molijn A, Kleter B, Quint W, van Doorn LJ. Molecular diagnosis of human papillomavirus (HPV) infections. J Clin Virol. 2005;32 Suppl 1:S43-51.

Odida M, de Sanjose S, Quint W, Bosch XF, Klaustermeier J, Weiderpass E. Human papillomavirus type distribution in invasive cervical cancer in Uganda. BMC Infect Dis. 2008;8:85.

Ogembo RK, Gona PN, Seymour AJ, Park HS, Bain PA, Maranda L, et al. Prevalence of human papillomavirus genotypes among African women with normal cervical cytology and neoplasia: a systematic review and meta-analysis. PLoS One. 2015;10(4):e0122488.

Ossel J, Litjens RJ, Reijans M, Brink AA, Ummelen M, Ramaekers FC, et al. Human papillomavirus typing by single tube multiplex amplification in real time (SMART): the PapillomaFinder(R) SMART 20 assay. J Clin Virol. 2014;61(4):540-7.

Pirog EC. Cervical Adenocarcinoma: Diagnosis of human papillomavirus-positive and human papillomavirus-negative tumors. Arch Pathol Lab Med. 2017;141(12):1653-67.

Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. J Gen Virol. 2012;93(Pt 9):1853-68.

Riemer AB, Keskin DB, Zhang G, Handley M, Anderson KS, Brusic V, et al. A conserved E7-derived cytotoxic T lymphocyte epitope expressed on human papillomavirus 16-transformed HLA-A2+ epithelial cancers. J Biol Chem. 2010;285(38):29608-22.

Salazar KL, Zhou HS, Xu J, Peterson LE, Schwartz MR, Mody DR, et al. Multiple human papilloma virus infections and their impact on the development of high-risk cervical lesions. Acta Cytol. 2015;59(5):391-8.

Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. Lancet. 2007;370(9590):890-907.

Skead, G. Detection early cervical carcinoma by immunohistochemistry: Can MET meet the challenge. MMed Thesis (2016), Division of Anatomical Pathology, Department of Pathology, UCT.

Snietura M, Chelmecka-Wiktorczyk L, Pakulo S, Kopec A, Piglowski W, Drabik G, et al. Vertically transmitted HPV-dependent squamous cell carcinoma of the external auditory canal: Case report of a child. Strahlenther Onkol. 2017;193(2):156-61.

Steinau M, Patel SS, Unger ER. Efficient DNA extraction for HPV genotyping in formalin-fixed, paraffin-embedded tissues. J Mol Diagn. 2011;13(4):377-81.

van Aardt MC, Dreyer G, Pienaar HF, Karlsen F, Hovland S, Richter KL, et al. Unique human papillomavirus-type distribution in South African women with invasive cervical cancer and the effect of human immunodeficiency virus infection. Int J Gynecol Cancer. 2015;25(5):919-25.

van den Brule AJ, Pol R, Fransen-Daalmeijer N, Schouls LM, Meijer CJ, Snijders PJ. GP5+/6+ PCR followed by reverse line blot analysis enables rapid and high-throughput identification of human papillomavirus genotypes. Journal of clinical microbiology. 2002;40(3):779-87.

van Dyne EA, Henley SJ, Saraiya M, Thomas CC, Markowitz LE, Benard VB. Trends in human papillomavirus-associated cancers - United States, 1999-2015. MMWR Morb Mortal Wkly Rep. 2018;67(33):918-24.

Vinh-Hung V, Bourgain C, Vlastos G, Cserni G, De Ridder M, Storme G, et al. Prognostic value of histopathology and trends in cervical cancer: a SEER population study. BMC Cancer. 2007;7:164.

Wilfinger WW, Mackey K, Chomczynski P. Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. Biotechniques. 1997;22(3):474-6, 8-81.

Zhong H, Liu Y, Talmor M, Wu B, Hui P. Deparaffinization and lysis by hydrothermal pressure (pressure cooking) coupled with chaotropic salt column purification: a rapid and efficient method of DNA extraction from formalin-fixed paraffin-embedded tissue. Diagn Mol Pathol. 2013;22(1):52-8.

Zohoncon TM, Ouedraogo TC, Brun LVC, Obiri-Yeboah D, Djigma WF, Kabibou S, et al. Molecular epidemiology of high-risk human papillomavirus in high-grade cervical intraepithelial neoplasia and in cervical cancer in Parakou, Republic of Benin. Pak J Biol Sci. 2016;19(2):49-56.

APPENDIX 1

Buffers and reagents

1. EDTA (ethylenediaminetetraacetic acid) stock solution (0.5 M). 93.06 g EDTA (Sigma, St. Louis, MO, USA) was dissolved in 490 ml dH_2O, titrated to pH 8.0 and made up to a final volume of 500 ml. This solution was autoclaved and kept at room temperature (RT).
2. 5X Tris-borate-EDTA (TBE) buffer stock solution: 54 g Tris, 27.5 g boric acid and 20 ml of 0.5 M EDTA stock solution were dissolved in 980 ml of dH_2O and made up to a final volume of 1000 ml without pH titration. This solution was autoclaved and kept at RT . One part 5X TBE stock solution was diluted with 4 parts water to yield a 1X TBE working solution.
3. HotStarTaq® Master Mix Kit (2X) was from QIAGEN (Valencia, CA, USA). This solution, when diluted 1:1, contains final concentrations of 2.5 units HotStarTaq® polymerase, 1.5 mM $MgCl_2$ and 200 μM of each dNTP. The solution was stored at -20ºC until use. When required, it was fully thawed, gently mixed and kept on ice. This product requires an initial activation step of 15 min at 95°C before the PCR cycling is commenced.
4. Sterile water (Adcock Ingram, Johannesburg, South Africa). This was aliquoted into 1.5 ml Eppendorf tubes and kept at -20°C until required.
5. 25 mM $MgCl_2$ stock solution (QIAGEN, Valencia, CA, USA). This stock solution was diluted according to desired final concentration with the understanding that a baseline of 1.5 mM $MgCl_2$ was present in the HotStarTaq® Master Mix. The stock solution was kept at -20°C until required.
6. 1.5% agarose gel: Agarose (Bioline, London, UK) . 1.3 g of agarose was added to 80 ml 1XTBE in a 250 ml Erlenmeyer flask, microwaved until boiling and the agarose

completely dissolved. The solution was cooled, stock ethidium bromide solution added (5 μl) and poured into a Perspex mould and allowed to set at RT. 12 well gel combs were used. A 1.5% agarose concentration was chosen for optimal separation of target DNA amplicon sizes.

APPENDIX 2

Extracted FFPE DNA concentrations of cases

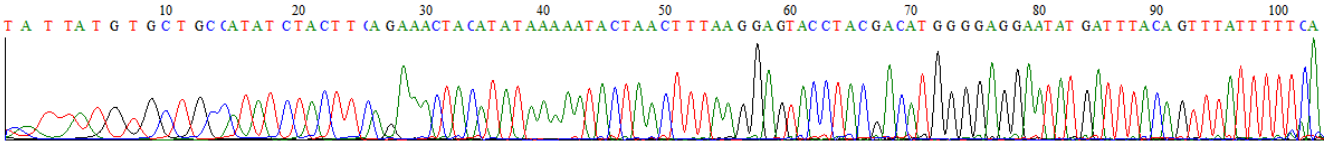
Case number	DNA concentration (ng/µl)	A_{260}/A_{280}
3	309.5	1.88
4	791.4	1.88
8	606.4	1.81
12	776.6	1.80
15	685.5	1.67
19	473.7	1.90
20	360.1	1.87
23	494.8	1.91
26	774.8	1.78
30	641.7	1.88
33	521.0	1.94
34	425.6	1.91
35	387.8	1.92
37	342.5	1.87
38	527.2	1.92
39	309.5	1.85
42	136.3	1.88
44	496.3	1.89
51	359.5	1.67
54	420.2	1.93

55	335.7	1.93
56	91.3	1.90
57	385.8	1.90

APPENDIX 3

Case 20

Sanger sequencing electropherogram:



Cleaned DNA sequence in FASTA format (5'-3'):

ATATAAAAATACTAACTTTAAGGAGTACCTACGACATGGGGAGGAATATGATTTACAGT

TTATT

BLAST alignment:

Human papillomavirus type 16 K1753 DNA, complete genome
Sequence ID: LC456606.1Length: 7905Number of Matches: 1
Range 1: 6699 to 6762GenBankGraphicsNext MatchPrevious Match

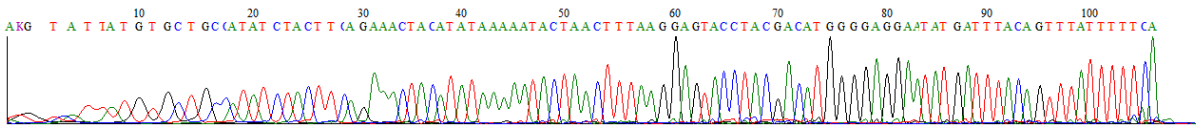
Score	Expect	Identities	Gaps	Strand
119 bits(64)	9e-24	64/64(100%)	0/64(0%)	Plus/Plus

```
Query  1     ATATAAAAATACTAACTTTAAGGAGTACCTACGACATGGGGAGGAATATGATTTACAGTT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6699  ATATAAAAATACTAACTTTAAGGAGTACCTACGACATGGGGAGGAATATGATTTACAGTT  6758

Query  61    TATT  64
             ||||
Sbjct  6759  TATT  6762
```

Case 23

Sanger sequencing fluorescence trace:



Cleaned DNA sequence in FASTA format (5'-3'):

TATCTACTTCAGAAACTACATATAAAAATACTAACTTTAAGGAGTACCTACGACATGGGG

AGGAATATGATTTACAGTTTATTTTT

BLAST alignment:

Human papillomavirus type 16 K1753 DNA, complete genome
Sequence ID: LC456606.1Length: 7905Number of Matches: 1
Range 1: 6680 to 6765GenBankGraphicsNext MatchPrevious Match

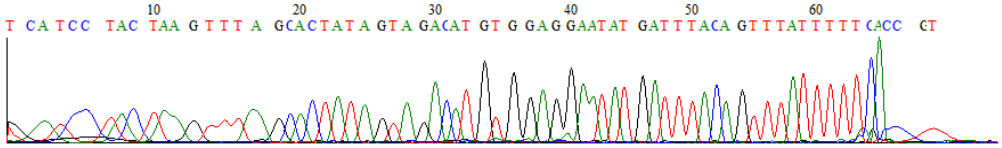
Score	Expect	Identities	Gaps	Strand
159 bits(86)	8e-36	86/86(100%)	0/86(0%)	Plus/Plus

```
Query  1     TATCTACTTCAGAAACTACATATAAAAATACTAACTTTAAGGAGTACCTACGACATGGGG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6680  TATCTACTTCAGAAACTACATATAAAAATACTAACTTTAAGGAGTACCTACGACATGGGG  6739

Query  61    AGGAATATGATTTACAGTTTATTTTT  86
             |||||||||||||||||||||||||
Sbjct  6740  AGGAATATGATTTACAGTTTATTTTT  6765
```

Case 33

Sanger sequencing fluorescence trace:



Cleaned DNA sequence in FASTA format (5'-3'):

AGCACTATAGTAGACATGTGGAGGAATATGATTTACAGTTTATTTTTCA

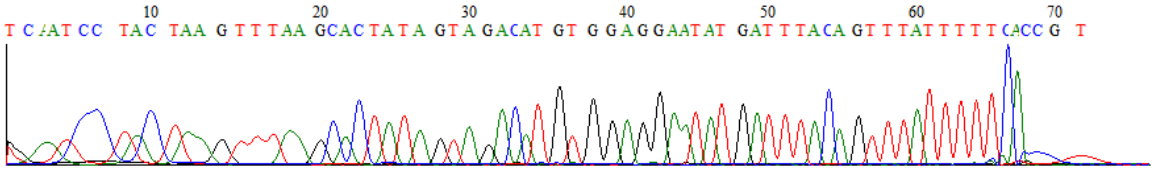
Human papillomavirus type 45 isolate 14Uk35 minor capsid protein L2 (L2) and major capsid protein L1 (L1) genes, complete cds
Sequence ID: KU049757.1Length: 2905Number of Matches: 1
Range 1: 2466 to 2514GenBankGraphicsNext MatchPrevious Match

Score	Expect	Identities	Gaps	Strand
91.6 bits(49)	1e-15	49/49(100%)	0/49(0%)	Plus/Plus

```
Query  1     AGCACTATAGTAGACATGTGGAGGAATATGATTTACAGTTTATTTTTCA  49
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2466  AGCACTATAGTAGACATGTGGAGGAATATGATTTACAGTTTATTTTTCA  2514
```

Case 34

Sanger sequencing fluorescence trace:



Cleaned DNA sequence in FASTA format (5'-3'):

ATCCTACTAAGTTTAAGCACTATAGTAGACATGTGGAGGAATATGATTTACAGTTTATTT

TTCA

BLAST alignment:

Human papillomavirus type 45 isolate 14Uk35 minor capsid protein L2 (L2) and major capsid protein L1 (L1) genes, complete cds

Sequence ID: KU049757.1Length: 2905Number of Matches: 1

Range 1: 2451 to 2514GenBankGraphicsNext MatchPrevious Match

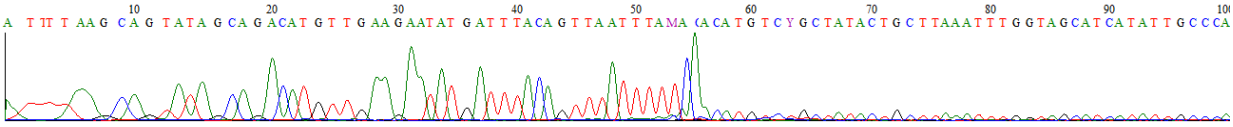
Score	Expect	Identities	Gaps	Strand
119 bits(64)	1e-23	64/64(100%)	0/64(0%)	Plus/Plus

```
Query  4     ATCCTACTAAGTTTAAGCACTATAGTAGACATGTGGAGGAATATGATTTACAGTTTATTT  63
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2451  ATCCTACTAAGTTTAAGCACTATAGTAGACATGTGGAGGAATATGATTTACAGTTTATTT  2510

Query  64    TTCA  67
             ||||
Sbjct  2511  TTCA  2514
```

Case 55

Sanger sequencing fluorescence trace:



Cleaned DNA sequence in FASTA format (5'-3'):

AAGCAGTATAGCAGACATGTTGAAGAATATGATTT

Human papillomavirus type 18 isolate 294_08 L1 gene, partial cds
Sequence ID: MH028425.1Length: 1323Number of Matches: 1
Range 1: 1015 to 1049GenBankGraphicsNext MatchPrevious Match

Score	Expect	Identities	Gaps	Strand
65.8 bits(35)	2e-08	35/35(100%)	0/35(0%)	Plus/Plus

```
Query  1     AAGCAGTATAGCAGACATGTTGAAGAATATGATTT  35
             |||||||||||||||||||||||||||||||||||
Sbjct  1015  AAGCAGTATAGCAGACATGTTGAAGAATATGATTT  1049
```