UNIVERSITY OF THE
FREE STATE
UNIVERSITEIT VAN DIE
VRYSTAAT
YUNIVESITHI YA
FREISTATA

UFS
UV

MOLECULAR ASSAYS FOR DETECTING HUMAN PAPILLOMAVIRUS IN HEAD AND NECK SQUAMOUS CELL CARCINOMA

TUMELO ROBERT SEKEE

JULY 2016

HEALTH SCIENCES
GESONDHEIDSWETENSKAPPE

UFS·UV

MOLECULAR ASSAYS FOR DETECTING HUMAN PAPILLOMAVIRUS IN HEAD AND NECK SQUAMOUS CELL CARCINOMA

TUMELO ROBERT SEKEE

B.Sc. (Honours)

Submitted in fulfilment of the requirements in respect of the MMedSc Virology degree qualification completed in the Department of Medical Microbiology and Virology in the Faculty of Health Sciences at the University of the Free State

Supervisor: Professor Felicity Jane Burt

Department of Medical Microbiology and Virology

Faculty of Health Sciences

University of the Free State

Co-supervisor: Dr Dominique Goedhals

Department of Medical Microbiology and Virology

Faculty of Health Sciences

University of the Free State

The financial assistance of the National Research Foundation and the Poliomyelitis Research Foundation is hereby acknowledged.

University of the Free State, Bloemfontein, South Africa

11 JULY 2016

Table of Contents

DECLARATIONS

I, Tumelo Robert Sekee declare that the master's research dissertation that I herewith submit at the University of the Free State, is my independent work and that I have not previously submitted it for a qualification at another institution of higher education.

I, Tumelo Robert Sekee hereby declare that I am aware that the copyright is vested in the University of the Free State.

I, Tumelo Robert Sekee hereby declare that all royalties as regards intellectual property that were developed during the course of and in connection with the study at the University of the Free State will accrue to the University.

Signature:_____

PRESENTATIONS AND PUBLICATIONS

Presentations

Sekee TR, Goedhals D, Seedat RY, Burt FJ. Polymerase chain reaction for the detection of human papillomaviruses in head and neck cancers. 47th Faculty Research Forum, University of the Free State 28-29 August 2014. Oral presentation.

Sekee TR, Goedhals D, Seedat RY, Burt FJ. The screening of human papillomaviruses in head and squamous cell carcinoma biopsies using polymerase chain reaction. PathReD (Pathology Research and Development Congress) Emperors Palace Johannesburg, South Africa 14-16 April 2015. Poster presentation.

Sekee TR, Goedhals D, Seedat RY, Burt FJ. Molecular assays for detecting HPV in HNSCC (3 minutes thesis competition) University of the Free State 27th May 2015. Oral presentation.

Sekee TR, Goedhals D, Seedat RY, Burt FJ. Screening of human papillomavirus (HPV) from patients with confirmed head and neck squamous cell carcinoma (HNSCC) in a small cohort study. 48th Faculty of Health Sciences, University of the Free State. Faculty Research Forum 27-28 August 2015. Oral presentation.

Sekee TR, Goedhals D, Seedat RY, Burt FJ. Screening of human papillomavirus (HPV) from patients with confirmed head and neck squamous cell carcinoma (HNSCC). 4th Annual Free State provincial health research day (12-13 November 2015). Oral presentation

Sekee TR, Goedhals D, Seedat RY, Burt FJ. Molecular assays for detecting human papillomavirus in head and neck squamous cell carcinoma. Virology Africa 2015. Radisson Blu Hotel, Cape Town (30 November-3 December 2015). Poster presentation.

Sekee TR, Goedhals D, Seedat RY, Burt FJ. Preparation of transcribed RNA for use as a positive control for detection of transcriptionally active human papillomaviruses. 49th Faculty Research Forum, University of the Free State 25-26 August 2016. Oral presentation

Sekee TR, Goedhals D, Seedat RY, Burt FJ. Detection of human papillomavirus in tissue biopsies from patients with head and neck squamous cell carcinoma in the Free State province, South Africa using E6 multiplex hemi-nested type specific PCR. 6th European congress in Virology Hamburg, Germany 19-22 October 2016. (**Abstract accepted**).

ACKNOWLEDGEMENTS

My heavenly father for taking care of me throughout this journey.

"I am the light of the world. Whoever follows me will never work in darkness, but will have the light of life.''-John 8:12

My supervisor, **Professor Felicity Jane Burt** who never gave up on me and gave me guidance through this project.

My co-supervisor **Doctor Dominique Goedhals** for her valuable feedback and input on this project.

Professor Riaz Seedat from the Department of Otorhinolaryngology from the University of the Free State, Faculty of the Health Sciences for providing the tissue biopsies used in the project.

My colleagues, **Natalie Viljoen, Danelle van Jaarsveldt, Yuri Munsamy, Atang Bulani, Carina Combrink, Deborah Damane, Maxwell Sokhela** and **Thomas Thipih.**

Armand Bester for his valuable input on Bioinformatics.

The National Health Laboratory Services Research Trust foundation and National Research Foundation for funding the project.

The National Health Laboratory Services for allowing me to do this project by making use of their facilities.

The Poliomyelitis Research Foundation and National Research Foundation for providing financial assistance over the period of two years (2014 and 2015).

The postgraduate bursary from School of Medicine, University of the Free State for financial assistance in 2015 and 2016.

My family, **Medupi Senoge, Ipeleng Sekee, Boipelo Sekee, Omphile Sekee** , **Kelebogile Sekee and Tshwane Sekee.**

My friends **Isaac Wool Booi, Zikholie Stephen Mxaka, Zanele Sekano, Mpho Samuel Mahamotsa and Isaac Moroeroe** who were always encouraging me to push hard.

ABSTRACT

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common malignancy worldwide and is traditionally associated with alcohol and tobacco. However over the past few decades there has been a decrease in smoking and drinking but still an increase in incidence of HNSCC with reports across South America, Europe and Asia. The increase in incidence is now attributed to human papillomavirus (HPV), an etiological agent of cervical cancer**.** HPV belongs to the *Papillomaviridae* family and over 150 HPV types have been identified. HPVs can be grouped into three groups based on the association with cancer; high risk HPV (HR-HPV) types which are associated with cancer, low risk HPV (LR-HPV) types which are not associated with cancer and possible cancer causing HPV types. Little is known about the association between HPV and HNSCC in South Africa (SA) with few studies conducted in Northern Transvaal and to our knowledge none in the Free State. Additionally, there is no standardized method that can be used for the detection of HPV in HNSCC. Therefore the aims of this study were to investigate molecular assays that can be used for detection of HPV types circulating in the Free State province, SA and to develop a method that can be used to determine transcriptionally active HPV. Three molecular assays were compared by screening for HPV DNA in a total of 74 tissue biopsies from patients with confirmed head and neck tumours. A nested polymerase chain reaction (PCR) that targets part of the L1 region, an E6 multiplex hemi-nested type specific PCR using type specific primers for HPV types -6,-11, -16, -18, -31, -33, -45, -58 and -84 that target the E6 region and the Roche linear array (LA) that target part of the L1 region. To investigate the performance of the Roche LA assay, the PCR targeting the L1 region was repeated on selected samples using modified primers PGMY11/09 and GP5+/6+ (nested PCR). A total of 4/74 (5.4%) samples tested positive for HPV DNA by nested PCR and sequencing analysis revealed HPV types -11, -16, -18 and -31. A total of 5/74 (6.8%) samples tested positive by the E6 multiplex hemi-nested type specific PCR which included the four HPV types already genotyped by nested PCR and an additional HPV type -45. Using the LA assay 60/74 (81.1%) samples tested positive for HPV DNA; 57/60 samples were positive for HPV type -84, one sample positive for HPV type -45 and two samples were positive for co-infections (-16/84 and -18/84). A conventional PCR was used to screen 10/57 samples that tested positive for HPV type -84 and all the 10 samples tested negative. Due to the fact that HR-HPV types are known to be carcinogenic, four samples from this study that tested positive for HR-HPV

types -16, -18, -31 and -45 were tested for transcriptionally active HPV infection by developing an E6 HnRT-PCR. All samples tested negative for HPV E6mRNA using the E6 HnRT-PCR. In conclusion the E6 multiplex hemi-nested type specific PCR detected all five HPV types in the study whereas nested PCR did not detect HPV type -45 and the Roche LA did not detect HPV types -11 and -31. Therefore the E6 multiplex hemi-nested type specific PCR will be used to screen additional samples for HPV DNA in tissue biopsies from head and neck tumours in our laboratory. However there is a limitation that needs to be kept in mind when working with the E6 multiplex hemi-nested type specific PCR and expanding the primers to include other HR-HPV types would be applicable.

Keywords: Human papillomavirus, head and neck squamous cell carcinoma, PCR, RT-PCR, L1, E6, integration, mRNA, alcohol, tobacco.

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

NOTE: Figures 1.1, 2.1a, 3.1a and 3.1b were self-constructed.

LIST OF TABLES

Chapter 2

Chapter 3

LIST OF ABBREVIATIONS

Ad-Adenovirus

AJCC-American Joint committee on cancer

BLAST-Basic Local Alignment Search Tool

Bp-Base pairs

cDNA-Complementary deoxyribonucleic acid

CR-Conserved region

CUP-Carcinoma of unknown primary

DNA-Deoxyribonucleic acid

EDTA-Ethylenediaminetetraacetic acid

EU-European Union

FDA-Food and Drug Administration

gDNA-Genomic deoxyribonucleic acid

hc2-Hybrid Capture 2

HNSCC-Head and neck squamous cell carcinoma

HNCs-Head and neck cancers

HnRT-PCR-Hemi-nested reverse transcriptase polymerase chain reaction

HPV-Human papillomavirus

HR-HPV-High risk human papillomavirus

ICTV-International Committee of Taxonomy of Viruses

IARC-International Committee on Taxonomy of Viruses

IPTG-Isopropyl-β-D-1-thiogalactopyranoside

ISH-*In situ* hybridization

kDA-Kilodalton

LA-Linear array

LB-Luria Bertani

LB/amp-Luria Bertani containing ampicillin at a final concentration of 100µg/ml

LCR-Long control region

LR-HPV-Low risk human papillomavirus

MCHA-Microplate colorimetric hybridization assay

mRNA-Messenger RNA

NFW-Nuclease free water

NGS-Next generation sequencing

NISH-Non-isotopic *in situ* hybridization

NTP-Nucleotide triphosphate

OCC-Oral cavity cancers

OPC-Oropharyngeal cancers

ORF-Open reading frame

Ori-Origin of replication

OSCC-Oral squamous cell carcinoma

PCR-Polymerase chain reaction

PCR-RFLP-Polymerase chain reaction restriction fragment length polymorphism

PV-Papillomavirus

qPCR-Real time polymerase chain reaction

RNA-Ribonucleic acid

RPM-Rotations per minute

RRP-Recurrent respiratory papillomatosis

RT-PCR-Reverse transcriptase polymerase chain reaction

SA-South Africa

SCC-Squamous cell carcinoma

SOC-Super optimal broth with catabolite repression

Tag-Tumour antigen

TAE-Tris acetate ethylenediaminetetraacetic acid

TNM-Tumour, node and metastases

U-Units

URR-Upstream regulatory region

VLP-Virus like particles

X-gal-5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

CHAPTER 1: LITERATURE REVIEW

1.1. Introduction

Human papillomavirus (HPV) belong to the *Papillomaviridae* family. There are five genera, namely *Alphapapillomaviruses, Betapapillomaviruses, Gammapapillomaviruses, Mupapillomaviruses and Nupapillomaviruses* (de Villiers *et al*., 2004; Bernard *et al*., 2010; Bzhalava *et al*., 2015). These genera are differentiated from each other based on their deoxyribonucleic acid (DNA) sequence, different replication characteristics and disease association (Doorbar *et al*., 2012).

Two genera, namely the *Alphapapillomaviruses* and the *Betapapillomaviruses* are comprised of the cutaneous HPVs which target the skin of the hands and feet (Burd, 2003), with the Alphapapillomavirus genus also containing the mucosal HPVs which infect the lining of the mouth, throat, respiratory tract or anogenital tract (Burd, 2003; Chung and Gillison 2009; Cubie, 2013).

1.2. Classification

Initially papillomaviruses (PVs) were grouped together into one family with the polyomaviruses and the family was known as the *Papovaviridae* (de Villiers *et al*., 2004). They were placed into one family based on the characteristics that both families shared including non-enveloped capsids and DNA genomes that are circular (de Villiers *et al*., 2004). However it was later found that they differed in terms of their genome organization, the size of their genomes, their amino acid sequence and nucleotide sequences. Currently the International Committee on Taxonomy of Viruses (ICTV) distinguishes them as two different families, *Papillomaviridae* and *Polyomaviridae* (de Villiers *et al*., 2004). PVs can be distinguished from other viruses based on their circular double stranded DNA genome and in that they are non-enveloped (IARC, 2007). PVs have an L1 open reading frame (ORF) which is conserved within the genome and which has been used for the past 15 years for the identification of new PV types (de Villiers *et al*., 2004). HPV that have differences in nucleotide identity of less than 2% within the L1 late gene are defined as variants. Nevertheless, nucleotide variability among variants differs across viral genes and can be as high as 5% in the upstream regulatory region (URR) or

long control region (LCR). Mostly, viral variants arise by nucleotide substitutions in a few restricted positions within the entire genome (Betiol *et al*., 2013). For identification purposes a small part within the HPV genome is sequenced, for instance 400 base pairs (bp) within the LCR region or 450bp within the E6 gene. A new PV type is recognized if the complete genome has been cloned and the DNA sequence differs by more than 10% from the closest known type (IARC, 2007). Differences between 2% and 10% homology define a subtype. A consequence of this redefinition was that the traditional subtypes (e.g. HPV-6a, HPV-6b and HPV-6c) had to be eliminated, as they showed less than 2% sequence diversity. Originally the term subtype had a different definition and it was used when isolates that were different but came from the same type differed partially in their restriction enzyme cleavage patterns, such as HPV 2a, HPV 2b, and HPV 2c. It later became clear that these subtypes would rather fall under the category variants and this principle has been applied to numerous HPV types. Previously the terms supergroups or major branches were used to identify higher-order clusters of HPV types for instance the genital PVs. A term genus is now used instead of supergroups or major branches. Various genera share nucleotide sequence similarity of 60% within the L1 ORF (de Villiers *et al*., 2004). Earlier terms like groups, subgroups or minor groups were used to identify lower-order clusters of HPV types like HPV -6, -11, -44 and -55. The new term species was introduced for these taxa (de Villiers *et al*., 2004). HPV with nucleotide similarity of between 60% and 70% within a genus are known as species (de Villiers *et al*., 2004).

1.3. Viral genome

HPV is a small icosahedral virus that is non-enveloped, 50-60 nanometers in size and has a circular double stranded DNA genome of approximately 8000 bp (IARC, 2007). There are three regions within the HPV genome; a non-coding URR, the early region and the late region. The non-coding URR is 1000bp in length and is also known as the LCR (Burd, 2003). The URR or LCR region has core promoter p97, enhancer and silencer sequences which are involved in the regulation of DNA replication by controlling the transcription of the ORFs. This region has the highest degree of variation within the genome of the virus and it functions in regulation of gene expression and replication (Rampias *et al*., 2014). The early region is 4000bp in size and encodes for proteins E1, 2, 4, 5, 6, and 7 and these proteins play a role in viral replication, viral gene expression and transformation (Burd,

2003, Rampias *et al*, 2014). The third and last region is the 3000bp late region which encodes two viral structural proteins, L1 and L2 (Burd, 2003, Rampias *et al*., 2014). Figure 1 shows an illustration of the HPV genome.

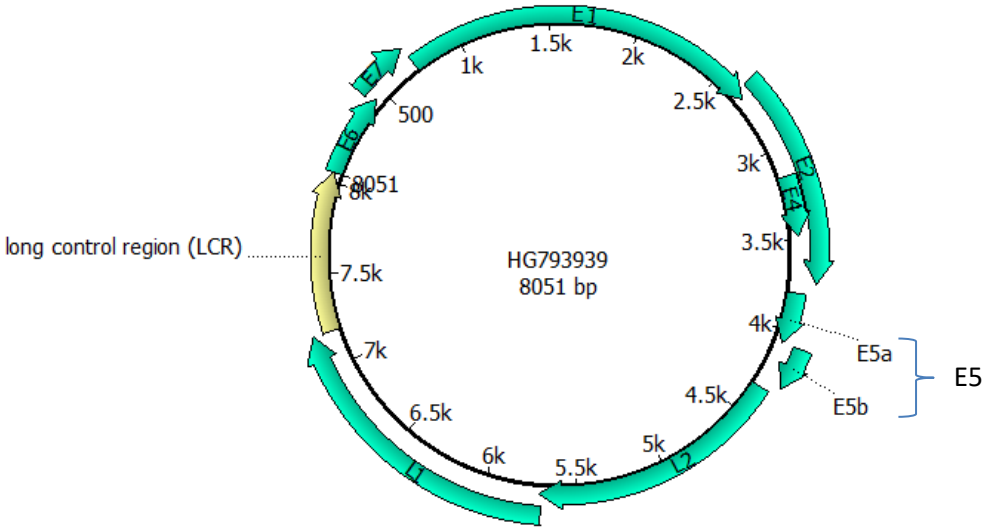


Figure 1.1 HPV genome schematic depicting the three regions found within the genome, the long control region, the early region (E1, 2, 4, 5, 6 and 7) and the late region (L1 and L2) of HPV type -6 isolate 131 (GenBank accession number: HG793939.1).

L1 protein

The L1 protein is the major capsid protein of the PVs (Favre *et al*., 1997). The PV virion has an exterior surface that is particularly knobby. The L1 protein is the primary structural element of infectious virions that contains 360 copies of the protein organized into 72 capsomeres (Doorbar, 2006). The L1 has two types of termini which are arranged as extended invading arms that form the floor between the capsomere knobs, namely N and C termini. The structure of the C-terminus of L1 is mostly established (Buck *et al*., 2013**).** The L1 ORF region is well conserved within the HPV genome and has been used for genotyping and identification of new HPV genotypes (Al-Shabanah *et al*., 2013) as well as construction of phylogenetic trees (Bernard *et al*., 2010). The L1 protein is involved in mediating efficient virus infectivity (Doorbar, 2006).

L2 protein

L2 is the minor capsid protein and its role includes assembly of the PV and facilitating efficiency of virus infectivity (Doorbar, 2006; Wang and Roden, 2013). A single L2 protein

may be present in the centre of pentavalent capsomers at the vertices of the virion (Doorbar, 2006). Despite the paucity of L2 in the virion, this minor capsid protein has recently been shown to have many functions. It contributes to the binding of virions to the cell receptor(s), facilitates virion uptake and transport to the nucleus, delivers the viral DNA to replication centres and helps the packaging of the viral DNA into capsids. By virtue of the presence of a neutralization epitope common to L2 proteins of many PVs, may be instrumental in conferring immunity across different types of PVs (IARC, 2007).

E1 protein

The E1 protein is 73 kilodalton (kDA) (IARC, 2007), a hexameric DNA helicase and the only enzyme and the most conserved protein encoded by PVs (Bergvall *et al*., 2013). The size of the protein ranges from 600 to 650 amino acids depending on the type of PV. This protein can be divided into three segments which have different functions. The three segments are as follows; an N-terminal regulatory region which optimizes reproduction *in vivo* but nonetheless is dispensable *in vitro*, a central origin binding domain which is also known as the DNA binding domain which recognizes a specific site in the origin of DNA replication (ori) and lastly the C terminal enzymatic domain which is sufficient for self-assembling into hexamers that display ATPase activity and are capable of unwinding short DNA duplexes (Bergvall *et al*., 2013). It is involved in several functions; in the initiation and catalysis of viral DNA synthesis and it must first recognize a specific segment of the viral genome known as the ori (Bergvall *et al*., 2013). For optimal function of the ori, a palindromic E1-binding region and an AT-rich sequence are required.

E2 protein

E2 proteins are sequence specific DNA binding proteins and the gene encodes a product of around 40-45 kDA, depending on the PV (IARC, 2007). These proteins are the main regulator of viral gene transcription; binds the viral transcriptional promoter as a dimer is involved in viral DNA replication and interacts with and recruits E1 to the origin (IARC, 2007). They bind 0.012kb motifs which are located inside the URR of the viral genomes. These proteins are expressed at two stages of the virus life cycle; the early and late stages. E2 consists of a preserved N terminal which is the transactivation domain of 200 amino acids that are linked to a C terminal DNA binding/dimerization domain of about 100 amino acids. The hinge which is a flexible linker sequence connects these two

domains (McBride, 2013). According to McBride (2013) these proteins are multi-functional and involved in many viral processes, mostly associated with transcription and regulation of the viral genome. The proteins regulate viral DNA transcription, play an important role in cell transformation, initiating and inhibiting apoptosis, transcriptional regulation, and in the modulation of the immortalizing and transformation potential of HPV (Morshed *et al*., 2014). The inactivation of the E2 protein results in the development of tumours by promoting the expression of E6 and E7 oncogenes while the active E2 inhibits the two oncoproteins (E6 and E7) thus resulting in an increase in p53 expression and apoptosis of the infected cells (Morshed *et al*., 2014).

E4 protein

The HPV E4 gene is located in the E region and overlaps with E2 but is transcribed in a different reading frame. The E4 is a cytoplasmic protein disturbing the structural framework of the keratin (Morshed *et al*., 2014) and is heterogeneous protein with the major form being a fusion product with a 5-amino acid sequence from the N-terminus of E1 and which is sometimes detected in the cell nucleus. The functions of E4 have been suggested to play a role in facilitating and supporting viral genome amplification, the regulation of late gene expression, the control of virus maturation and the mediation of virus release (IARC, 2007).

E5 protein

This protein is not encoded by all PVs. According to DiMaio and Petti (2013) the E5 gene is situated at the 3' end of the early region of the viral genome and is expressed from a spliced messenger ribonucleic acid (mRNA) that initiates up stream of the E2 gene. The E5 protein is roughly 40 to 85 amino acids in length and these amino acids are hydrophobic and grouped into one or more putative amino transmembrane domains (DiMaio and Petti, 2013). It is assumed that these proteins do not have intrinsic enzyme activity rather they act by modulating the activity of multiple cellular proteins (DiMaio and Petti, 2013). The E5 is small in size, hydrophobic and does not have large globular domains, so cannot mediate specific protein interactions. The protein makes use of another mechanism to occupy their target proteins (DiMaio and Petti, 2013). E5 protein is involved in cell transformation and participates in viral DNA replication. This protein

also allows for the infected cell to avoid being recognized by the immune system (Morshed *et al*., 2014).

E6 protein

According to Howie *et al* (2009) the E6 proteins consist of about 150 amino acids which contain two zinc like fingers joined by an inter domain linker of 36 amino acids, flanked by short amino (N) and carboxyl (C) terminal domains of variable lengths. E6 is an important oncogene in HPV associated neoplasias which target a number of cellular pathways, one of which is the blocking of the p53 tumour suppressor protein which leads to inhibition of apoptotic signalling that would under normal conditions eliminate HPV infected cells (Howie *et al*., 2009). The E6 oncogene protein connects to the p53 protein, leading to its proteolytic degradation and this may result in an uncontrolled replication of infected cells (Morshed *et al.,* 2014). The p53 tumour suppressor protein which is one of the well-studied interacting proteins of the E6 gene is a DNA site specific transcription factor, which forms part of the most important signalling regulators within the cell resulting from genotoxic or cytotoxic stress (Howie *et al*., 2009). The p53 suppressor protein is involved in inhibiting the growth of cells, arresting the cell cycle at several points and under certain conditions it activates the apoptotic mechanism that then leads to cell death (Ashcroft and Vousden, 1999).

E7 protein

E7 is an oncogene protein (Lajer and Von Buchwald, 2010) and is comprised of around 100 amino acid residues. The E7 oncogene protein has an amino terminus which has sequence similarity to the portion of the conserved region (CR) 1 and the entire CR2 of the adenovirus (Ad) E1A. It also contains sequence similar to simian vacuolating virus 40 large tumour antigen (T Ag) (McLaughlin-Drubin and Munger, 2009). E7 protein does not share any extensive similarity with cellular proteins, even though the E7 protein has some sequence motifs, particularly the LXCXE sequence which is also found in cellular proteins (Roman and Munger, 2013). The E7 oncogene protein plays a central role in HPV-dependent malignant transformation and causes the impairment of the control of cell cycle regulation and cell maturation. During malignant transformation, the E7 oncogene protein binds and inactivates the pRb protein preventing it from binding to the E2F transcription factor and thereby promoting cell cycle progression. This functional

inactivation of pRb results in a reciprocal overexpression of p16 tumour suppressor protein $p16^{INK4A}$ (Lajer and Von Buchwald, 2010; Morshed *et al*., 2014)

1.4. Replication

PVs are species-specific and these DNA viruses have a particular tropism for squamous epithelial cells and replicating within the nucleus of the squamous epithelia (Howley and Lowy, 2007). The early and the late phases which separate the PVs reproductive infection in the host cells are associated with the epithelial cell's state of differentiation (Howley and Lowy, 2007). For lesion formation, the basal stem cells when infected are associated with formation of a persistent lesion, although it has been suggested that for high risk types this might not be required, as they can stimulate cell proliferation (Doorbar *et al*., 2012). Infection of the basal epithelial cells most probably occurs by exposure to the virus via abrasions or microwounds (Howley and Lowy, 2007, Lazarczyk *et al*., 2009; Doorbar *et al*., 2012). Specific binding of the PV virions to the alpha 6 integrin subunit receptor (Evander *et al*., 1997) as well as the interaction of the HPV virions with heparin and cell surface glycosaminoglycans on human keratinocytes (Joyce *et al*., 1999) ensures that the virus gains entry into the host cell. It is then that the PV virions are taken up by receptor mediated endocytosis (Acheson, 2007; Howley and Lowy, 2007). The viral capsid is disassembled within the endosome, followed by migration of the viral genome across the cytoplasm and into the nucleus with the assistance of the L1 major capsid protein (Howley and Lowy, 2007, Doorbar *et al*., 2012). The transcription of the PV is strictly controlled by the state of differentiation of infected squamous epithelia (Howley and Lowy, 2007). In the nucleus of the undifferentiated basal epithelia, the genome of the virus is maintained at low numbers of approximately 100 copies per cell (Acheson, 2007; Longworth and Laimins, 2004; Ai *et al*., 2000). As basal cells differentiate to keratinocytes, there is a burst of viral DNA replication referred to as vegetative replication (Longworth and Laimins, 2004). The L1 and L2 viral capsid proteins are expressed and virion assembly occurs for production of progeny virion particles. The virus particles are released upon shedding and death of the epithelial cell at the surface of the lesions (Acheson, 2007; Howley and Lowy, 2007).

1.5. Diseases associated with HPV

The mucosal group within the Alphapapillomavirus genus is divided into three groups, based on whether they cause malignancy (Roman and Munger, 2013; Cubie, 2013). The LR-HPV types that are non-malignant are -6; -11; -40; -42; -43; -44; -54; -61; -62; -71; -72; -81; -83 and -84 (Abreu *et al*., 2012), whereas HR-HPV types -16, -18, -26; -31; -33, -35; -39; -45; -51; -52; -53; -56; -58; -59; -66; -68; -70; -73; -82 and -85 causes malignancy (Abreu *et al*., 2012). HPV types -68 and -73 are defined by the World Health Organisation (WHO) as being possible cancer-causing (Doorbar *et al*., 2012).

HPV infection by different strains can infect any area of the skin or mucous membrane. Different strains are linked to different skin diseases which range from common warts to tumours (Ljubojevic and Skerlev, 2014). HPV types -16, -18, -31 and -45 are mostly associated with cervical cancer (Burd, 2003) while HPV type -2 frequently causes common warts which are characterized by multiple irregular, rough nodules which show different patterns at different sites of trauma particularly on fingers, but also on other frequently rubbed and abraded skin like hands, elbows and knees (Cubie, 2013). HPV types -3, -10 and -28 are known to cause plane warts which are small and less rough showing as flat-topped papules, flesh coloured or lightly pigmented, especially on light exposed areas of the face and back of the hands, usually in multiple crops (Cubie, 2013). HPV type -4 is associated with punctate lesions most often seen on palms of the hands (Cubie, 2013). Persistent and florid warts which are seen in fishmongers and meat handlers due to the skin chronically macerated due to moisture and cold are usually associated with HPV type -7 (Cubie, 2013). LR-HPV types -6 and -11 infections within the larynx can lead to the development of recurrent respiratory papillomatosis (RRP) and genital warts (Ljubojevic and Skerlev, 2014).

1.6. Transmission

There are different ways in which HPV can be transmitted, sexually or non-sexually. It occurs primarily by skin-to-skin contact (Burd, 2003). Genital HPV is transmitted sexually, which occurs by direct contact with infected tissue. A condom does not protect an individual from being exposed to HPV because it does not necesarily cover the infected tissue (Burd, 2003). Non sexual transmission includes vertical, horizontal, perinatal, autoinoculation and fomite transmission (Burd, 2003; Syrjänen, 2010a). Vertical

transmission occurs from mother to child and it has been suggested that it occurs via contact with vaginal and cervical mucosa during delivery and horizontal transmission during infancy (Erickson *et al*., 2013). Perinatal transmission certainly occurs; it appears that the only serious consequence of this transmission is recurrent laryngeal papillomatosis, which is fortunately extremely rare (Burd, 2003). Autoinoculation is also possible when an individual scratches one site of the body which is infected with HPV and touches another site of the body (Syrjänen, 2010a). HPV is known to be very resistant to desiccation and heat, so fomite transmission can also occur such as by prolonged exposure to shared contaminated clothing (Burd, 2003).

1.7. Diagnosis and detection

HPV cannot be grown in conventional cell cultures and serology has only limited accuracy (Torres *et al*., 2012). Therefore the accurate diagnosis of HPV infection relies on molecular assays (Torres *et al*., 2012). Currently nucleic acid hybridization assays, signal amplification assays and nucleic acid amplification are used for identification (Torres *et al*., 2012).

1.7.1. Nucleic acid hybridization assays

This type of assay includes methods like Southern blotting, *in situ* hybridization and dot blot hybridization (Snijders *et al*., 2010). Nucleic acid hybridization assays are used for the detection of HPV infection from samples taken from the cervix. There is an advantage to using these assays as they generate high quality information, however there are some drawbacks, they have low sensitivity, they are time consuming procedures and relatively large amounts of pure DNA is required (Snijders *et al*., 2010).

1.7.2. Signal amplification assays

Two such assays are in use namely, the Digene® HPV test (Hologic, Inc., Marlborough, MA, USA) using Hybrid Capture®2(hc2) (hc2, Digene Corp., USA) technology and Cervista® HPV HR assays (Hologic, Inc., Marlborough, MA, USA) (Poljak *et al*., 2016; Torres *et al*., 2012). These two assays are used for diagnostic purposes in the US and have been approved by the Food and Drug Administration (FDA). Both techniques detect concurrently 13 HPV genotypes (HPV-16, -18, -31, -33, -45, -51, -52, -56, -58, -59 and -68); Cervista HPV HR test further includes HPV-66. HC2 is an in vitro nucleic acid hybridization

assay with signal-amplification and Cervista is based on the Invader Chemistry®, which detects specific nucleic acid sequences using two isothermal reactions simultaneously (Torres *et al*., 2012; Poljak *et al*., 2016)

1.7.3. Nucleic acid amplification methods

Methods include polymerase chain reaction (PCR), polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP), real time polymerase chain reaction (qPCR), the Abbott real time HR-HPV, the linear array (LA) (Roche Molecular Diagnostics, Pleasanton, CA, USA), clinical arrays® HPV (Genomica SAU, Madrid, Spain) and microplate colorimetric hybridization assay (MCHA) (Boehringer Mannheim, Germany) (Snijders *et al*., 2010) just to mention a few.

The Abbot Real-Time HR-HPV test is a new assay which can detect 12 HPV genotypes (-31, -33, - 35, -39, -45, -51, -52, -56, -58, -59, -66 and -68) and can detect individual HPV types -16/-18 (Kocjan *et al*., 2011) . Real-time is performed on a fully automated nucleic acid preparation instrument and the real-time PCR instrument using a modified GP5+/GP6+ primer mix. Additionally co-amplification of a 136-bp region of human beta-globin is used as an internal process control for sample adequacy, DNA extraction and amplification (Kocjan *et al*., 2011).

The LA uses both PCR and reverse blot hybridization. LA can detect 37 high- and low-risk HPV genotypes, i.e. -6, -11, -16, -18, -26, -31, -33, -35, -39, -40,-42, -45, -51, -52, -53, -54, -55, -56, -58, -59, -61, -62, -64, -66, -67, -68, -69, -70, -71, -72, -73 (MM9), -81, -82 (MM4), -83 (MM7), -84 (MM8), IS39, and CP6108 (Torres *et al*., 2012). Furthermore it targets the L1 region and uses biotinylated primers, PGMY 09/11 to amplify a 450bp fragment within the L1 region. Human beta-globin is co-amplified which ensures that the DNA extraction was successful (Torres *et al*., 2012).

The clinical arrays® HPV kit, detection and genotyping of HPV can be performed. The DNA extraction method is a modified procedure using absorption columns. The kit uses biotinylated primers that amplify a region of 451bp within the L1 region. To check for both the PCR procedure and DNA integrity, the kit includes a human cystic-fibrosis transmembrane conductance regulator (CFTR) gene and control plasmids. This allows the detection of 35 HR-HPV genotypes including (-16, -18, -26; -31; -33, -35; -39; -45; -51; -52;

-53; -56; -58; -59; -66; -68; -70; -73; -82 and -85) and the LR-HPV types (-6; -11; -40; -42; -43; -44; -54; -61; -62; -71; -72; -81; -83 and -84). The kit can also identify co-infections and simple infection (Otero-Motta *et al*., 2011).

The MCHA is a PCR based method on the amplification of the L1 region with an expected fragment size of 150bp using consensus primers GP5+/6+, followed by colorimetric hybridization to six type specific probes on microcell plates and detection through a colorimetric assay (Barcellos *et al*., 2011). The GP5+/6+ primers are designed for the most conserved viral region and have been used widely for the detection of a broad spectrum of HPV types. It can identify six HR-HPV types (-16, -18, -31, -33, -39 and -45) (Barcellos *et al*., 2011).

1.8. Vaccines

There are currently three types of vaccines that are used for the protection against HPV, namely the quadrivalent (Gardasil®), bivalent (Cervarix®) and the most recent 9-valent (Gardasil®-9) formulations (Serrano *et al*., 2012; Printz, 2015). These three vaccines are prepared from virus-like particles (VLP) that are a non-infectious protein shells derived from the L1 major capsid protein (Levin *et al*., 2010; Serrano *et al*., 2012). These empty viral capsids self-assemble into VLP and effectively mimic a natural HPV viral infection, but are not infectious, given that they do not contain any DNA (Garland and Smith, 2010)

The Gardasil® vaccine protects against two LR-HPV types (-6 and -11) and two HR-HPV types (-16 and -18) and is given to girls and young women aged from nine to 26 years (Serrano *et al*., 2012). The vaccine is also given to boys and younger men, aged between nine to 25 years (Center for Disease Control, 2010b). The vaccine protects against genital warts and cervical cancer in women and genital cancer and anal cancer in boys and men (Center for Disease Control, 2010b; Serrano *et al*., 2012). It is given intramuscularly in a series of three vaccinations at months zero, one to two and month six in children aged 14 years or older, and according to a two-dose regimen at zero and six months in children aged nine to 13 (Garland and smith, 2010; World Health Organization, 2014).

The Cervarix® vaccine protects against two HR-HPV types, -16 and -18 (Serrano *et al*., 2012). The vaccine therefore protects about 70% of cervical cancers (Serrano *et al*., 2012). The vaccine is given to girls and women aged nine to 25 years. It is administered

in three doses, at zero months, one to two months and six months (Garland and Smith, 2010).

Lastly, Gardasil®-9 is a recombinant vaccine which prevents five additional HPV types (-31, -33, -45, -52 and -58) to the already known HPV types prevented by Gardasil® (-6, -11, -16 and -18) (Printz, 2015). It has the potential to prevent approximately 90% of cervical cancers as well as 90% of the vulvar, vaginal and anal cancers that are caused by HPV (Printz, 2015). The five additional HPV types covered by Gardasil®-9 account for approximately 20 % of cervical cancers (Printz, 2015). The vaccine is administered as three injections (at months zero, one to two and lastly month six) and the maximum benefit is obtained by individuals who are vaccinated before becoming infected with the 9 HPV types. Gardasil®-9 is recommended for both sexes and is administered between nine and 15 years for males and nine to 26 years for females (Printz, 2015).

The VLP stimulate type specific neutralizing antibodies against the above mentioned HPV types (Levin *et al*., 2010). HPV vaccines (Cervarix® and Gardasil®) were first introduced in SA in 2008 in the private health care sector (Botha and Richter, 2015). In SA, Cervarix® was introduced nationwide in March 2014 for primary school girls using a two-dose regimen.

At present there is no data available evaluating the effectiveness of HPV vaccination for preventing oral cancers (Munoz *et al*., 2006). There are clinical trials that are underway where HPV vaccines are also being used in HNSCCs associated with HPV (Pai and Westra, 2009).

1.9. Head and Neck cancer

Head and neck cancer include a variety of tumours that are characterized by several different histological and etiological types in various anatomical sites. Traditionally, head and neck cancer has been grouped into categories tumour such as nasal cavity and paranasal sinuses, lip and oral cavity, pharynx, larynx, major salivary glands, thyroid and carcinoma of unknown primary (CUP) (Johansen and Eriksen, 2016). Anatomical sites of the head and neck cancer are shown in Figure 1.2.

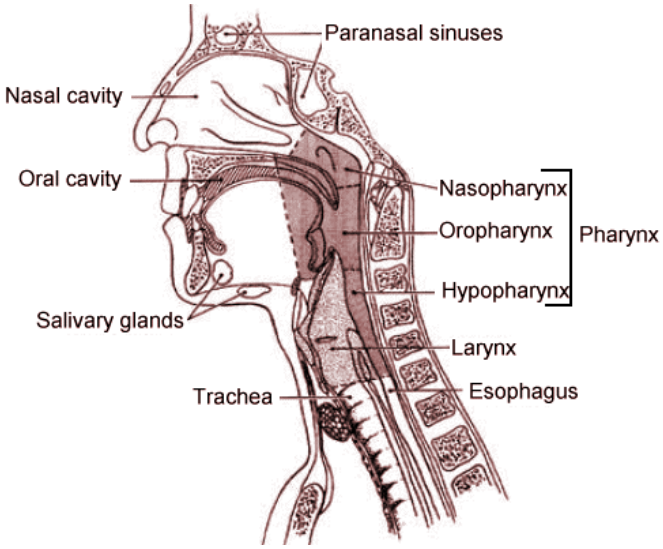


Figure 1.2. Anatomical sites of the head and neck [Source: Adapted from web site: (http://www.aboutcancer.com)].

1.9.1. Epidemiology of head and neck cancer

The incidence of oral cavity cancers (OCC) has declined in recent years in most parts of the world, consistent with declines in tobacco use (Chaturvedi *et al*., 2013). In contrast, oropharyngeal cancers (OPC) incidence has increased over the last 20 years in several countries, including Australia, Canada, Denmark, the Netherlands, Norway, Sweden, the United States, and the United Kingdom (Chaturvedi *et al*., 2013). In Stockholm, Sweden, with tonsillar cancer there was a 2.8 fold increase between 1970 and 2002, despite the decrease in incidence of smoking (Chaturvedi *et al*., 2013). Men are three times more at risk than women of acquiring head and neck cancer. In 1998 in the European Union (EU), 42 109 cases of cancer of the pharynx and oral cavity were reported by the EUCAN with 15 744 fatalities among males and 11 447 cases with 4 434 fatalities in females for a total of 53 556 cases with 20178 deaths (Pannone *et al*., 2011). The occurrence of HNSCCs differs from region to region. It accounts for 3 to 4% of all cancer diagnoses in North America and the EU (Pannone *et al*., 2011). In Africa and South East Asia, approximately 8-10% of all cancers are HNSCC (Pannone *et al*., 2011).

1.9.2. Risk factors for head and neck cancer

The two most important risk factors associated with head and neck cancer are alcohol use and tobacco (Syrjänen, 2005; Gavid *et al*., 2013). However recent studies have shown that HPV, an etiological agent associated with cervical cancer is also associated with

HNSCC, with approximately 25% of tissue biopsies from HNSCC, being positive for HPV (Furniss *et al*., 2009; Kumar *et al*., 2015).

1.9.2.1. Alcohol and tobacco

The incidence of HNSCC has increased over the last 30 years. Patients with HNSCC usually present with advanced metastatic disease leading to higher mortality. As SCC develops in the epithelium of the upper aerodigestive tract, repeated exposure to tobacco (smoked, chewed or taken as snuff) and alcohol are proved to be the primary risk factors related to HNSCC. The reason is the upper aerodigestive tract is the first part of the body that is exposed to these harmful carcinogenic components. The carcinogen may cause multiple neoplastic lesions in the area that causes initiation or progression of the HNSCC (Burke *et al*., 2014).

Tobacco smoking is well-established as a risk factor for HNSCC and this risk is correlated with intensity and duration of smoking (Pai and Westra, 2009; Burke *et al*., 2014; Kumar *et al*., 2015). Smoking cessation reduces but does not eliminate the risk of cancer development (Pai and Westra, 2009). In patients older than 50 years with HNSCC, association with tobacco smoking plays an important role in the development of HNSCC probably through immune suppression (Kumar *et al*., 2015). Environmental exposure to tobacco smoke (passive smoking) also appears to increase the risk of developing HNSCC, even for individuals who have never actively smoked (Pai and Westra, 2009). This increased risk for HNSCC related to tobacco smoking is largely due to the genotoxic effects of carcinogens in tobacco smoke including nitrosamines and polycyclic hydrocarbons (Pai and Westra, 2009). Case control studies in the United States of America have shown that individuals who had a history or a moderate (16-25 cigarettes per day) to heavy (40 cigarettes a day) smoking for 20 years are at risk for developing HNSCC (Burke *et al*., 2014). Cigarette smoke generates particular matter, gaseous extracts and water solubles. Major classified mutagenic and carcinogenic components of cigarettes are nicotine, tar, ammonia carbon monoxide, carbon dioxide, formaldehyde, acrolein, acetone, benzopyrenes, hydroxyquinone nitrogen oxide and cadmium (Kumar *et al*., 2015).

Heavy alcohol consumption is recognized as an independent risk factor of HNSCC particularly for hypopharyngeal cancer (Pai and Westra, 2009; Burke *et al*., 2014; Kumar *et al*., 2015). The risk of HNSCC is three times higher among individuals who consume alcohol than non-drinkers (Burke *et al*., 2014).

According to Furniss and colleagues (2009), individuals who consume 15-30 or more than 30 drinks per week are at an increased risk for HNSCC (2 fold and 3.5 fold respectively) than those who consume less than 5 drinks per week (Furniss *et al*., 2009). The International Agency for Research on Cancer of the WHO has categorized alcohol as a group one carcinogen (Kumar *et al*., 2015).

1.9.2.2. HPV

It is documented that HPV also plays a role in the pathogenesis of a subset of HNSCCs (Fakhry *et al*., 2008). The involvement of HPV in the carcinogenesis in oral and oropharyngeal cancer was first suggested in 1983 by Syrjänen who noted that 40% of the tumours in a study that they were doing had histological and morphological similarities with lesions associated with HPV (Syrjänen *et al*., 1983) and soon thereafter other authors supported what Syrjänen primarily proposed. This was established on the following evidence; the well assessed broad epithelial-tropism of HPV, the resemblances in morphologies amongst HPV found in the oropharyngeal and genital epithelia, the ability to immortalize human oral keratinocytes *in vitro*, the strongly established etiological role of HR-HPV in cervical SCC and finally the detection of HR-HPV genotypes in samples of oral squamous cell carcinoma (Pannone *et al*., 2011). Different types of HR-HPVs have been associated with the pathogenesis of HNSCCs. These types of HR-HPV infect different anatomical sites, for example HPV types -16, -31, and -33 have been found in the tonsilar oropharynx with other types (-35 and -45) have also been detected (Kreimer *et al*., 2005; Rampias *et al*., 2014; Wang *et al*., 2012). Besides those types of HR-HPVs mentioned above HPV type -18 has also been detected (Pannone *et al*., 2011). LR-HPVs especially HPV types -6 and -11 have also been identified in cancers of the head and neck and they have been found in some oral cavity, tonsilar and laryngeal cancers (Pannone *et al*., 2011).

1.10. Human papillomavirus and head and neck cancer in South Africa

The most common HNSCC that are registered in the cancer registries of South Africa (SA) are cancers of the oesophagus, gums, tongue, pharynx and oral cavity (Boy *et al*., 2006). Little research has been done in SA to determine HPV prevalence and its association with HNSCC. In 1985, Hille did a study to determine the role of HPV in oesophageal carcinoma in black males. A total of 24 samples were screened by histology looking for morphological manifestations of HPV infection and eight tested positive by this method. In 1986, Hille and colleagues did a study to screen for HPV infection in oesophageal carcinoma in black SA males in which a total of 70 oesophageal samples were screen by histological examination and immunohistochemical staining. From 70 biopsy specimens, 23 tested positive by histological screening (morphological manifestations of HPV infection) and seven of those positive by histology also demonstrated the presence of HPV antigen. In 1991, Williamson and colleagues did a study to investigate for the presence of HPV DNA in oesophageal biopsies from patients with and without cancer of the oesophagus by nested PCR targeting the L1 region. A total of 14 oesophageal carcinoma biopsies and 41 non-cancerous oesophageal biopsies were tested. A total of 10 oesophageal carcinoma biopsies tested positive for HPV DNA and six of the 41 non-cancerous biopsies also tested positive for HPV DNA.

In 1994, Togawa and colleagues did a study to detect HPV in oesophagus SCC by using a radioactive nested PCR. A total of 72 samples were tested of which 18 were from SA. Three of the 18 samples tested positive for HPV type -18.

In 1995, two studies were carried out to screen for HPV DNA in oesophageal cancer and oral squamous carcinoma (OSCC) by Cooper and Van Rensburg *et al*. respectively. Cooper screened 48 archival formalin fixed paraffin wax-embedded biopsy specimens using PCR and non-isotopic *in situ* hybridization (NISH). A total of 23 samples harboured HPV DNA within the nuclei using NISH and HPV positive cancers were distributed as HPV types -16 (84%); -18 (12%) and -6 (4%). Using PCR, six out of nine samples tested positive for HPV DNA. Matsha *et al*. (2002) did a study screening for HPV DNA in oesophageal cancer using nested PCR targeting the L1 region, in which 50 samples were screened and 23 tested positive for HPV types -11,- 16, -52, -39 and two unknowns.

Van Rensburg also did a retrospective study in 1995 looking at the prevalence of HPV DNA in oral squamous carcinoma (OSCC) in the west of the Transvaal in a rural black population by using *in situ* hybridization. A total of 66 samples were taken from patients and only one tested positive for HPV type -18. In 1996 Van Rensburg did another study analysing 146 samples from OSCC from black South Africans using E6 type specific PCR and a prevalence of 1.6% for HPV types -18 and -11 was reported in the study.

In 2006 a study was done by Boy to determine the prevalence of HPV in samples of OSCC using qPCR, ISH (conventional and signal amplification). Seven samples came up positive for HPV types -18 from 59 samples tested. Paquette *et al*. (2013) carried out a study which provided evidence that alpha-9 HPV infections are a major etiological factor for oropharyngeal carcinoma in black South Africans using three methods, PCR, *in situ* hybridization and $p16^{INK4a}$(surrogate marker). A total of 51 samples from 41 patients samples were tested and 48/51 samples came up positive for HPV types -16, -18, -31 and/or -33 with multiple co-infections of HPV types -16 and -18 or HPV types -16 and -31.

Davidson *et al.* (2014) did a pilot study on the prevalence of oral and oropharyngeal human papillomavirus in a sample of SA men. Seven samples came up positive for different HPV types with two samples from those seven having HR-HPV types -16 and -68.

In summary, studies investigating the role of HPV in HNSCC in South African populations are limited and have largely examined oesophageal cancer samples. These studies are also limited in their geographical range, focusing primarily on population in the area currently known as Gauteng, while the array of laboratory assays utilised in these studies makes comparison of the findings difficult.

1.11. Problem identification, aim and objectives

Cancer of the upper aerodigestive tract (hypopharynx, oral cavity and the lower part of the pharynx) has been traditionally associated with alcohol and smoking. However over the past decades despite a decrease in smoking rates, there has been an increase in the incidence of HNSCC around the world, with the increasing incidence having been reported across America, Europe and Asia. The increasing incidence of HNSCC is now attributed to HPV infection, a well-known cause of cervical cancer. HPV types -6, -11, -16, -18, -31, -33 and -58 are the most prevalent HPV types found in the HNSCC with other HPV types being rarely detected. Anatomical regions of the head and neck where these HPV types are most prevalent is as follows: HR-HPV types -16, -18 and -31 have been detected in tonsillar oropharynx, LR-HPV types -6 and -11 have been detected in the oral cavity and laryngeal cancers, and two HR-types -33 and- 58 have frequently been detected in the oral cavity. Since 1985 there have been a number of studies conducted on HNSCC and HPV in SA. However limited information is available with few studies having been done on this topic compared to studies conducted in Asia, Europe and South America. The South African studies were mostly performed in the northern parts of the country, while in the Free State province there is no information on the association between HPV and HNSCC. There is no standardised method that can be used to screen for HPV DNA in tissue biopsies from patients with HNSCC. A selection of methods that have been used previously include ISH, ICC using HPV-specific serum antibodies and molecular methods (PCR, RT-PCR and qRT-PCR). Although PCR is regarded as the gold standard for detecting HPV DNA in HNSCC samples there is a wide range of primers sets used in-house and commercial assays making a comparison of methods and results difficult. In this study commercial and in house assays were compared using a reasonably sized cohort of samples.

Aims of the study:

1. To investigate molecular assays that can be used to detect HPV DNA in tissue biopsies from patients with confirmed head and neck tumours.
2. To develop molecular assay to detect transcriptionally active HPV DNA in tissue biopsies.

Objectives

- To confirm the application of molecular assays for detection of HPV in biopsy material.
- To identify the HPV types present in the head and neck tumours in the Free State by determining the nucleotide sequence of purified PCR products.
- To develop molecular assays that will detect replicating/transcriptionally active HPV.

CHAPTER 2: MOLECULAR ASSAYS FOR DETECTION OF HUMAN PAPILLOMAVIRUS IN PATIENTS WITH CONFIRMED HEAD AND NECK TUMOURS

2.1. Introduction

Cancer of the head and neck is ranked sixth on the list of cancers in the world. Squamous cell carcinomas (SCC) account for 90% of head and neck cancers (HNCs) (Kermani *et al*., 2012). The major risk factors for SCC were previously alcohol and tobacco (Snow and Laudadio, 2010). However some patients do not have any obvious risk factors and in recent years, both epidemiologic and molecular evidence have established a strong link between HPV and the upper aerodigestive tract cancers (Psyrri and Dimaio, 2007; Kermani *et al*., 2012).

There are difficulties with detection of the virus (HPV) using cell culture and serology techniques as HPV cannot be easily cultured *in vitro* and serological assays cannot distinguish between past and current infection (Molijn *et al*., 2005). Molecular methods are therefore frequently used to screen for HPV DNA. Molecular assays include Southern blot, ISH, PCR based assays with and without probe detection of products, reverse transcriptase PCR (RT-PCR) for replicating HPV and next generation sequencing (NGS). In addition expression of p16 has been used frequently as a surrogate marker for the presence of HPV (El-Naggar and Westra, 2011). Each of the above mentioned molecular methods has its own advantages and disadvantages. Detection of E6/E7 oncogene expression is considered to be the gold standard for identification of HPV associated cancers (Larque *et al*., 2014), however amplification of HPV DNA using PCR is a frequently used method and the preferred method for routine detection of HPV and can be designed to target small fragments of DNA and hence has useful application for screening paraffin embedded tissues for the presence of HPV in retrospective studies (Venceslau *et al*., 2014). Many commercial assays and in-house assays used for the detection HPV DNA are based on the amplification of the L1 region which is reasonably well conserved between different HPV types. Most assays were originally designed for detection of HPV associated with cervical cancers and were based on a pair of nested consensus primer pairs designated MY11/09 and GP5+/6+ that amplify a region within the conserved region of the major viral capsid L1 gene (Manos *et al*., 1989; de Roda Husman *et al*., 1995). The MY 11/09 primers amplify a region of approximately 450bp within the L1 gene and the

GP5+/6+ amplify a region of approximately 140bp within the region targeted by the MY 11/09 primers (Saini *et al*., 2009) (Figure 2.1.).

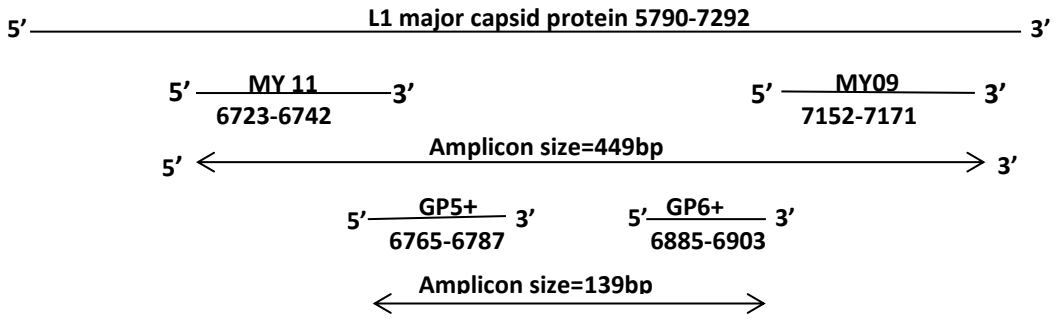


Figure 2.1. Diagram illustrating position of primers that target a region of the L1 gene. Positions shown are relative to HPV type -6a isolate 103C.6 (GenBank accession number KU298878.1). MY11/09 primers target the region between nucleotides 6723 to 7171 and GP5+/6+ are situated at nucleotide positions 6765 to 6903.

Although these primers have been subsequently modified by various researches to include type specific primers the original design has been used routinely within the laboratory of the Department of Medical Microbiology and Virology for the detection of HPV DNA in biopsy material from patients with RRP for detection of HPV types -6 and -11.

The incidence of viral integration in HPV associated HNSCC is not clear. During viral integration into the host DNA, the major viral capsid L1 gene may be disrupted leading to false negative results using assays based on amplification of the L1 capsid gene, thus leading to underestimation of the true prevalence of HPV. Therefore primers that target other regions may be preferred (Husnjak *et al*., 2000; Torrente *et al*., 2011) and type specific primers that amplify the E6 or E7 genes may be more effective, because they remain stable during viral integration (Torrente *et al*., 2011).

Commercial assays are a significant expense and target the L1 gene which sometimes gets disrupted during viral integration and have potential to give false negative results in HNSCC. In addition the targeted region may be larger than recommended for testing archived tissues in which DNA is fragmented due to fixing tissues in buffered formalin or similar. Hence, to investigate the incidence of HPV in patients with confirmed head and

neck cancers, a suitable molecular assay was required that would be sensitive, suitable for HR-types, could detect co-infection if required, targeted a small region so that the assay could have future application for screening large numbers of archived tissues and was cost-effective for screening large numbers. Many laboratories use the Roche Linear Array assay however the cost is prohibitive for large studies within the laboratory of the Department of Medical Microbiology and Virology. Therefore in this study fresh biopsy tissues from histologically confirmed head and neck tumours were screened for HPV DNA using three assays, an in-house nested PCR with MY11/09 and GP5+/6+ primer pairs, the Roche Linear Array assay and a novel in-house type specific multiplex PCR targeting the E6 gene (E6 hemi-nested type specific PCR).

2.2. Aim

The aim of this study was to investigate molecular assays that can be used to target HPV DNA in biopsy material from patients with confirmed head and neck tumours from the Free State, Bloemfontein SA using molecular assays that target the L1 and E6 genes respectively.

2.3. Objectives

1. To screen a total of 74 biopsy tissues from patients with histologically confirmed head and neck tumours using three molecular assays; a nested PCR, E6 multiplex hemi-nested type specific PCR and Roche Linear Array.
2. To compare the results obtained using the three molecular assays.
3. To identify genotypes of HPV identified in the biopsy samples.

2.4. Materials and methods

2.4.1. Sample collection

A total of 74 patients who were admitted to the Universitas Academic Hospital in Bloemfontein during the two and a half years study period (January2014-June2016) with histologically confirmed tumours of the oropharynx, oral cavity, larynx, hypopharynx or paranasal sinuses were enrolled in the study. Informed consent was obtained from each patient on enrolment by Professor Seedat from the Department of Otorhinolaryngology, Faculty of the Health Sciences, University of the Free State and a tissue biopsy collected from each patient was provided by Professor Seedat. Tissue biopsies were stored in

RNAlater® for preservation of the RNA. On submission of tissues to the research laboratory, the samples were assigned a laboratory number designated VBD followed by the year of collection. A portion of the biopsy was used for the extraction of DNA and the remaining segment was stored in RNAlater® at -20^{0}C for extraction of RNA. The study was approved by the Ethics Committee of the Faculty of Health Sciences, University of the Free State (137/2013B) (Appendix A).

2.4.2. DNA Extraction

DNA was extracted from fresh tissue biopsies using the QIAamp DNA Mini Kit from QIAGEN (QIAGEN, Valencia, CA, USA) according to the manufacturer's instructions. Briefly the tissue was lysed by the addition of 180µl of buffer ATL and 20µl of proteinase K. The sample was incubated at 56^{0}C until the tissue had completely lysed with vortexing to ensure that efficient lysis was achieved. The sample was centrifuged at 14 000xg for 10 seconds. A 200µl aliquot of AL buffer was added to the sample and incubated at 70^{0}C for 10 minutes and 200µl of 70% ethanol was added to the sample, vortexed and centrifuged at 14 000xg for 10 seconds. The mixture was transferred to a QIAamp mini spin column and centrifuged at 6 000xg for one minute. Wash steps were carried out using AW1 and AW2 buffers provided in the kit to remove residual contaminants. The DNA was eluted in 200µl of AE buffer. For short term storage the DNA was stored at -20^{0}C and for long term storage an aliquot was frozen at -80^{0}C. The DNA extraction was performed in a class two biosafety cabinet in a biosafety level two laboratory.

2.4.3. PCR detection of HPV DNA

Three molecular methods, a nested PCR targeting the L1 region, a multiplex hemi-nested PCR using type specific primers that target the E6 gene and the Roche Linear Array assay were used to screen for the presence of HPV DNA in tissue biopsies.

2.4.3.1. Preparation of positive controls

Nested PCR: DNA extracted from a biopsy obtained from a patient diagnosed with RRP and previously genotyped as HPV type -11 in an unrelated study was used as a positive control. E6 multiplex PCR: The multiplex hemi-nested PCR was designed to be performed as two reactions, targeting LR-HPV types and HR-HPV types in separate reactions. For the low risk reaction, HPV type -6, identified in an unrelated study from a

patient with RRP, was used as a positive control. Clinical samples were available for use as positive controls for HPV types -16, -18, -31 and -45. In the absence of clinical controls and to confirm that the E6 primers would amplify HPV types -33 and -58, two controls were prepared by using HPV types -33 and -58 gene fragments synthesized by GenScript and supplied in pUC 57. The vector map for pUC 57 plasmid and partial E6 gene of HPV types -33 and -58 are shown in Appendix D. To prepare DNA template for use as a positive control, PCR was performed using the forward primer targeting the HPV types -33 or -58 gene (HPV 33F, HPV 58F) and M13 reverse primer which targets a region of the pUC 57 plasmid (5'AGCGGATAACAATTTCACACA3') downstream of the inserted gene. PCR components are shown in Table 2.1. The reactions were cycled on a GeneAmp® PCR system instrument model 9700 as follows: initial denaturation, 95^0C for two minutes for one cycle, followed by 30 cycles of denaturation at 95^0C for 30 seconds, annealing 59^0C for 30 seconds and elongation at 72^0C for one minute and final elongation at 72^0C for five minutes for one cycle and lastly the sample was held at 4^0C indefinitely.

Table 2.1. Plasmid PCR for preparing the positive control used in the high risk PCR reaction

PCR components	**Final volume**	**Final concentration**
5x Green Go Taq® flexi buffer	10µ	1X
$MgCl_2$ solution, 25mM	4µl	2mM
PCR nucleotide mix, 10mM each	1µl	0.2mM
HPV 33F/58F (forward primer) (20pmol/µl)	1µl	0.4µM
M13 pUC (reverse primer) (20pmol/µl)	1µl	0.4µM
Go Taq® G2 Hot Start Polymerase (5U/µl)	0.25µl	1.25U
Plasmid DNA	1µl	-
Nuclease free water (NFW)	31.75µl	-
Total	**50µl**	-

2.4.3.2. Nested PCR targeting the L1 region

The nested PCR had two pairs of consensus primers, MY11/09 and GP5+/6+ that target the L1 capsid gene (Manos *et al*., 1989). In the first round of amplification, MY11/09

primers target nucleotides 6723-7171 relative to the nucleotide sequence of HPV type -6a, isolate 103C.6 (GenBank Accession number: KU298878.1) (Figure 2.1) and in the nested reaction, GP5+/6+ primers target nucleotides 6765-6903 relative to the nucleotide sequence of HPV type -6a isolate 103C.6 (Figure 2.1). The primer sequences are provided in Table 2.2.

Table 2.2. Consensus primers used in nested PCR targeting the L1 region of the HPV genome (Manos *et al*., 1989; de Roda Husman *et al*., 1995)

Primers	Nucleotide sequence	Forward /Reverse	*T_m	%GC	Amplicon size
MY 11	5'CGCCAGGGACATAACAATGG3'	Forward	64^0C	55	450bp
MY 09	5'CGTCCAAAAGGAAACTGATC3'	Reverse	60^0C	45	
GP5+	5'TTTGTTACTGTGGTAGATACTAC3'	Forward	61^0C	39	139bp
GP6+	5'GAAAAATAAATTGTAAATC3'	Reverse	46^0C	16	

*T_m and %GC calculated using Biomath from Promega (www.**promega**.com/a/apps/**biomath**/?calc=tm).

The PCR reaction was performed using Go Taq®G2 Hot Start Polymerase (Promega, Madison, USA) according to the manufacturer's instructions. PCR components are shown in Table 2.3. Negative and positive controls were included in each PCR run. For the negative control, a 5µl volume of NFW was added to the reaction instead of template. For the positive control a volume of 5µl template of a known positive sample (HPV type -11) was added. The beta-globin gene was amplified concurrently with each sample and it served as an internal control for DNA integrity. The beta-globin primers are shown in Table 2.4. The reactions were cycled using the following cycling conditions: initial denaturation, 95^0C for two minutes for one cycle, followed by 30 cycles of denaturation at 95^0C for 30 seconds, annealing 47^0C for 30 seconds and elongation at 72^0C for one minute and final elongation at 72^0C for five minutes for one cycle and lastly held at 4^0C indefinitely.

Table 2.3. PCR components for nested PCR using MY11/09 primers

PCR components	Volume	Final concentration
5XGreen Go Taq® flexi buffer	10µl	1X
$MgCl_2$ solution, 25mM	4µl	2.0mM
PCR nucleotide mix, 10mM each	1µl	0.2mM each dNTP
MY 11 primer (20pmol/µl)	1µl	0.4µM
MY 09 primer (20pmol/µl)	1µl	0.4µM
GoTaq ®G2 Hot Start Polymerase (5U/µl)	0.25µl	1.25 units (U)
Template DNA	5µl	-
NFW	27.75µl	-
Total	**50µl**	**-**

Table 2.4. Beta globin primers used as internal control (Das *et al*., 1993)

Primers	Nucleotide sequence	Forward/Reverse	*T_m	*%GC	Amplicon size
GH20	5'GAAGAGCCAAGGACAGGTAC3'	Forward	63^0C	55	268 bp
PC04	5'CAACTTCATCCACGTTCACC3'	Reverse	62^0C	50	

*T_m and %GC calculated using Biomath from Promega (www.**promega**.com/a/apps/**biomath**/?calc=tm).

All samples that were negative or inconclusive from first round were further tested using a pair of internal nested primers designated GP5+/6+ (second round) which amplify 139bp within the region amplified by the MY11/09 primers. The nested reactions were prepared as described in Table 2.3 replacing the MY11/09 primers with GP5+/6+ primers. A 1µl aliquot of the PCR amplicon from the first round was used as a template for the second round PCR. The PCR was done as described previously, using the following cycling conditions: initial denaturation, 95^0C for two minutes for one cycle, followed by 30 cycles of denaturation at 95^0C for 30 seconds, annealing 43^0C for 30 seconds and elongation at 72^0C for one minute and final elongation at 72^0C for five minutes for one cycle and lastly samples were held at 4^0C indefinitely.

2.4.3.3. E6 multiplex hemi-nested type specific PCR targeting the E6 gene

Primers that target the E6 gene of the HPV genome were designed based on an alignment of DNA sequences retrieved from GenBank for HPV types -6, -11, -16, -18, -31, -33, -45 and -58. HPV types were selected based on outcomes from previous publications (Togawa *et al*., 1994; Cooper *et al*., 1995; Matsha *et al*., 2002; Kreimer *et al*., 2005; Paquette *et al*., 2013). Sequence data for a total of 15 isolates for each HPV type were retrieved from GenBank (source of isolates shown in Appendix B). The exception was HPV type -84 as there was only one complete sequence therefore multiple sequences could not be aligned. Primers for HPV type -84 were designed after the screening of HPV DNA using LA and were included in the low risk reaction. DNA sequences for each HPV type were aligned using Clustal Omega version 1.2.1. To confirm the specificity of each primer, the oligonucleotide sequences were analysed using Basic Local Alignment Search Tool (BLAST) analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi) from NCBI. The aligned data files for each HPV type indicating the position of the primers are provided in Appendix C. Primers designed for the first round and second round amplifications are shown on Tables 2.5 and 2.6 respectively.

Table 2.5. Primer pairs designed for HPV types based on the E6 gene used in the first round of the E6 multiplex hemi-nested type specific PCR

Primer	Nucleotide sequence (5' to 3' direction)	*T_m	*%GC	Amplicon size (bp)
HPV 6F	5'CCTCCACGTCTGCAACGACCA3'	68.9^{0}C	61.9	174bp
HPV 6R	5'AGGCTGCATATGGATAGCCGG3'	66.8^{0}C	57.1	
HPV 11F	5'ATGGAAAGTAAAGATGCCTCCACGT3'	67.0^{0}C	44.0	200bp
HPV 11R	5'CAACAGGCACACGCTGCAAG3'	67.1^{0}C	60.0	
HPV 16F	5'AGGACCCACAGGAGCGAC3'	66.3^{0}C	66.7	147bp
HPV 16R	5'TGCATAAATCCCGAAAAGCAAAGTC 3'	65.3^{0}C	40.0	
HPV 18F	5'ATGGCGCGCTTTGAGGATCC3'	68.0^{0}C	60.0	191bp
HPV 18R	5'GCAGCATGCGGTATACTGTCT3'	65.2^{0}C	52.4	
HPV 31F	5'CGGCATTGGAAATACCCTACGA 3'	65.1^{0}C	50.0	141bp

HPV 31R	5'GCACACACTCCGTGTGGTGTG3'	68.0^{0}C	61.9	
HPV 33F	5' GAGAGGGAAATCCATTTGGAATATG3'	62.6^{0}C	40.0	178bp
HPV 33R	5'TCTTGAGGACACAAAGGTCTTTG3'	63.7^{0}C	43.5	
HPV 45F	5'GGCGCGCTTTGACGATCCAAAG3'	68.7^{0}C	59.1	136bp
HPV 45R	5'TTGATATACCTCTGTGCGTTCC3'	62.9^{0}C	45.5	
HPV 58F	5'ATGTTCCAGGCACAGAGGAGAAAC3'	67.1^{0}C	50.0	195bp
HPV 58R	5' CACTTTACATACTGCAAATGGATTC3'	61.8^{0}C	36.0	
HPV 84F	5'CAACGGACGCTACCGCCCCACC3'	72.8^{0}C	72.7	192bp
HPV 84R	5'CATTTCATGCACACCCCATGAGGG3'	68.1^{0}C	54.2	

* depicts the annealing temperature and GC content calculated using OligoAnalyzer 3.1.

*T_m and %GC content calculated based on the oligoanalyzer 3.1 (https://www.idtdna.com/calc/analyzer).

Table 2.6. Forward primer designed for second round of the E6 multiplex hemi-nested type specific PCR

Primer	**Nucleotide sequence (5' to 3' direction)**	***T_m**	***%GC**	**Amplicon size (bp)**
HPV 06F2	5'GCAAGAATGCACTGACCACTGCAG3'	68.4^{0}C	54.2	90bp
HPV 11F2	5'CTTTGCACACTCTGCAAATTCAG3'	63.8^{0}C	43.5	133bp
HPV 16F2	5'CCACAGTTATGCACAGAGCTGCAA3'	67.8^{0}C	50.0	117bp
HPV 18F2	5'GTGCACGGAACTGAACACTTCACT3'	67.6^{0}C	50.0	141bp
HPV 31F2	5'CTGCAAAGGTCAGTTAACAGAAAC3'	63.2^{0}C	41.7	96bp
HPV 33F2	5'CTGTGTTTGCGGTTTTTATCTAAAC3'	62.5^{0}C	36.0	149bP
HPV 45F2	5'CCCTACAAGCTACCAGATTTG3'	61.2^{0}C	59.1	107bp
HPV 58F2	5'GTCAGGCGTTGGAGACATCTGTGC3'	69.1^{0}C	58.3	149bp
HPV 84F2	5'CGAGGTGGAGTTCGACCTACG3'	66.0^{0}C	61.1	142bp

*T_m and %GC content calculated based on the OligoAnalyzer 3.1 (https://www.idtdna.com/calc/analyzer).

For the E6 multiplex hemi-nested type specific PCR, two multiplex PCR reactions were performed to accommodate differences in the T_m of the primer pairs. One multiplex PCR reaction included primers specific for LR-HPV types -6, -11 and -84. HPV DNA type -6, which was isolated in an unrelated study from a patient with RRP, was used as a positive control. The second multiplex PCR reaction included type specific primers for HR-HPV types -16, -18, -31, -33, -45 and -58. A negative control was included in each run and a partial gene of HPV types -33 and -58 which were synthesised and supplied in pUC 57 plasmid by GenScript (USA Inc., New Jersey, USA) were used as positive controls.

PCR components for the first round of the multiplex hemi-nested PCR for both the high risk and low risk reactions are shown in Tables 2.7 and 2.8 respectively. PCR was performed using the following cycling conditions: initial denaturation, 95^0C for two minutes for one cycle, followed by 30 cycles of denaturation at 95^0C for 30 seconds, annealing 59^0C for 30 seconds and elongation at 72^0C for one minute and final elongation at 72^0C for five minutes for one cycle and lastly samples were held at 4^0C indefinitely.

Table 2.7. PCR components for the first round E6 multiplex hemi-nested type specific PCR using HR-HPV primers

PCR component	Volume	Final concentration
5x Green Go Taq® flexi buffer	10 µl	1X
$MgCl_2$ solution, 25mM	4µl	2mM
PCR nucleotide mix, 10mM each	1µl	0.2mM each dNTP
HPV 16F primer (20pmol/µl)	1µl	0.4µM
HPV 16R primer (20pmol/µ)	1µl	0.4µM
HPV 18F primer (20pmol/µl)	1µl	0.4µM
HPV 18R primer (20pmol/µl)	1µl	0.4µM
HPV 31F primer (20pmol/µl)	1µl	0.4µM
HPV 31 R primer (20pmol/µl)	1µl	0.4µM
HPV 33 F primer (20pmol/µl)	1µl	0.4µM
HPV 33 R primer (20pmol/µl)	1µl	0.4µM
HPV 45 F primer (20pmol/µl)	1µ	0.4µM
HPV 45 R primer (20pmol/µl)	1µl	0.4µM

HPV 58 F primer (20pmol/µl)	1µl	0.4µM
HPV 58 R primer (20pmol/µl)	1µl	0.4µM
Go Taq®G2 Hot Start Polymerase (5U/µl)	0.25µl	1.25U
Template	5µl	-
NFW	17.75µl	-
Total	**50µl**	**-**

Table 2.8. PCR components for the first round of the E6 multiplex hemi-nested type specific PCR using LR-HPV primers

PCR component	**Volume**	**Final concentration**
5x Green GoTaq®flexi buffer	10 µl	1X
$MgCl_2$ solution, 25mM	4µl	2mM
PCR nucleotide mix, 10mM each	1µl	0.2Mm each dNTP
HPV 6F primer (20pmol/µl)	1µl	0.4µM
HPV 6R primer (20pmol/µl)	1µl	0.4µM
HPV 11F primer (20pmol/µl)	1µl	0.4µM
HPV 11R primer (20pmol/µl)	1µl	0.4µM
HPV 84F primer (20pmol/µl)	1µl	0.4µM
HPV 84R primer (20pmol/µl)	1µl	0.4µM
Go Taq®G2 Hot Start Polymerase (5U/µl)	0.25µl	1.25U
Template	5µl	-
NFW	23.75µl	-
Total	**50µl**	-

Samples that tested negative after the first round were further tested (hemi-nested). PCR components for both reactions are shown in Table 2.9. A 1µl aliquot of PCR amplicon from the first round was used as template in the second round of the multiplex hemi-nested PCR reaction.

PCR was done with the cycling conditions as follows: initial denaturation, 95^0C for two minutes for one cycle, followed by 30 cycles of denaturation at 95^0C for 30 seconds, annealing 60.5^0C (the melting temperature was modified to accommodate the T_m of the primers used in the second round) for 30 seconds and elongation at 72^0C for one minute and final elongation at 72^0C for five minutes for one cycle and lastly samples were held at 4^0C indefinitely.

Table 2.9. PCR components for second round of the E6 multiplex hemi-nested type specific PCR. Table shows primers for both the high risk and LR-HPV reactions (reactions were performed separately as shown by asterix)

PCR Components	Volume	Final concentration
5X Green Go Taq®flexi buffer	10µl	1x
$MgCl_2$ solution, 25mM	4µl	2mM
PCR nucleotide mix, 10mM each	1µl	0.2mM each dNTP
*HPV 06F2 primer (20pmol/µl)	1µl	0.4µM
*HPV 06R primer (20pmol/µl)	1µl	0.4µM
*HPV 11F2 primer (20pmol/µl)	1µl	0.4µM
*HPV 11R primer (20pmol/µl)	1µl	0.4µM
HPV 16F2 primer (20pmol/µl)	1µl	0.4µM
HPV 16R primer (20pmol/µ)	1µl	0.4µM
HPV 18F2 primer (20pmol/µl)	1µl	0.4µM
HPV 18R primer (20pmol/µl)	1µl	0.4µM
HPV 31F2 primer (20pmol/µl)	1µl	0.4µM
HPV 31R primer (20pmol/µl)	1µl	0.4µM
HPV 33F2 primer (20pmol/µl)	1µl	0.4µM
HPV 33R primer (20pmol/µl)	1µl	0.4µM
HPV 45F2 primer (20pmol/µl)	1µl	0.4µM
HPV 45R primer (20pmol/µl)	1µl	0.4µM
HPV 58F2 primer (20pmol/µl)	1µl	0.4µM
HPV 58R primer (20pmol/µl)	1µl	0.4µM
*HPV 84F primer (20pmol/µl)	1µl	0.4µM

*HPV 84R primer (20pmol/µl)	1µl	0.4µM
Go Taq®G2 Hot Start Polymerase (5U/µl)	0.25µl	1.25U
Template (First round PCR amplicon)	1µl	-
NFW	To a volume of 50µl	-
Total	50µl	-

*Primers used for the second round of the E6 multiplex hemi-nested type specific PCR for the low risk reaction.

2.4.3.4. Comparison between molecular assays developed in-house and the commercial assay

The two in-house molecular assays were compared with a commercially available assay from Roche, the Linear Array (LA). LA is an assay that is based on four major steps, specimen preparation, PCR amplification of target DNA using specific primers, hybridization of amplified products to oligonucleotide probes and detection of probe-bound amplification products by colorimetric determination. The LA-assay utilizes pooled, non-degenerate biotin-labelled primers (PGMY11/09) designed to amplify 37 individual genital HPV which amplify approximately 450bp of the L1 major capsid protein.

All samples that were tested using the two in-house assays were tested using the LA HPV genotyping test according to manufacturer's instructions (Roche Molecular Systems, New Jersey, USA). For the amplification step, the working master mix was prepared by adding 125µl of magnesium chloride to the vial of HPV Master Mix (MMX) which consists of Amplitaq® Gold DNA polymerase, biotinylated PGMY09/PGMY11 HPV primers, dNTPs, an additional pair of primers that target the human beta-globin gene (used as an internal control), AmpErase (uracil-N-glycosylase) enzyme, and deoxyuridine triphosphate (dUTP). A volume of 50µl of the master mix was added to each reaction tube and additional 50µl of DNA template was added to each reaction tube to a final volume of 100µl. A 50 µl aliquot of the positive and negative control template provided in the kit were added to the designated reaction tubes.

PCR was done on Proflex™ PCR system. The cycling conditions were as follows: hold at 50^0C for two minutes, initial denaturation at 95^0C for nine minutes, followed for 40 cycles of denaturation at 95^0C for 30 seconds, annealing 55^0C for one minute, elongation at 72^0

C for one minute and final elongation at 72^0C for five minutes and lastly hold at 72^0C indefinitely. The tubes were removed within four hours of the start of the final hold program and a volume of 100 µl of denaturation solution (DN) was added. The denatured PCR amplicons were stored at 4^0C until genotyping.

For genotyping of denatured PCR amplicons, the amplicon was analyzed by using the LA genotyping strip provided in the detection kit. The genotyping process was performed as follows; a 4ml volume of pre-warmed working hybridization solution was added to the 24-well tray containing the labelled strip. A 100µl denatured amplicon was added to the appropriate well containing pre-warmed working hybridization solution and the tray was placed in a 55^0C shaking water bath and hybridized for 30 minutes at a shaking speed of 60 RPM. After 30 minutes, working hybridization solution was removed and replaced with 4ml of working ambient wash buffer. The strips were rinsed in wash buffer followed by incubation in pre-warmed working stringent buffer at 55^0C for 15 minutes at a shaking speed of 60 RPM. The strips were subsequently incubated in 4ml of working conjugate at room temperature (30^0C) at a shaking speed of 60 RPM on the orbital shaker for 30 minutes. After 30 minutes, the working conjugate was removed from the wells and the strips washed twice in ambient wash buffer for 10 minutes with shaking. The strips were exposed to citrate buffer for 5 minutes. The buffer was removed and the strips reacted with substrate at room temperature for 5 minutes with shaking. The strips were washed with distilled water and dried for an hour prior to interpretation of results.

2.4.4. Performance of the Roche Linear Array assay

The results of the LA test indicated a high proportion of HPV type -84 present in the samples tested and hence it was decided to repeat some of the samples using a conventional assay and modified MY11 and MY09 primers. The concern was that either the MY11/09 primers had not detected HPV type -84 or the results were non-specific. Hence the accuracy of the LA assay results was assessed by performing a nested PCR using PGMY11/09 primers that amplify a region of 450bp within the L1 region in the first round and GP5+/6+ primers (Table 2.2) that amplify 139bp within the PGMY11/09 primers. Primers for PGMY11/09 primers are shown in Table 2.10. They target the same region as MY11/09 but are designed to increase specificity and are used in the LA assay.

Table 2.10: Modified PGMY11/09 primers used in the LA assay (Gravitt *et al*., 2000)

Primer	Nucleotide sequences (5'to 3' direction)	*T_m	%GC	Amplicon size
PGMY11-A	5'GCACAGGGACATAACAATGG3'	62.0^{0}C	50.0	450bp
PGMY11-B	5' GCGCAGGGCCACAATAATGG3'	66.8^{0}C	60.0	
PGMY11-C	5'GCACAGGGACATAATAATGG3'	59.4^{0}C	45.0	
PGMY11-D	5'GCCCAGGCGCACAACAATGG3'	69.3^{0}C	65.0	
PGMY11-E	5'GCTCAGGGTTTAAACAATGG3'	60.4^{0}C	45.0	
PGMY11-F	5'CGTCCCAAAGGAAACTGATC3'	61.6^{0}C	50.0	
PGMY11-G	5'CGACCTAAAGGAAACTGATC3'	58.9^{0}C	45.0	
PGMY11-H	5'CGTCCAAAAGGAAACTGATC3'	60.0^{0}C	45.0	
PGMY11-I	5'GCCAAGGGGAAACTGATC3'	61.1^{0}C	55.6	
PGMY11-J	5'CGTCCCAAAGGATACTGATC3'	60.6^{0}C	50.0	
PGMY11-K	5'CGTCCAAGGGGATACTGATC3'	62.2^{0}C	55.0	
PGMY11-L	5'CGACCTAAAGGGAATTGATC3'	59.0^{0}C	45.0	
PGMY11-K	5'CGTCCAAGGGGATACTGATC3'	62.2^{0}C	55.0	
PGMY11-M	5'CGACCTAGTGGAAATTGATC3'	59.0^{0}C	45.0	
PGMY11-N	5'CGACCAAGGGGATATTGATC3'	60.7^{0}C	50.0	
PGMY11-P	5'GCCCAACGGAAACTGATC3'	61.6^{0}C	55.6	
PGMY11-Q	5'CGACCCAAGGGAAACTGGTC3'	65.3^{0}C	60.0	
PGMY11-R	5'CGTCCTAAAGGAAACTGGTC3'	60.9^{0}C	50.0	
HMB01	5'GCGACCCAATGCAAATTGTT3'	63.6^{0}C	45.0	

*T_m and %GC content were calculated based on the OligoAnalyzer 3.1 (https://www.idtdna.com/calc/analyzer).

PCR reaction was performed using Go Taq® G2 Hot Start Polymerase (Promega, Madison, USA) according to the manufacturer's instructions. PCR components are shown in Table 2.11. Negative and positive controls were included in the PCR run. For the negative control, a 5µl volume of NFW was added to the reaction instead of template. HPV type - 16 that is supplied with the LA HPV genotyping kit was used as a positive control. The reactions were cycled using the following cycling conditions: initial denaturation, 95^{0}C for two minutes for one cycle, followed by 40 cycles of denaturation at 95^{0}C for 30 seconds,

annealing 55^{0}C for 30 seconds and elongation at 72^{0}C for one minute and final elongation at 72^{0}C for five minutes for one cycle. Lastly samples were held at 4^{0}C indefinitely.

Table 2.11: PCR components for PGMY11/09 PCR

PCR components	Volume	Final concentration
5X Green GoTaq®flexi buffer	10µl	1X
$MgCl_2$ solution, 25mM	4µl	2mM
PCR nucleotide, 10mM each	1µl	0.2mM each dNTP
PGMY11-A (20pmol/µl)	1µl	0.4µM
PGMY11-B (20pmol/µl)	1µl	0.4µM
PGMY11-C (20pmol/µl)	1µl	0.4µM
PGMY11-D (20pmol/µl)	1µl	0.4µM
PGMY11-E (20pmol/µl)	1µl	0.4µM
PGMY09-F (20pmol/µl)	1µl	0.4µM
PGMY09-G (20pmol/µl)	1µl	0.4µM
PGMY09-H (20pmol/µl)	1µl	0.4µM
PGMY09-I (20pmol/µl)	1µl	0.4µM
PGMY09-J (20pmol/µl)	1µl	0.4µM
PGMY09-K (20pmol/µl)	1µl	0.4µM
PGMY09-L (20pmol/µl)	1µl	0.4µM
PGMY09-M (20pmol/µl)	1µl	0.4µM
PGMY09-N (20pmol/µl)	1µl	0.4µM
PGMY09-P (20pmol/µl)	1µl	0.4µM
PGMY09-Q (20pmol/µl)	1µl	0.4µM
PGMY09-R (20pmol/µl)	1µl	0.4µM
HMB01 (20pmol/µl)	1µl	0.4µM
Go Taq®G2 Hot Start Polymerase (5U/µl)	0.25µl	1.25U
Template	5µl	-
NFW	11.75µl	-
Total	50µl	-

Negative samples by the first round were subjected to the second round using the GP5+/6+ primers. A 1µl volume of the PCR amplicon from the first round was used as a template. PCR components for the second round were the same as the first round PCR except that PGMY11/09 primers were replaced with GP5+/6+ primers and 1µl of template was used instead of 5µl volume of the template. The reactions were cycled using the following cycling conditions: initial denaturation, 95^0C for two minutes for one cycle, followed by 30 cycles of denaturation at 95^0C for 30 seconds, annealing 43^0C for 30 seconds and elongation at 72^0C for one minute and final elongation at 72^0C for five minutes for one cycle. Lastly samples were held at 4^0C indefinitely.

2.4.5. Agarose gel electrophoresis

Electrophoresis of 5µl aliquots of PCR amplicons was done using a 1% or 2.5% Seakem® LE agarose gel (Lonza, Maine, USA), depending on the predicted size of the product, prepared in Tris-acetate-EDTA (TAE) buffer (pH 8.0) containing ethidium bromide at a final concentration of 0.5µg/ml (Sigma, Missouri, USA) (Sambrook and Russel, 2001). Larger DNA fragments were analysed by electrophoresis using a 1% agarose gel and smaller DNA fragments were separated using a 2.5% agarose gel. O'GeneRuler™ 100bp DNA ladder SM 1173 (Fermentas, Illinois, USA) comprising of DNA fragments from 100 to 10 000bp and Lonza SimplyLoad™ 20bp DNA ladder comprising of DNA fragments from 20bp-500bp (Lonza, Rocklands, USA) were used to estimate the size of PCR amplicons. Gel electrophoresis was performed using BioRad PowerPac Basic Systems (BioRad, California, USA) at 100V, 400A for 35 minutes for a 1% agarose gel and 100V, 400A for 50 minutes for a 2.5% agarose gel. PCR amplicons were visualised under a UV transilluminator (UVItec, Cambridge, UK).

2.4.6. Purification of PCR product

Wizard ®SV Gel and PCR Clean-Up System was used for purification of PCR amplicons according to manufacturer's instructions (Promega, Wisconsin, USA). This system is based on the ability of the DNA to bind to silica membranes in the presence of chaotropic salts and removes excess nucleotides, primers and enzymes. Following electrophoresis, the band of interest was excised from the 1% or 2.5% agarose gel and transferred to a tube. An equal amount of membrane binding solution (supplied in the kit) was added to the tube containing the excised gel slice and the mixture was incubated at 56^0C until the

gel was completely dissolved. After incubation, the melted gel and solution were transferred to a minicolumn and centrifuged at 14 000xg for one minute. Isolated PCR amplicon was washed using membrane wash solution, this was done twice to get rid of excess nucleotides. The DNA was eluted in a 30µl NFW, centrifuged at 14 000xg for one minute and stored at -20^{0}C for downstream application.

2.4.7. DNA concentration

DNA concentration was determined using a NANODROP 2000 spectrophotometer (Thermo Scientific, lllinois, USA). The purity was determined from the 260nm:280nm ratio of absorbance.

2.4.8. Sequencing

Determination of nucleotide sequence of the amplicon was performed using the Big Dye Terminator sequencing ready reaction kit according to manufacturer's instructions (Applied Biosystems, Foster City, CA, USA). Sequencing reaction components are shown in Tables 2.12 and 2.13 respectively. Reactions were cycled using the following cycling conditions: initial denaturation, 96^{0}C for one minute for one cycle, followed by 25 cycles for denaturation, 96^{0}C for 10 seconds, annealing, 50^{0}C for five seconds and extension, 60^{0}C for four minutes. Lastly samples were held at 4^{0}C indefinitely.

Table 2.12. Sequencing reaction components

Components	**Volume**
Ready reaction	1µl
Sequencing primer (0.8pmol/µl)	4µl
Dilution buffer	2µl
Template DNA	1-3ng
NFW	to a total of 10µl

Table 2.13. Control sequencing reaction

Components	**Volume**
Ready reaction	1µl

Control sequencing primer(0.8pmol/µl) (M13)	4µl
Dilution buffer	2µl
Control sequencing plasmid (pGem-3z(f)t)	1µl
NFW	2µl
Total control reaction	**10µl**

For post reaction clean-up, EDTA/ethanol precipitation was used. A solution of 0.5M EDTA (pH 8.0) was diluted to 125mM with NFW. A 5µl aliquot of 125mM EDTA and 60µl absolute ethanol were added to a 1.5ml microcentrifuge tube. The sequence reaction volume was adjusted to 20µl reaction by adding 10µl NFW. The diluted sequencing reaction was added to the tube containing 125mM EDTA and 60µl absolute ethanol. The tube was vortexed for five seconds. Precipitation was allowed by leaving the reaction at room temperature for 15 minutes. The samples were centrifuged at 14 000xg for 10 minutes at 4^0C. The supernatant was completely aspirated without disturbing the pellet. A volume of 500µl of 70% ethanol was added to each tube. The reaction tubes were centrifuged at 14 000xg for 10 minutes at 4^0C. The supernatant was completely aspirated without disturbing the pellet. The reaction tubes were incubated at 37^0C for two hours until completely dry. Lastly the samples were stored at 4^0C in the dark until submission for electrophoresis at the Department of Microbial, Biochemical and Food Biotechnology, University of the Free State, Bloemfontein SA.

2.4.9. Genotyping and aligning

Nucleotide sequence data from positive PCR amplicons were edited using Chromas Pro version 1.6 and aligned using Clustal Omega version 1.2.1. The HPV genotypes were determined by comparison with nucleotide sequence data retrieved from GenBank and by BLAST analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.5. Results

2.5.1. Patient data

A total of 74 patients with confirmed head and neck tumours were included in the study. All samples received a laboratory number preceded by VBD on submission. The cohort was comprised of 64 males and 10 females. The age of patients ranged from one year to 85 years with a mean age of 59.8 years. Cancerous lesions were confirmed by histology in the Department of Anatomical Pathology, University of the Free State, and defined as follows; 45 moderately differentiated SCC, two carcinoma *in situ*, 12 well differentiated SCC, one diffuse large B-cell lymphoma, one anaplastic large cell lymphoma, one Burkitt's lymphoma, nine poorly differentiated SCC, two inverted papillomas and one neuroendocrine carcinoma. The location of the carcinoma varied as follows; 15 located in the oropharynx, 44 located in the larynx, one was identified from the maxillary sinus; two from tonsils, one located in the nasal cavity, five from the hypopharynx, one from the nasopharynx, one from the parapharyngeal space, three from the parapharyngeal sinuses and one located in the tongue. Further details for each patient are provided in Table 2.14.

2.5.2. Optimization of PCR

Nested PCR and E6 multiplex hemi-nested type specific PCR were optimized using positive controls and after testing a tenfold dilutions of the positive controls, each control was subsequently used at a 1:1000 dilution.

2.5.3. Nested PCR targeting the L1 region

All tissue biopsies (n=74) tested positive for the beta-globin gene (Appendix E). A total of 4/74 samples were positive for HPV DNA by nested PCR targeting the L1 capsid gene, namely VBD 13/14, VBD 47/14, VBD 10/10 and VBD 17/15.

In the first round of the nested PCR using MY11/09 primers which amplify a 450bp region within the L1 region, two samples tested positive (VBD 13/14 and VBD 17/15) (Figure 2.2A). Sample VBD 47/14 was weakly positive and therefore could not be genotyped. PCR amplicons were excised from the gel and purified using Wizard®SV Gel and PCR Clean-Up System prior to sequence determination. To confirm the presence of the band after DNA purification, a 5µl aliquot of the purified PCR amplicon was separated by

electrophoresis and visualised under a UV transilluminator. Lanes 4 and 5 show purified samples at 450bp (Figure 2.2B).

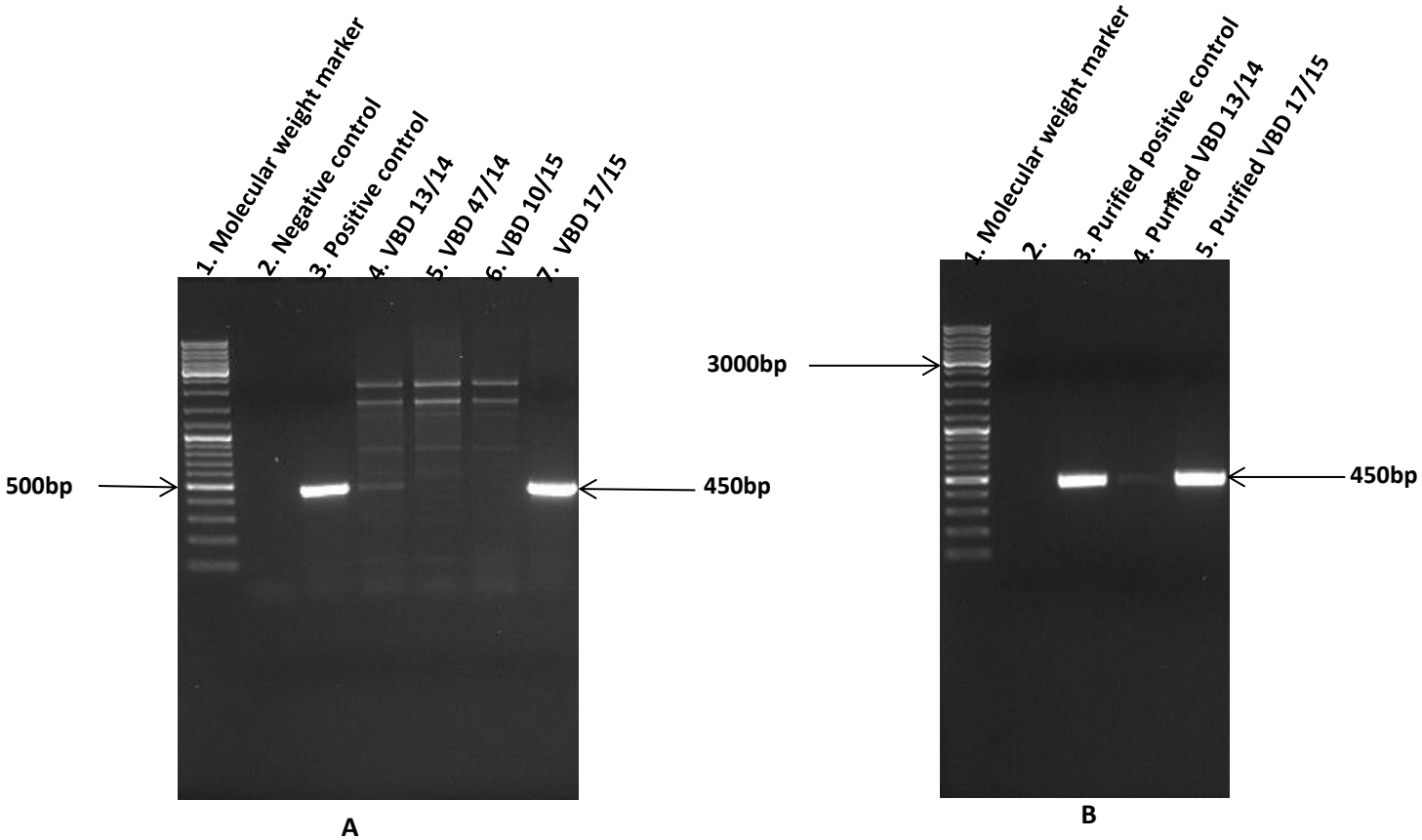


Figure 2.2. A 1% agarose gel electrophoresis analysis showing results for first round of the nested PCR under UV transilluminator before and after purification. 2.2A: Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: Negative control; Lane 3: Positive control; Lane 4: VBD 13/14; Lane 5: VBD 47/14; Lane 6: VBD 10/15 and Lane 7: VBD 17/15. **2.2B**: Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: Empty; Lane 3: Positive control; Lane 4: VBD 13/14; Lane 5: VBD 17/15.

Samples that tested negative or inconclusive by first round were further amplified using the primers GP5+/6+ that target 139bp within the MY11/09 primer region. Two samples, namely VBD 47/14 and VBD 10/15 gave bands of predicted size (Figure 2.3A).

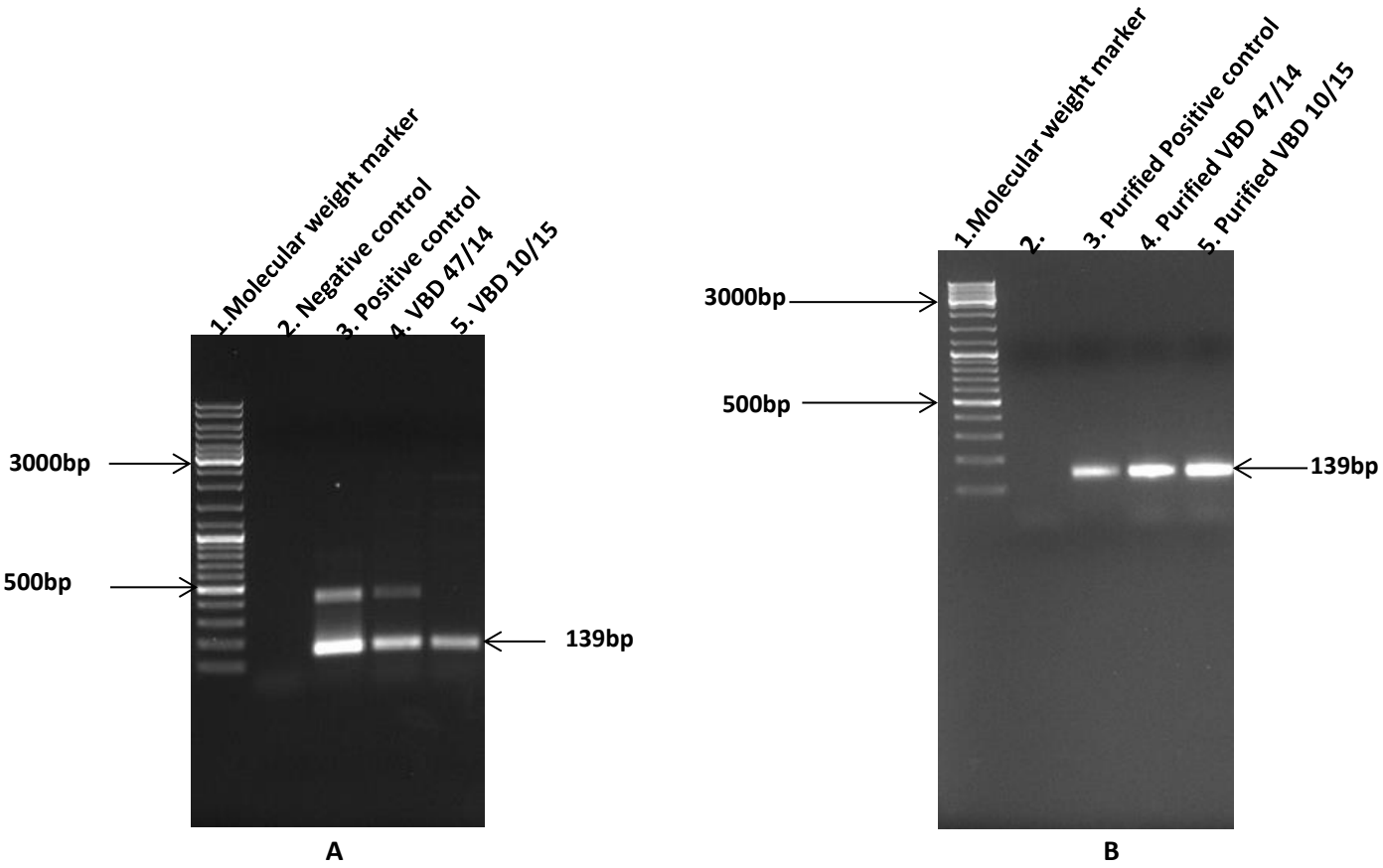


Figure 2.3. A 1% agarose gel electrophoretic images showing PCR products for second round of the nested PCR under a UV transilluminator before and after purification. **2.3A**: Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: Negative control; Lane 3: Positive control; Lane 4: VBD 47/14; Lane 5: VBD 10/15. **2.3B:** Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: Empty; Lane 3: Positive control; Lanes 4-5 shows positive samples.

As previously, positive PCR amplicons were excised from the gel and purified using Wizard® SV Gel and PCR Clean-Up System prior to sequence determination and the presence of the purified PCR amplicon after purification confirmed by electrophoresis (Figure 2.3B).

2.5.4. E6 multiplex hemi-nested type specific PCR targeting the E6 gene

Two positives control (HPV types -33 and -58) were prepared using a synthesised gene supplied in pUC 57 plasmid. DNA for use as template control was prepared by amplification of the gene using a type specific primer (HPV 33F/HPV 58F) forward primer and M13 reverse primer located on the plasmid downstream of the inserted gene in the first round and in the second round PCR, primers (HPV33F2/HPV58F2) and M13 reverse

primer were used. A volume of 5µl aliquot of the PCR amplicon was analysed by electrophoresis and visualised under a UV transilluminator (Figures 2.4A and B).

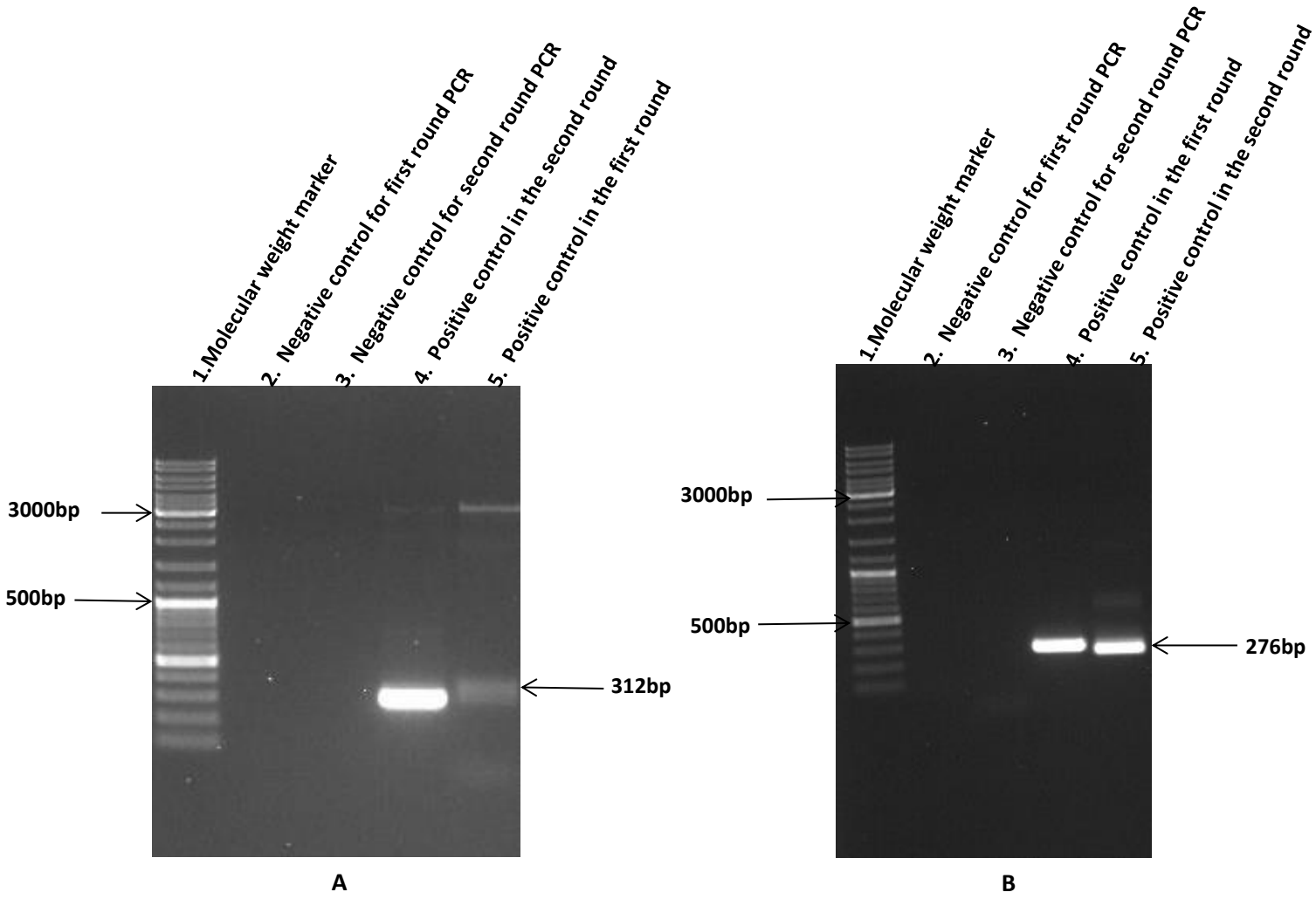
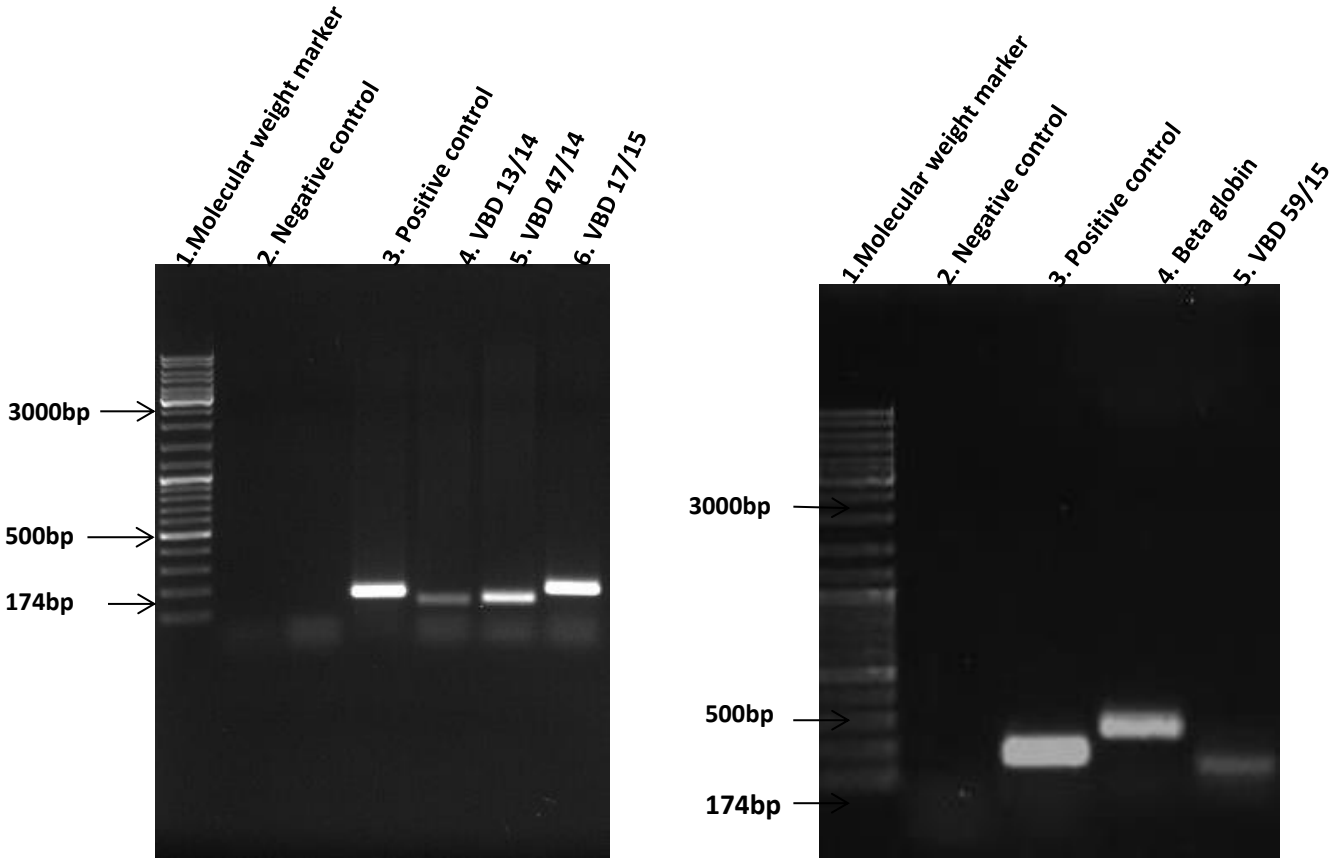


Figure 2.4. A 1% agarose gel electrophoresis analysis depicting PCR product for positive controls for E6 multiplex hemi nested PCR. 2.4A (HPV type -58): Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control for first round PCR; Lane 3: Negative control for second round PCR; Lane 4: Positive control in the second round (276bp); Lane 5: Positive control in the first round (312bp). **2.4B (HPV type -33):** Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control for first round PCR; Lane 3: Negative control for second round PCR; Lane 4: Positive control in the first round (305bp); Lane 5: Positive control in the second round (276bp)

All samples were tested using the E6 multiplex hemi-nested type specific PCR. A total of 5/74 samples were positive for HPV DNA. In the first round, types specific primers were used which amplify the E6 region with expected band sizes of interest ranging from 141bp-200bp. Products were visualised as previously after electrophoresis. Figures 2.5A and 2.5B shows results for the first round PCR for the high risk reaction.



A B

Figure 2.5. A 1% agarose gel electrophoresis analysis depicting PCR products for the first round of the E6 multiplex hemi-nested type specific PCR for HR-HPV types. **2.5A**: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lane 3: Positive control-174bp; Lane 4: VBD 13/14-141bp; Lane 5: VBD 47/14-147bp and lane 6: VBD 17/15-191bp. **2.5B**: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lane 3: Positive control-174bp; Lane 4: Beta globin; Lane 5: VBD 59/15-136bp.

As previously, positive PCR amplicons were excised from the gel and purified prior to sequence determination (Figures 2.6A and 2.6B).

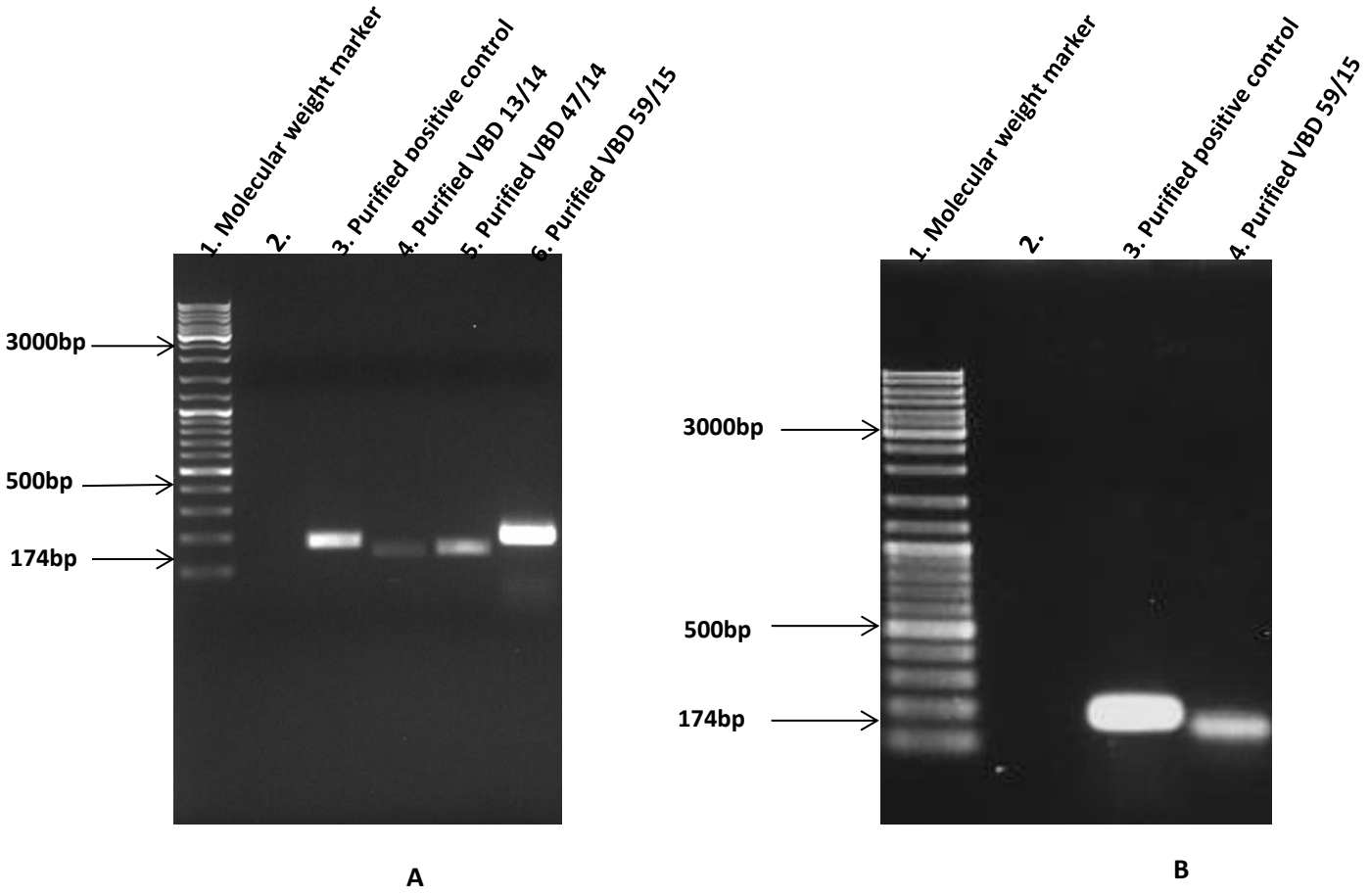


Figure 2.6. A 1% agarose gel electrophoresis analysis depicting purified PCR products for the first round of the E6 multiplex hemi-nested type specific PCR for HR-HPV types visualized under a UV transilluminator. **2.6A**: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Empty; Lane 3: Purified positive control-174bp; Lane 4: VBD 13/14-141bp; Lane 5: VBD 47/14-147bp; Lane 6: VBD 17/15-191bp. **2.6B**: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Empty; Lane 3: Positive control-174bp; Lane 4: VBD 59/15-136bp.

Negative samples were further tested using the hemi nested PCR which amplify regions ranging from 90bp-149bp. Positive reactions and purified amplicons are shown in Figures 2.7A-2.7B.

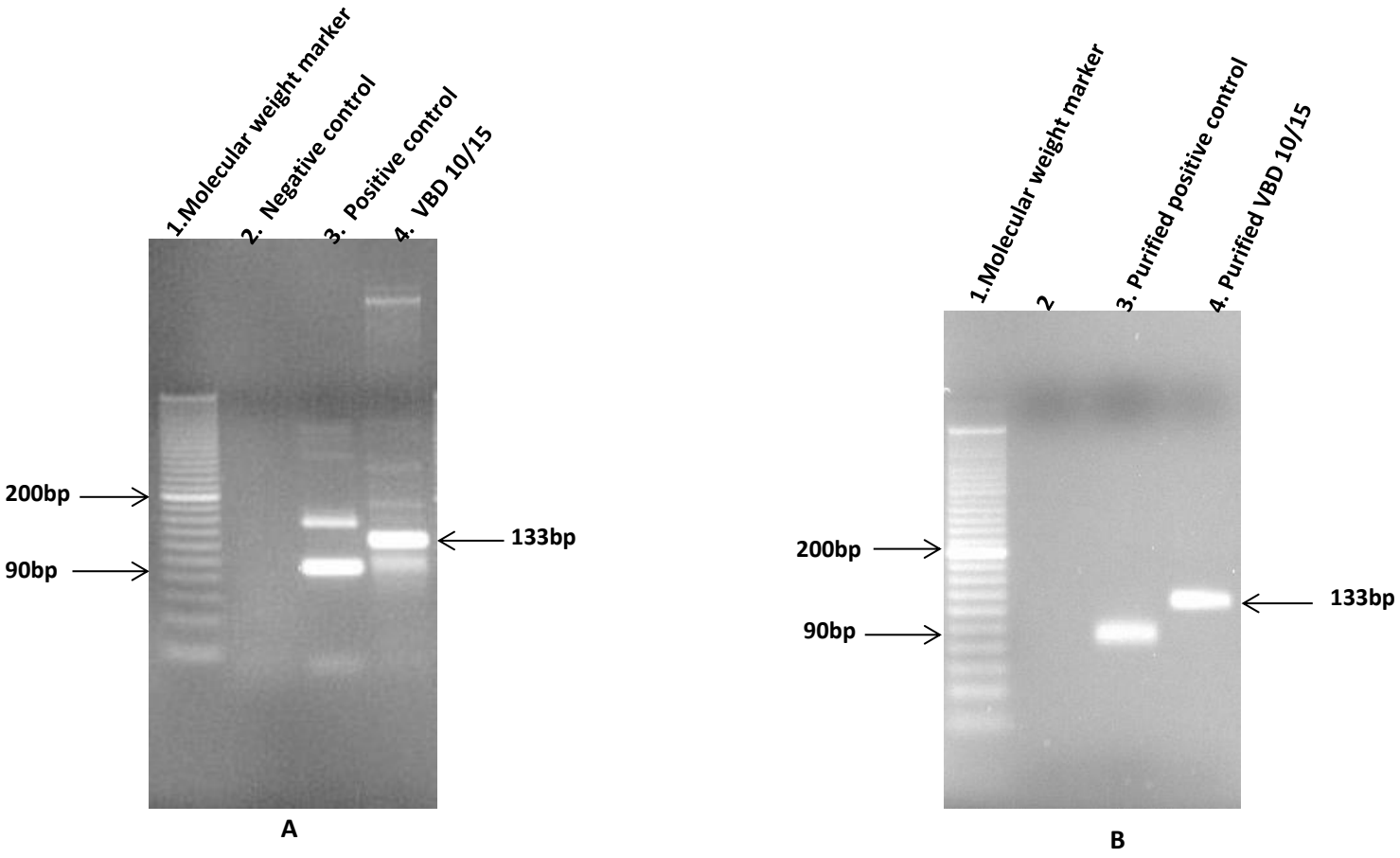


Figure 2.7. A 2.5% agarose gel electrophoresis analysis of the second round E6 multiplex hemi-nested type specific PCR for low risk PCR before and after purification. **2.7A:** Lane 1: Lonza SimplyLoad™ 20bp DNA ladder; Lane 2: negative control; Lane 3: Positive control; Lane 4: VBD 10/15. **2.7B:** Lane 1: Lonza SimplyLoad™ 20bp DNA ladder; Lane 2: Empty; Lane 3: Positive control; Lane 4: Purified VBD 10/15.

2.5.5. Linear array

A total of 74 samples were tested for HPV DNA using the LA and 74/74 samples tested positive for the beta-globin gene. A total of 57/74 samples tested positive for HPV type -84. In addition VBD 47/14 was positive for HPV types -16/-84 and VBD 17/15 was positive for HPV types -18/-84. Detailed results for LA are shown in Table 2.14.

2.5.6. Performance of the assay

An unexpected high number of samples were positive for HPV type -84, 10/74 samples were randomly selected and were tested using nested PCR (PGMY11/09 and GP5+/6+ primers) and they tested negative for HPV type -84. One sample, VBD 59/15 that tested

positive for HPV type -45 by the LA was also tested positive by the nested PCR. Figures 2.8A and 2.8B shows first round PCR amplicon for HPV type -45.

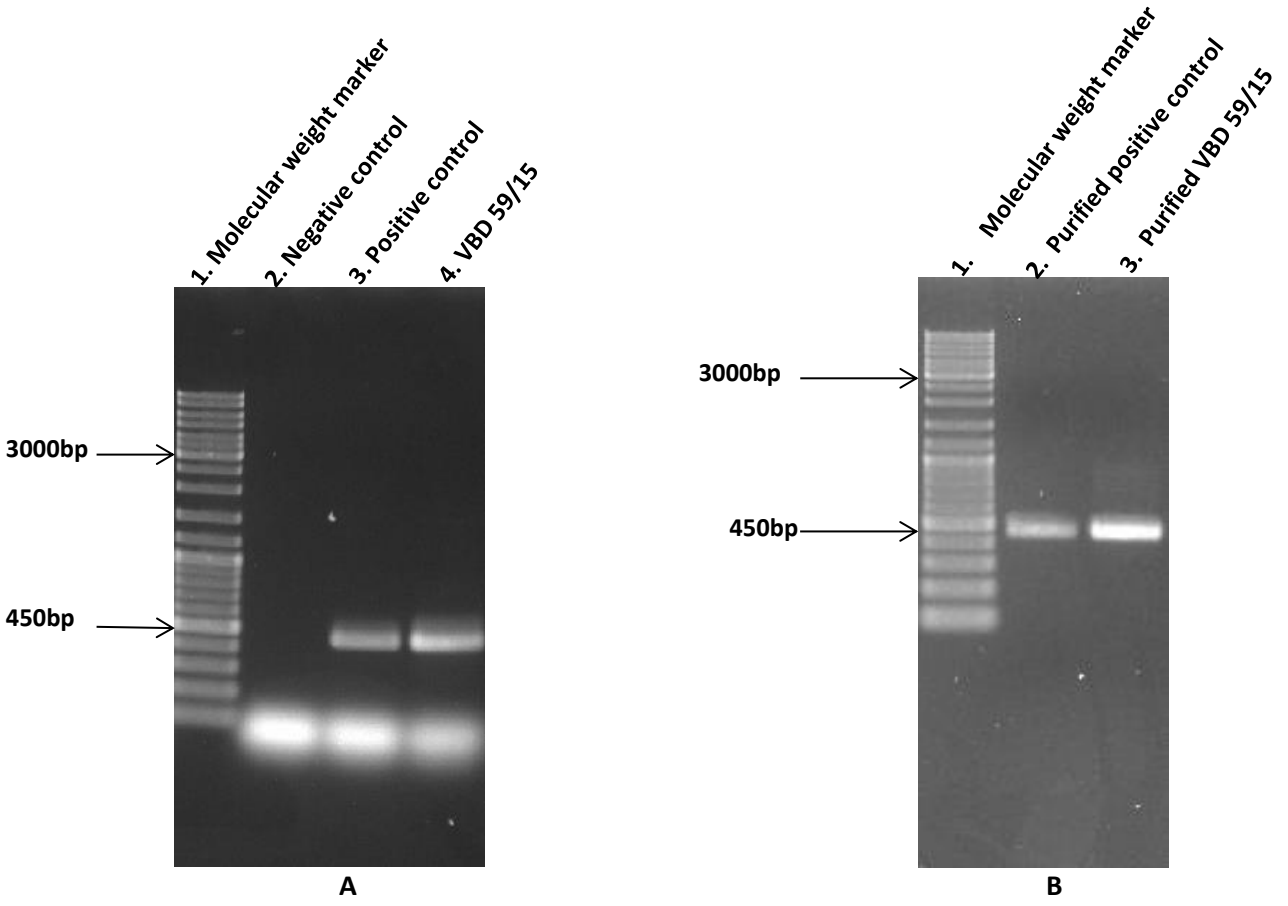


Figure 2.8. A 1% agarose gel electrophoresis analysis depicting results for sample VBD 59/15 before and after purification visualized under a UV transilluminator. **2.8A**: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lane 3: Positive control; Lane 4: VBD 59/15. **2.8B**: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Positive control; Lane 3: Purified VBD 59/15.

2.5.7. Genotyping of HPV types

To identify HPV types, the nucleotide sequence was determined for all amplicons that were within the predicted sizes. Table 2.14 shows the results for the four PCR methods performed. All samples were initially tested by the E6 multiplex hemi-nested PCR, but HPV type -84 was not included in the low risk reaction. Once HPV type -84 was genotyped by LA, it was included in the low risk reaction and ten samples were randomly

selected tested. Samples highlighted in yellow in the table below under the LA assay are the 10 samples that were selected and retested using the E6 multiplex hemi-nested type specific PCR and nested PCR (PGMY11/09 and GP5+/6+) after they tested positive by the LA. Partial nucleotide sequences for HPV types genotyped using the nested PCR (MY11/09 and GP5+/6+ primers); nested PCR (PGMY11/09 and GP5+/6+ primers) and the E6 multiplex hemi-nested type specific PCR are shown in Appendix F.

Table 2.14. Patient information and genotyping results for nested PCR, E6 multiplex hemi-nested PCR, Roche Linear array and PGMY11/09 PCR. HPV types -84 indicated in bold are the ten samples selected for retesting using nested PCR (PGMY11/09 and GP5+/6+) and E6 multiplex hemi-nested type specific PCR.

VBD no.	Sex	Age	Anatomical site	Histology	Beta-globin	Nested PCR (MY11/09 and GP5+/6+)	E6 multiplex hemi-nested PCR	Roche Linear array	Nested PCR (PGMY11/09 AND GP5+/6+)
06/14	M	53	Oropharynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A
11/14	M	60	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A
12/14	M	60	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A
13/14	M	76	Larynx	Carcinoma-*in situ*	Positive	31	31	84	NT
24/14	M	60	Oropharynx	Moderately-differentiated SCC	Positive	Negative	Negative	**84**	Negative
30/14	M	59	Oropharynx	Moderately differentiated SCC	Positive	Negative	Negative	Negative	N/A
31/14	M	69	Larynx	Well-differentiated SCC	Positive	Negative	Negative	**84**	Negative
32/14	M	70	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	**84**	Negative
33/14	M	69	Larynx	Carcinoma-*in situ*	Positive	Negative	Negative	Negative	N/A
39/14	M	62	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	**84**	Negative
40/14	M	60	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A
43/14	F	69	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	**84**	Negative
44/14	M	81	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	**84**	Negative
45/14	M	01	Larynx	Well-differentiated SCC	Positive	Negative	Negative	**84**	Negative
46/14	M	44	Maxillary sinus	Burkitt's lymphoma	Positive	Negative	Negative	**84**	Negative
47/14	M	54	Tonsil	Moderately-differentiated SCC	Positive	16	16	16/84	NT
48/14	F	43	Nose	Anaplastic large cell lymphoma	Positive	Negative	Negative	**84**	Negative
49/14	F	48	Tonsil	Diffuse large B-cell lymphoma	Positive	Negative	Negative	**84**	Negative
50/14	F	59	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
52/14	F	64	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
53/14	F	69	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
54/14	M	65	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A

55/14	M	71	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
56/14	F	59	Oropharynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
57/14	M	52	Larynx	Well-differentiated SCC	Positive	Negative	Negative	84	NT
59/14	M	50	Tongue	Poorly-differentiated SCC	Positive	Negative	Negative	84	NT
01/16	M	63	Larynx	Well-differentiated SCC	Positive	Negative	Negative	Negative	N/A
04/15	M	70	Hypopharynx	Poorly-differentiated SCC	Positive	Negative	Negative	Negative	N/A
09/15	M	55	Hypopharynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A
10/15	F	56	Larynx	Poorly-differentiated SCC	Positive	11	11	Negative	N/A
12/15	M	20	Oropharynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
13/15	M	64	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
17/15	M	36	Nasopharynx	Poorly-differentiated SCC	Positive	18	18	18/84	NT
18/15	M	58	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
19/15	M	55	Larynx	Well-differentiated SCC	Positive	Negative	Negative	84	NT
21/15	M	58	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
22/15	M	79	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A
23/15	M	79	Parapharyngeal space	Poorly-differentiated SCC	Positive	Negative	Negative	84	NT
24/15	M	28	Paranasal sinuses	Neuroendocrine carcinoma	Positive	Negative	Negative	84	NT
25/15	M	74	Larynx	Poorly-differentiated SCC	Positive	Negative	Negative	84	NT
26/15	M	52	Larynx	Well-differentiated SCC	Positive	Negative	Negative	84	NT
27/15	M	76	Oropharynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
28/15	M	63	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
30/15	M	62	Oropharynx	Well-differentiated SCC	Positive	Negative	Negative	84	NT
31/15	M	59	Oropharynx	Well-differentiated SCC	Positive	Negative	Negative	84	NT
32/15	M	50	Larynx	Well-differentiated SCC	Positive	Negative	Negative	84	NT
33/15	M	62	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
35/15	M	50	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
36/15	M	50	Hypopharynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
37/15	M	63	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
38/15	M	68	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
39/15	M	62	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
40/15	M	60	Hypopharynx	Well-differentiated SCC	Positive	Negative	Negative	84	NT

45/15	M	65	Larynx	Poorly-differentiated SCC	Positive	Negative	Negative	84	NT
46/15	F	66	Paranasal sinuses	Inverted papilloma	Positive	Negative	Negative	84	NT
56/15	M	55	Oropharynx	Well-differentiated SCC	Positive	Negative	Negative	84	NT
57/15	M	60	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
58/15	M	71	Oropharynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A
59/15	F	71	Larynx	Moderately-differentiated SCC	Positive	Negative	45	45	45
61/15	M	85	Oropharynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
63/15	M	53	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A
64/15	M	57	Oropharynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
65/15	M	55	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
66/15	M	65	Larynx	Poorly-differentiated SCC	Positive	Negative	Negative	84	NT
67/15	M	33	Paranasal sinuses	Inverted papilloma	Positive	Negative	Negative	84	NT
71/15	M	60	Hypopharynx	Well-differentiated SCC	Positive	Negative	Negative	84	NT
72/15	M	55	Oropharynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
73/15	M	53	Larynx	Well-differentiated SCC	Positive	Negative	Negative	84	NT
74/15	M	59	Oropharynx	Poorly-differentiated SCC	Positive	Negative	Negative	84	NT
76/15	M	48	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
77/15	M	58	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	84	NT
01/16	M	63	Larynx	Well-differentiated SCC	Positive	Negative	Negative	Negative	N/A
02/16	M	75	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A
03/16	M	72	Larynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A
04/16	M	62	Oropharynx	Moderately-differentiated SCC	Positive	Negative	Negative	Negative	N/A

Abbreviations: M-Male; F-Female; N/A-Not applicable; SCC-Squamous cell carcinoma; NT-Not tested

2.6. Discussion

In recent years an association between HPV and HNSCC has been described (Gillison *et al*., 2000; Mork *et al*., 2001; Paquette *et al*., 2013). Although there are several molecular assays that have been described to investigate HPV in head and neck biopsies, there is no standardised assay and the frequently used commercial assays are not cost effective for large studies. The aim of this chapter was to initially investigate and compare various molecular based assays for detection of HPV DNA and then use these assays to screen tissue biopsies from patients with histologically confirmed HNSCC in the Free State, SA. The assay needed to be sensitive and cost effective for future studies involving large numbers of samples. In-house assays were compared with the Roche LA.

Two molecular assays were employed to screen for HPV DNA in 74 tissue biopsies from patients with confirmed head and neck tumours. The L1 nested PCR uses two pairs of published consensus primers (MY11/09 and GP5+/6+) targeting a region of the L1 gene of HPV. This assay is currently used to screen for HPV DNA in tissue biopsies from patients with RRP in our laboratory. Due to the fact that there is a possibility that the L1 gene may be disrupted during viral integration into the host genome, resulting in false negatives results (Paquette *et al*., 2013), primers that target the E6 gene were designed (Morris, 2005). The primers were designed based on the alignment of nucleotide sequences of the E6 gene isolates for HPV types -6, -11, -16, -18, -31, -33, -45, -58 and -84. To confirm specificity, BLAST analysis was performed on each primer designed. Types that were selected for inclusion were initially based on previous reports of types that were commonly associated with head and neck tumours. A total of 4/74 (5.4%) tissue biopsies tested positive for HPV DNA using the nested PCR targeting the L1 region, two samples were positive for HPV DNA by first round PCR and two additional samples were positive by second round PCR. Sequencing analysis revealed HPV types -11, -16, -18 and -31. Using the E6 multiplex hemi-nested type specific PCR five HPV types (-6, -11, -16, -18, -31 and -45) were identified. It was anticipated that this assay may have limitations and would be extended to include other subtypes based on the outcome of the LA assay. All samples were tested for HPV DNA using the Roche LA assay and a total of 60/74 (81.1%) tested positive, 57/60 were positive for HPV type -84, one positive for HPV type -45 and two co-infections (-16 and -84; -18 and -84). Based on the LA assay results an additional

type specific primer pair was designed to amplify HPV type -84. A total of 10 from the 57 samples that tested positive for HPV type -84 using the LA were tested using type specific primers for HPV type -84 and modified primers PGMY11/09 and GP5+/6+ primers (termed nested PCR). All 10 samples tested negative.

In a previous study performed in SA, HPV was identified frequently in samples collected from the oropharynx by Paquette *et al*., in 2013 where he screened 51 samples from the oropharynx and 48/51 (94.1%) harboured HPV DNA. In this study 57.8% of the tissue biopsies were from the larynx and only 17 (23.4%) of the samples were from the oropharynx with one sample testing positive for HPV type -16. The different location of biopsies available for this study compared with the previous study may contribute to the different prevalence rates. In addition most of the previous studies conducted in SA were looking for HPV prevalence in the oesophagus and there were no samples from the oesophagus in this study. To our knowledge this is one of the few studies that screened for the presence of HPV DNA in head and neck tumours from anatomical sites of the head and neck besides the oesophagus.

In conclusion, based on the study, nested PCR (MY11/09 and GP5+/6+) that is currently used in-house to screen for HPV DNA in tissue biopsies form patients with RRP may not be suitable to screen for HPV DNA in tissue biopsies from patients with head and neck tumours as it did not detect HPV type -45 and that the modified primers designed to have less mismatches with HR-types would be more suitable. The Roche LA assay did not detect HPV types -11 and -31 whereas the E6 multiplex hemi-nested type PCR detected all the five HPV types isolated in this study. Primers used in the Roche LA were aligned with both HPV types -31 and -11 and a minimum of two mismatches were observed which does not account for the negative result. The in house E6 multiplex hemi-nested type specific PCR detected five HPV types in the study, performed better than the Roche LA assay and therefore could have application in future studies. However there are limitations that need to be kept in mind when using this assay; there are 15 HR-HPV types and future applications should expand the primer sets to include additional HR-types.

CHAPTER 3: DEVELOPMENT OF A MOLECULAR ASSAY FOR THE DETECTION OF TANSCRIPTIONALLY ACTIVE HPV IN HEAD AND NECK TUMOURS

3.1. Introduction

SCC of the head and neck has been associated with HR-HPV types from the mucosal group in the alphapapillomavirus genus (Smith *et al*., 2010; Bishop *et al*., 2012; Larque *et al*., 2014) and it was first suspected in 1983 with histo-pathological observations which were consistent with HPV infection in oral cancers (Mirghani *et al*., 2014). Nowadays it is widely accepted that HPV-related HNSCC differ substantially those HPV unrelated (Lassen, 2010). Identification of HPV-related HNSCC is important due to the fact that patients respond better to treatment compared to patients with HNSCC unrelated to HPV (Sedaghat *et al*., 2009; Ang *et al*., 2010; Lassen, 2010; Lassen *et al*., 2011; Andersen *et al*., 2014). The development of this type of carcinoma is attributed to two viral oncoproteins found within the HPV genome, the E6 and the E7 which are regulated by the E2 gene, the site of viral integration (Panwar *et al*., 2014). HNSCC develops when the E6 and E7 bind to the two proteins involved in regulating cell growth, the p53 suppressor protein and pRB protein respectively (Figure 3.1). The two proteins (p53 and pRB), under normal conditions function by regulating cell growth and as a result allowing the repair of damaged DNA and they also promote apoptosis in response to cell damage by possibly inactivating the E2F transcription factor (Panwar *et al*., 2014). The E2 gene is often the site of the integration, resulting in the disruption of the E2 gene and subsequent derepression of E6 and E7 oncogenes (Ruttkay-Nedecky *et al*., 2013). The E6 oncoprotein binds to the p53 protein thus disrupting the p53 pathway which is involved in responding to the host cell's to DNA damage and once the E6 oncoprotein bind to the p53 protein it leads to the chromosomal mutations and instability within the infected cells (Braakhuis *et al*., 2004; Panwar *et al*., 2014) and the E7 oncoprotein binds to the second protein, the pRB protein thus leading to the phosphorylation state and therefore inactivating the protein (Ruttkay-Nedecky *et al*., 2013) (Figure 3.1).

The detection of HPV E6 or E7 mRNA by PCR is considered the gold standard in identifying transcriptionally active HPV infection. The identification of transcriptionally active HPV types suggests that HPV might be the likely cause of cancer (Braakhuis *et al*., 2004; Boscolo-Rizzo *et al*., 2013; Larque *et al*., 2014).

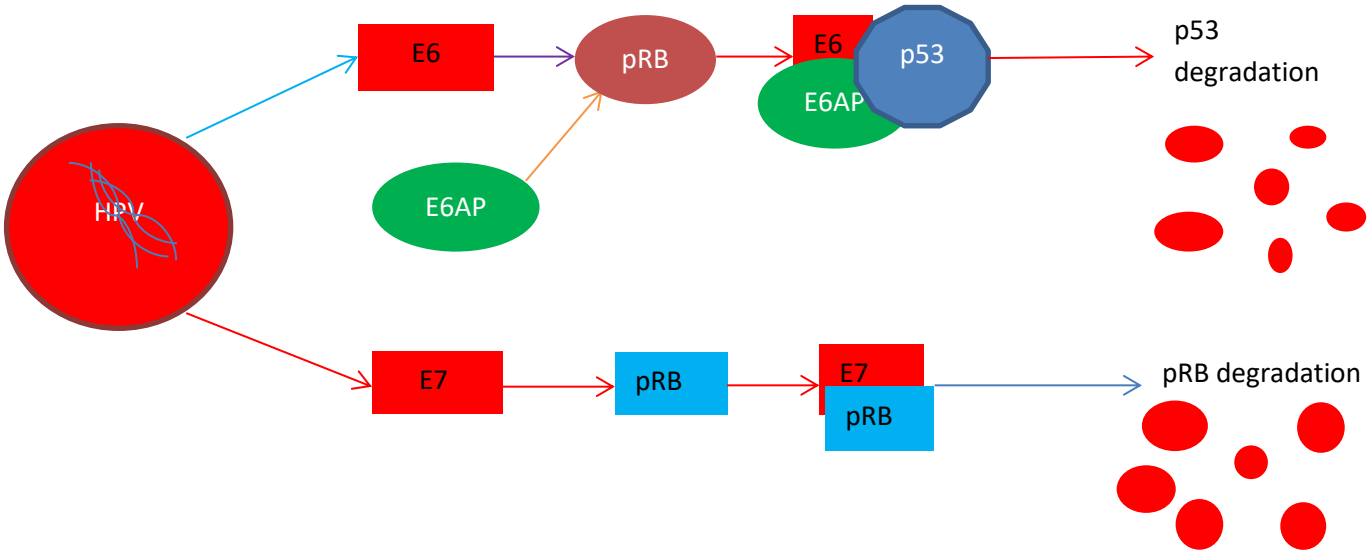


Figure 3.1. Mechanism used by the two oncoproteins, the E6 and the E7 allowing abnormal cell growth.

3.2. Aim

The aim of this chapter was to develop a hemi-nested RT-PCR (HnRT-PCR) targeting HPVE6 mRNA that can be used for the detection of transcriptionally active HPV types in patients with confirmed HNSCC in the Free State, Bloemfontein, SA**.**

3.3. Objectives

1. To prepare positive controls for optimization of HnRT-PCR.

2. To validate the assay (HnRT-PCR) by screening four samples that were positive for HR-HPV types -16, -18, -31 and -45, for replicating/transcriptionally active HPV infection.

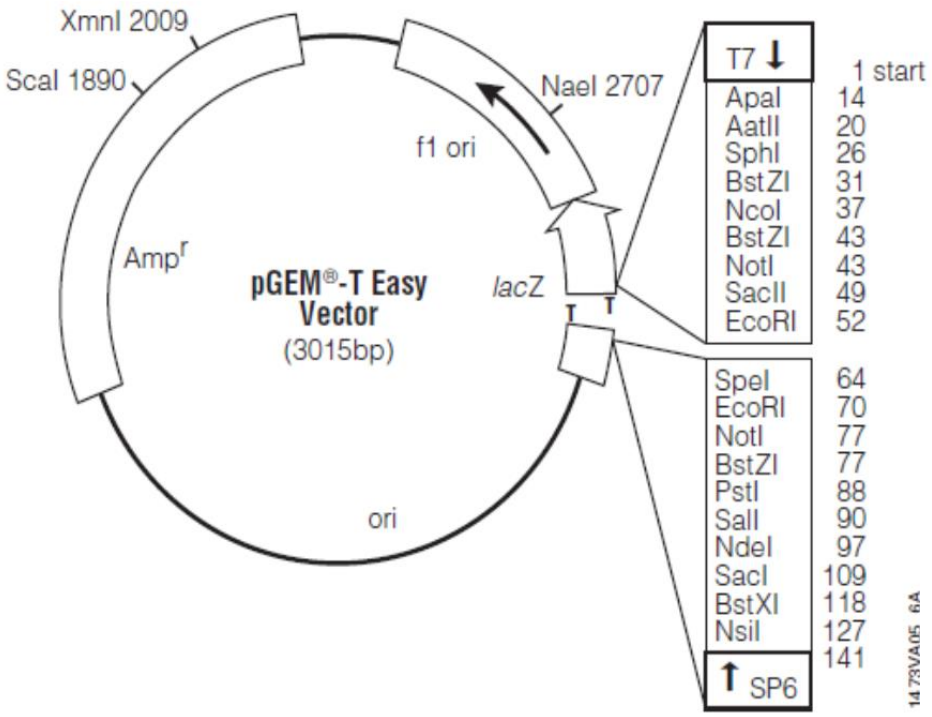
3.4. Materials and methods

3.4.1. RNA controls

Based on the results obtained from the previous chapter, an HnRT-PCR for HR-HPV types -16, -18, -31 and -45 was developed. In order to optimize the HnRT-PCR, four RNA controls were prepared that represent HPV types -16, -18, -31 and -45.

3.4.1.1. Cloning of partial E6 gene into pGEM®-T easy bacterial vector

HPV DNA was amplified from samples previously genotyped as HPV types -16, -18, -31 and -45 respectively. Positive PCR amplicons for each sample were purified using Wizard®SV Gel and PCR Clean-Up System previously described in 2.4.6. The DNA concentration and purity of PCR amplicons were determined using a NANODROP spectrophotometer. Each PCR amplicon was ligated into pGEM®-T easy vector (Promega, Madison, USA) by TA cloning using T4 DNA ligase. The pGEM®-T easy vector is a linearized vector with a single 3'-terminal thymidine at both ends; it has a high copy number and contains T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the alpha peptide coding region of the enzyme beta galactosidase. Figure 3.2. Shows the vector map and multiple cloning sites of the pGEM®-T easy vector and Appendix G Shows sequence reference points. The ligation reaction components are provided in Table 3.1.



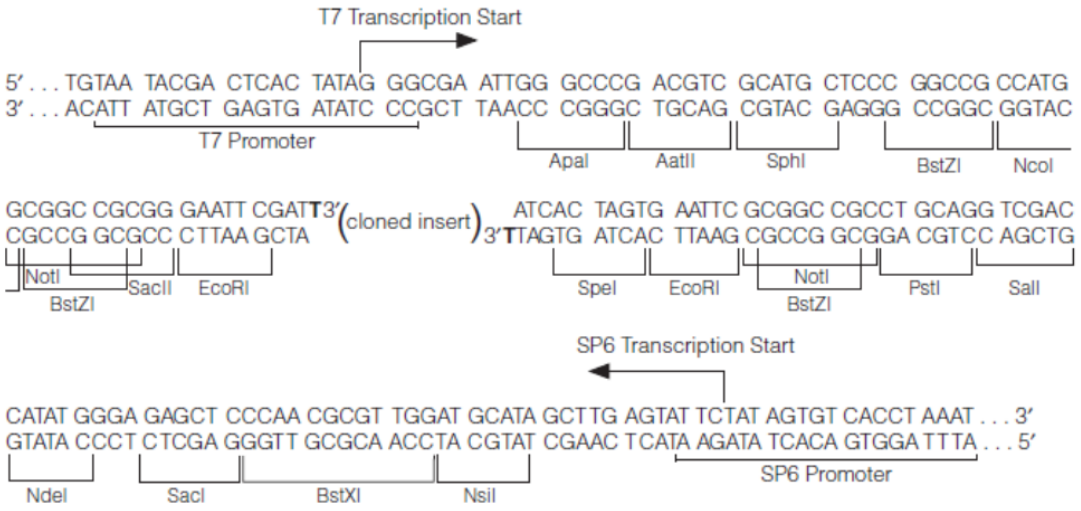


Figure 3.2. Vector map and sequence reference points of pGEM®-T easy vector (Promega, Madison, USA).

Table 3.1. Ligation reaction components

Reaction components	**Insert (HPV types -16, -18, -31 and -45).**	**Positive control**	**Background control**
2x Rapid ligation buffer, T4 DNA ligase	5µl	5µl	5µl
pGEM T easy vector (50ng)	1µl	1µl	1µl
PCR amplicon (Insert)	3µl (1-3ng)	-	-
Control insert DNA	-	2µl	-
T4 DNA ligase (3 Weiss units/µl)	1µl	1µl	1µl
NFW	-	1µl	3µl
Total	**10µl**	**10µl**	**10µl**

Ligation reactions were incubated at 4^0C overnight.

3.4.1.2. Transformation of chemically competent JM 109 cells

A volume of 50µl JM 109 competent cells with transformation efficiency greater than 10^8cfu/µg (Promega, Madison, USA) were transferred into the ligation reaction tubes (Table 3.1) and incubated on ice for 20 minutes. Cells were heat-shocked at exactly 42^0C for 50 seconds; this was done so that the cells can take up the DNA plasmid. A volume of

950µl super optimal broth with catabolite repression (SOC) media was added to the ligation reaction. The ligation/transformation reaction was incubated for 1.5 hours at 37^0C shaking at approximately 150rpm. A total volume of 100µl of each transformation culture was plated in duplicate on Luria Bertani (LB) plates containing ampicillin (amp) at a final concentration of 100µg/ml, isopropyl β-D-thionalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indoyl-β-D-galactopyranosidase (X-gal) plates. The X-Gal/IPTG plates were prepared by applying a volume of 40µl of X-gal stock solution at a final concentration of 20mg/ml and 4µl volume of a 200mg/ml of IPTG. X-gal and IPTG were spread over the entire surface of the plate and incubated at 37^0C until the fluid was no longer visible.

3.4.1.3. Confirmation of positive transformants

Blue/white colony selection was used for screening of transformants. The beta galactosidase converts the colourless substrate X-gal to produce blue colonies. The LacZ gene contains the multiple cloning and A/T cloning sites. The gene will be disrupted in positive transformants, therefore beta galactosidase will no longer be produced and X-gal can no longer be metabolised to produce blue colonies. Colonies containing positive transformants will thus be white but do require confirmation.

Three white colonies were selected for each ligation reaction from the LB/ampicillin/IPTG/X-gal plates and were designated 13/14A, 13/14B and 13/14C for HPV type -31; 47/14A, 47/14B and 47/14C for HPV type -16; 17/15A, 17/15B and 17/15C for HPV type -18, lastly 59/15A, 59/15B and 59/15C for HPV type -45. Each colony was inoculated into a 5ml of LB/amp and grown overnight at 37^0C with a shaking rotor at 200rpm. The following day, plasmid DNA was purified from each culture using PureYield™Plasmid Miniprep system (Promega, Madison, USA) following the centrifugation protocol; briefly 1.5ml bacterial culture was transferred into a microcentrifuge tube and centrifuged at 14 000xg for 30 seconds, this was done twice to obtain a high yield of purified plasmid DNA. A volume of 600µl TE buffer (pH 8.0) was added to the cell pellet and resuspended completely. A 100µl aliquot of cell lysis buffer and 350µl neutralization solution were added to the bacterial culture. The reaction mixture was inverted six times and centrifuged at 14 000xg for three minutes. The supernatant was transferred to a Pure Yield™ Minicoloumn and centrifuged at 14 000xg

for 15 seconds. Wash steps using endotoxin removal and column wash solution were performed to get rid of residual debris. Lastly plasmid DNA was eluted in 30µl elution buffer. The DNA concentration of the plasmid DNA was measured by using a NANODROP spectrophotometer (Table 3.12). The eluted plasmid DNA was stored at -20^{0}C. Conformation and correct orientation of positive transformants were confirmed by restriction digestion and plasmid DNA PCR respectively.

To verify that the insert was successfully ligated into pGEM®-T easy vector, the vector was linearized using a restriction enzyme, *Not1* which recognizes two sites located within the multiple cloning site of the pGEM®-T easy vector at positions 43 and 77 (Figure 3.2). Reaction components are shown in Table 3.2. For the negative control, a 1µl volume of NFW was used instead of plasmid. The restriction digestion reaction was incubated at 37^{0}C for two hours, the products were separated by electrophoresis on 1% agarose gel at 80V for 60 minutes and visualised under a UV transilluminator.

Table 3.2. Reaction components of the restriction digestion using *Not1* restriction enzyme

Reaction components	**Volume**
Not 1 (10U/µl)	1µl
Plasmid	1µl
10x restriction enzyme, Buffer D	2µl
NFW	16µl
Total	**20µl**

A positive transformant for each HPV type was selected and the partial HPV E6 gene region was amplified using T7 forward primer which targets a site present on the pGEM®-T easy vector (Table 3.3) and an HPV type specific reverse primer, HPV 16R, HPV18R, HPV31R and HPV 45R) for HPV types -16, -18, -31 and -45 respectively, downstream of the inserted gene. This reaction was performed to confirm the correct orientation of the gene of interest. PCR components are shown in Table 3.4.

Table 3.3. Primers that flank the multiple cloning site of the pGEM®-T easy vector

Primer name	Forward/reverse	Nucleotide sequence	*T_m	*%GC	Length
T7 primer	Forward	5'TAATACGACTCACTATAGG3'	56^0C	40	20bp
SP6 primer	Reverse	5'ATTTAGGTGACACTATAG3'	50^0C	32	18bp

*T_m and GC content calculated using Promega biomath.

(www.**promega**.com/a/apps/**biomath**/?calc=tm).

Table 3.4. PCR components for plasmid DNA PCR

PCR Components	Volume	Final concentration
5x Green GoTaq®flexi buffer	10µl	1X
$MgCl_2$ solution, 25mM	4µl	2mM
PCR nucleotide mix, 10mM each	1µl	0.2mM each dNTP
T7 forward primer (20pmol/µl)	1µl	0.4µM
*HPV type specific primer reverse primer (20pmol/µl)	1µl	0.4µM
GoTaq®G2 Hot Start Polymerase (5U/µl)	0.25µl	1.25U
DNA plasmid	1µl	-
NFW	31.75µl	-
Total	**50µL**	**-**

*Type specific reverse primers shown below

HPV 16R (5'TGCATAAATCCCGAAAAGCAAAGTC3')

HPV 18R (5'GCAGCATGCGGTATCTGTCT3')

HPV 31R (5'GCACACACTCCGTGTGGTGTG3')

HPV 45R (5'TTGATATACCTCTGTGCGTTCC3')

The reaction was cycled using the following cycling conditions: initial denaturation for one minute at 95^0C followed by 25 cycles of denaturation at 95^0C for one minute, annealing temperature at 50^0C for one minute, extension 72^0C for one minute and extension at 72^0C for five minutes and samples were held at 4^0C indefinitely. The PCR amplicons were visualised on a 1% agarose gel as previously (described in 2.4.5.) and purified using Wizard® SV Gel and PCR Clean-Up System as described in 2.4.6. The DNA concentration and purity of the PCR amplicon was measured using a NANODROP

spectrophotometer already described in 2.4.7. The nucleotide sequence of each amplicon was determined using the T7 forward primer (0.8pmol/µl) and HPV type specific reverse primer (0.8pmol/µl) using the Big® Dye Terminator V3.1 cycle sequencing kit (described in 2.4.8). The sequence reaction was cleaned using EDTA/ethanol precipitation. Samples were stored at 4^0C until submission to the Department of Microbial, Biochemical and Food Biotechnology, University of the Free State, Bloemfontein in SA. To confirm the genotypes of each positive control, nucleotide sequences were edited using Chromas Pro version 1.6, aligned using Clustal Omega version 1.2.1 and sequence data retrieved from GenBank and a BLAST analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was performed.

3.4.1.4. RNA transcript

DNA templates for transcribing RNA for application as positive controls to optimize the RT-PCR for detection of replicating HPV were prepared by amplification of the pGEM®-T easy constructs (pGEM.HPV16; pGEM.HPV18; pGEM.HPV31 and pGEM.HPV45) using GoTaq®G2 hot start polymerase (Promega, Madison, USA) and primers targeting regions flanking the HPV gene of interest. A reverse primer SP6 (Table 3.3) targeting a site located on the pGEM®-T easy vector and an HPV type specific forward primer were used. PCR components are shown in Table 3.5. PCR reaction was performed using the following cycling conditions; 95^0C for two minutes followed by 30 cycles at 95^0C for 30 seconds, 49^0C for 30 seconds, 72^0C for one minute and incubation at 72^0C for five minutes and samples were held at 4^0C indefinitely.

Table 3.5. PCR components for preparation of the RNA transcript

PCR components	Volume	Final concentration
5x Green GoTaq® flexi buffer	10µl	1X
$MgCl_2$ solution, 25Mm	4µl	2mM
PCR nucleotide mix, 10mM each	1µl	0.2mM each dNTP
*Type specific forward primer (20pmol/µl)	1µl	0.4µM
SP6 reverse primer (20pmol/µl)	1µl	0.4µM
GoTaq® G2 Hot Start Polymerase (5U/µl)	0.25µl	1.25U
DNA template	5µl	-

NFW	27.75µl	-
Total	**50µl**	**-**

*Type specific primers shown below.

HPV 16F (5'AGGACCCACAGGAGCGAC3')

HPV18F (5'ATGGCGCGCTTTGAGGATCC3')

HPV31F (5'CGGCATTGGAAATACCCTACGA 3')

HPV45F (5'GGCGCGCTTTGACGATCCAAAG3')

The amplicon was purified using Wizard® SV Gel and PCR Clean-Up System described in 2.4.6. The DNA concentration and purity was measured using a NANODROP spectrophotometer. The nucleotide sequence of each amplicon was confirmed using SP6 reverse primer (0.8pmol/µl) and the HPV type specific forward primer (0.8pmol/µl) and the Big® Dye Terminator V3.1 cycle sequencing kit (described in 2.4.8).

RNA was transcribed from the DNA template using SP6 MEGAscript®Kit (Ambion Inc., Texas, USA) according to manufacturer's instructions. The reaction components are shown in Table 3.6.

Table 3.6. Reaction components for RNA transcription

Components	**Volume**
ATP (5mM)	1µl
GTP (5mM)	1µl
CTP (5Mm)	1µl
UTP (5mM)	1µl
10xReaction buffer	2µl
SP6 enzyme mix (20 U/µl)	2µl
*Plasmid DNA	1-3ng
NFW	8µl
Total	to a total of 20µl

*Plasmid DNA for each reaction for RNA transcription. For HPV types -16, -18, -31 and -45.

The reaction mixture was incubated at 37^0C for 16 hours. The RNA transcript was purified using SV Total RNA Isolation System (Promega, Madison, USA) according to manufacturer's instructions. Briefly the 20µl transcript reaction was transferred into a microcentrifuge tube containing 175µl RNA lysis buffer. A volume of 350µl of dilution buffer was added to the lysate and mixed thoroughly by inverting four times and the lysate was incubated at 70^0C for three minutes. The lysate was centrifuged for 10 minutes at 14 000 x g. A 200µl aliquot of 95% ethanol was added to the cleared lysate, mixed and centrifuged at 14 000 x g for one minute through a spin column. The column was washed with 600µl RNA wash solution. DNA was removed by incubation with 50 µl of DNase incubation mix (40µl yellow core buffer, 5µl 0.09M $MnCl_2$ and 5µl of DNase 1 enzyme) at 25^0C for 15 minutes. The enzyme activity was stopped by using DNase stop solution and the spin column washed to remove potential inhibitors. The RNA was eluted in a 50µl NFW and stored at -80^0C for downstream application. Due to the fact that HPV is a DNA virus and positive RNA controls were prepared from the DNA, it was necessary to check for DNA contamination in the RNA. Therefore a reaction without reverse transcription (RT) was run in parallel for each sample for control for DNA contamination, using RNA as a template instead of cDNA.

3.4.2. Optimization of RT-PCR

Purified RNA was diluted 1:1000 and was used as a template to optimize the RT-PCR. A one step method using Superscript®III one-step RT-PCR system with Platinum®Taq High Fidelity (Invitrogen, Waltham, Massachusetts, USA) was compared with a two-step RT-PCR using Superscript™III reverse transcriptase (Invitrogen, Waltham, Massachusetts, USA) and GoTaq®hot start polymerase.

The following mix was prepared for the one-step RT-PCR reaction; 25µl of 2x reaction mix, 5µl and 10µl template (RNA) for separate reactions, 2µl HPV type specific forward primer (10pmol/µl), 2µl HPV type specific reverse primer (10pmol/µl), 1µl Superscript®III RT/platinum®Taq high fidelity enzyme mix, add NFW to a final reaction volume of 50µl. The reaction was cycled as follows: 60^0C for 15 minutes, 94^0C for two minutes and 40 cycles at 94^0C for 15 minutes, 60.5^0C for 30 seconds and 68^0C for one minute, the reaction was incubated at 68^0C for five minutes and held at 4^0C indefinitely.

The two step reaction was performed as follows: 1µl HPV type specific forward primer (2pmol/µl), 1µl dNTPS (10mM), 5µl template (RNA), 6µl NFW, 4µl of 5X first-stranded buffer, 1µl DTT (0.1M), 0.5µl Rnase inhibitor (2U/µl), 1µl Superscript™RT enzyme (200U/µl) to a final total reaction volume of 20µl. The RNA was transcribed at 65^0C for five minutes, 4^0C for two minutes, 50^0C for 60 minutes, 85^0C for five minutes and 4^0C indefinitely. The complementary deoxyribonucleic acid (cDNA) was amplified as follows: 10µl 5x green GoTaq flexi buffer, 4µl $MgCl_2$ solution (25mM), 1µl PCR nucleotide mix (10mM), 1µl HPV type specific forward primer (20pmol/µl), 1µl HPV type specific reverse primer (20pmol/µl), 0.25µl GoTaq®G2 hot start polymerase (5U/µl), 5µl cDNA and 27.75µl NFW to a final reaction volume of 50µl. The reaction was cycled at 95^0C for two minutes followed by 30 cycles of 95^0C for 30 seconds, 60.5^0C for 30 seconds, 72^0C for one minute. The sample was incubated at 72^0C and held at 4^0C indefinitely. The limit of detection of this assay was not determined and this could have influenced the negative results found in four samples.

3.4.3. Hemi nested RT-PCR (HnRT-PCR) for detection of HPVE6 mRNA

3.4.3.1. Samples

Samples that tested positive for HR-HPV types -16, -18, -31 and -45 respectively were tested for HPV E6 mRNA expression using HnRT-PCR.

3.4.3.2. RNA extraction

RNA was extracted from tissue biopsies stored in RNAlater® using RNeasy® Plus Universal mini kit from Qiagen according to manufacturer's instructions (QIAGEN, Valencia, CA, USA). Briefly tissue biopsies were homogenised in 900µl of QIAzol lysis reagent. A volume of 100µl genomic deoxyribonucleic acid (gDNA) eliminator solution was added to the homogenate and 180µl of chloroform. Wash steps were performed to get rid of contaminating debris. Purified RNA was eluted in 50µl RNase-free water and stored at -80^0C for downstream applications. To check for DNA contamination in RNA samples, a reaction without RT was run in parallel (RT-control) with each sample.

3.4.3.3. Primers

Type specific primer pairs for HPV types -16, -18, -31 and -45 (described in Chapter 2) were used to amplify HPV E6mRNA from four samples that tested positive for HR-HPV types.

3.4.3.4. First strand cDNA synthesis

On comparison of the one step and two step protocols, the two-step assay was sensitive and produced less primer dimers. The samples were therefore tested using a two-step protocol. RNA extracted from tissue biopsies was reverse transcribed using Superscript™III reverse transcriptase according to manufacturer's instructions. Components of the reaction mix are showed in Tables 3.7 and 3.8 respectively.

Table 3.7. RNA and primer mixture for cDNA synthesis

Components	**Volume**
Forward primer (2pmol/µl)	1µl
PCR nucleotide mix (10mM)	1µl
H_2O (distilled)	6µl
RNA	5µl
Total	**13µl**

Table 3.8. Master Mix for cDNA synthesis

Components	**Volume**
5X first-stranded buffer	4µl
DTT (0.1M)	1µl
RNase inhibitor (2U/µl)	0.5µl
Superscript™RT enzyme (200U/µl)	1µl
H_2O (distilled)	0.5µl
Total	**7µl**

First strand cDNA was synthesized using the following reaction conditions: 65^0C for five minutes, 4^0C for two minutes, 50^0C for 60 minutes, 85^0C for five minutes and lastly

samples were held at 4^0C indefinitely. First strand cDNA was used as a template in the first round of the HnRT-PCR, reaction components are shown in Table 3.9.

Table 3.9. PCR components for first round HnRT-PCR

Components	**Volume**	**Final concentration**
5X Green GoTaq® flexi buffer	10µl	1X
$MgCl_2$ solution, 25mM	4µl	2mM
PCR nucleotide mix, (10mM each)	1µl	0.2mM each dNTP
*Type specific forward primer (20pmol/µl)	1µl	0.4µM
*Type specific reverse primer (20pmol/µl)	1µl	0.4µM
GoTaq®G2 Hot Start Polymerase (5U/µl)	0.25µl	1.25U
*Template (first strand cDNA)	5-10µl	-
NFW	to a total of 50µl	-
Total	**50µl**	

*Type specific forward primer

HPV 16F (5'AGGACCCACAGGAGCGAC3')

HPV 18F (5'ATGGCGCGCTTTGAGGATCC3')

HPV 31F (5'CGGCATTGGAAATACCCTACGA3')

HPV 45F (5'GGCGCGCTTTGACGATCCAAAG3')

*Type specific reverse primer

HPV 16R (5'TGCATAAATCCCGAAAAGCAAAGTC3')

HPV 18R (5'GCAGCATGCGGTATACTGTCT3')

HPV 31R (5'GCACACACTCCGTGTGGTGTG3')

HPV 45R (5'TTGATATACCTCTGTGCGTTCC3')

*5µl and 10µl of template (cDNA) were added into the reaction to test for mRNA. The amplification reaction was performed as follows: 95^0C for two minutes and followed by 30 cycles of 95^0C for 30 seconds, 62^0C for 30 seconds and 72^0C for one minute. The samples were incubated at 72^0C for five minutes and held at 4^0C indefinitely. PCR products were visualised on a 1% agarose gel under a UV transilluminator (described in 2.4.5).

Each sample was then tested in a hemi-nested PCR reaction using the type specific reverse primer HPV16R, HPV18R, HPV31R or HPV45R and a type specific forward primer described in chapter two (HPV16F2, HPV18F2, HPV31F2 or HPV45F2). A 1µl aliquot of each first round PCR product was used as a template in the HnRT-PCR with the following cycling conditions: 95^0C for two minutes followed by 30 cycles of 95^0C for 30 seconds, 62^0C for 30 seconds, 72^0C for one minute and one final extension at 72^0C for five minutes. The samples were held at 4^0C indefinitely. PCR products were visualised on a 2.5% agarose gel under a UV transilluminator (described in 2.4.5). PCR reaction components are shown in Table 3.10.

Table 3.10. PCR components for second round HnRT-PCR

Components	**Volume**	Final concentration
5X Green GoTaq®flexi buffer	10µl	1X
$MgCl_2$ solution, 25mM	4µl	2mM
PCR nucleotide mix (10mM each)	1µl	0.2mM
*Type specific forward primer (20pmol/µl)	1µl	0.4µM
*Type specific reverse primer (20pmol/µl)	1µl	0.4µM
GoTaq® Hot Start Polymerase (5U/µl)	0.25µl	1.25U
Template (first round PCR amplicon)	1µl	-
NFW	31.75µl	-
Total	**50µl**	-

*Type specific forward primers

HPV 16F2 (5'CCACAGTTATGCACAGAGCTGCAA3')

HPV 18F2 (5'GTGCACGGAACTGAACACTTCACT3')

HPV 31F2 (5'CTGCAAAGGTCAGTTAACAGAAAC3')

HPV 45F2 (5'CCCTACAAGCTACCAGATTTG3')

*Type specific reverse primers

HPV 16R (5'TGCATAAATCCCGAAAAGCAAAGTC3')

HPV 18R (5'GCAGCATGCGGTATACTGTCT3')

HPV 31R (5'GCACACACTCCGTGTGGTGTG3')

HPV 45R (5'TTGATATACCTCTGTGCGTTCC3')

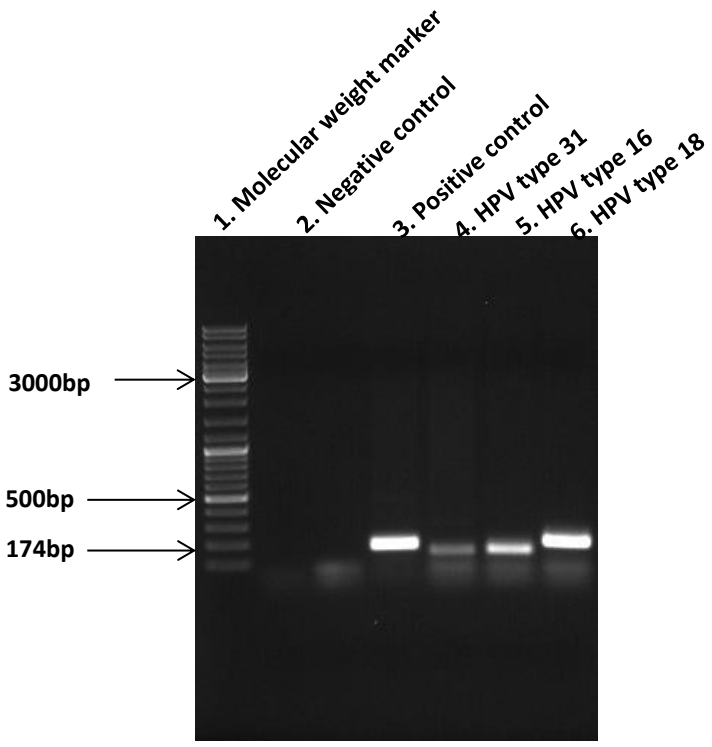
3.5. Results

3.5.1. Preparation of RNA controls

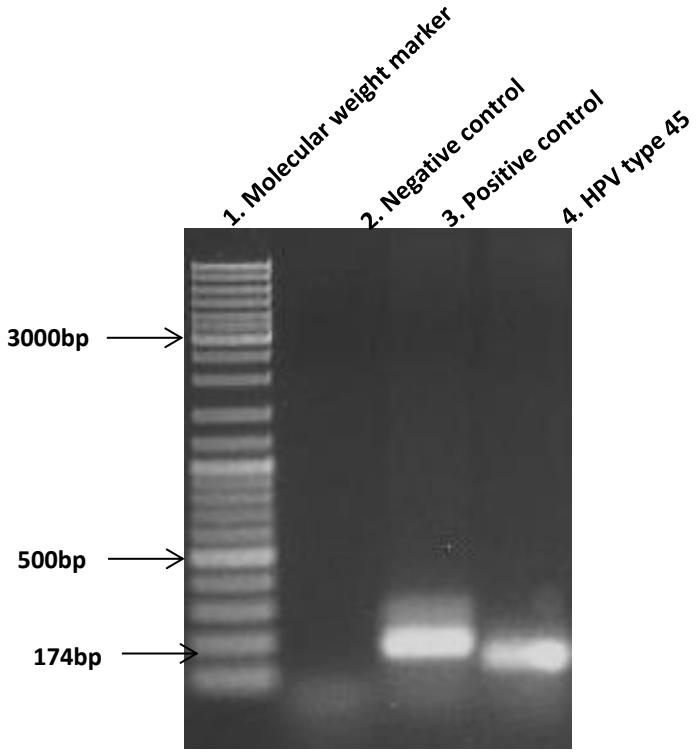
A region of the HPV E6 gene, for each of the HPV types -16, -18, -31 and -45, was amplified (Figure 3.3A-B) and ligated into pGEM®-T easy. PCR amplicons were excised from the 1% agarose gel and purified using Wizard®SV Gel and PCR Clean-Up System (Figures 3.3C-D) and the DNA concentration and purity for each purified PCR amplicon was determined (Table 3.11).

Table 3.11. DNA concentrations of the purified PCR amplicons used for cloning in the pGEM®- T easy vector

HPV types	DNA concentrations
31	11.9ng/µl
16	30.1ng/µl
18	172.3ng/µl
45	123.4ng/µl



A



B

Figure 3.3. A 1% agarose gel image showing results for the first round of the E6 hemi-nested type specific PCR under UV transilluminator. 3.3A: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lane 3: Positive control-174bp; Lane 4: HPV type 31-141bp; Lane 5: HPV type 16-147bp; Lane 6: HPV type 18-195bp. **3.3B:** Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lane 3: Positive control- 174bp; Lane 4: HPV type 45-136bp.

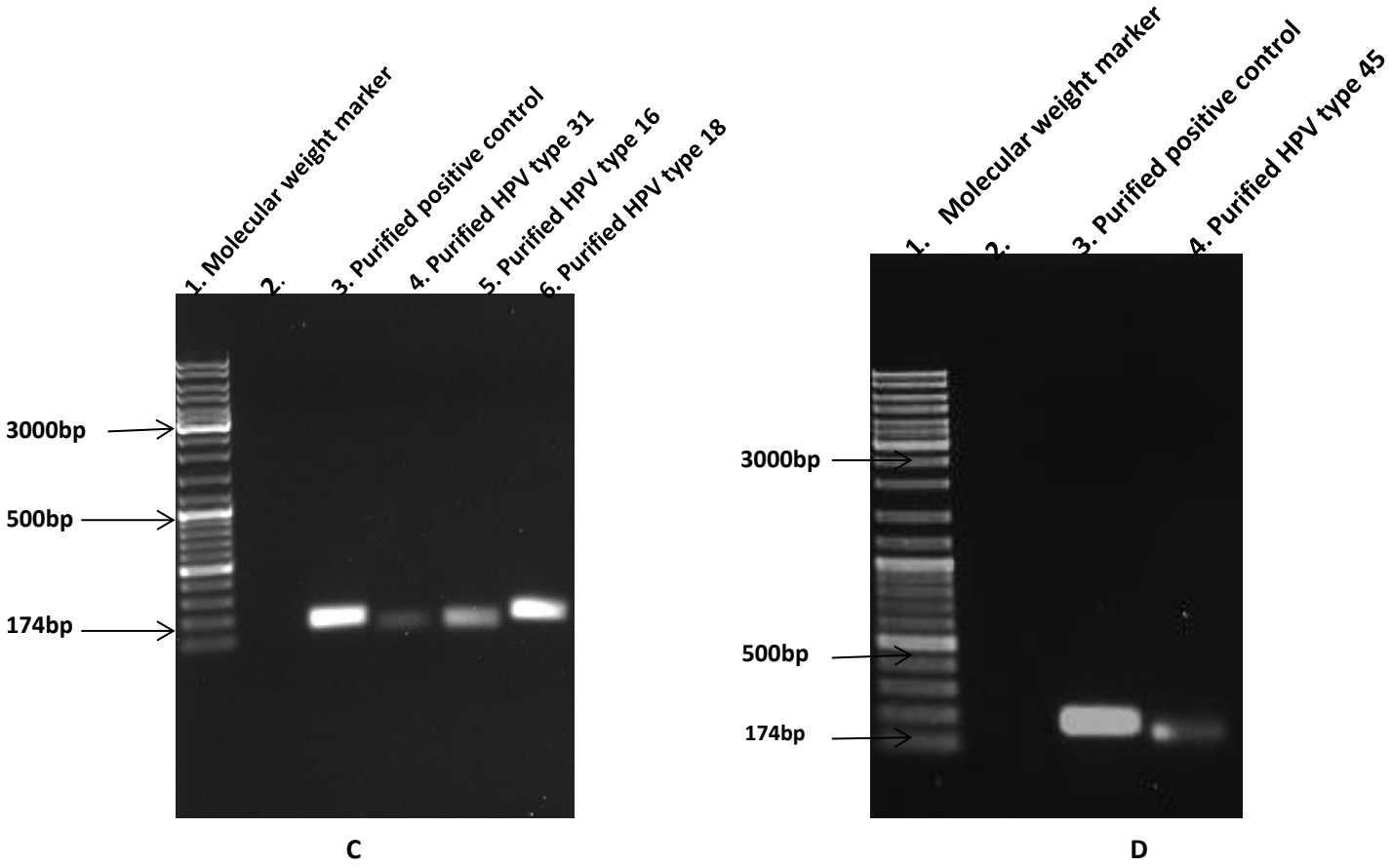


Figure 3.3 continues. 1% Agarose gel images showing results after PCR amplicons purification for the first round of the E6 hemi-nested type specific PCR under UV transilluminator. 3.3C: Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173, Lane 2: Empty; Lane 3: Positive control-174bp; Lane 4: HPV type 31-141bp; Lane 5: HPV type 16-147bp; Lane 6: HPV type 18-195bp. **3.3D:** Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Empty; Lane 3: Positive control; Lane 4: HPV type 45-136bp.

3.5.2. Confirmation of positive transformants

Three colonies from each ligation reaction were selected and cultured overnight in LB broth. Plasmid preparations were purified by using PureYield™Plasmid Miniprep system and DNA concentrations were measured (Table 3.12). Confirmation of positive transformants was determined using restriction enzyme digestion and orientation of the inserted gene confirmed by performing PCR using plasmid DNA from the selected colonies**.**

Table 3.12. DNA concentrations of purified colonies from different HPV types

HPV types	Colonies		
	A	**B**	**C**
31	179.6ng/µl	128.4ng/µl	181.9ng/µl
16	294.9ng/µl	286.5ng/µl	241.9ng/µl
18	151.8ng/µl	322.3ng/µl	195.4ng/µl
45	516.7ng/µl	383.1ng/µl	393.4ng/µl

3.5.2.1. Restriction digestion

Plasmid DNA purified from each culture was digested using *Not 1* restriction enzyme. The 20µl reaction volume was separated by electrophoresis and visualised under a UV transilluminator. The expected sizes were 171bp for HPV type -31; 177bp for HPV type -16; 221bp for HPV type -18; 166bp for HPV type -45 and 3015bp for pGEM®-T easy vector (Figures 3.4A-3.4D).

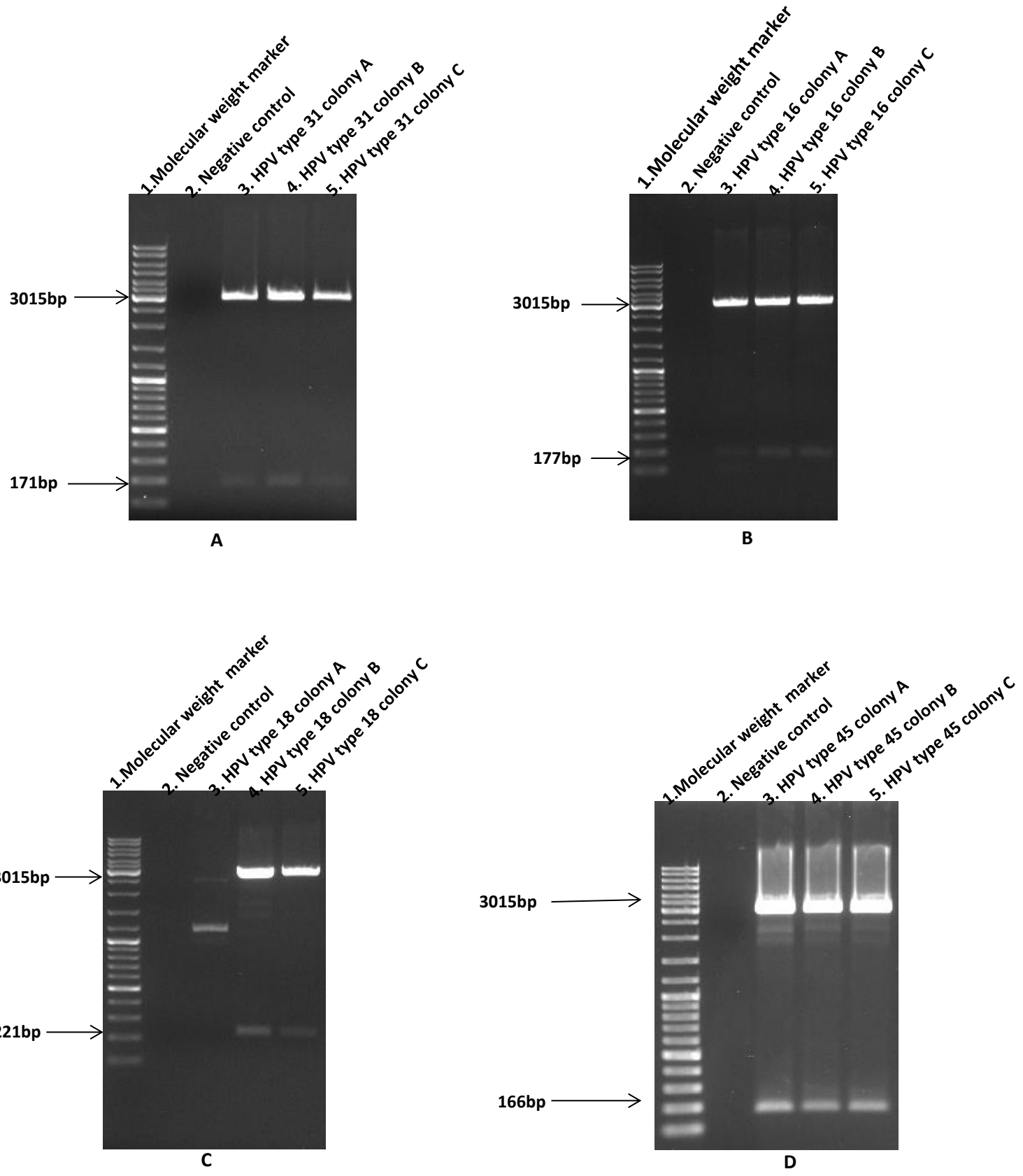
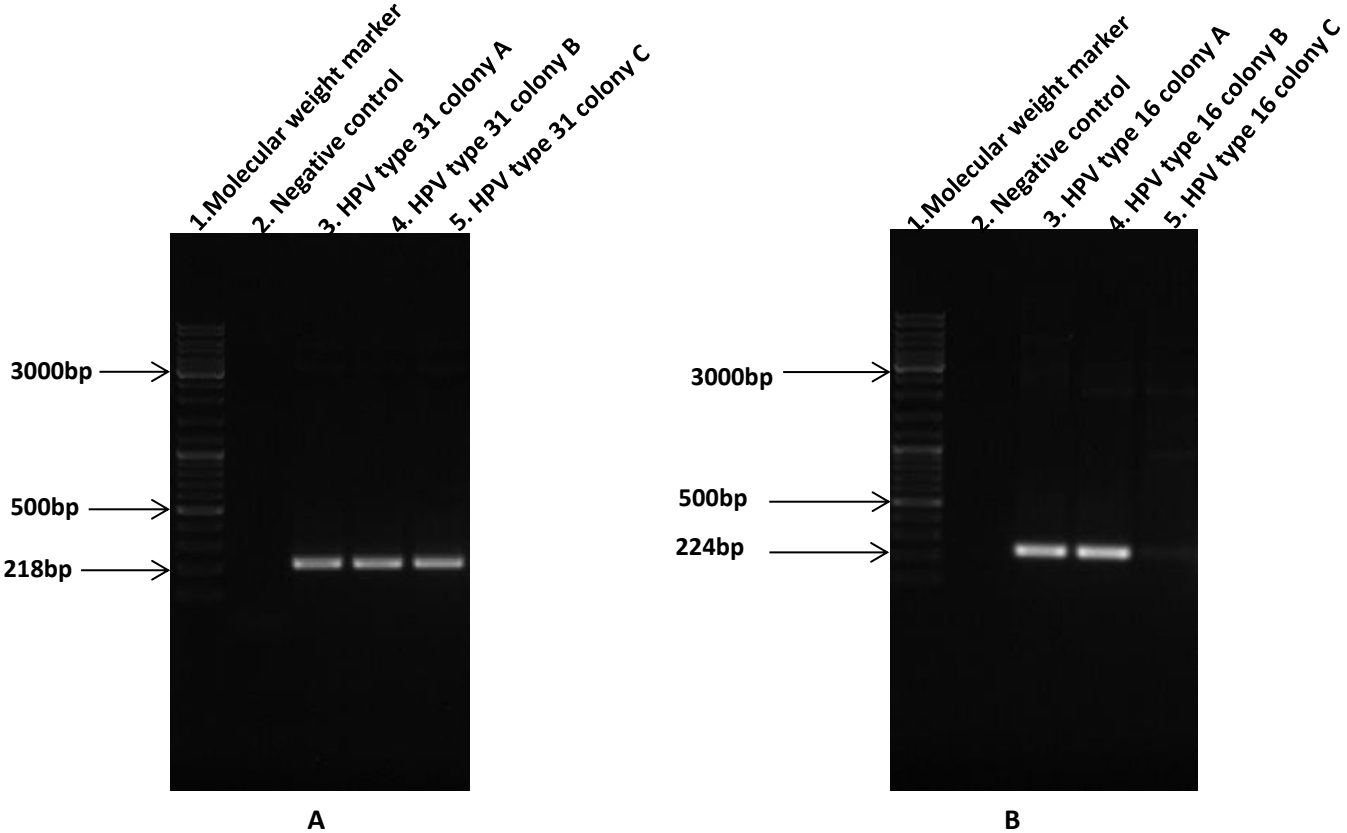


Figure 3.4. A 1% agarose gel electrophoresis analysis of restriction digestion. 3.4A: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control, lane 3 to lane

5 show expected band sizes at 171bp for HPV type -31 and pGEM®-T easy vector at 3015bp. **3.4B**: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control, lane 3 to lane 5 shows expected band sizes at 177bp for HPV type -16 and pGEM®-T easy vector at 3015bp. **3.4C**: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control, lanes 4 and 5 show expected band sizes at 221bp for HPV type -18 and pGEM®-T easy vector at 3015bp. **3.4D**: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control, lane 3 to lane 5 show expected band sizes at 166bp for HPV type -45 and pGEM®-T easy vector at 3015bp

3.5.2.2. Plasmid DNA PCR for confirmation of correct orientation

PCR was performed using plasmid DNA from each of the three purified colonies selected for HPV types -16, -31, -45 and two colonies for HPV type -18. The PCR was performed using T7 forward primer located in the plasmid upstream of the inserted partial gene and an HPV type specific reverse primer (HPV16R, HPV18R, HPV 31R and HPV 45R respectively). A 5µl aliquot of the PCR amplicon was separated by electrophoresis and visualised (Figures 3.5A-D)**.**



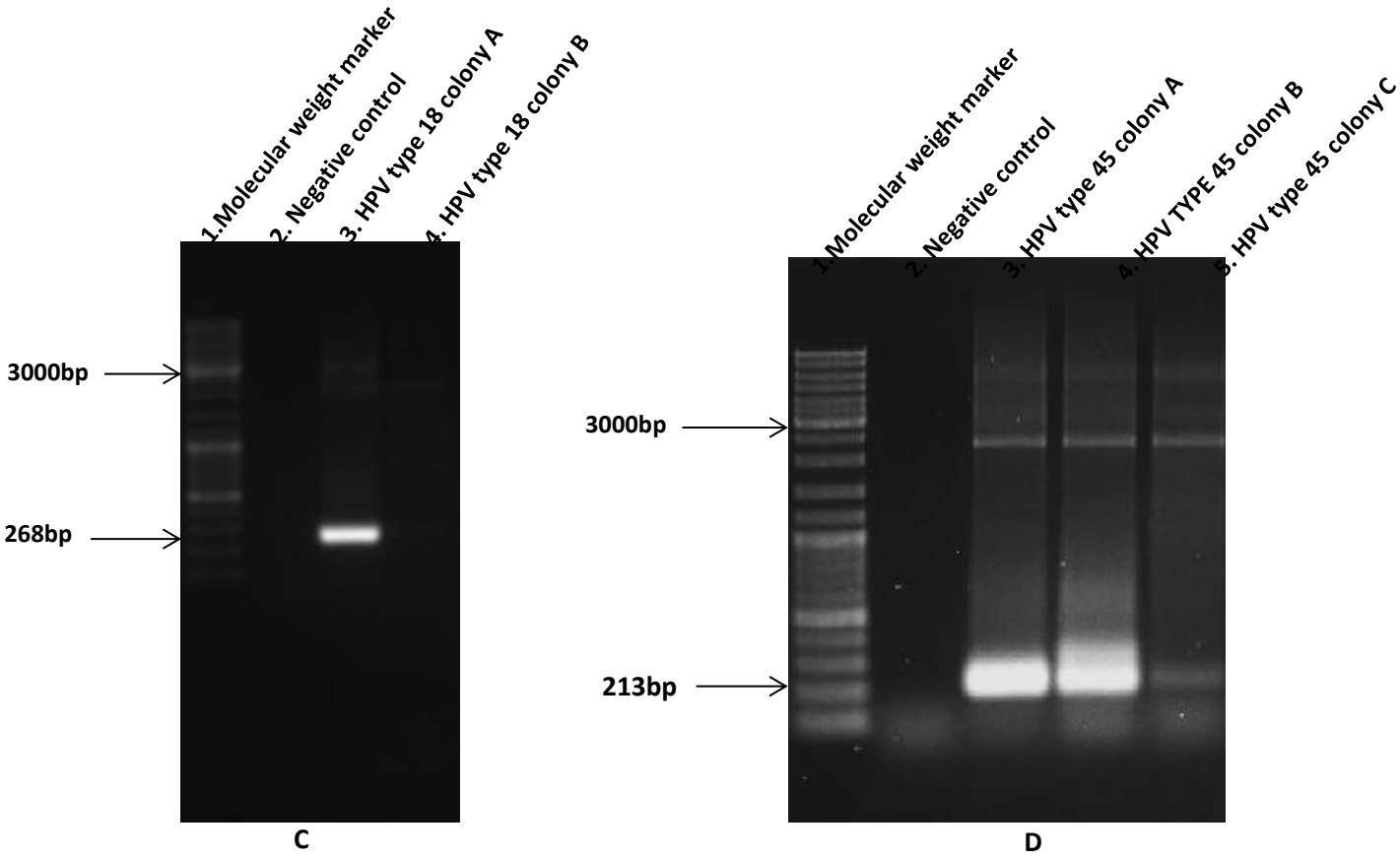


Figure 3.5 A 1% agarose gel electrophoresis analysis for results for colony PCR. 3.5A: HPV type -31 Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lanes 3-5 expected band sizes at 218bp. **3.5B: HPV type -16** Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lanes 3-5 expected band sizes at 224bp. **3.5C: HPV type -18** Lane 1: O'GeneRuler™ DNA ladder plus, SM 1173; Lane 2: Negative control; Lanes 3 and 4 expected band sizes at 268bp. **3.5D: HPV type -45** Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lanes 3-5 expected band sizes at 213bp.

Clones in which the gene was shown to be in the correct orientation by PCR were excised from a 1% agarose gel and purified using Wizard ®SV Gel and PCR Clean-Up System according to the manufacturer's instructions (described in 2.4.6) (Figure 3.6A-B).

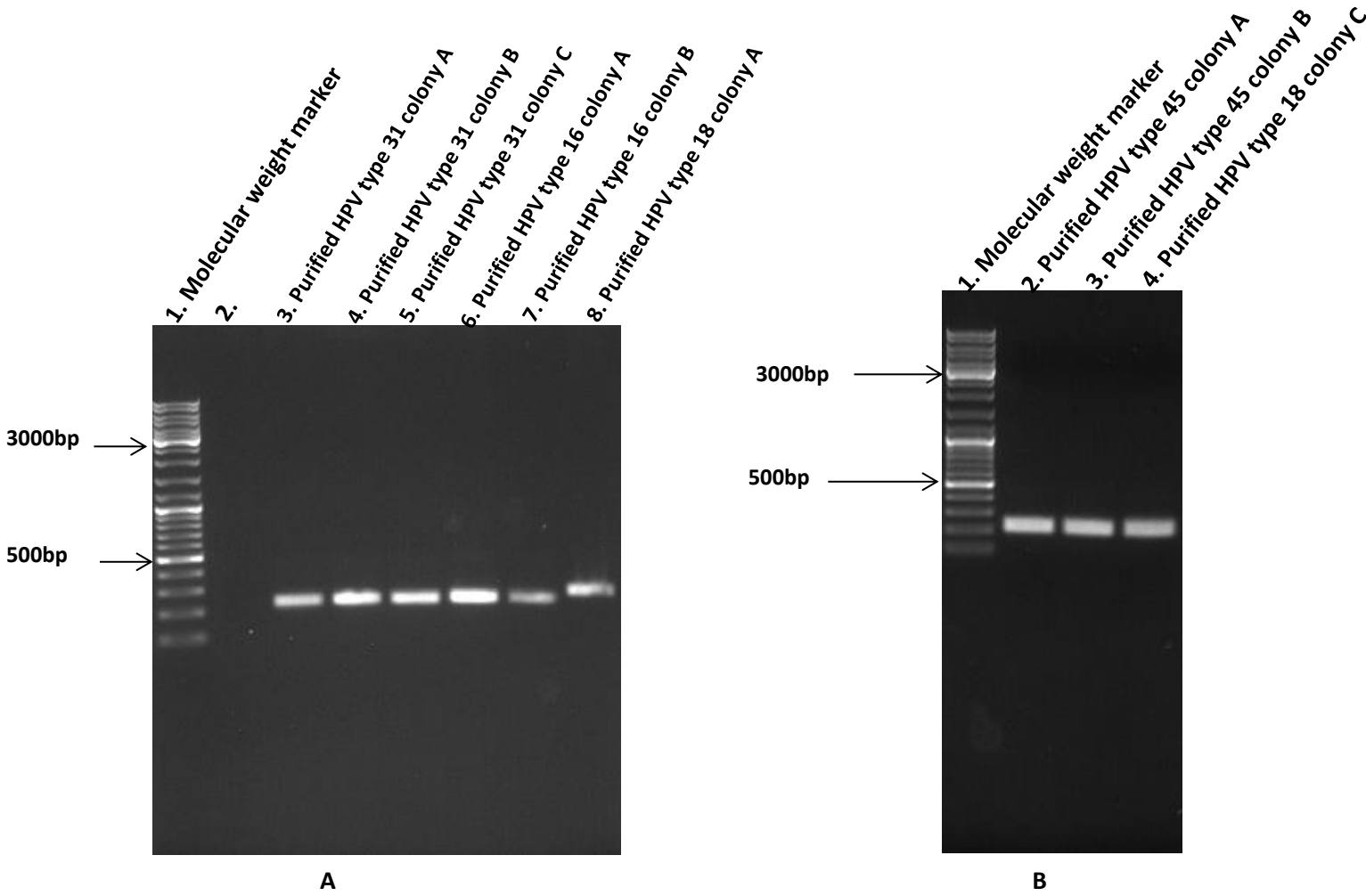


Figure 3.6. A 1% agarose gel electrophoresis results showing purified PCR Clones. 3.6A: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Empty; Lanes 3-5 PCR clones for HPV type -31 at 218bp; Lane 6-7 PCR clones for HPV type -16 at 224bp and lane 8: PCR clone for HPV type -18 at 268bp. **3.6B**: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lanes 3-5 PCR clones for HPV type -45 at 213bp.

Final confirmation of the plasmids was performed by determining the sequence of the inserted gene as described previously described in section 2.3.7. Nucleotide sequences are in Appendix H.

3.5.3. RNA transcription

The positively transformed constructs with the gene inserted in a correct orientation, were designated pGEM.HPV16; pGEM.HPV18; pGEM.HPV31 and pGEM.HPV45. To prepare a DNA template to be transcribed using SP6 MEGAscript®Kit, Plasmid PCR was

performed on one construct for each HPV type using SP6 and HPV type specific reverse primer (HPV 16R, HPV18R, HPV 31R and HPV 45R) (Figure 3.7A-D).

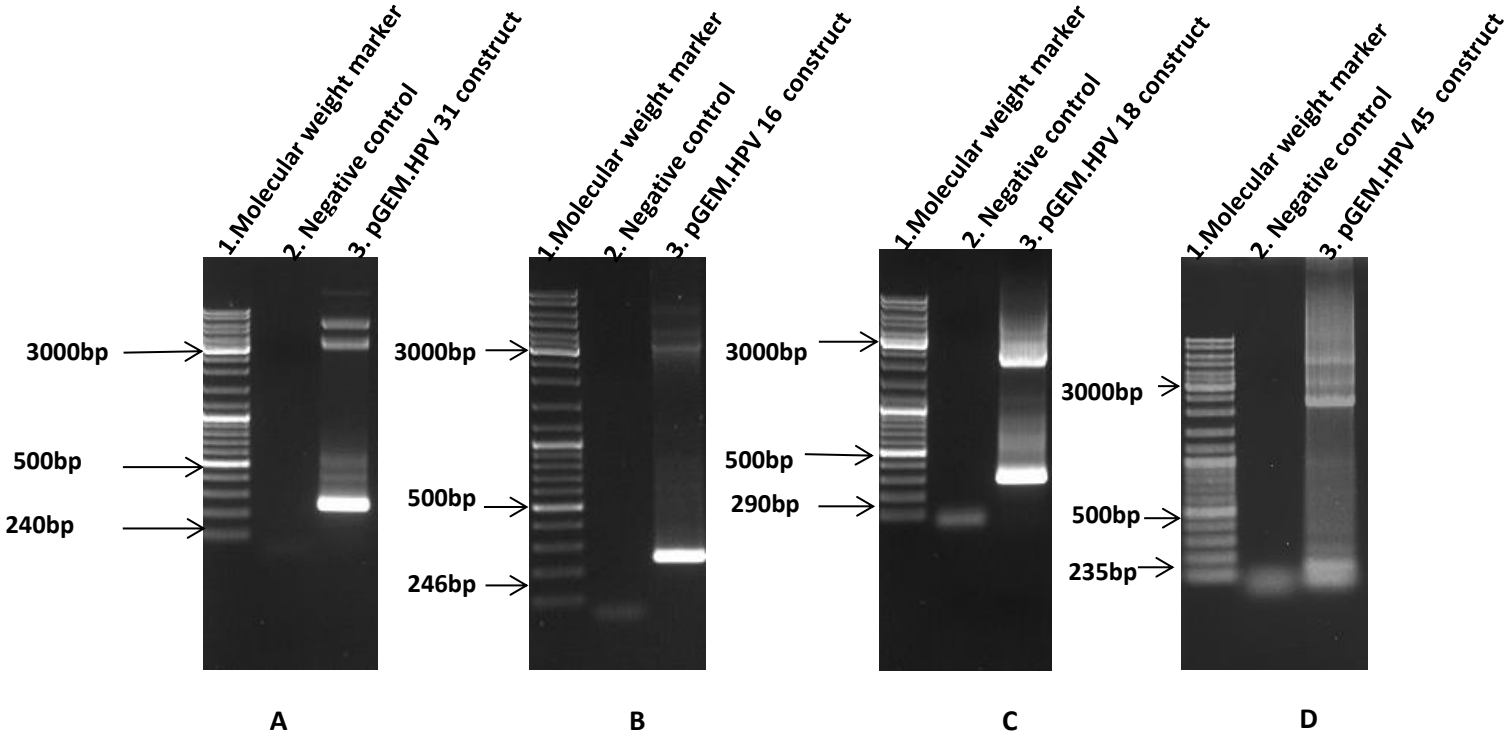


Figure 3.7. An image of a 1% agarose gel electrophoresis results of plasmid constructs: 3.7A: pGEM.HPV 31 construct**:** Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lane 3: Positive DNA template. **3.7B:** pGEM.HPV 16 construct**:** Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lane 3: Positive DNA template. **3.7C:** pGEM.HPV 18 construct**:** Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lane 3: Positive DNA template. **3.7D:** pGEM.HPV 45 construct**:** Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Negative control; Lane 3: Positive DNA template.

Positive PCR amplicons were purified using Wizard®SV Gel and PCR Clean-Up System according to manufacturer's instructions (described in 2.4.6) (Figures 3.8A-B).

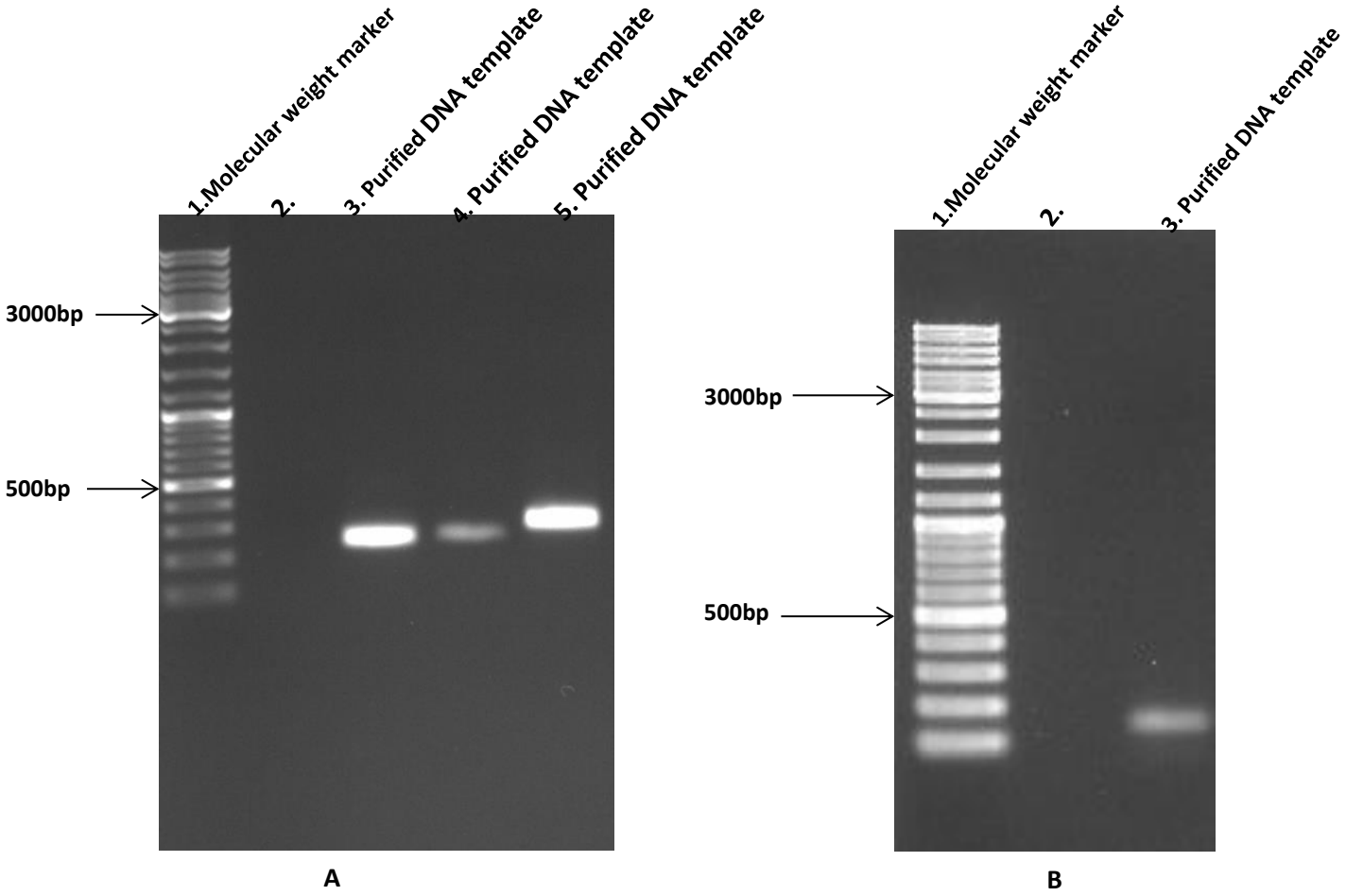


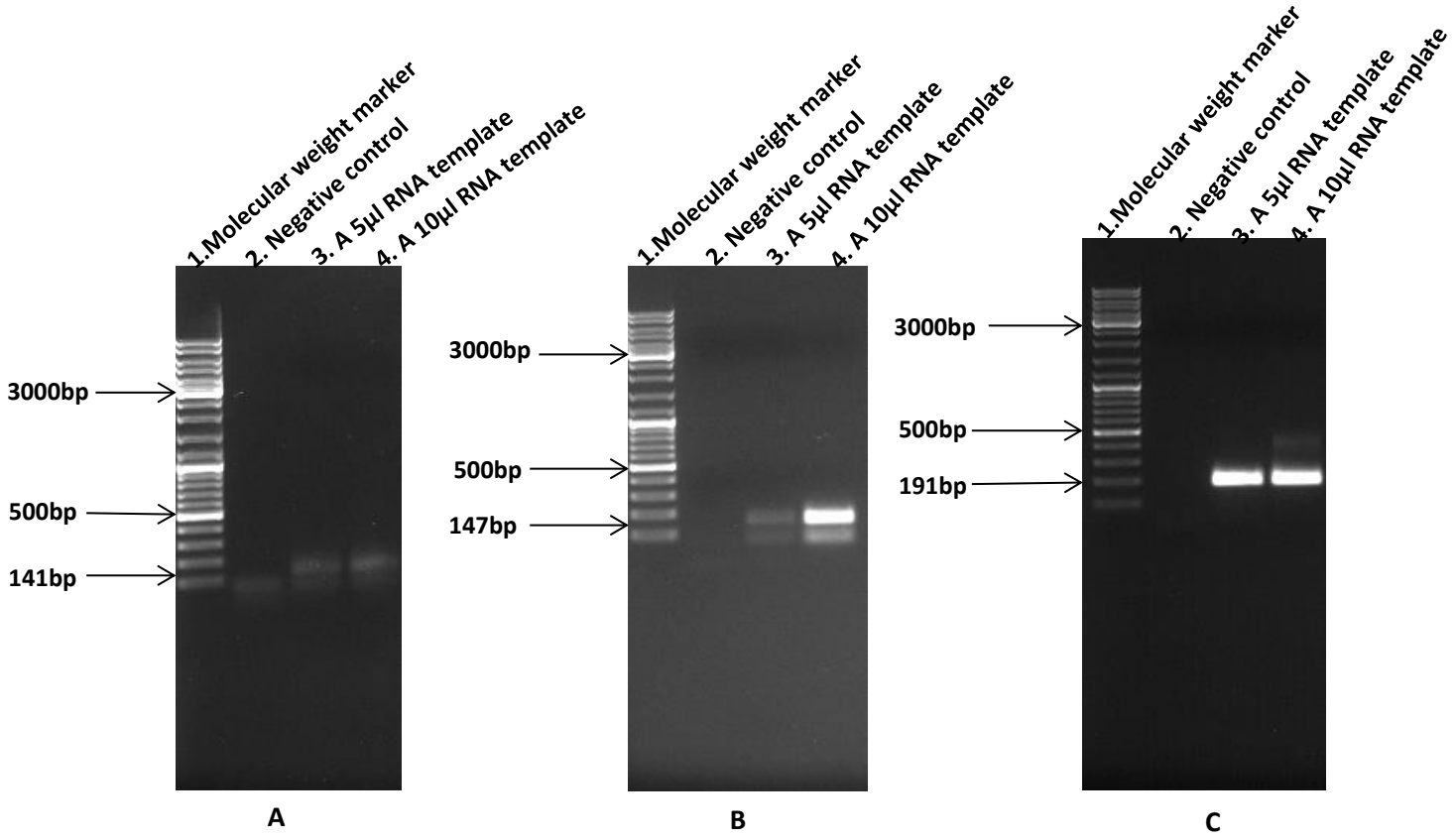
Figure 3.8. A 1% agarose gel electrophoresis results for purified PCR amplicons of plasmid constructs. 3.8A: Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Empty; Lane 3: DNA template from pGEM.HPV 16 construct; Lane 3: DNA template from pGEM.HPV 16 construct; Lane 4: DNA template from pGEM.HPV 18 construct **3.8B:** Lane 1: O'GeneRuler™ DNA ladder Plus, SM 1173; Lane 2: Empty; Lane 3: DNA template from pGEM.HPV 45 construct.

3.5.4. Optimization of RT-PCR

Transcribed RNA was purified by SV Total RNA Isolation system and was used as template in HnRT-PCR. Two methods were compared, one step method using Superscript®III one step RT-PCR system with Platinum®Taq High Fidelity and a two-step Superscript™III reverse Transcriptase. Purified RNA was diluted ten-fold to 1:1000 to determine the limit of detection.

3.5.4.1. Superscript®III one step RT-PCR system with Platinum®Taq High Fidelity

First strand cDNA was synthesized using Superscript®III one-step RT-PCR system with Platinum®Taq High Fedility enzyme and diluted RNA was used as a template. There were two reactions, one with a volume of 5µl diluted RNA was used as the template and the other with reaction 10µl volume of diluted RNA was used as a template. A 5µl RT-PCR product for each reaction was visualised after separation by electrophoresis. See below Figures 3.9A-D.



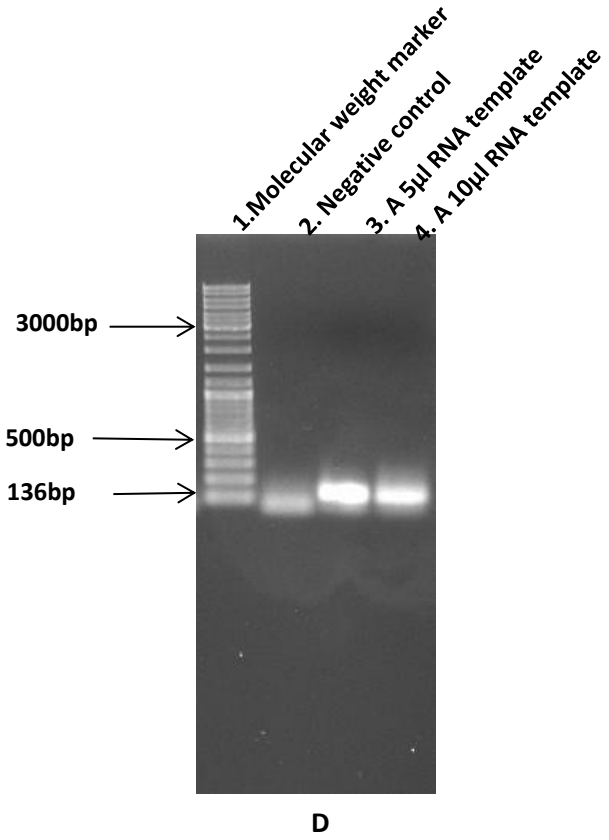


Figure 3.9. A 1% agarose gel images depicting results for RNA controls for RT-PCR using superscript®III one-step RT-PCR system with platinum®Taq high fidelity kit. 3.9A: HPV type -31 PCR; Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: Negative control; Lane 3: A volume of 5µl RNA; Lane 4: A volume of 10µl of RNA. **3.9B: HPV type -16 PCR;** Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: Negative control; Lane 3: A volume of 5µl RNA; Lane 4: A volume of 10µl RNA. **3.9C: HPV type -18 PCR;** Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: Negative control; Lane 3: A volume of 5µl RNA; Lane 4: A volume of 10µl RNA**. 3.9D: HPV type -45 PCR;** Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: Negative control; Lane 3: A volume of 5µl RNA; Lane 4: A volume of 10µl RNA.

3.5.4.2. Superscript™III reverse transcriptase

Superscript™III reverse transcriptase was used to synthesize first strand cDNA from transcribed RNA. First strand cDNA served as a template for PCR, two reactions were performed, one reaction a volume of 5µl of first strand cDNA was used as the template and the other reaction a 10µl volume of first strand cDNA was used as a template. A

volume of 5µl RT-PCR product for both reactions were separated by electrophoresis and visualised using a UV transilluminator (Figure3.10A-D).

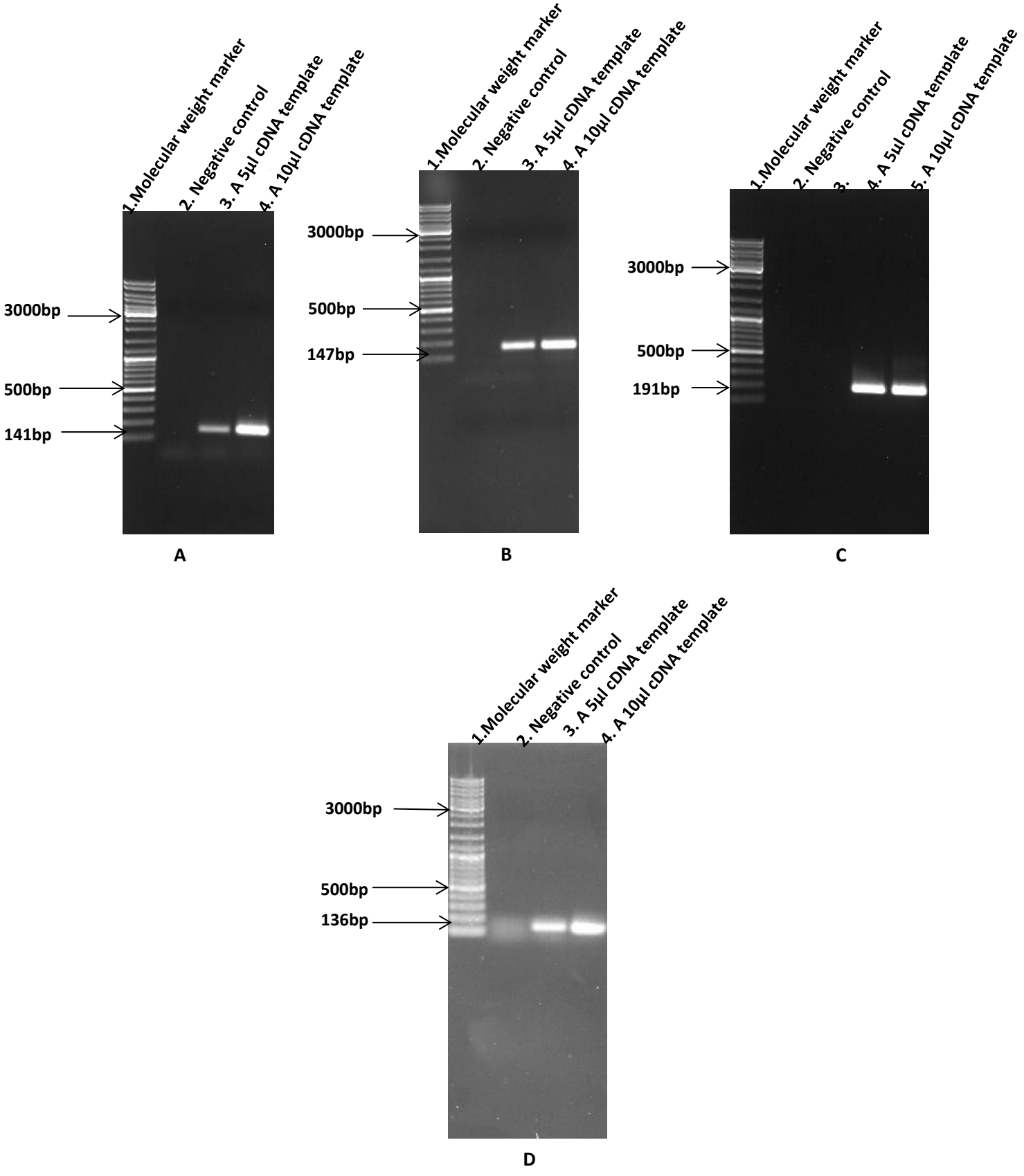


Figure 3.10. A 1% agarose gel depicting results for RNA controls for RT-PCR using superscript™III reverse transcriptase. 3.10A: HPV type -31 PCR: Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: Negative control; Lane 3: A volume of 5µl cDNA and lane 4: A volume of 10µl of cDNA. **3.10B: HPV type -16 PCR**: Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: Negative control; Lane 3: A volume of 5µl cDNA and lane 4: A volume of 10µl of cDNA. **3.10C: HPV type -18 PCR:** Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: negative control; Lane 3: Empty; Lane 4: A volume of 5µl cDNA and lane 5: A 10µl volume of cDNA. **3.10D: HPV -45 PCR**: Lane 1: O'GeneRuler™ DNA ladder Plus SM 1173; Lane 2: negative control; Lane 3: A volume of 5µl cDNA and lane 4: A 10µl volume of cDNA.

3.5.5. HPV E6 mRNA detection

A total of four samples that were genotyped as HPV types -16, -18, -31 and -45 were tested for replicating/transcriptionally active mRNA. All samples tested negative by first round and were further tested by a hemi-nested PCR. Again they were all negative.

3.6. Discussion

The detection of HPV DNA in tissue biopsies is insufficient to implicate a role for HPV in causing cancer (Kreimer *et al*., 2005). Transcriptionally active or replicating HPV or evidence of integration of HPV into the host genome is more suggestive of an aetiological role for HPV. Over the past few years, the clinical importance of detecting HPV associated with HNSCC has been emphasized. E6/7 mRNA detected in a tissue biopsy is clinically and prognostically important in that patients respond better to treatment (Larque *et al.*, 2014). To detect replicating HPV, RT-PCR assays are required that can detect mRNA (Bishop *et al*., 2012). Therefore an RT-PCR was developed to specifically test the samples in which DNA was amplified for the detection of transcriptionally active or replicating HPV. In the previous chapter, the specificity of the in house E6 primers to recognise targeted HPV sequences and amplify these regions was confirmed using either clinical samples that were identified in the study or alternatively using synthesised genes. However RNA controls were not available hence transcribed RNA was prepared to optimize the RT-PCR. In order to develop and optimize the assay, transcribed RNA was

prepared for use as controls. The transcribed RNA was prepared for each HR-HPV type genotyped (-16, -18, -31 and -45) by ligating PCR amplicons into pGEM®-T easy cloning vector. Templates for transcription of RNA were prepared by amplifying the target genes, including SP6 promotor site. The SP6 promoter site was used to transcribe RNA and positive RNA controls were confirmed by RT-PCR. PCR amplicons from the first round of the HnRT-PCR were sequenced for confirmation.

RNA was extracted from each of four samples that were positive for HPV DNA. The biopsies were stored in RNAlater® to preserve RNA at the time of collection. All samples tested negative for HPVE6 mRNA thus suggesting they are not transcriptionally active. In conclusion although all samples suggest that they are not transcriptionally active using this developed method, the assay can be used in our laboratory to detect transcriptionally active HPV types.

CHAPTER 4: CONCLUSION

Evidence suggesting an involvement of HPV in squamous cell carcinoma was first reported in 1983 by Syrjänen and colleagues. In a study investigating 40 biopsies from oral squamous cell carcinomas, certain histological features suggesting HPV infection were identified in some samples. HPV antigens were subsequently detected in epithelial cells from five papillomas lesions, two inverted lesions and one flat lesion using immunohistochemistry. The results suggested HPV might be associated with carcinogenesis (Syrjänen *et al*., 1983). This concept is now well accepted and a growing body of evidence is supporting approximately 20% of oral cancers and 40-60% of oropharyngeal cancers are caused by HPV (Syrjänen, 2010b; Rautava and Syrjänen, 2012; Krupar *et al*., 2014).

HPV belongs to the *Papillomaviridae* family and over 150 HPV types have been identified (Burk *et al*., 2011). The DNA genome is divided into three regions, the regions encoding the early proteins (E1, E2, E4, E5, E6 and E7) and late proteins (L1 and L2) and the noncoding region, URR/LCR (Graham, 2010; Miller *et al*., 2012). The virus is mainly transmitted via sexual contact but non-sexual transmission has been recorded, including vertical, fomite, autoinoculation and horizontal transmission. HPVs cause a wide range of diseases from laryngeal papillomatosis, genital warts, cervical cancer, anogenital cancer and head and neck cancer (Smith *et al*., 2010). Currently there are three vaccines used to protect against HPV; Gardasil® protects against HPV types -6, -11, -16 and -18 is given to both females and males and it prevents genital warts, cervical, vaginal and vulva cancer in females and genital warts in males (D'Souza and Dempsey, 2011). The second vaccine is Cervarix which protects against HPV types -16 and -18 and is administered to females only (D'Souza and Dempsey, 2011). The third vaccine is Gardasil®9 which contains HPV -6, -11, -16 and -18 VLP similar to Gardasil® but in addition it contains HPV -31, -33, -45, -53 and 58 VLPs (Petrosky *et al*., 2015).

Since 1985 there have been a number of studies describing HPV and HNSCC in SA (Hille *et al*., 1984; Hille *et al*., 1986; Williamson *et al*., 1991; Togawa *et al*., 1994; Van Rensburg *et al*., 1995; Cooper *et al*., 1995; Van Rensburg *et al*., 1996; Matsha *et al*., 2002; Boy *et al*., 2006; Paquette *et al*., 2013; Davidson *et al*., 2014). Most of these studies were performed in the Northern Transvaal (now referred to as Gauteng) with a highly variable

detection rate probably related to the different techniques that were used by different groups. The early studies were based on histological examination of tissues to look for morphological evidence of HPV infection. Subsequently, molecular methods were employed particularly ISH and PCR. HPV DNA was detected in 1/66 (Van Rensburg *et al*., 1995) and 0/59 (Van Rensburg *et al*., 1996) samples using ISH and PCR respectively. In more recent studies, PCR amplified HPV DNA in 10/14 samples (Williamson *et al*., 1991); 23/50 biopsies (Matsha *et al*., 2000) and 48/51 (Paquette *et al*., 2013) tissue biopsies. Studies conducted thus far in SA used a variety of methods to screen for HPV DNA in tissue biopsies from HNSCC including PCR and *in situ* hybridization. The application of a range of tests makes it difficult to determine prevalence for HPV in SA however the outcome has indicated that ISH perhaps has a low detection rate and that PCR is a more appropriate technique.

There are a wide variety of methods currently being used worldwide for the detection of HPV; these include *in situ* hybridization, nucleic acid hybridization with signal amplification, microarray, Southern blotting, type specific PCR, real time PCR and RT-PCR (Robinson *et al*., 2010; Entiauspe *et al*., 2013). In addition to in house methods, there are commercially available assays however the commercial assays are validated for HPV detection in cervical cancers and have not been developed specifically for HNSCC. There is currently no standardised method described that can be used for the detection of HPV DNA from tissue biopsies from patients with confirmed head and neck tumours, but PCR is considered the gold standard as the DNA-target is selectively amplified (Entiauspe *et al*., 2013). A standardised method that can be used for screening of HPV in HNSCC would contribute to normalising epidemiological data and studies investigating prevalence of HPV. The aims of this study were to investigate molecular assays that could be used to screen for HPV DNA in tissue biopsies from patients with confirmed head and neck tumours in the Free State province, SA and to develop a molecular assay that could be used to detect transcriptionally active HPV in HNSCC.

Three molecular assays were compared. The first was the Roche LA that uses PGMY11/09 primers that amplify approximately 450bp of the L1 region. A second was a nested PCR that targets the L1 region of the HPV genome using two sets of published consensus primers, the MY11/09 primers which amplify a region of approximately 450bp

of the L1 gene and GP5+/6+ primers which amplify 139bp region within the MY11/09 primers. Lastly the in-house E6 multiplex hemi-nested type specific PCR that uses type specific primers for HPV types -6, -11, -16, -18, -31, -33, -45, -58 and -84 to amplify different sizes of the E6 oncogene that range from 90bp-200bp. The beta-globin gene was used as an internal control and was co-amplified with each sample to check the DNA integrity. The Roche LA assay is designed for detecting HPV associated with cervical cancers but has been used for head and neck tumours. In our facility the pair of primers targeting the L1 gene have been used routinely for detection of HPV in patients with laryngeal papillomatosis and hence it was a natural progression to try these primers for head and neck biopsies. However it was taken into consideration that the LR-HPV type is associated with recurrent laryngeal papillomas and that this study was focused more on detection of HR-HPV types. Taking this into consideration, and the fact that the L1 gene may be disrupted if integration occurs, type specific primers were designed targeting the E6 gene.

A total of 74 tissue biopsies from patients with confirmed head and neck tumours were screened for the presence of HPV DNA. All samples tested positive for the beta-globin gene. A total of 4/74 (5.4%) samples tested positive for HPV using nested PCR, with two samples showing a band at the expected size in the first round and two additional sample showed a band of expected size in the second round (nested PCR). Sequencing analysis revealed three HR-types (-16, -18 and -31) and a one LR-HPV type (-11). Due to the fact that the L1 may be disrupted during viral integration thus giving false negative results, primers targeting the E6 oncogene were designed as the oncogene is believed to be retained during viral integration (Rusan and Ovesen, 2012). The occurrence of viral integration is not clear however it was considered appropriate to compare an assay targeting a region that could be disrupted with an assay that targets a more stable region. Positive controls for designing in house assays were prepared using either known positive samples or if not available, then using synthesized genes covering the region targeted by the primers. All 74 samples that were tested using nested PCR were tested using the E6 multiplex hemi-nested type specific PCR. A total of 5/74 (6.8%) samples tested positive with sequence analysis confirmation of the HPV types. The positive samples included one of each of the following HPV types -11, -16, -18, -31 and -45. Four samples tested

positive for HR-types -16, -18, -31 and -45 in the first round and the fifth sample was positive by second round PCR (HPV type -11). The two conventional assays developed in-house (nested PCR and E6 multiplex hemi-nested type specific PCR) were compared with a commercial assay, Roche LA assay. A total of 57/74 samples were positive for HPV type -84, two samples had co-infections, HPV type -16 and -84 and HPV type -18 and -84 and one sample was HPV type -45 positive. The detection of HPV type -84 in most of the samples was unexpected. To investigate the performance of the LA and to obtain genetic sequence data for further analysis, an additional PCR was introduced based on the modifications of the MY11/09 primer pair. A number of samples were tested using modified primers PGMY11/09 which were reportedly more specific for a number of HPV types including HPV type -84 however the modified primers also failed to amplify HPV type -84 from any of the samples tested.

The Roche LA assay did not detect HPV types -11 and -31 which were detected by both the nested PCR and the E6 multiplex hemi-nested type specific PCR. Primers with more than five mismatches tend to be less efficient (Gravitt *et al*., 2000). Primer sequences used in the Roche LA assay were aligned with nucleotide sequences of HPV types -11 and -31. A total of four mismatches were observed between PGMY11/09 primers and HPV type -31 and two mismatches with HPV type -11 which does not account for the negative results. DNA integrity was confirmed using a reference hence it is difficult to determine why the LA assay did not detect HPV DNA in the samples.

The Roche LA did however detect HPV type -45 which was negative by the nested PCR. The MY11/09 and GP5+/6+ primers were originally designed to amplify HPV type -45 and mismatches were taken into account in the design (Gravitt *et al*., 2000). However there are four mismatches between HPV type -45 and the nested GP5+/6+ primers which might have contributed to a negative result using the nested PCR.

In summary the Roche LA assay detected HPV type -84 in a large number of samples and this result could not be confirmed using the nested PCR with conventional or modified primers. Additional primers also introduced to target the E6 of HPV type -84 were also negative. The inability to reproduce these results using an in house assay does perhaps suggest that the result may be erroneous, but will further be investigated before being

discarded as a non-specific. In retrospect, modified primers are likely a more suitable pair of primers and detected HPV type -45 and hence will replace the MY11/09 primer pair in a long term. The detection of HPV using molecular assays is a complex topic due to co-infections, variations of nucleotide sequences of type specific and probably also viral loads influencing which type may be amplified. The application of more than one assay is probably an option for accurate detection. The Roche LA assay is cost prohibitive for screening large numbers as well as cumbersome to perform. The results in this study will influence the assay selected for testing large number of paraffin embedded tissues for HPV DNA and hence the inclusion of an assay targeting a smaller region (<200bp) and a different gene. The in house hemi nested PCR targeting the E6 gene with type specific primers was able to detect all HPV positive samples in this study. At this stage there may be limitations in the assay as it is type specific for low risk HPV types -6, -11 and -84 and high risk types-16, -18, -31, -33, -45 and -58. With the exception of HPV type -84 results, the Roche LA assay did not indicate that any other types were missed in the study. Submission of fresh biopsies for testing will continue and the replacement of conventional MY11/09 with modified PGMY11/09 used to compliment the E6 hemi nested PCR should provide data that can be used to expand the type specific primer selection as required.

Recent epidemiological data indicate that HR-HPV types are associated with subset of HNSCC (Jung *et al*., 2010) and that HPV-positive HNSCC patients have a better prognosis than patients with HPV-negative HNSCC (Rautava *et al*., 2012). The oncogenic potential of HPV involves the expression of the E6 and E7 viral oncoproteins which disrupt the p53 and pRB signalling pathways and therefore in order to address the role of HPV in malignancies it is necessary to determine whether it is actively transcribed (Jung *et al*., 2010). The presence of HPV DNA in the tissue biopsy does not necessarily mean that they are involved in causing HNSCC or are transcriptionally active as the last requires active transcription of the E6 and E7 oncogenes (Winder *et al*., 2009; Morbini *et al*., 2015). Therefore an E6 HnRT-PCR was developed to determine whether samples that tested positive for HR-HPV types -16, -18, -31 and -45 were transcriptionally active or replicating. In the absence of HPV RNA controls and in order to adequately optimize RT-PCR, four RNA controls were prepared by *in vitro* transcription. RNA controls prepared *in vitro* from

a sequence homogenous to the target of the concerned RT-PCR indicate the quality of the whole RT-PCR and primers (Chen *et al*., 2006). PCR products derived from the E6 gene of HPV types -16, -18, -31 and -45 were cloned into for pGEM®-T easy vector to prepare template for *in vitro* transcription. Positive transformants were confirmed by restriction digestion using *Not* 1 restriction enzyme that flanks the multiple cloning site of the pGEM®-T easy vector and by plasmid DNA PCR using T7 forward primer which targets a site present on the pGEM®-T easy vector and an HPV type specific reverse primer (HPV 16R; HPV 18R HPV 31R and HPV 45R respectively) downstream of the inserted gene. Positive PCR amplicons (termed constructs) were confirmed by sequencing. DNA template for transcribing RNA was prepared by amplification of type specific constructs using the relevant forward primer for each HPV types and a reverse primer that targeted the SP6 site located on the pGEM®-T easy vector. Positive PCR amplicons were sequenced for confirmation prior to transcription of RNA. The RT-PCR assay was optimized using transcribed RNA and specific HPV primers. Four samples that tested positive for HR-HPV types were tested for E6mRNA by extracting RNA from tissue biopsies stored at -20^{0}C in RNAlater®. As HPV is a DNA virus, it is necessary to ensure that no viral DNA was present in the RNA sample, therefore each sample was checked for DNA contamination by performing a PCR using RNA instead of cDNA as a template (RT-control). All samples tested negative for DNA contamination and HPV E6mRNA thus suggesting that they are not transcriptionally active.

There are limitations in this study, including samples size and the fact that HPV is most prevalent in the oropharynx while most of the samples in this study came from sites with low HPV prevalence (oral cavity, larynx maxillary sinuses, hypopharynx and nasal cavity). Additionally integrity of the RNA was not checked however tissue biopsies were stored in RNAlater® to prevent RNA degradation. In the future studies confirming the integrity of the RNA should be considered.

As mentioned previously, our laboratory has previously used MY11/09 primer pair for detecting HPV DNA in patients with recurrent laryngeal papillomas. The method of DNA extraction from tissues and amplification used in this study was the same as that used for laryngeal papillomas (Seedat *et al*., 2010; Combrinck *et al*., 2013). Interestingly a nested PCR is not usually required for detection of LR-HPV in laryngeal papillomas (nested PCR has been performed in 6/224 samples) suggesting a difference in viral loads. We have

previously determined that the viral loads in patients with HPV type -11 infections were lower than viral load in HPV type -6 infections (personal communication, Professor R Seedat, unpublished data) and that the disease severity is higher in patients with HPV type -11. The relationship if any, between viral load for HR-types, disease severity or carcinogenesis is not known.

The study has indicated that the primer specific for hemi nested PCR targeting the E6 has application for future studies. Patients with HPV associated tumours have improved prognosis compared to patients that are HPV negative. In addition, identification of HPV types -31, -33, and -45 suggest that the polyvalent vaccine which protects against nine HR-HPV types could help to reduce the burden of disease caused by HPV types other than HPV types -16 and -18.

In conclusion the association of HPV with HNSCC is well-known in other parts of the world. In contrast in SA little is known about the association with few studies conducted and no standardised method that could be used to screen for HPV DNA in tissue biopsies from patients with confirmed HNSCC. Therefore there is an urgent need for a cost-effective method that can be used for screening of HPV in HNSCC. Three methods were investigated to find the most appropriate to be used in the future in our laboratory. The three methods showed discordant results, nested PCR detected 4/74 positive samples for HPV DNA, Roche LA detected 3/74 positive samples for HPV DNA and 5/74 sample tested positive by E6 multiplex hemi-nested type specific PCR. The E6 multiplex hemi-nested PCR was the most sensitive of the three assays and it will be used in our laboratory to screen additional samples for HPV. However there are limitations to using the assay that need to be kept in mind. There are more than 15 HR-HPV types and more than 15 LR-HPV types, while the assay is designed to detect six HR-HPV types and three LR-HPV types. It may be prudent to expand the assay to include other HPV types.

An E6 HnRT-PCR was developed to detect transcriptionally active HPV types from the four samples that tested positive for HR-HPV types -16, -18, -31 and -45 and all samples tested negative for replicating HPV E6mRNA suggesting that they are not transcriptionally active. Based on the positive results obtained in 5/74 (6.8%) samples in this study it shows that there is a low HPV prevalence in head and neck tumours in the Free State province and this study will add to the limited information available in SA. The study also highlights

the need for more studies in other provinces because to the best of our knowledge this is the first study to look for an association between HPV and HNSCC outside Gauteng.

Appendix A: Ethics approval letter

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV
Research Division
Internal Post Box G40
☎(051) 401-7795
Fax (051) 4444359

E-mail address: EthicsFHS@ufs.ac.za

Ms M Marais

2015-02-09

REC Reference nr 230408-011
IRB nr 00006240

MR TR SEKEE
DEPT OF MEDICAL MICROBIOLOGY AND VIROLOGY
FACULTY OF HEALTH SCIENCES
UFS

Dear Mr Sekee

ECUFS NR 137/2013B
PROJECT TITLE: MOLECULAR ASSAYS FOR DETECTING HUMAN PAPILLOMAVIRUS IN HEAD AND NECK SQUAMOUS CELL CARCINOMA

1. You are hereby kindly informed that the study was approved at the Ethics Committee meeting held on 20 January 2015.

   - ***The above study is a sub-project and forms part of the ECUFS 137/2013A study***

2. Committee guidance documents: Declaration of Helsinki, ICH, GCP and MRC Guidelines on Bio Medical Research. Clinical Trial Guidelines 2000 Department of Health RSA; Ethics in Health Research: Principles Structure and Processes Department of Health RSA 2004; Guidelines for Good Practice in the Conduct of Clinical Trials with Human Participants in South Africa, Second Edition (2006); the Constitution of the Ethics Committee of the Faculty of Health Sciences and the Guidelines of the SA Medicines Control Council as well as Laws and Regulations with regard to the Control of Medicines.

3. Any amendment, extension or other modifications to the protocol must be submitted to the Ethics Committee for approval.

4. The Committee must be informed of any serious adverse event and/or termination of the study.

5. All relevant documents e.g. signed permission letters from the authorities, institutions, changes to the protocol, questionnaires etc. have to be submitted to the Ethics Committee before the study may be conducted (if applicable).

6. A progress report should be submitted within one year of approval of long term studies and a final report at completion of both short term and long term studies.

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

UFS
UV

7. Kindly refer to the ECUFS reference number in correspondence to the Ethics Committee secretariat.

Yours faithfully

DR SM LE GRANGE
CHAIR: ETHICS COMMITTEE

Cc Prof RY Saadat
Prof FJ Burt
Dr D Goedhals

Appendix B: Details of isolates for which sequence data which was retrieved from GenBank and used for designing primers for E6 multiplex hemi-nested type specific PCR. For HPV 84 there were no 15 isolates, therefore E6 region was retrieved from the complete sequence.

Isolate	Country	GenBank Accession number
• **HPV 6 E6 region**		
LP 18	Brazil	KC 285855
LP 6	Slovenia	FM 897019.1
LP 83	Slovenia	FM 897019.1
LP 16	Brazil	KC 285853.1
AC 93	Australia	KC 333888.1
GW 6788B	Australia	KC 300188.1
VBD 44/08	South Africa	JN 573163.1
VDB 46/08	South Africa	JN 573164.1
VBD 04/09	South Africa	JN 573168.1
VBD 07/09	South Africa	JN 573171.1
VBD 09/09	South Africa	JN 573165.1
VBD 12/09	South Africa	JN 573166.1
VBD 77/09	South Africa	JN 573169.1
VBD 80/09	South Africa	JN 573170.1
VBD 02/10	South Africa	JN 573167.1
• **HPV 11 E6 region**		
LP 19	Brazil	KC 285862.1
LP 20	Brazil	KC 285860.1
LP 22	Brazil	KC 285857.1
LP 23	Brazil	KC 285858.1
LP 25	Brazil	KC 285862.1
RRP 63871	Australia	KC 329891.1
RRP 76510	Australia	KC 329893.1
RRP 57861	Australia	KC 329889.1
GW 9399A	Australia	KC 329884.1
RRP 75380	Australia	KC 329892.1
M3	Slovenia	FN 870435.1
A50	Slovenia	FN 870436.1
A57	Slovenia	FN 870437.1
A32	Slovenia	FN 870438.1
A51	Slovenia	FN 870439.1
• **HPV 16 E6 protein**		
IR-32	Iran	KM 058604.1
IR-0	Iran	KM 058574.1

Appendix B continues

NA 1 (PWH-Q39)	China	AF 486325.1
CG-1	China	JX 412112.1
JO7-130	Japan	AB 663707.1
JO6-008	Vietnam, Japan and Philliphens	AB 66388.1
NA 1-Z6 (2)	Morocco	KC 904935.1
NA 1-Z6	Morocco	KC 904934.1
NA 1-b/r (4)	Morocco	KC 904926.1
NA 1-d/G350	Morocco	KC 904928.1
NA 1	Morocco	KC 904911.1
AF2-Z2	Morocco	KC 904897.1
AF 1	Morocco	KC 904870.1
E-C442/G350 (41)	Morocco	KC 904866.1
E-T350/r	Morocco	KC 904798.1
• **HPV 18 E6 region**		
Bsb-27	Brazil	EF 661654.1
Bsb-48	Brazil	EF 661655.1
Bsb-82	Brazil	EF 661656.1
Bsb-206	Brazil	EF 661657.1
WZ 82	China	JX 412196.1
WZ 81	China	JX 412195.1
PE 23	China	JX 412188.1
P299	Italy	EF422109.1
PE 56	China	JX 412193.1
PE 26	China	JX 412190.1
375	Spain	KF 700152.1
289	Spain	KC 700154.1
123	Spain	KC 700141.1
231	Spain	KC 700145.1
211	Spain	KC 700144.1
• **HPV 31 E6 region**		
B 848	Greece	KC662562.1
B 7857	Greece	KC 662561.1
B 5057	Greece	KC 662560.1
B 6394	Greece	KC 662559.1
118	Spain	KF 700160.1
7	Spain	KF 700159.1
12	Spain	KF 700158.1
73	Spain	KF700157.1
BR 1692	USA	DQ507308.1
MR 9917	USA	DQ507309.1
TL 2069	USA	DQ507306.1
BR1213	USA	DQ057305.1
9	Canada	GQ 369946.1
3	Canada	GQ 369943.1

Appendix B continues

8	Canada	GQ 369945.1
• **HPV 33 E6 region**		
2	Canada	GQ 374537.1
1	Canada	GQ 374549.1
3	Canada	GQ 374550.1
291	Spain	KF 700165.1
149	Spain	KC 700161.1
14	Spain	KC 700162.1
346	Spain	KC 700163.1
100	Spain	KF 700164.1
B8454	Greece	KC 662567.1
B2659	Greece	KC 662566.1
B597	Greece	KC 662565.1
B4083	Greece	KC 662564.1
B3200	Greece	KC 662563.1
P396	Italy	EF 422127.1
P232	Italy	EF 422126.1
• **HPV 45 E6 region**		
B3587	Greece	KC 662573.1
B3730	Greece	KC 662572.1
B8463	Greece	KC 662571.1
B2408	Greece	KC 662570.1
92.070	Algeria	JQ976848.1
91.971	Algeria	JQ976847.1
91.841	Algeria	JQ 976846.1
91.811	Algeria	JQ 976845.1
91.471	Algeria	JQ 976844.1
91.401	Algeria	JQ976843.1
91.631	Algeria	JQ 976842.1
90.391	Algeria	JQ 976841.1
90.331	Algeria	JQ 976840.1

Appendix B continues

82.091	India	JQ 9766839.1
71.751	Peru	JQ 976833.1
• **HPV 58 E6 region**		
USA_990989	USA	KC 190291.1
THA_269	Thailand	KC 190290.1
ITA_52PA	Italy	KC 190288.1
ZWE_064436	Zimbabwe	KC 190288.1
ZWE_050364	Zimbabwe	KC 190279.1
ZWE_044103	Zimbabwe	KC 190278.1
ZWE_063739	Zimbabwe	KC 190274.1
ZJ 11	China	HM 004108.1
ZJ 8	China	HM 004107.1
ZJ 4	China	HM 004106.1
ZJ 14	China	HM 004105.1
JS049297	China	EU 999968.1
CAN 420	Canada	KC 190286.1
HK 262	Hong Kong	KC 190285.1
ARG_P2005	Argentina	KC 190283.1
• **HPV 84 E6 region**		
N/A	USA	AF293960.1

Appendix C: Alignment of partial sequence data for E6 region used for design of each primer pair used in the E6 multiplex hem-nested type specific PCR. Sequence data was aligned using Clustal Omega version 1.2.1. Primers are highlighted as follows: red-forward primer used in the first round, purple-forward primer used in a hemi-nested PCR, turquoise-reverse primer. All nucleotide sequences are in a 5' to 3' direction and conserved regions are shown by an asterix. Refer to appendix B for accession numbers. Nt-nucleotide

1. HPV 06 E6 region-450bp

Forward primer used in the first round (nt position 17-37)

Forward primer used in the second round (nt position 101-124)

Reverse primer (nt position 170-190)

```
Isolates        ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
VBD 44/08       ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
VBD 46/08       ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
VBD 80/09       ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
VBD 77/09       ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
VBD 04/09       ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
VBD 02/10       ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
VBD 12/10       ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
VBD 09/09       ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
LP 83           ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
LP 16           ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
VBD 07/09       ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
AC 93           ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
LP 6            ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
GW 6788B        ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
LP 8            ATGGAAAGTG CAAATGCCTC CACGTCTGCA ACGACCATAG ACCAGTTGTG
                ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        60         70         80         90        100
VBD 44/08       CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
VBD 46/08       CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
VBD 80/09       CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
VBD 77/09       CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
VBD 04/09       CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
VBD 02/10       CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
VBD 12/10       CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
VBD 09/09       CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
LP 83           CAAGTCGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
LP 16           CAAGTCGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
VBD 07/09       CAAGTCGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
AC 93           CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
LP 6            CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
GW 6788B        CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
```

```
LP 8             CAAGACGTTT AATCTATCTA TGCATACGTT GCAAATTAAT TGTGTGTTTT
                 **** ***** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         110        120        130        140        150
VBD 44/08        GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
VBD 46/08        GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
VBD 80/09        GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
VBD 77/09        GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
VBD 04/09        GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
VBD 02/10        GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
VBD 12/10        GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
VBD 09/09        GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
LP 83            GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
LP 16            GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
VBD 07/09        GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
AC 93            GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
LP 6             GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
GW 6788B         GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
LP 8             GCAAGAATGC ACTGACCACT GCAGAGATTT ATTCATATGC ATATAAACAG
                 ********** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         160        170        180        190        200
VBD 44/08        CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
VBD 46/08        CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
VBD 80/09        CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
VBD 77/09        CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
VBD 04/09        CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
VBD 02/10        CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
VBD 12/09        CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
VBD 09/09        CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
LP 83            CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
LP 16            CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
VBD 07/09        CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
AC 93            CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
LP 6             CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
GW 6788B         CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
LP 8             CTAAAGGTCC TGTTTCGAGG CGGCTATCCA TATGCAGCCT GCGCGTGCTG
                 ********** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         210        220        230        240        250
VBD 44/08        CCTAGAATTT CATGGAAAAA TCAACCAATA TAGACACTTT GATTATGCTG
VBD 46/08        CCTAGAATTT CATGGAAAAA TCAACCAATA TAGACACTTT GATTATGCTG
VBD 80/09        CCTAGAATTT CATGGAAAAA TCAACCAATA TAGACACTTT GATTATGCTG
VBD 77/09        CCTAGAATTT CATGGAAAAA TCAACCAATA TAGACACTTT GATTATGCTG
VBD 04/09        CCTAGAATTT CATGGAAAAA TCAACCAATA TAGACACTTT GATTATGCTG
VBD 02/10        CCTAGAATTT CATGGAAAAA TCAACCAATA TAGACACTTT GATTATGCTG
VBD 12/09        CCTAGAATTT CATGGAAAAA TCAACCAATA TAGACACTTT GATTATGCTG
VBD 09/09        CCTAGAATTT CATGGAAAAA TCAACCAATA TAGACACTTT GATTATGCTG
LP 83            CCTAGAATTT CATGGAAAAA TCAACCAATA TAGACACTTT GATTATGCTG
LP 16            CCTAGAATTT CATGGAAAAA TCAACCAATA TAGACACTTT GATTATGCTG
VBD 07/09        CCTAGAATTT CATGGAAAAA TCAACCAATA TAGACACTTT GATTATGCTG
AC 93            CCTAGAATTT CATGGAAAAA TTAACCAATA TAGACACTTT GATTATGCTG
LP 6             CCTAGAATTT CATGGAAAAA TTAACCAATA TAGACACTTT GATTATGCTG
GW 6788B         CCTAGAATTT CATGGAAAAA TTAACCAATA TAGACACTTT GATTATGCTG
LP 8             CCTAGAATTT CATGGAAAAA TTAACCAATA TAGACACTTT GATTATGCTG
                 ********** ********** * ******** ********** **********
```

```
                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            260        270        280        290        300
VBD 44/08           GATATGCAAC AACTGTCGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
VBD 46/08           GATATGCAAC AACTGTCGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
VBD 80/09           GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
VBD 77/09           GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
VBD 04/09           GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
VBD 02/10           GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
VBD 12/09           GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
VBD 09/09           GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
LP 83               GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
LP 16               GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
VBD 07/09           GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
AC 93               GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
LP 6                GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
GW 6788B            GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGACGTG
LP 8                GATATGCAAC AACTGTTGAA GAAGAAACTA AACAAGACAT TTTAGGCGTG
                    ********** ********** ********** ********** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            310        320        330        340        350
VBD 44/08           CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
VBD 46/08           CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
VBD 80/09           CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
VBD 77/09           CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
VBD 04/09           CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
VBD 02/10           CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
VBD 12/09           CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
VBD 09/09           CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
LP 83               CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
LP 16               CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
VBD 07/09           CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
AC 93               CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
LP 6                CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
GW 6788B            CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
LP 8                CTAATTCGGT GCTACCTGTG TCACAAACCG CTGTGTGAAG TAGAAAAGGT
                    ********** ********** ********** ********** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            360        370        380        390        400
VBD 44/08           AAAACATATA CTAACCAAGG CACGGTTTAT AAAGCTAAAT TGTACGTGGA
VBD 46/08           AAAACATATA CTAACCAAGG CACGGTTTAT AAAGCTAAAT TGTACGTGGA
VBD 80/09           AAAACATATA CTAACCAAGG CACGGTTTAT AAAGCTAAAT TGTACGTGGA
VBD 77/09           AAAACATATA CTAACCAAGG CACGGTTTAT AAAGCTAAAT TGTACGTGGA
VBD 04/09           AAAACATATA CTAACCAAGG CACGGTTTAT AAAGCTAAAT TGTACGTGGA
VBD 02/10           AAAACATATA CTAACCAAGG CACGGTTTAT AAAGCTAAAT TGTACGTGGA
VBD 12/09           AAAACATATA CTAACCAAGG CACGGTTTAT AAAGCTAAAT TGTACGTGGA
VBD 09/09           AAAACATATA CTAACCAAGG CACGGTTTAT AAAGCTAAAT TGTACGTGGA
LP 83               AAAACATATA CTAACCAAGG CACGGTTCAT AAAGCTAAAT TGTACGTGGA
LP 16               AAAACATATA CTAACCAAGG CACGGTTCAT AAAGCTAAAT TGTACGTGGA
VBD 07/09           AAAACATATA CTAACCAAGG CACGGTTCAT AAAGCTAAAT TGTACGTGGA
AC 93               AAAACATATA CTAACCAAGG CACGGTTCAT AAAGCTAAAT TGTACGTGGA
LP 6                AAAACATATA CTAACCAAGG CACGGTTCAT AAAGCTAAAT TGTACGTGGA
GW 6788B            AAAACATATA CTAACCAAGG CACGGTTCAT AAAGCTAAAT TGTACGTGGA
LP 8                AAAACATATA CTAACCAAGG CACGGTTCAT AAAGCTAAAT TGTACGTGGA
                    *********  ********** ******* ** ********** **********
```

```
                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            410        420        430        440        450
VBD 44/08           AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
VBD 46/08           AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
VBD 80/09           AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
VBD 77/09           AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
VBD 04/09           AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
VBD 02/10           AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
VBD 12/09           AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
VBD 09/09           AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
LP 83               AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
LP 16               AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
VBD 07/09           AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
AC 93               AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
LP 6                AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
GW 6788B            AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
LP 8                AGGGTCGCTG CCTACACTGC TGGACAACAT GCATGGAAGA CATGTTACCC
                    ********** ********** ********** ********** **********

                    ...

VBD 44/08           TAA
VBD 46/08           TAA
VBD 80/09           TAA
VBD 77/09           TAA
VBD 04/09           TAA
VBD 02/10           TAA
VBD 12/09           TAA
VBD 09/09           TAA
LP 83               TAA
LP 16               TAA
VBD 07/09           TAA
AC 93               TAA
LP 6                TAA
GW 6788B            TAA
LP 8                TAA
                    ***
```

2. HPV 11 E6 region-453bp

Forward primer used in the first round (nt position 1-25)

Forward primer used in the second round (nt position 64-90)

Reverse primer (nt position 180-200)

```
Isolates            ....|....| ....|....| ....|....| ....|....| ....|....|
                            10         20         30         40         50
RRP 57861           ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCCATAG ACCAGTTGTG
GW 9399A            ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCCATAG ACCAGTTGTG
M 3                 ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCCATAG ACCAGTTGTG
A 50                ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCCATAG ACCAGTTGTG
LP 19               ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCCATAG ACCAGTTGTG
RRP 76510           ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCCATAG ACCAGTTGTG
RRP 75380           ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCCATAG ACCAGTTGTG
GW 63871            ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCCATAG ACCAGTTGTG
A 57                ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCCATAG ACCAGTTGTG
A 32                ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCCATAG ACCAGTTGTG
A 51                ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCCATAG ACCAGTTGTG
```

```
LP 25            ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCTATAG ACCAGTTGTG
LP 23            ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCTATAG ACCAGTTGTG
LP 22            ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCTATAG ACCAGTTGTG
LP 20            ATGGAAAGTA AAGATGCCTC CACGTCTGCA ACATCTATAG ACCAGTTGTG
                 ********** ********** ********** ***** **** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                          60         70         80         90        100
RRP 57861        CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
GW 9399A         CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
M 3              CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
A 50             CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
LP 19            CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
RRP 76510        CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
RRP 75380        CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
GW 63871         CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
A 57             CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
A 32             CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
A 51             CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
LP 25            CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
LP 23            CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
LP 22            CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
LP 20            CAAGACGTTT AATCTTTCTT TGCACACTCT GCAAATTCAG TGCGTGTTTT
                 ********** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         110        120        130        140        150
RRP 57861        GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
GW 9399A         GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
M 3              GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
A 50             GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
LP 19            GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
RRP 76510        GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
RRP 75380        GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
GW 63871         GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
A 57             GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
A 32             GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
A 51             GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
LP 25            GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
LP 23            GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
LP 22            GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
LP 20            GCAGGAATGC ACTGACCACC GCAGAGATAT ATGCATATGC CTATAAGAAC
                 ********** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         160        170        180        190        200
RRP 57861        CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
GW 9399A         CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
M3               CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
A 50             CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
LP 19            CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
RRP 76510        CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
RRP 75380        CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
GW 63871         CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
A 57             CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
A 32             CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
A 51             CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
LP 25            CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
LP 23            CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
LP 22            CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
```

```
LP 20            CTAAAGGTTG TGTGGCGAGA CAACTTTCCC TTTGCAGCGT GTGCCTGTTG
                 ********** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         210        220        230        240        250
RRP 57861        CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
GW 9399A         CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
M 3              CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
A 50             CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
LP 19            CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
RRP 76510        CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
RRP 75380        CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
GW 63871         CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
A 57             CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
A 32             CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
A 51             CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
LP 25            CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
LP 23            CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
LP 22            CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
LP 20            CTTAGAACTG CAAGGGAAAA TTAACCAATA TAGACACTTT AATTATGCTG
                 ********** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         260        270        280        290        300
RRP 57861         CATATGCACC TACAGTAGAA GAAGAAACTA ATGAAGATAT TTTAAAAGTG
GW9399B           CATATGCACC TACAGTAGAA GAAGAAACTA ATGAAGATAT TTTAAAAGTG
M 3               CATATGCACC TACAGTAGAA GAAGAAACTA ATGAAGATAT TTTAAAAGTG
A 50              CATATGCACC TACAGTAGAA GAAGAAACTA ATGAAGATAT TTTAAAAGTG
LP 19             CATATGCACC TACAGTAGAA GAAGAAACTA ATGAAGATAT TTTAAAAGTG
RRP 76510         CATATGCACC TACAGTAGAA GAAGAAACTA ATGAAGATAT TTTAAAAGTG
RRP 75380         CATATGCACC TACAGTAGAA GAAGAAACTA ATGAAGATAT TTTAAAAGTG
GW 63871          CATATGCACC TACAGTAGAA GAAGAAACTA ATGAAGATAT TTTAAAAGTG
A 57              CATATGCACC TACAGTAGAA GAAGAAACTA ATGAAGATAT TTTAAAAGTG
A 32              CATATGCACC TACAGTAGAA GAAGAAACTA ATGAAGATAT TTTAAAAGTG
A 51              CATATGCACC TACAGTAGAA GAAGAAACTA ATGAAGATAT TTTAAAAGTG
LP 25             CATATGCACC TACAGTAGAA GAAGAAACCA ATGAAGATAT TTTAAAAGTG
LP 23             CATATGCACC TACAGTAGAA GAAGAAACCA ATGAAGATAT TTTAAAAGTG
LP 22             CATATGCACC TACAGTAGAA GAAGAAACCA ATGAAGATAT TTTAAAAGTG
LP 20             CATATGCACC TACAGTAGAA GAAGAAACCA ATGAAGATAT TTTAAAAGTG
                  ********** ********** ******** * ********** **********

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                          310        320        330        340
350
RRP 57861         TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
GW 9399A          TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
M 3               TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
A 50              TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
LP 19             TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
RRP 76510         TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
RRP 75380         TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
GW 63871          TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
A 57              TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
A 32              TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
A 51              TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
LP 25             TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
LP 23             TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
LP 22             TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
LP 20             TTAATTCGTT GTTACCTGTG TCACAAGCCG TTGTGTGAAA TAGAAAAACT
                  ********** ********** ********** ********** **********
```

```
                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            360        370        380        390
400
RRP 57861           AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
GW 9399A            AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
M 3                 AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
A 50                AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
LP 19               AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
RRP 76510           AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
RRP 75380           AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
GW 63871            AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
A 57                AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
A 32                AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
A 51                AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
LP 25               AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
LP 23               AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
LP 22               AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
LP 20               AAAGCACATA TTGGGAAAGG CACGCTTCAT AAAACTAAAT AACCAGTGGA
                    ********** ********** ********** ********** **********


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            410        420        430        440
450
RRP 57861           AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
GW 9399A            AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
M 3                 AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
A 50                AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
LP 19               AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
RRP 76510           AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
RRP 75380           AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
GW 63871            AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
A 57                AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
A 32                AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
A 51                AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
LP 25               AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
LP 23               AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
LP 22               AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGAAGA CTTGTTACCC
LP 20               AGGGTCGTTG CTTACACTGC TGGACAACAT GCATGGGAGA CTTGTTACCC
                    ********** ********** ********** ****** *** **********

                    ...

RRP 57861           TAA
GW 9399A            TAA
M 3                 TAA
A 50                TAA
LP 19               TAA
RRP 76510           TAA
RRP 75380           TAA
GW 63871            TAA
A 57                TAA
A 32                TAA
A 51                TAA
LP 25               TAA
LP 23               TAA
LP 22               TAA
LP 20               TAA
                    ***
```

3. HPV 16 E6 region-477bp

Forward primer used in the first round (nt position 29-46)

Forward primer used in the second round (nt position 58-82)

Reverse primer (nt position 161-175)

```
Isolates          ....|....| ....|....| ....|....| ....|....| ....|....|
                           10         20         30         40         50
NA1               ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
EC-C442/G350(41)  ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
IR 0              ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
CG-1              ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
JO7-130           ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
IR-32             ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
JO6-008           ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
ET350/r           ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
AF2-Z2            ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
AF1               ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
NA1-z6            ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
NA1(PWH-Q39)      ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
NA1-Z(2)          ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
NA1-b/r (4)       ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
NA1-d/G350        ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG AGCGACCCAG
                  ********** ********** ********** ********** **********

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                           60         70         80         90        100
NA1               AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
EC-C442 G350(41)  AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
IR-0              AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
CG-1              AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA CATGAGATAA
JO7-130           AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA CATGAGATAA
IR-32             AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
JO6-008           AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
ET 350/r          AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
AF2-Z2            AAAGTTACCA GATTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
AF1               AAAGTTACCA GATTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
Na 1-z6           AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
NA 1 (PWH-Q39)    AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
NA1-Z(2)          AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
NA1-b/r (4)       AAAGTTACCA CATTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
NA1-d/G350        AAAGTTACCA CATTTATGCA CAGAGCTGCA AACAACTATA CATGATATAA
                  **********  * ******* ********** ********** ***** ****

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                          110        120        130        140        150
NA1               TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
EC-C442/G350 (40) TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
IR-0              TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
CG-1              TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
JO7-130           TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
IR-32             TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
JO6-008           TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
ET350/r           TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
AF2-Z2            TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
```

```
AF1                 TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
NA 1-z6             TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
NA 1 (PHW-Q39)      TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
Na-Z (2)            TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
Na1-b/r (4)         TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
Na1-d/G350          TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG TGAGGTATAT
                    ********** ********** ********** ********** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                          160        170        180        190        200
NA1                 GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
EC-C442/G350 (40)   GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
IR-0                GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
CG-1                GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
JO7-130             GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
IR-32               GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
JO6-008             GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
ET350/r             GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
AF2-Z2              GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
AF1                 GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
NA 1-z6             GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
NA 1 (PHW-Q39)      GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
Na-Z (2)            GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
Na 1-b/r (4)        GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
Na 1-d/G350         GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG GGAATCCATA
                    ********** ********** ********** ********** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                          210        220        230        240        250
NA1                 TGCAGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
EC-C442/G350 (40)   TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
IR-0                TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
CG-1                TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
JO7-130             TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
IR-32               TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
JO6-008             TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
E-T350/r            TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
AF2-Z2              TGGAGTGTGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
AF1                 TGCAGTGTGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
Na 1-z6             TGCAGTGTGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
NA (PHW-Q39)        TGCAGTGTGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
Na-Z (2)            TGCAGTGTGT GAGAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
Na 1-b/r (4)        TGCAGTGTGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
Na 1-d/G350         TGCAGTGTGT GATAAATGTT TAAAGTTTTA TTCTAAAATT AGTGAGTATA
                    **  ** *** ** ******* ********** ********** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                          260        270        280        290        300
NA1                 GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA GCAATACAAC
EC-C442/G350 (40)   GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA GCAATACAAC
IR-0                GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA GCAATACAAC
CG-1                GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA GCAATACAAC
JO7-130             GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA GCAATACAAC
IR-32               GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA GCAATACAAC
JO6-008             GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA GCAATACAAC
ET-350/r            GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA GCAATACAAC
AF2-Z2              GATATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA GCAATACAAC
AF1                 GATATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA GCAATACAAC
NA 1-z6             GATATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA GCAATACAAC
NA (PHW-Q39)        GATATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA GCAATACAAC
Na-Z (2)            GATATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA GCAATACAAC
```

```
Na 1-b/r (4)        GATATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA GCAATACAAC
Na 1-d/G350         GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA GCAATACAAC
                    ** ******* ******* ** ********** ********** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           310        320        330        340        350
NA1                 AAACCGTTGT GTGATTTGTT GATTAGGTGT ATTAACTGTC AAAAGCCACT
EC-C442/G350 (40)   AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
IR-0                AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
CG-1                AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
JO7-130             AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
IR-32               AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
JO6-008             AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
ET-350/r            AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
AF2-Z2              AAACCGTTGT GTGATTTGTT GATTAGGTGT ATTAACTGTC AAAAGCCACT
AF1                 AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
Na 1-z6             AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
NA (PHW-Q39)        AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
Na-Z (2)            AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
Na 1-b/r (4)        AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
Na 1-d/G350         AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC AAAAGCCACT
                    ********** **********  ********* ********** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           360        370        380        390        400
NA1                 GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
EC-C442/G350 (40)   GTGTCCTGAC GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
IR-0                GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
CG-1                GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
JO7-130             GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
IR-32               GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
JO6-008             GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
E-T350/r            GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
AF2-Z2              GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
AF1                 GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
Na 1-z6             GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
NA (PHW-Q39)        GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
Na 1-Z (2)          GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
Na 1-b/r (4)        GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
Na 1-d/G350         GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA AGATTCCATA
                    *********  ********** ********** ********** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           410        420        430        440        450
NA1                 ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
EC-C442/G350 (40)   ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
IR-0                ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
CG-1                ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
JO7-130             ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
IR-32               ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
JO6-008             ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
E-T350/r            ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
AF2-Z2              ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
AF1                 ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
Na 1-z6             ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
NA (PHW-Q39)        ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
Na 1-Z (2)          ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
Na 1-b/r (4)        ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
Na 1-d/G350         ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG CAGATCATCA
                    ********** ********** ********** ********** **********
```

```
                        ....|....| ....|....| ....|..
                               460        470
NA1                     AGAACACGTA GAGAAACCCA GCTGTAA
EC-C442/G350 (40)       AGAACACGTA GAGAAACCCA GCTGTAA
IR-0                    AGAACACGTA GAGAAACCCA GCTGTAA
CG-1                    AGAACACGTA GAGAAACCCA GCTGTAA
JO7-130                 AGAACACGTA GAGAAACCCA GCTGTAA
IR-32                   AGAACACGTA GAGAAACCCA GCTGTAA
JO6-008                 AGAACACGTA GAGAAACCCA GCTGTAA
E-T350/r                AGAACACGTA GAGAAACCCA GCTGTAA
AF2-Z2                  AGAACACGTA GAGAAACCCA GCTGTAA
AF1                     AGAACACGTA GAGAAACCCA GCTGTAA
Na 1-z6                 AGAACACGTA GAGAAACCCA GCTGTAA
Na (PHW-Q39)            AGAACACGTA GAGAAACCCA GCTGTAA
Na 1-Z (2)              AGAACACGTA GAGAAACCCA GCTGTAA
Na 1-b/r (4)            AGAACACGTA GAGAAACCCA GCTGTAA
Na 1-d/G350             AGAACACGTA GAGAAACCCA GCTGTAA
                        ********** ********** *******
```

4. HPV 18 E6 region-477bp

Forward primer used in the first round (nt position 1-20)

Forward primer used in the second round (nt position 51-74)

Reverse primer (nt position 172-191)

```
Isolates                ....|....| ....|....| ....|....| ....|....| ....|....|
                               10         20         30         40         50
289                     ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
Bsb-82                  ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
WZ 81                   ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
PE 23                   ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
P299                    ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
PE 26                   ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
PE 56                   ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
Bsb-27                  ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
Bsb-48                  ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
WZ 82                   ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
123                     ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
231                     ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
Bsb-206                 ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
375                     ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
211                     ATGGCGCGCT TTGAGGATCC AACACGGCGA CCCTACAAGC TACCTGATCT
                        ********** ********** ********** ********** **********

                        ....|....| ....|....| ....|....| ....|....| ....|....|
                               60         70         80         90        100
289                     GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
Bsb-82                  GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
WZ 81                   GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
PE 23                   GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
P299                    GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
PE 26                   GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
PE 56                   GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
```

```
Bsb-27          GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
Bsb-48          GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
WZ 82           GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
123             GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
231             GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
Bsb-206         GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
375             GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
211             GTGCACGGAA CTGAACACTT CACTGCAAGA CATAGAAATA ACCTGTGTAT
                ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       110        120        130        140        150
289             ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTCAAA
Bsb-82          ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTCAAA
WZ 81           ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
PE 23           ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
P299            ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
PE 26           ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
PE 56           ATTGCAAGAC AGTTTTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
Bsb-27          ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
Bsb-48          ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
WZ 82           ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
123             ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
231             ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
Bsb-206         ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
375             ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
211             ATTGCAAGAC AGTATTGGAA CTTACAGAGG TATTTGAATT TGCATTTAAA
                ********** *** ****** ********** ********** ****** ***

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
289             GATTTATTTG TAGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
Bsb-82          GATTTATTTG TAGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
WZ 81           GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
PE 23           GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
P299            GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
PE 26           GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
PE 56           GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
Bsb-27          GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
Bsb-48          GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
WZ 82           GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
123             GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
231             GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
Bsb-206         GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
375             GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
211             GATTTATTTG TGGTGTATAG AGACAGTATA CCGCATGCTG CATGCCATAA
                ********** * ******** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
289             ATGTATAGAT TTCTATTCTA GAATTAGAGA ATTAAGATAT TATTCAGACT
Bsb-82          ATGTATAGAT TTCTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
WZ 81           ATGTATAGAT TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
PE 23           ATGTATAGAT TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
P299            ATGTATAGAT TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
PE 26           ATGTATAGAT TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
PE 56           ATGTATAGAT TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
Bsb-27          ATGTATAGAT TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
Bsb-48          ATGTATAGAT TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
WZ 82           ATGTATAGAT TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
123             ATGTATAGAT TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
```

```
231             ATGTATAGAT TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
Bsb-206         ATGTATAGAT TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
375             ATGTATAGAG TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
211             ATGTATAGAG TTTTATTCTA GAATTAGAGA ATTAAGACAT TATTCAGACT
                ********* ** ******* ********** ******* ** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      260        270        280        290        300
289             CTGTGTATGG AGACACATTA GAAAAACTAA CTAACACTGG GTTATACAAT
Bsb-82          CTGTGTATGG AGACACATTA GAAAAACTAA CTAACACTGG GTTATACAAT
WZ 81           CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
PE 23           CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
P299            CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
PE 26           CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
PE 56           CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
Bsb-27          CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
Bsb-48          CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
WZ 82           CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
123             CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
231             CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
Bsb-206         CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
375             CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
211             CTGTGTATGG AGACACATTG GAAAAACTAA CTAACACTGG GTTATACAAT
                ********** *********  ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      310        320        330        340        350
289             TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
Bsb-82          TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
WZ 81           TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
PE 23           TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
P299            TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
PE 26           TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
PE 56           TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
Bsb-27          TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
Bsb-48          TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
WZ 82           TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
123             TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
231             TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
Bsb-206         TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
375             TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
211             TTATTAATAA GGTGCCTGCG GTGCCAGAAA CCGTTGAATC CAGCAGAAAA
                ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      360        370        380        390        400
289             ACTTAGACAC CTTAATGAAA AACGACGATT CCACAAAATA GCTGGGCACT
Bsb-82          ACTTAGACAC CTTAATGAAA AACGACGATT CCACAAAATA GCTGGGCACT
WZ 81           ACTTAGACAC CTTAATGAAA AACGACGATT TCACAACATA GCTGGGCACT
PE 23           ACTTAGACAC CTTAATGAAA AACGACGATT TCACAACATA GCTGGGCACT
P299            ACTTAGACAC CTTAATGAAA AACGACGATT TCACAACATA GCTGGGCACT
PE 26           ACTTAGACAC CTTAATGAAA AACGACGATT TCACAACATA GCTGGGCACT
PE 56           ACTTAGACAC CTTAATGAAA AACGACGATT CCACAACATA GCTGGGCACT
Bsb-27          ACTTAGACAC CTTAATGAAA AACGACGATT CCACAACATA GCTGGGCACT
Bsb-48          ACTTAGACAC CTTAATGAAA AACGACGATT CCACAACATA GCTGGGCACT
WZ 82           ACTTAGACAC CTTAATGAAA AACGACGATT CCACAACATA GCTGGGCACT
123             ACTTAGACAC CTTAATGAAA AACGACGATT CCACAACATA GCTGGGCACT
231             ACTTAGACAC CTTAATGAAA AACGACGATT CCACAACATA GCTGGGCACT
Bsb-206         ACTTAGACAC CTTAATGAAA AACGACGATT CCACAACATA GCTGGGCACT
375             ACTTAGACAC CTTAATGAAA AACGACGATT CCACAACATA GCTGGGCACT
211             ACTTAGACAC CTTAATGAAA AACGACGATT CCACAACATA GCTGGGCACT
```

```
                  ********** ********** **********  ***** *** **********

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                         410        420        430        440        450
289               ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAGAGACTC
Bsb-82            ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAGAGACTC
WZ 81             ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAACGACTC
PE 23             ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAACGACTC
P299              ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAACGACTC
PE 26             ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAACGACTC
PE 56             ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAAAGACTC
Bsb-27            ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAAAGACTC
Bsb-48            ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAAAGACTC
WZ 82             ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAAAGACTC
123               ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAAAGACTC
231               ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAAAGACTC
Bsb-206           ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAAAGACTC
375               ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAAAGACTC
211               ATAGAGGCCA GTGCCATTCG TGCTGCAACC GAGCACGACA GGAAAGACTC
                  ********** ********** ********** ********** ***  *****

                  ....|....| ....|....| ....|..
                         460        470
289               CAACGACGCA GAGAAACACA AGTATAA
Bsb-82            CAACGACGCA GAGAAACACA AGTATAA
WZ 81             CAACGACGCA GAGAAACACA AGTATAA
PE 23             CAACGACGCA GAGAAACACA AGTATAA
P299              CAACGACGCA GAGAAACACA AGTATAA
PE 26             CAACGACGCA GAGAAACACA AGTATAA
PE 56             CAACGACGCA GAGAAACACA AGTATAA
Bsb-27            CAACGACGCA GAGAAACACA AGTATAA
Bsb-48            CAACGACGCA GAGAAACACA AGTATAA
WZ 82             CAACGACGCA GAGAAACACA AGTATAA
123               CAACGACGCA GAGAAACACA AGTATAA
231               CAACGACGCA GAGAAACACA AGTATAA
Bsb-206           CAACGACGCA GAGAAACACA AGTATAA
375               CAACGACGCA GAGAAACACA AGTATAA
211               CAACGACGCA GAGAAACACA AGTATAA
                  ********** ********** *******
```

5. HPV 31 E6 region-452bp

Forward primer used in the first round (nt position 50-70)

Forward primer used in the second round (nt position 95-119)

Reverse primer (nt position 161-175)

```
Isolates          ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
B6394             ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
73                ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
12                ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
TL 2069           ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
3                 ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
MR 9917           ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
7857              ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
```

```
B5057          ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
9              ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
8              ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
B848           ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
118            ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
7              ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
BR 1692        ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
BR 1213        ATGTTCAAAA ATCCTGCAGA AAGACCTCGG AAATTGCATG AACTAAGCTC
               ********** ********** ********** ********** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            60         70         80         90        100
B6394          GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
73             GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
12             GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
TL 2069        GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
3              GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
MR 9917        GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
B7857          GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
B5057          GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
9              GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
8              GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
B848           GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
118            GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
7              GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
BR 1692        GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
BR 1213        GGCATTGGAA ATACCCTACG ATGAACTAAG ATTGAATTGT GTCTACTGCA
               ********** ********** ********** ********** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           110        120        130        140        150
6394           AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT CACAGATTTA
73             AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT CACAGATTTA
12             AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
TL 2069        AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
3              AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
MR 9917        AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
B7857          AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
B5057          AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
9              AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
8              AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
B848           AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
118            AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
7              AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
BR 1692        AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
BR 1213        AAGGTCAGTT AACAGAAACA GAGGTATTAG ATTTTGCATT TACAGATTTA
               ********** ********** ********** **********  **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           160        170        180        190        200
B6394          ACAATAGTAT ATAGGGACGA CACACCACAC GGAGTGTGTG CAAAATGTTT
73             ACAATAGTAT ATAGGGACGA CACACCACAC GGAGTGTGTG CAAAATGTTT
12             ACAATAGTAT ATAGGGACGA CACACCACAC GGAGTGTGTA CAAAATGTTT
TL 2069        ACAATAGTAT ATAGGGACGA CACACCACAC GGAGTGTGTA CAAAATGTTT
3              ACACTAGTAT ATAGGGACGA CACACCATAC GGAGTGTGTA CAAAATGTTT
MR 9917        ACAATAGTAT ATAGGGACGA CACACCATAC GGAGTGTGTA CAAAATGTTT
B7857          ACAATAGTAT ATAGGGACGA CACACCATAC GGAGTGTGTA CAAAATGTTT
B5057          ACAATAGTAT ATAGGGACGA CACACCATAC GGAGTGTGTA CAAAATGTTT
9              ACAATAGTAT ATAGGGACGA CACACCATAC GGAGTGTGTA CAAAATGTTT
8              ACAATAGTAT ATAGGGACGA CACACCATAC GGAGTGTGTA CAAAATGTTT
B848           ACAATAGTAT ATAGGGACGA CACACCATAC GGAGTGTGTA CAAAATGCTT
```

```
118             ACAATAGTAT ATAGGGACGA CACACCATAC GGAGTGTGTA CAAAATGTTT
7               ACAATAGTAT ATAGGGACGA CACACCATAC GGAGTGTGTA CAAAATGTTT
BR 1692         ACAATAGTAT ATAGGGACGA CACACCATAC GGAGTGTGTA CAAAATGTTT
BR 1213         ACAATAGTAT ATAGGGACGA CACACCATAC GGAGTGTGTA CAAAATGTTT
                *** ****** ********** ******* ** *********  ******* **

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
B6394           AAGATTTTAT TCTAAAGTAA GTGAATTTAG ATGGTATAGA TATAGTGTGT
73              AAGATTTTAT TCTAAAGTAA GTGAATTTAG ATGGTATAGA TATAGTGTGT
12              AAGATTTTAT TCAAAAGTAA GTGAATTTAG ATGGTATAGA TATAGTGTGT
TL 2069         AAGATTTTAT TCAAAAGTAA GTGAATTTAG ATGGTATAGA TATAGTGTGT
3               AAGATTTTAT TCTAAAGTAA GTGAATTTAG ATGGTATAGA TATAGTGTGT
MR 9917         AAGATTTTAT TCTAAAGTAA GTGAATTTAG ATGGTATAGA TATAGTGTGT
B7857           ACGATTTTAT TCTAAAGTGA GTGAATTTAG ATGGTATAGA TATAGTGTGT
B5057           AAGATTTTAT TCTAAAGTGA GTGAATTTAG ATGGTATAGA TATAGTGTGT
9               AAGATTTTAT TCTAAAGTGA GTGAATTTAG ATGGTATAGA TATAGTGTGT
8               AAGATTTTAT TCTAAAGTAA GTGAATTTCG ATGGTATAGA TATAGTGTGT
B848            AAGATTTTAT TCTAAAGTAA GTGAATTTAG ATGGTATAGA TATAGTGTGT
118             AAGATTTTAT TCTAAAGTAA GTGAATTTAG ATGGTATAGA TATAGTGTGT
7               AAGATTTTAT TCTAAAGTAA GTGAATTTAG ATGGTATAGA TATAGTGTGT
BR 1692         AAGATTTTAT TCTAAAGTAA GTGAATTTAG ATGGTATAGA TATAGTGTGT
BR 1213         AAGATTTTAT TCTAAAGTAA GTGAATTTAG ATGGTATAGA TATAGTGTGT
                * ******** ** ***** * ******** * ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
B6394           ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTGTTA
73              ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTGTTA
12              ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTGTTA
TL 2069         ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTGTTA
3               ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTATTA
MR 9917         ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTATTA
B7857           ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTATTA
B5057           ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTATTA
9               ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTATTA
8               ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTATTA
B848            ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTATTA
118             ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTATTA
7               ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTATTA
BR 1692         ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTATTA
BR 1213         ATGGAACAAC ATTAGAAAAA TTGACAAACA AAGGTATATG TGATTTATTA
                ********** ********** ********** ********** ****** ***

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
B6394           ATTAGGTGTA TAACGTGTCA AAGACCGTTG TGTCCAGAAG AAAAACAAAG
73              ATTAGGTGTA TAACGTGTCA AAGACCGTTG TGTCCAGAAG AAAAACAAAG
12              ATTAGGTGTA TAACGTGTCA AAGACCGTTG TGTCCAGAAG AAAAACAAAG
TL 2069         ATTAGGTGTA TAACGTGTCA AAGACCGTTG TGTCCAGAAG AAAAACAAAG
3               ATTAGGTGTA TAACGTGTCA AAGACCGTTG TGTCCAGAAG AAAAACAAAG
MR 9917         ATTAGGTGTA TAACGTGTCA AAGACCGTTG TGTCCAGAAG AAAAACAAAG
B7857           ATTAGGTGTA TAACGTGTCA GAGACCGTTG TGTCCAGAAG AAAAACAAAG
B5057           ATTAGGTGTA TAACGTGTCA GAGACCGTTG TGTCCAGAAG AAAAACAAAG
9               ATTAGGTGTA TAACGTGTCA GAGACCGTTG TGTCCAGAAG AAAAACAAAG
8               ATTAGGTGTA TAACGTGTCA GAGACCGTTG TGTCCAGAAG AAAAACAAAG
B848            ATTAGGTGTA TAACGTGTCA GAGACCGTTG TGTCCAGAAG AAAAACAAAG
118             ATTAGGTGTA TAACGTGTCA GAGACCGTTG TGTCCAGAAG AAAAACAAAG
7               ATTAGGTGTA TAACGTGTCA GAGACCGTTG TGTCCAGAAG AAAAACAAAG
BR 1692         ATTAGGTGTA TAACGTGTCA GAGACCGTTG TGTCCAGAAG AAAAACAAAG
BR 1213         ATTAGGTGTA TAACGTGTCA GAGACCGTTG TGTCCAGAAG AAAAACAAAG
```

```
                  ********** **********  ********* ********** **********

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                         360        370        380        390        400
B6394             ACATTTGGAT AAAAAGAGAC GATTCCACAA CATAGGAGGA AGGTGGACAG
73                ACATTTGGAT AAAAAGAGAC GATTCCACAA CATAGGAGGA AGGTGGACAG
12                ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
TL 2069           ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
3                 ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
MR 9917           ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
B7857             ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
B5057             ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
9                 ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
8                 ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
B848              ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
118               ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
7                 ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
BR 1692           ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
BR 1213           ACATTTGGAT AAAAAGAAAC GATTCCACAA CATAGGAGGA AGGTGGACAG
                  ********** ******* ** ********** ********** **********

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                         410        420        430        440        450
B6394             GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
73                GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
12                GACGTTGCAT AGCATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
TL 2069           GACGTTGCAT AGCATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
3                 GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
MR 9917           GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
B7857             GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
B5057             GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
9                 GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
8                 GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
B848              GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
118               GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
7                 GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
BR 1692           GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
BR 1213           GACGTTGCAT AGTATGTTGG AGAAGACCTC GTACTGAAAC CCAAGTGTAA
                  ********** ** ******* ********** ********** **********


B6394               --
73                  --
12                  --
TL 2069             --
3                   --
MR 9917             --
B7857               AC
B5057               --
9                   --
8                   --
B848                AC
118                 --
7                   --
BR 1692             --
BR 1213             --
```

6. HPV 33 E6 region-464bp

Forward primer used in the first round (nt position 164-188)

Forward primer used in the second round (nt position 193-217)

Reverse primer (nt position 319-341)

```
Isolates        ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
2               ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
3               ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
149             ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
B8454           ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
B2659           ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
B597            ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
P232            ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
100             ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
1               ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
B3200           ATGTTTCAAG ACACTGAGGA AAAACCACGC ACATTGCATG ATTTGTGCCA
291             ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
14              ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
346             ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
B4083           ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
P396            ATGTTTCAAG ACACTGAGGA AAAACCACGA ACATTGCATG ATTTGTGCCA
                ********** ********** *********  ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        60         70         80         90        100
2               AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
3               AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
149             AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
B8454           AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
B2659           AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
B597            AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
P232            AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
100             AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
1               AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
B3200           AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
291             AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
14              AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
346             AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
B4083           AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
P396            AGCATTGGAG ACAACTATAC ACAACATTGA ACTACAGTGC GTGGAATGCA
                ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       110        120        130        140        150
2               AAAAACCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
3               AAAAAACTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
149             AAAACCCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
B8454           AAAACCCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
B2659           AAAACCCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
B597            AAAACCCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
P232            AAAACCCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
100             AAAAACCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
1               AAAAACCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
B3200           AAAAACCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
```

```
291              AAAAACCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
14               AAAAACCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
346              AAAAACCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
B4083            AAAAACCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
P396             AAAAACCTTT GCAACGATCT GAGGTATATG ATTTTGCATT TGCAGATTTA
                 ****  **** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
2                ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
3                ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
149              ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
B8454            ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
B2659            ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
B597             ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
P232             ACAGTTGTAT ATAGAGAGGG AAATTCATTT GGAATATGTA AACTGTGTTT
100              ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
1                ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
B3200            ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
291              ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
14               ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
346              ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
B4083            ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
P396             ACAGTTGTAT ATAGAGAGGG AAATCCATTT GGAATATGTA AACTGTGTTT
                 ********** ********** **** ***** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
2                GCGGTTTTTA TCTAAACTTA GTGAATATAG ACATTATAAT TATTCTTTAT
3                GCGGTTTTTA TCTAAACTTA GTGAATATAG ACATTATAAT TATTCTTTAT
149              GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
B8454            GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
B2659            GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
B597             GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
P232             GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
100              GCGGTTCTTA TCTAAAATTA GTGAATATAA ACATTATTAT TATTCTGTAT
1                GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
B3200            GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
291              GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
14               GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
346              GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
B4083            GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
P396             GCGGTTCTTA TCTAAAATTA GTGAATATAG ACATTATAAT TATTCTGTAT
                 ****** *** ****** *** *********  ******* ** ****** ***

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
2                ATGGAAATAC ATTAGAACAA ACAGTTAACA AACCTTTAAA TGAAATATTA
3                ATGGAAATAC ATTAGAACAA ACAGTTAACA AACCTTTAAA TGAAATATTA
149              ATGGACATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
B8454            ATGGACATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
B2659            ATGGACATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
B597             ATGGACATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
P232             ATGGACATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
100              ATGGAAATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
1                ATGGAAATAC ATTAGAACAA ACAGTTAAAA AACCCTTAAA TGAAATATTA
B3200            ATGGAAATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
291              ATGGAAATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
14               ATGGAAATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
346              ATGGAAATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
B4083            ATGGAAATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
```

```
P396            ATGGAAATAC ATTAGAACAA ACAGTTAAAA AACCTTTAAA TGAAATATTA
                ***** **** ********** ******** * **** ***** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
2               ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
3               ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
149             ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
B8454           ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
B2659           ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
B597            ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
P232            ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
100             ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
1               ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
B3200           ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
291             ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
14              ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
346             ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
B4083           ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
P396            ATTAGGTGTA TTATATGTCA AAGACCTTTG TGTCCTCAAG AAAAAAAACG
                ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
2               ACATGTGGAT TTAAACAAAC GATTTCATAA TATTTCGGGT CGTTGGGCAG
3               ACATGTGGAT TTAAACAAAC GATTTCATAA TATTTCGGGT CGTTGGGCAG
149             ACATGTGGAT TTAAACAAAC GTTTTCATAA TATTTCGGGT CGTTGGGCAG
B8454           ACATGTGGAT TTAAACAAAC GTTTTCATAA TATTTCGGGT CGTTGGGCAG
B2659           ACATGTGGAT TTAAACAAAC GTTTTCATAA TATTTCGGGT CGTTGGGCAG
B597            ACATGTGGAT TTAAACAAAC GTTTTCATAA TATTTCGGGT CGTTGGGCAG
P232            ACATGTGGAT TTAAACAAAC GTTTTCATAA TATTTCGGGT CGTTGGGCAG
100             ACATGTGGAT TTAAACAAAC GATTTCATAA TATTTCGGGT CGTTGGGCAG
1               ACATGTGGAT TTAAACAAAC GATTTCATAA TATTTCGGGT CGTTGGGCAG
B3200           ACATGTGGAT TTAAACAAAC GATTTCATAA TATTTCGGGT CGTTGGGCAG
291             ACATGTGGAT TTAAACAAAC GATTTCATAA TATTTCGGGT CGTTGGGCAG
14              ACATGTGGAT TTAAACAAAC GATTTCATAA TATTTCGGGT CGTTGGGCAG
346             ACATGTGGAT TTAAACAAAC GATTTCATAA TATTTCGGGT CGTTGGGCAG
B4083           ACATGTGGAT TTAAACAAAC GATTTCATAA TATTTCGGGT CGTTGGGCAG
P396            ACATGTGGAT TTAAACAAAC GATTTCATAA TATTTCGGGT CGTTGGGCAG
                ********** ********** * ******** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       410        420        430        440        450
2               GGCGCTGTGC GGTGTGTTGG AGGTCTCGAC GTAGAGAAAC TGCACTGTGA
3               GGCGCTGTGC GGTGTGTTGG AGGTCTCGAC GTAGAGAAAC TGCACTGTGA
149             GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
B8454           GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
2B659           GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
B597            GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
P232            GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
100             GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
1               GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
B3200           GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
291             GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
14              GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
346             GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
B4083           GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
P396            GGCGCTGTGC GGCGTGTTGG AGGTCCCGAC GTAGAGAAAC TGCACTGTGA
                ********** ** ******* ***** **** ********** **********
```

```
                    ....|....| ....
                           460
2                   ---------- ----
3                   ---------- ----
149                 ---------- ----
B8454               CGTGTAAAAA CGCC
B2659               CGTGTAAAAA CGCC
B597                CGTGTAAAAA CGCC
P232                ---------- ----
100                 ---------- ----
1                   ---------- ----
B3200               CGTGTAAAAA CGCC
291                 ---------- ----
14                  ---------- ----
346                 ---------- ----
B4083               CGTGTAAAAA CGCC
P396                ---------- ----
```

7. HPV 45 E6 region-477bp

Forward primer used in the first round (nt position 3-24)

Forward primer used in the second round (nt position 31-51)

Reverse primer (nt position 117-138)

```
Isolates    ....|....| ....|....| ....|....| ....|....| ....|....|
                   10         20         30         40         50
B3587       ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
B3730       ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
B8463       ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
B2408       ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
92.070      ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
91.971      ATGGCGCGCT TTGACGATCC AACGCAACGA CCCTACAAGC TACCAGATTT
91.841      ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
91.811      ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
91.471      ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
91.401      ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
91.631      ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
90.391      ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
90.331      ATGGCGCGCT TTGACGATCC AACGCAACGA CCCTACAAGC TACCAGATTT
82.091      ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
71.751      ATGGCGCGCT TTGACGATCC AAAGCAACGA CCCTACAAGC TACCAGATTT
            ********** ********** ** ******* ********** **********

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           60         70         80         90        100
B3587       GTGCACAGAA TTGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
B3730       GTGCACAGAA TTGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
B8463       GTGCACAGAA TTGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
B2408       GTGCACAGAA TTGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
92.070      GTGCACAGAA ACGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
91.971      GTGCACAGAA TTGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
91.841      GTGCACAGAA ACGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
91.811      GTGCACAGAA TTGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
91.471      GTGCACAGAA ACGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
91.401      GTGCACAGAA TTGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
91.631      GTGCACAGAA TTGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
90.391      GTGCACAGAA ACGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
```

```
90.331          GTGCACAGAA TTGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
82.091          GTGCACAGAA TTGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
71.751          GTGCACAGAA ACGAATACAT CACTACAAGA CGTATCTATT GCCTGTGTAT
                ********** ******** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      110        120        130        140        150
B3587           ATTGCAAAGC AACATTGGAA CGCACAGAGG TATATCAATT TGCCTTTAAA
B3730           ATTGCAAAGC AACATTGGAA CGCACAGAGG TATATCAATT TGCTTTTAAA
B8463           ATTGCAAAGC AACATTGGAA CGCACAGAGG TATATCAATT TGCTTTTAAA
B2408           ATTGCAAAGC AACATTGGAA CGCACAGAGG TATATCAATT TGCTTTTAAA
92.070          ATTGCAAAGC AACATTGGAA CGCAGAGAGG TATATCAATT TGCTTTTAAA
91.971          ATTGCAAAGC AACATTGGAA CGCACAGAGG TATATCAATT TGCTTTTAAA
91.841          ATTGCAAAGC AACATTGGAA CGCAGAGAGG TATATCAATT TGCTTTTAAA
91.811          ATTGCAAAGC AACATTGGAA CGCACAGAGG TATATCAATT TGCTTTTAAA
91.471          ATTGCAAAGC AACATTGGAA CGCAGAGAGG TATATCAATT TGCTTTTAAA
91.401          ATTGCAAAGC AACATTGGAA CGCACAGAGG TATATCAATT TGCTTTTAAA
91.631          ATTGCAAAGC AACATTGGAA CGCACAGAGG TATATCAATT TGCTTTTAAA
90.391          ATTGCAAAGC AACATTGGAA CGCAGAGAGG TATATCAATT TGCTTTTAAA
90.331          ATTGCAAAGC AACATTGGAA CGCACAGAGG TATATCAATT TGCTTTTAAA
82.091          ATTGCAAAGC AACATTGGAA CGCACAGAGG TATATCAATT TGCTTTTAAA
71.751          ATTGCAAAGC AACATTGGAA CGCAGAGAGG TATATCAATT TGCTTTTAAA
                ********** ********** **** ***** ********** *** ******

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      160        170        180        190        200
B3587           GATTTATGTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
B3730           GATTTATGTA TAGTGTATAG AGACTGTATA GCCTATGCTG CATGCCATAA
B8463           GATTTATGTA TAGTGTATAG AGACTGTATA GCCTATGCTG CATGCCATAA
B2408           GATTTATGTA TAGTGTATAG AGACTGTATA GCCTATGCTG CATGCCATAA
92.070          GATTTATGTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
91.971          GATTTATTTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
91.841          GATTTATGTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
91.811          GATTTATGTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
91.471          GATTTATGTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
91.401          GATTTATGTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
91.631          GATTTATGTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
90.391          GATTTATGTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
90.331          GATTTATTTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
82.091          GATTTATGTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
71.751          GATTTATGTA TAGTGTATAG AGACTGTATA GCATATGCTG CATGCCATAA
                ********** ********** ********** ** ******* **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      210        220        230        240        250
B3587           ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
B3730           ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
B8463           ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
B2408           ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
92.070          ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
91.971          ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
91.841          ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
91.811          ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
91.471          ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
91.401          ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
91.631          ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
90.391          ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
90.331          ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
82.091          ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
71.751          ATGTATAGAC TTTTATTCCA GAATTAGAGA ATTAAGATAT TATTCAAACT
```

```
                ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
B3587           CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
B3730           CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
B8463           CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
B2408           CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
92.070          CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
91.971          CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
91.841          CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
91.811          CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
91.471          CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
91.401          CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
91.631          CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
90.391          CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
90.331          CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
82.091          CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
71.751          CTGTATATGG AGAGACACTG GAAAAAATAA CTAATACAGA GTTGTATAAT
                ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
B3587           TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
B3730           TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
B8463           TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
B2408           TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
92.070          TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
91.971          TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
91.841          TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
91.811          TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
91.471          TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
91.401          TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
91.631          TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
90.391          TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
90.331          TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
82.091          TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
71.751          TTGTTAATAA GGTGCCTGCG GTGCCAGAAA CCATTGAACC CAGCAGAAAA
                ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
B3587           ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAACATA GCTGGACAGT
B3730           ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAGCATA GCTGGGCAGT
B8463           ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAGCATA GCTGGGCAGT
B2408           ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAGCATA GCTGGGCAGT
92.070          ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAGCATA GCTGGACAGT
91.971          ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAGCATA GCTGGACAGT
91.841          ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAACATA GCTGGACAGT
91.811          ACGTAGACAC CTTAAGGACA AACGAAGATT TCACAGCATA GCTGGACAGT
91.471          ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAACATA GCTGGACAGT
91.401          ACGTAGACAC CTTAAGGACA AACGAAGATT TCACAGCATA GCTGGACAGT
91.631          ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAACATA GCTGGACAGT
90.391          ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAACATA GCTGGACAGT
90.331          ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAGCATA GCTGGGCAGT
82.091          ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAACATA GCTGGACAGT
71.751          ACGTAGACAC CTTAAGGACA AACGAAGATT CCACAACATA GCTGGACAGT
                ********** ********** **********  *** **** **********
```

```
                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           410        420        430        440        450
B3587               ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
B3730               ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
B8463               ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
B2408               ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
92.070              ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
91.971              ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
91.841              ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
91.811              ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
91.471              ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
91.401              ACCGAGGGCA GTGCAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
91.631              ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
90.391              ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
90.331              ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
82.091              ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
71.751              ACCGAGGGCA GTGTAATACA TGTTGTGACC AGGCACGGCA AGAAAGACTT
                    ********** *** ****** ********** ********** **********

                    ....|....| ....|....| ....|..
                           460        470
B3587               CGCAGACGTA GGGAAACACA AGTATAG
B3730               CGCAGACGTA GGGAAACACA AGTATAG
B8463               CGCAGACGTA GGGAAACACA AGTATAG
B2408               CGCAGACGTA GGGAAACACA AGTATAG
92.070              CGCAGACGTA GGGAAACACA AGTATAG
91.971              CGCAGACGTA GGGAAACACA AGTATAG
91.841              CGCAGACGTA GGGAAACACA AGTATAG
91.811              CGCAGACGTA GGGAAACACA AGTATAG
91.471              CGCAGACGTA GGGAAACACA AGTATAG
91.401              CGCAGACGTA GGGAAACACA AGTATAG
91.631              CGCAGACGTA GGGAAACACA AGTATAG
90.391              CGCAGACGTA GGGAAACACA AGTATAG
90.331              CGCAGACGTA GGGAAACACA AGTATAG
82.091              CGCAGACGTA GGGAAACACA AGTATAG
71.751              CGCAGACGTA GGGAAACACA AGTATAG
                    ********** ********** *******
```

8. HPV 58 E6 region-450bp

Forward primer used in the first round (nt position 1-25)

Forward primer used in the second round (nt position 47-70)

Reverse primer (nt position 170-190)

```
Isolates            ....|....| ....|....| ....|....| ....|....| ....|....|
                           10         20         30         40         50
ZWE_064436          ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
ITA_52PA            ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
ZWE_050364          ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
ZWE_044103          ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
ARG_P2005           ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
HK 262              ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
ZWE_063739          ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
CAN 420             ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
USA 990989          ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
```

```
THA_269           ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
ZJ 8              ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
ZJ 14             ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
ZJ 4              ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
ZJ 11             ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
JS 049297         ATGTTCCAGG ACGCAGAGGA GAAACCACGG ACATTGCATG ATTTGTGTCA
                  ********** ********** ********** ********** **********

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                        60         70         80         90        100
ZWE_064436        GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTGAATGCA
ITA 52PA          GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTGAATGCA
ZWE_050364        GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTGAATGCA
ZWE_044103        GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTGAATGCA
ARG_P2005         GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTGAATGCA
HK 262            GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTCAATGCA
ZWE_063739        GGCGTTGGAG ACATCTGTGC ATGAAATTGA ATTGAAATGC GTTGAATGCA
CAN 420           GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTCAATGCA
USA 990989        GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTGAATGCA
THA_269           GGCGTTGGAG ACATCTGTGC ATAAAATCGA ATTGAAATGC GTTGAATGCA
ZJ 8              GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTGAATGCA
ZJ 14             GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTGAATGCA
ZJ 4              GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTGAATGCA
ZJ 11             GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTGAATGCA
JS 049297         GGCGTTGGAG ACATCTGTGC ATGAAATCGA ATTGAAATGC GTTGAATGCA
                  ********** ********** ** **** ** ********** *** ******

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
ZWE_064436        AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTACATT TGCAGATTTA
ITA 52PA          AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTACATT TGCAGATTTA
ZWE_050364        AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTACATT TGCAGATTTA
ZWE_044103        AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTACATT TGCAGATTTA
ARG_P2005         AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTATATT TGCAGATTTA
HK 262            AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTGTATT TGCAGATTTA
ZWE_063739        AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTGTATT TGCAGATTTA
CAN 420           AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTGTATT TGCAGATTTA
USA 990989        AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTGTATT TGCAGATTTA
THA_269           AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTGTATT TGCAGATTTA
ZJ 8              AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTGTATT TGCAGATTTA
ZJ 14             AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTGTATT TGCAGATTTA
ZJ 4              AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTGTATT TGCAGATTTG
ZJ 11             AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTGTATT TGCAGATTTA
JS 049297         AAAAGACTTT GCAGCGATCT GAGGTATATG ACTTTGTATT TGCAGATTTA
                  ********** ********** ********** *****  *** *********

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
ZWE_064436        AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGTTT
ITA 52PA          AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGTTT
ZWE_50364         AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGTTT
ZWE_044103        AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGTTT
ARG_P2005         AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGTTT
HK 262            AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGTTT
ZWE_063739        AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGTTT
CAN 420           AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGTTT
USA 990989        AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGTTT
THA_269           AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGTTT
ZJ 8              AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGTTT
ZJ 14             AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGCTT
ZJ 4              AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGCTT
ZJ 11             AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGCTT
```

```
JS 049297          AGAATAGTGT ATAGAGATGG AAATCCATTT GCAGTATGTA AAGTGTGCTT
                   ********** ********** ********** ********** ******* **

                   ....|....| ....|....| ....|....| ....|....| ....|....|
                          210        220        230        240        250
ZWE_064436         ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
ITA 52PA           ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
ZWE_050364         ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
ZWE_044103         ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
ARG_P2005          ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
HK 262             ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
ZWE_063739         ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
CAN 420            ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
USA 90989          ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
THA_269            ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
ZJ 8               ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
ZJ 14              ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
ZJ 4               ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
ZJ 11              ACGATTGTTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
JS 049297          ACGATTGCTA TCTAAAATAA GTGAGTATAG ACATTATAAT TATTCGCTAT
                   ******* ** ********** ********** ********** **********

                   ....|....| ....|....| ....|....| ....|....| ....|....|
                          260        270        280        290        300
ZWE_064436         ATGGAGACAC ATTAGAACAA ACACTAAAAA AGTGTTTAAA GGAAATATTA
ITA 52PA           ATGGAAACAC ATTAGAACAA ACACTAAAAA AGCGTTTAGA GGAAATATTA
ZWE_050364         ATGGAGACAC ATTAGAACAA ACACTAAAAA AGCGTTTAGA GGAAATATTA
ZWE_044103         ATGGAGACAC ATTAGAACAA ACACTAAAAA AGCGTTTAGA GGAAATATTA
ARG_P2005          ATGGAGACAC ATTAGAACAA ACACTAAAAA AGTGTTTAGA TGAAATATTA
HK 262             ATGGAGACAC ATTAGAACAA ACACTAAACA AGTGTTTAAA TGAAATATTA
ZWE_063739         ATGGAGAAAC ATTAGAACAA ACACTAAAAA AGTGTTTAAA TGAAATATTA
CAN 420            ATGGAGAAAC ATTAGAACAA ACACTAAAAA AGTGTTTAAA TGAAATATTA
USA 990989         ATGGAGACAC ATTAGAACAA ACACTAAAAA AGTGTTTAAA TGAAATATTA
THA_269            ATGGAGACAC ATTAGAACAA ACACTAAAAA AGTGTTTAAA TGAAATATTA
ZJ 8               ATGGAGACAC ATTAGAACAA ACACTAAAAA AGTGTTTAAA TGAAATATTA
ZJ 14              ATGGAGACAC ATTAGAACAA ACACTAAACA AGTGTTTAAA TGAAATATTA
ZJ 4               ATGGAGACAC ATTAGAACAA ACACTAAAAA AGTGTTTAAA TGAAATATTA
ZJ 11              ATGGAGACAC ATTAGAACAA ACACTAAAAA AGTGTTTAAA TGAAATATTA
USA 049297         ATGGAGACAC ATTAGAACAA ACACTAAAAA AGTGTTTAAA TGAAATATTA
                   ***** * ** ********** ******** * ** ***** *  *********

                   ....|....| ....|....| ....|....| ....|....| ....|....|
                          310        320        330        340        350
ZWE_064436         ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
ITA 52PA           ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
ZWE_050364         ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
ZWE_044103         ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
ARG_P2005          ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
HK 262             ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
ZWE_063739         ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
CAN 420            ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
USA 990989         ATTAGATGTA TTATGTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
THA_269            ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
ZJ 8               ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
ZJ 14              ATTAGATGTA TTATTTGTCA AAAACCATTG TGTCCACAAG AAAAAAAAAG
ZJ 4               ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
ZJ 11              ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
JS 049297          ATTAGATGTA TTATTTGTCA AAGACCATTG TGTCCACAAG AAAAAAAAAG
                   ********** **** ***** ** ******* ********** **********
```

```
                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           360        370        380        390        400
ZWE_064436          GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
ITA 52PA            GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
ZWE_050364          GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
ZWE_044103          GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
ARG_P2005           GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
HK 262              GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
ZWE_063739          GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
CAN 420             GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
USA 990989          GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
THA_269             GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
ZJ 8                GCATGTGGAT TTAAACAAAC GGTTTCATAA TATTTCGGGT CGTTGGACAG
ZJ 14               GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
ZJ 4                GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
ZJ 11               GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
JS 049297           GCATGTGGAT TTAAACAAAA GGTTTCATAA TATTTCGGGT CGTTGGACAG
                    ********** *********  ********** ********** **********


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           410        420        430        440        450
ZWE_064436          GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
ITA 52PA            GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
ZWE_050364          GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
ZWE_044103          GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
ARG_P2005           GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
HK 262              GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
ZWE_063739          GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
CAN 420             GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
USA 990989          GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
THA_269             GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
ZJ 8                GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
ZJ 14               GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
ZJ 4                GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
ZJ 11               GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
JS 049297           GGCGCTGTGC AGTGTGTTGG AGACCCCGAC GTAGACAAAC ACAAGTGTAA
                    ********** ********** ********** ********** **********
```

9. HPV 84 E6 region-447bp

Forward primer used in the first round (nt position 6-27)

Forward primer used in the second round (nt position 54-77)

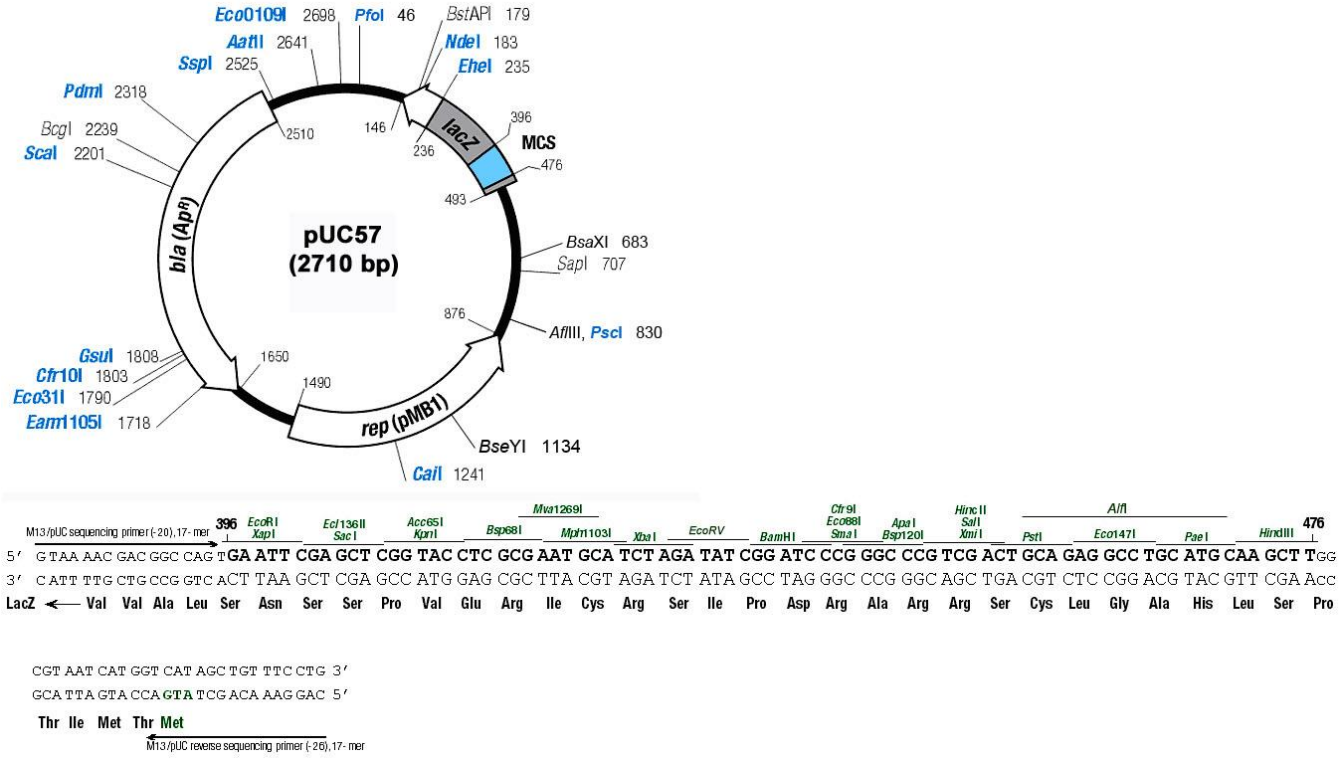
Reverse primer used (nt position 171-195)

```
Isolate             ....|....| ....|....| ....|....| ....|....| ....|....|
                           10         20         30         40         50
HPV 84E6            ATGCCCAACG GACGCTACCA CCCCACCAAT ATTTTTGTGC TGTGCCAGGA


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           60         70         80         90        100
HPV 84E6            ATACGAGGTG GAGTTCGACG ACCTACGATT AATTTGCATC TTTTGCAAGG


                    ....|....| ....|....| ....|....| ....|....| ....|....|
```

```
                        110        120        130        140        150
HPV 84E6          AAGAATTAAC GGAAGGCGAA GTGCTGGCCT TTGCAGTAAA GGAATTACTA

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
HPV 84E6          ATTGTTTGGA GGTATAATTT CCCTCATGGG GTGTGCATGA AATGCTTATG

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
HPV 84E6          CAGGGAAGCC AAAGTACGTG AGCTACGCCA CTGGGATTAC TCCAGCTTTG

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
HPV 84E6          GACCAACAGT GGAAGAAGAA ACAGGATTAC CACTTGCACA AATAAATATA

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
HPV 84E6          AGGTGCCACG CGTGCTGCAA GCCATTGTGC TATCAGGAAA AGGAGTATAT

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                        360        370        380        390        400
HPV 84E6          GGTGGAATTG CAGTTGCTAT TCCACAAAAT AGCTGGACAG TGGACAGGGA

                  ....|....| ....|....| ....|....| ....|....| ....|..
                        410        420        430        440
HPV 84E6          AGTGCTGCAA CTGTAGGGTA ACATGCGCGG CCAGACGCCA ACGTTAA
```

Appendix D: Vector map with sequence reference points of pUC 57 plasmid and partial E6 HPV 33 gene.

A: Vector map and multiple cloning sites of cloning vector pUC 57 (GenScript cooperation, New Jersey, USA).



B: Partial E6 HPV 33 gene synthesised into pUC 57 plasmid.

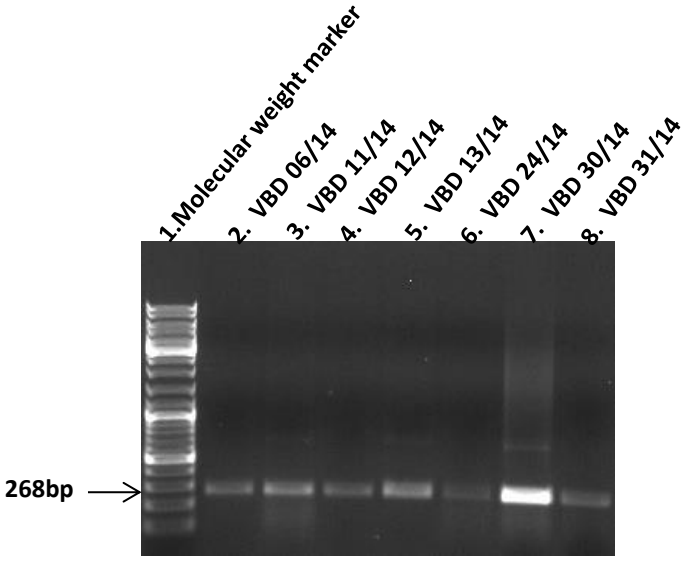
Nucleotide position 164-341 of the whole HPV type -33 gene.

5′GAGAGGGAAATCCATTTGGAATATGTAAACTGTGTTTGCGGTTTTTATCTAAACTTAGTGAATATAG
ACATTATAATTATTCTTTATATGGAAATACATTAGAACAAACAGTTAACAAACCTTTAAATGAAATATTAATT
AGGTGTATTATATGTCAAAGACCTTTGTGTCCTCAAGA3′

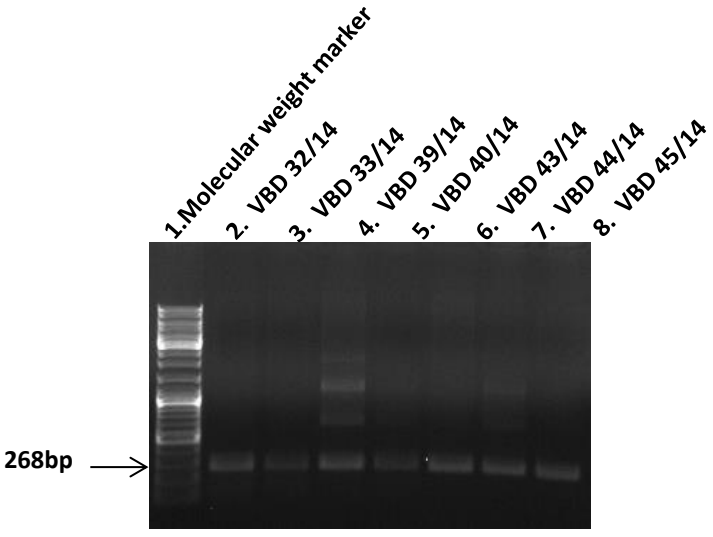
Nucleotide position 1-195 of the whole HPV type -58 gene

5′ATGTTCCAGGACGCAGAGGAGAAACCACGGACATTGCATGATTTGTGTCAGGCGTTGGAGACATCTGTGC
ATGAAATCGAATTGAAATGCGTTGAATGCAAAAAGACTTTGCAGCGATCTGAGGTATATGACTTTACATT
TGCAGATTTAAGAATAGTGTATAGAGATGGAAATCCATTTGCAGTATGTAAAGTG3′

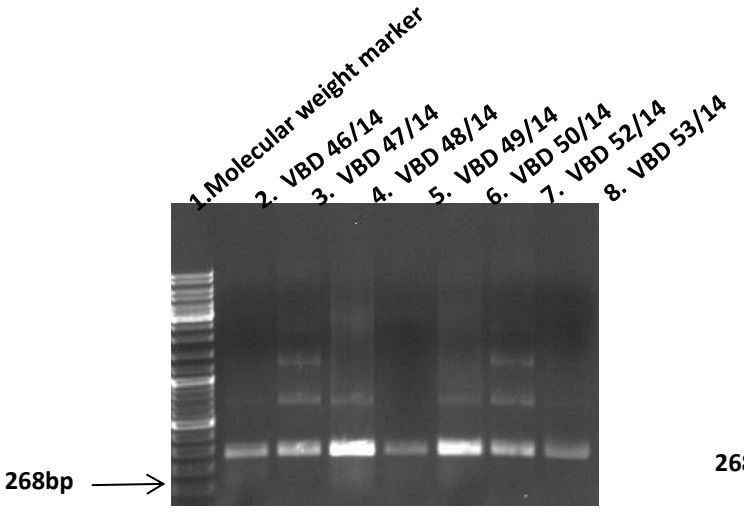
Appendix E: Beta-globin amplification results for 74 samples.



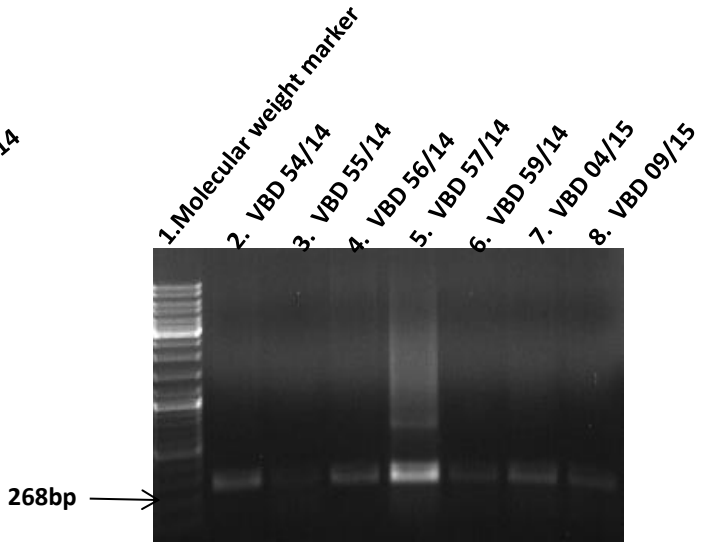
A



B

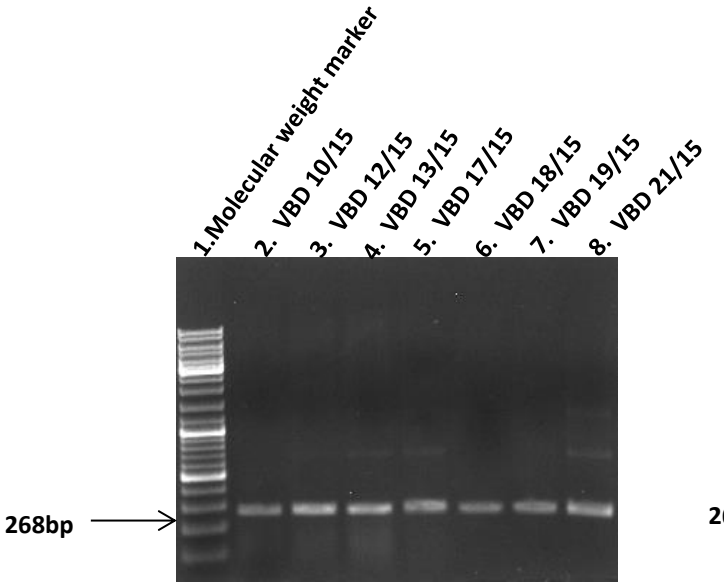


C

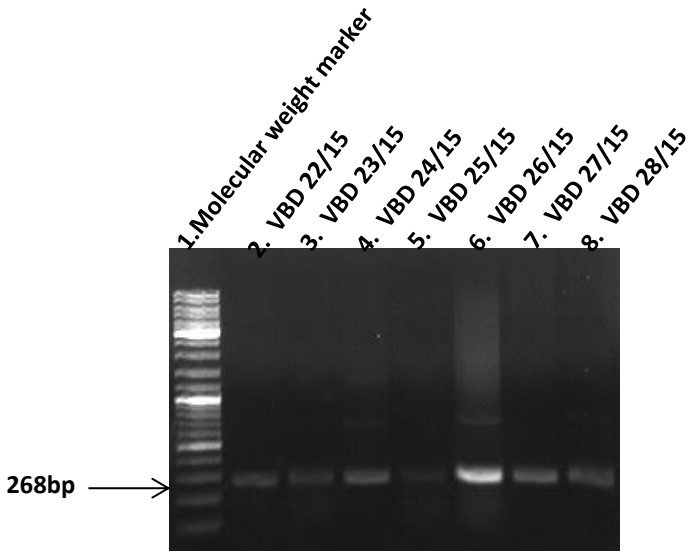


D

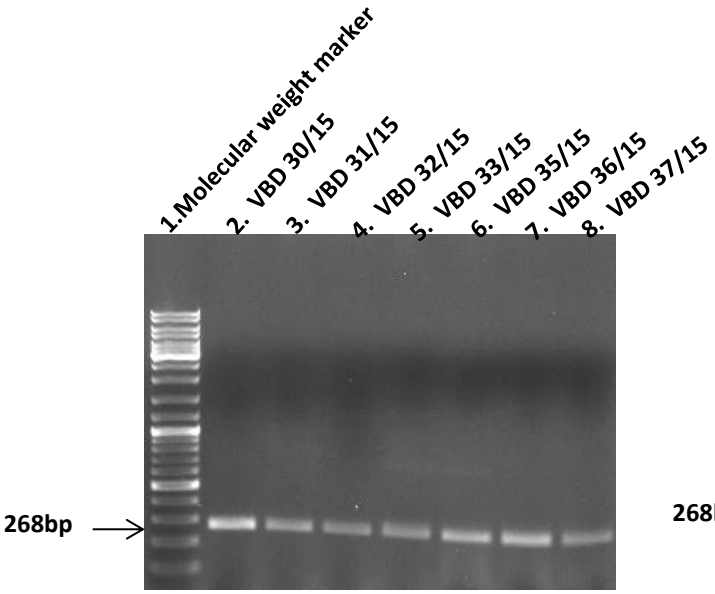
Appendix E continues



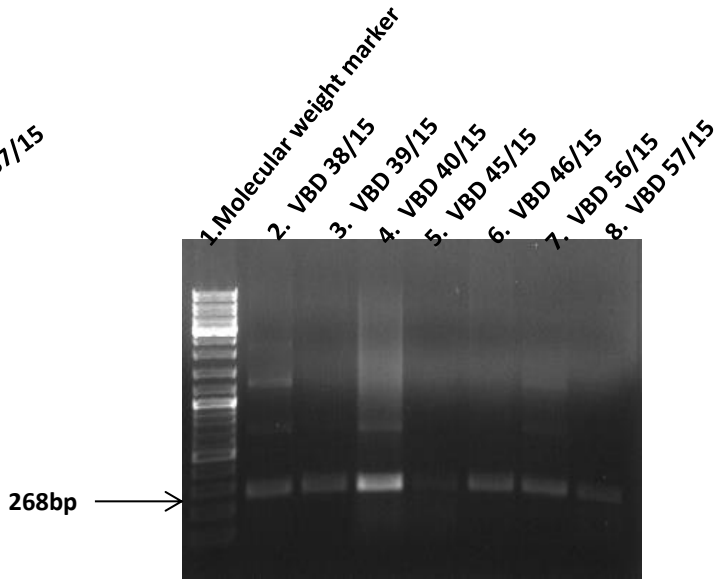
E



F

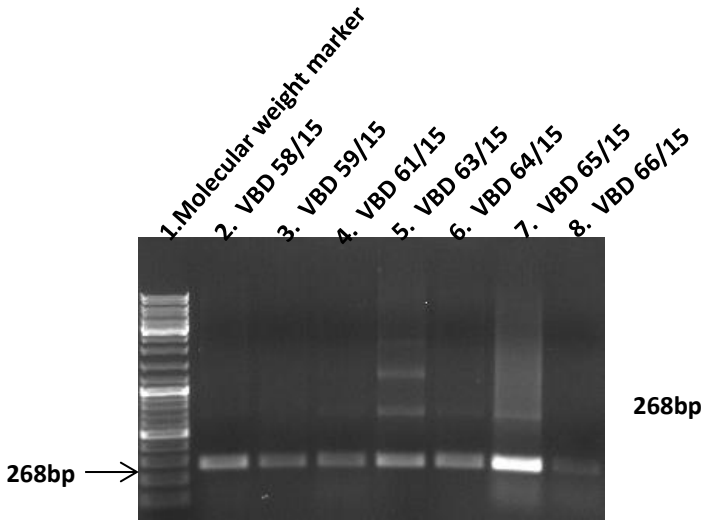


G

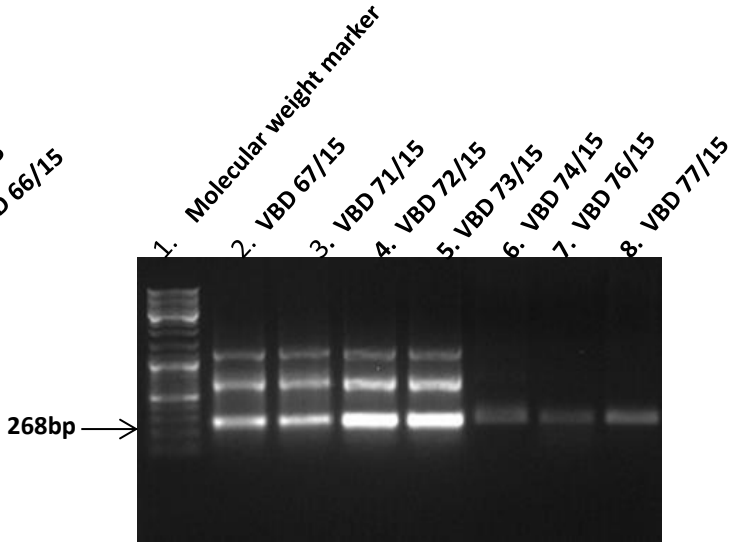


H

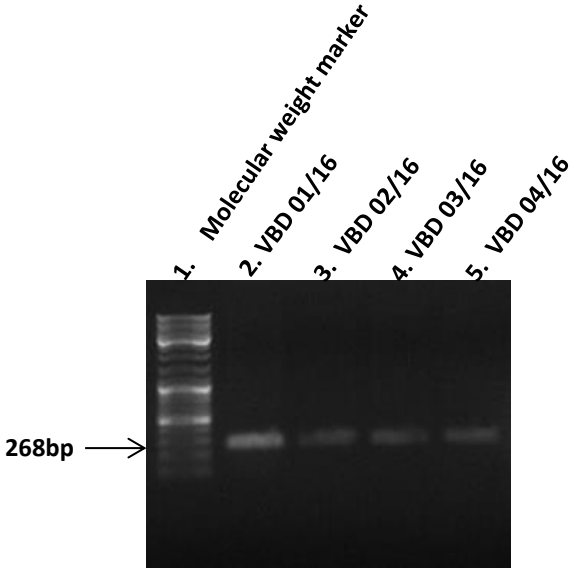
Appendix E continue



I



J



K

Appendix F: Nucleotide sequences for HPV types isolated in study.

Below are the nucleotide sequences obtained from each method using Sanger sequencing. Identity between the nucleotide from the samples in the study and nucleotide sequences obtained from GenBank is an extent to which two (nucleotide or amino acid) sequences have the same residues at the same positions in an alignment, often expressed as a %age. The Expectation value or Expect value represents the number of different alignments with scores equivalent to or better than that is expected to occur in a database search by chance. The lower the E value, the more significant the score and alignment. The query is an input sequence to which all of the entries in a database are to be compared (Altschul *et al*., 1990).

1. Nested PCR targeting the L1 region of the HPV genome

First round

VBD number 13/14-356bp

The HPV type obtained in the current study with a nucleotide sequence below has a 100% identity to HPV type 31 isolate 6857ENG from GenBank (accession number KJ754568.1). It has a query cover of 100% and an E value of 0.

Nucleotide sequence

5'ACCTCCCAAAATACATAATCTTTAAATGGATCTTCCTTAGGCTTTTGGGGGGCAGTTTTTTGACA
TGTAATGGCCTGTGAGGTTACAAACCTATAGGTATCCTCTAAAGAACCTGAGGGAGGTGTGGTTA
ATCCAAAATTCCAATCTTCCAAAATAGCAGGATTCATACTGTGAATATATGTCATTATGTCTGCAG
ATAATGTTATTTTGCATAACTGAAATATAAATTGTAAATCAAATTCCTCACCATGTCTTAAATACTC
TTTAAAATTACTACTTTTAAATGTAGTATCACTGTTTGCAATTGCAGCACAAACAGACATATTGGT
ACTGCGTGTGGTATCTACCACAGTAAC3'

VBD number 17/15-366bp

The HPV type obtained in the current study with a nucleotide sequence below has a 96% identity to HPV type 18 isolate CUBA-DIAG-1539-2011 from GenBank (accession number KC991261.1). It has a query cover of 87% and expectancy value of $6e^{-166}$.

Nucleotide sequence

5'CCAACATTCCAAAACTTTAACTTATCATAGGGGATCCTTATTTTCAGCCGGTGCAGCATCCTTTT
GACAGGTAATAGCAACAGATTGTACAAAACGATATGTATCCACCAAACTAGTAGTTGGCGGGGG
GGGAACACCAAAGTTCCAATCCTCTAAAATACTGCTATTCATACTATGAATATAGGACATAACATC
TGCAGTTAAAGTAATAGTACACAACTGAAAAATAAACTGCAAATCATATTCCTCAACATGTCTGCT
ATACTGCTTAAATTTGGTAGCATCATATTGCCCAGGTACAGGAGACTGTGTAGAAGCACATATTG
TTAAATTGGTACTGCGAGTGGTATCTACCACAGTAACAAA3'

Second round

VBD number 47/14- 123bp

The HPV type obtained in the current study with a nucleotide sequence below has a 96% identity to HPV type 16 isolate 6 (nt position 1-114) from GenBank (accession number KC706453). It has a query cover of 95% and expectancy value of $3e^{-44}$.

Nucleotide sequence

5'GTAAGGTTTCTGAAGTAGATATGGCAGCACATAATCGCACATATATTAGTACTGCGTGTAGTAT
CAACAACAGTAACAAATAGTTGGTTACCCCAACAAATGCCATTGTTATGACCCTGGGCA3'

VBD number 10/15-91bp

The HPV type obtained in the current study with a nucleotide sequence below has a 100% identity to HPV type 11 isolate 3 from GenBank (accession number AF548815.1). It has a query cover of 100% and expectancy value of $8e^{-39}$.

Nucleotide sequence

5'ATAATCTGAATTAGTGTATGTAGCAGATTTAGACACAGATGCACATAGTGTCATATTTGTACTG
CGTGTAGTATCTACCACAGTAACAAAA3'

2. E6 multiplex hemi-nested type specific type PCR

First round

VBD 13/14-50bp

The HPV type obtained in the current study with a nucleotide sequence below has a 100% identity to HPV type 31 isolate 34 (nt position 96-128) from GenBank (accession number KC700156.1). It has a query cover of 84% and expectancy value of $2e^{-13}$.

Nucleotide sequence

5’GTGTCTACCTGCAAAGGTCAGTTAACAGAAACAGAGGTATTAGATTTTGC3’

VBD 47/14-53bp

The HPV type obtained in the current study with a nucleotide sequence below has a 100% identity to HPV type 16 isolate HPV 16479 (nt position 895-947) from GenBank (accession number KP965162.1). It has a query cover of 100% and expectancy value of $4e^{-18}$.

Nucleotide sequence

5’CTGCAAACAACTATACATGATATAATATTAGAATGTGTGTACTGCAAGCAACA3’

VBD 17/15-104bp

The HPV type obtained in the current study with a nucleotide sequence below has a 100% identity to HPV type 18 isolate HPVHap 30 (nt 865-968) from GenBank (accession number KP965188.1). It has a query cover of 100% and expectancy value of $6e^{-49}$.

Nucleotide sequence

5’GATCTGTGCACGGAACTGAACACTTCACTGCAAGACATAGAAATAACCTGTGTATATTGCAAGA
CAGTATTGGAACTTACAGAGGTATTTGAATTTGCATTTAA3’

VBD 59/15-109bp

The HPV type obtained in the current study with a nucleotide sequence below has a 100% identity to HPV type 45 pop-variant IARC variant 16 from GenBank (accession number KF591384.1). It has a query cover of 75% and expectancy value of $1e^{-33}$.

Nucleotide sequence

5’CATGGCGGCCGCGAATTCACTAGTGATTGGCGCGCTTTGACGATCCAAAGCAACGACCCTACAAGCTAC
CAGATTTGTGCACAGAATTGAATACATCACTACAAGACGT3’

Second round

VBD 10/15-103bp

The HPV type obtained in the current study with a nucleotide sequence below has a 92% identity to HPV type 11 isolate 30 (nt position 97-200) from GenBank (accession number LN833190.1). It has a query cover of 98% and expectancy value of $1e^{-31}$

Nucleotide sequence

5’CTTTTGTAAGGATGCCTTGCCACCGCAGAGAATATGCATATGCCTATAAGAACCTAAAGGTTGT
GGGCGAGACAACTTTCCCCTTGCAGCGTGTGCCTGTTGA3’

3. Nested PCR (PGMY11/09 and GP5+/6+ primers)

First round PCR

VBD 59/15-363bp

The HPV type obtained in the current study with a nucleotide sequence below has a 100% identity to HPV type 45 isolate 14UK31 from GenBank (accession number KU049753.1). It has a query cover of 100% and expectancy value of 0.

Nucleotide sequence

5’GTCAACAGTCCAAAACTTTAATTTATCATATGGATCCTGCTTTTCTGGAGGTGTAGTATCCTTTT
GACAGGTAACAGCAACTGATTGCACAAAACGATATGTATCCACTAAACTTGTAGTAGGTGGTGG
AGGGACACCAAAATTCCAATTTTCTAATATACTACTATTCATACTATGGATATATGACATAACCTCT
GCAGTTAAAGTAATAGTGCACAACTGAAAAATAAACTGTAAATCATATTCCTCCACATGTCTACTA
TACTGCTTAAACTTAGTAGGGTCATATGTACTTGGCACAGGATTTTGTGTAGAGGCACATAATGTT
AAATTAGTACTGCGGGTAGTGTCTACTACAGTAAC3’

Appendix G: Sequence reference points for pGEM®- T easy cloning vector

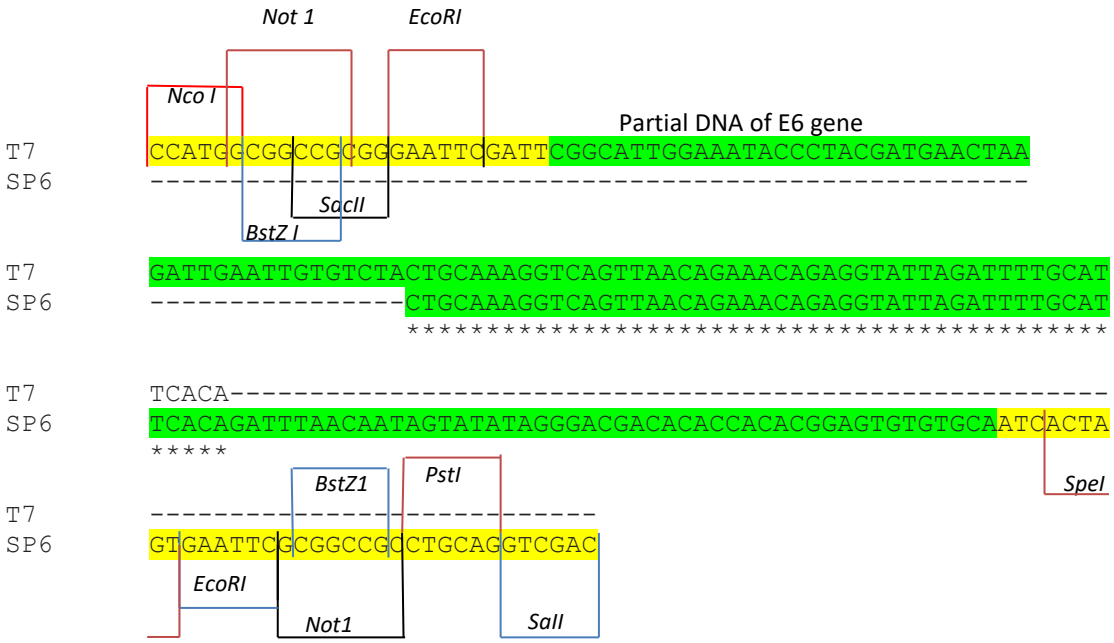
pGEM T easy sequence reference points	**Position on vector**
T7 RNA polymerase transcription initiation site	1
Multiple cloning region	10-128
SP6 RNA polymerase promoter (-17 to +3)	139-158
SP6 RNA polymerase transcription initiation site	141
pUC/M13 reverse sequencing primer binding site	176-197
lacZ start codon	188
Lac operator	200-216
β-lactamase coding region	1337-2197
Phage f1 region	2380-2835
Lac operon sequences	2836-2996; 166-395
pUC/M13 forward sequencing primer binding site	2949-2972
T7 RNA polymerase promoter (-17 to +3)	2995-3

Appendix H: Nucleotide sequences of E6 genes in HPV types -31, -18, -16 and -45 in pGEM®-T easy vector. The T7 and an HPV type specific reverse primer nucleotide sequence was aligned with the SP6 and HPV type specific forward primer nucleotide sequence.

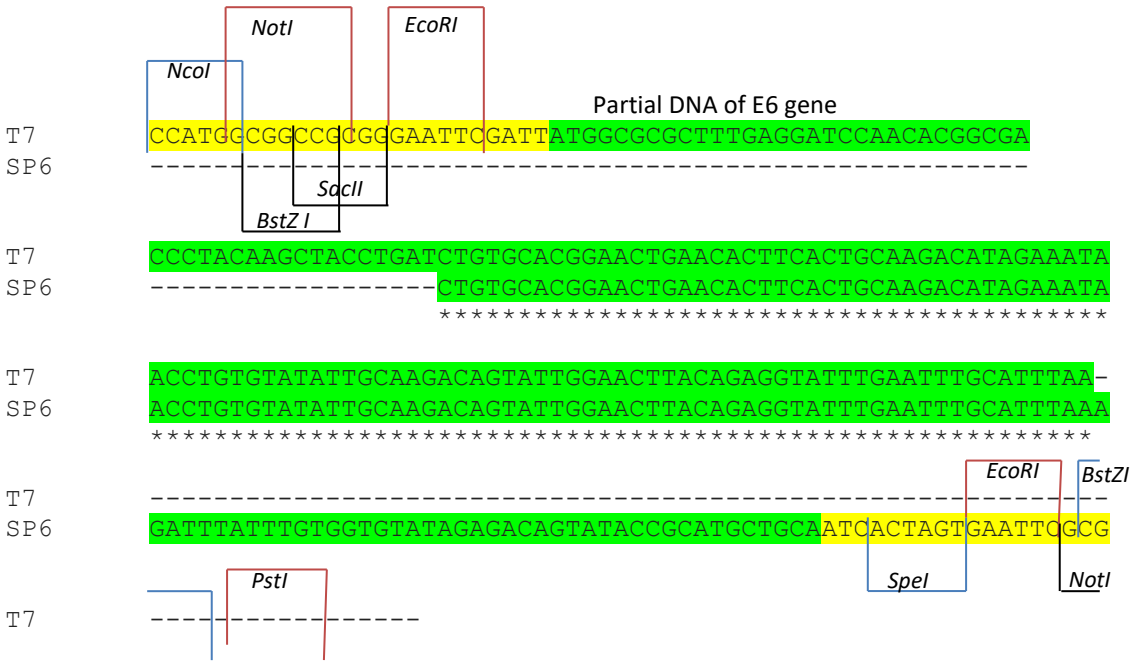
Yellow colour-multiple cloning site

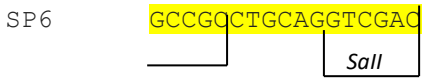
Green colour-Gene of interest

1. **HPV type -31**

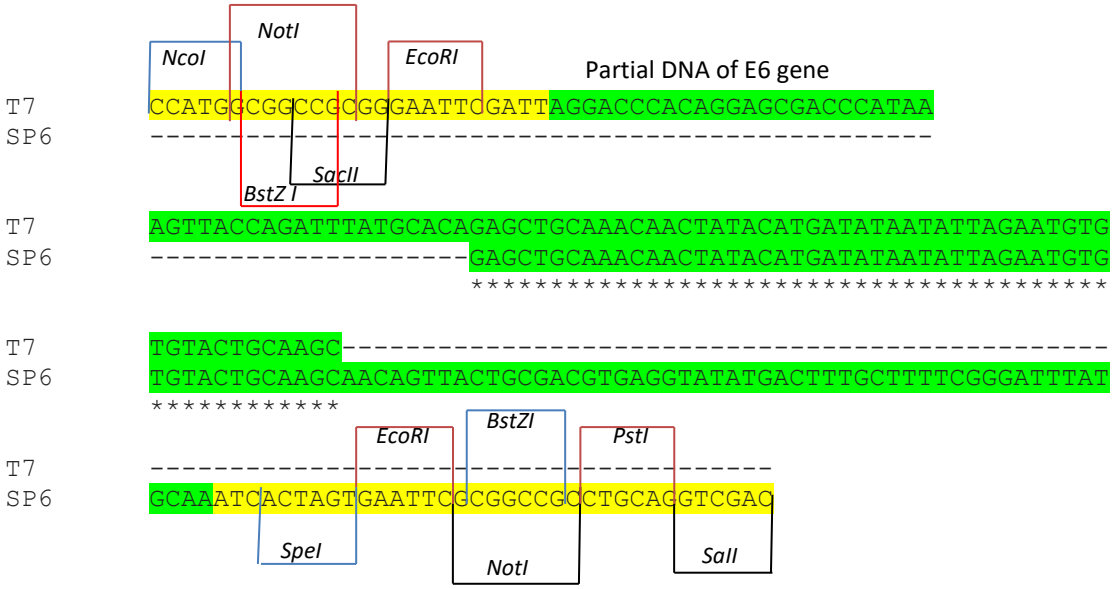


2. **HPV type -18**

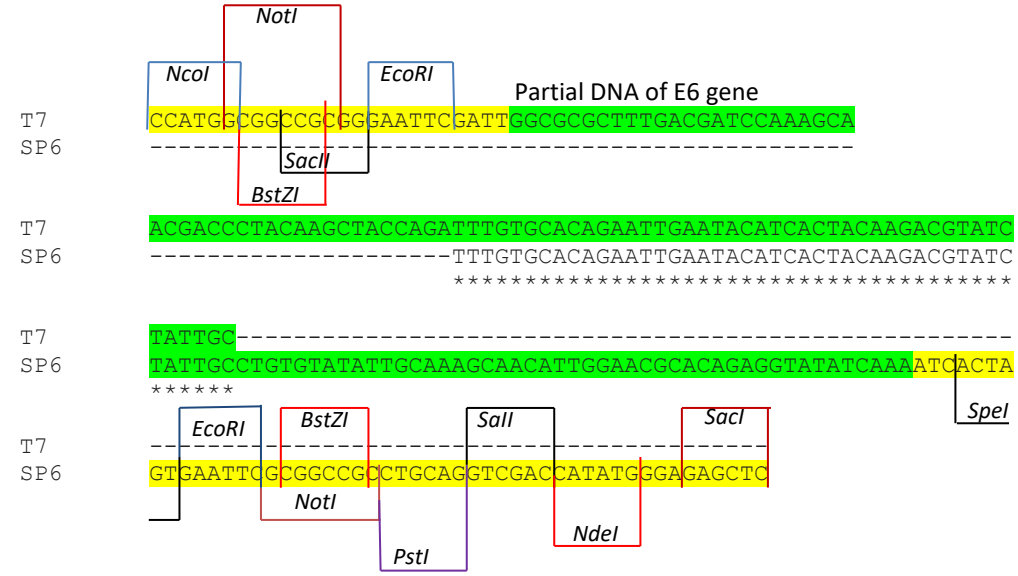




3. HPV type -16



4. HPV type -45



Appendix I: Media, buffers and solutions used.

1% Agarose gel

1. Weigh one gram of Seakem®LE agarose powder (Lonza, Maine, USA) and mix it with 100 ml of 1XTAE buffer pH 8.
2. Heat mixture in a microwave oven
3. Wait until cool and add ethidium bromide (Sigma, Missouri, USA) at final concentration of 0.5µg/ml.

2.5% Agarose gel

1. Weigh 2.5g of Seakem®LE agarose powder and mix it with 100ml of 1XTAE buffer pH 8.0.
2. Heat mixture in a microwave oven.
3. Wait until cool and add ethidium bromide at a final concentration of 0.6µg/ml.

Luria Bertani broth media

1. Mix 10g Bacto-Tryptone, 5g Bacto-Yeast extract and 10g NaCl in 900 ml H_20.
2. Adjust the pH to 7.0 with 10 M NaOH (approximately 200µl).
3. Adjust volume to 1 liter with H_2O.
4. Sterilize by autoclaving and store at room temperature.

SOC medium

1. Mix 20g Bacto-Tryptone, 5g Bacto-Yeast extract, 0.5g NaCl and 2.5ml 1 M KCl in 900 ml H_2O.
2. Adjust the pH to 7.0 with 10 M NaOH (approximately 100µl) and add H_2O to 990 ml.
3. Sterilize by autoclaving and store at room temperature.
4. Before use, add 10 ml sterile 1 M $MgCl_2$ and 20 ml sterile 1 M glucose.

50X TAE stock pH

1. Tris-Base-242g.
2. Acetate (100% acetic acid)-57.1 ml.

3. EDTA: 100ml (0.5M sodium EDTA)
4. Add dH_2O to one liter.

1XTAE (pH 8.0)

1. Dilute 20ml from 50X stock into 980 ml dH_2O.

REFERENCES

Abreu ALP, Souza RP, Gimenes F, Consolaro MEL. A review of methods for detect human papillomavirus infection. Virol J 2012; 9:264.

Acheson NH. Papillomaviruses. In Witt, K, ed. Fundamentals of molecular virology. Wiley; New Jersey 2007:114-122.

Ai W, Narahari J, Roman A. Yin yang 1 negatively regulates the differentiation-specific E1 promoter of human papillomavirus type 6. J Virol 2000; 74:5198-5205.

Al-Shabanah OA, Hafez MM, Hassan ZK, Sayed-Ahmed MM et al. Human papillomavirus genotyping and integration in ovarian cancer Saudi patients. Virol J 2013; 10:343.

Altschul SF, Gish W, Miller W, Myers et al. Basic local alignment search tool. J Mol Biol 1990; 215:403-410.

Andersen AS, Koldjaer S, Ovesen T, Rusan M. The interplay between HPV and host immunity in head and neck squamous cell carcinoma. Int J Cancer 2014; 134: 2755-2763.

Ang KK, Harris J, Wheeler R, Weber R et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Eng J Med 2010; 363:24-35.

Ashcroft M, Vousden KH. Regulation of p53 stability. Oncogene 1999; 18:7637-7643.

Barcellos RB, de Matos Almeida S, Sperhacke RD, Verza M et al. Evaluation of a novel microplate colorimetric hybridization genotyping assay for human papillomavirus. J Virol Methods 2011; 177:38-43.

Bergvall M, Melendy T, Archambault, J. The E1 proteins. Virology 2013; 445:35-56.

Bernard HU, Burk RD, Chen Z, Van Doorslaer K et al. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology 2010; 401:70-79.

Betiol J, Villa LL, Sichero L. Impact of HPV infection on the development of head and neck cancer. Braz J Med Biol Res 2013; 46:217-226.

Bishop JA, Ma XJ, Wang H, Luo Y et al. Detection of transcriptionally active high-risk HPV in patients with head and neck squamous cell carcinoma as visualized by a novel E6/E7 mRNA in situ Hybridization method. Am J Surg Pathol 2012; 36:1874-1882.

Boscolo-Rizzo PB, Delmistro AD, Bussu F, Rupato V et al. New insights into human papillomavirus-associated head and neck squamous cell carcinoma. Acta Otorhinolaryngol Ital 2013; 33:77-87.

Botha MH, Richter KL. Cervical cancer prevention in South Africa: HPV vaccination and screening both essential to achieve and maintain a reduction in incidence. S Afr Med J 2015; 105:33-34.

Boy S, Van Rensburg EJ, Engelbrecht S, Dreyer L et al. HPV detection in primary intra-oral squamous cell carcinomas – commensal, aetiological agent or contamination? J Oral Pathol Med 2006; 35:86–90.

Braakhuis BJ, Snijders PJ, Keune WJ, Meijer CJ et al. Genetic patterns in head and neck cancers that contain or lack transcriptionally active human papillomavirus. J Natl Cancer Inst 2004; 96:998-1006.

Buck CB, Day PM, Trus BL. The papillomavirus major capsid protein L1. Virology 2013; 445:169-174.

Burd EM. Human papillomavirus and cervical cancer. Clin Microbiol Rev 2003; 16:1-17.

Burk RD, Chen Z, Harari A, Smith BC et al. Classification and nomenclature system for human alphapapillomavirus variants general features, nucleotide landmarks and assignment of HPV 6 and HPV 11 isolates to variant lineages. Acta Dermatovenerol Alp Panonica Adriat 2011; 20:113-123.

Burke SC, Smith KV, Shamima S, Winkelman C. Prevalence of risk factors related to head and neck squamous cell carcinoma (HNSCC) among college students. Cancer and Oncology Research 2014; 2:7-16.

Bzhalava D, Eklund C, Dillner J. International standardization and classification of human papillomavirus types. Virology 2015; 476:341-344.

Centers for Disease Control Prevention (CDC). FDA licensure of quadrivalent human papillomavirus vaccine (HPV4, Gardasil) for use in males and guidance from the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 2010b; 59:630–632.

Chaturvedi AK, Anderson WF, Lortet-Tieulent J, Curado MP et al. Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. J Clin Oncol 2013; 31:4550-4559.

Chen JM, Guo LX, Sun CY, Sun YX. A stable and differentiable RNA positive control for reverse transcription-polymerase chain reaction. Biotechnol Lett 2006; 28:1787-1792.

Chung CH, Gillison ML. Human papillomavirus in head and neck cancer: its role in pathogenesis and clinical implications. Clin Cancer Res 2009; 15:6758-6762.

Combrinck CE, Seedat RY, Burt FJ. FRET-based detection and genotyping of HPV -6 and -11 causing recurrent respiratory papillomatosis. J Virol Methods 2013; 189:271-276.

Cooper K, Taylor L, Govind S. Human papillomavirus DNA in oesophageal carcinomas in South Africa. J Pathol 1995; 173:273-277. Cubie HA. Diseases associated with human papillomavirus infection. J Virol 2013; 445:21-34.

Cubie HA. Diseases associated with human papillomavirus infection. J Virol 2013; 445:21-34.

Cutts FT, Franceschi S, Goldie S, Castellsague X et al. Human papillomavirus and HPV vaccines: a review. Bull World Health Organ 2007; 85:719-726.

Das BC, Gapalkrishna V, Das DK, Sharma JK et al. Human papillomavirus DNA sequences in adenocarcinoma of the uterine cervix in Indian women. Cancer 1993; 72:147-153.

Davidson CL, Richter KL, Van der Linde M, Coetsee J et al. Prevalence of oral and oropharyngeal human papillomavirus in a sample of South African men: a pilot study. S Afr Med J 2014; 104:358-361.

de Freitas AC, Gurgel AP, Chagas BS, Coimbra EC et al. Susceptibility to cervical cancer: an overview. Gynecol Oncol 2012; 126:304–311.

de Roda Husman AM, Walboomers JM, van den Brule AJ, Meijer CJ et al. The use of general primers GP5 and GP6 elongated at their 3'ends with adjacent highly conserved sequences improves human papillomavirus detection by PCR. J Gen Virol 1995; 76:1057–62.

de Villiers EM, Fauquet C, Broker TR, Bernard HU et al. Classification of papillomaviruses. Virology 2004; 324:17-27.

DiMaio D, Petti LM. The E5 proteins. Virology 2013; 445:99-114.

Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. Clin Sci (Lond) 2006; 110:525–541.

Doorbar J, Quint W, Banks L, Bravo IG et al. The biology and life-cycle of human papillomaviruses. Vaccine 2012; 30 suppl 5:F55–70.

D'Souza G, Dempsey A. The role of HPV in head and neck cancer and review of the HPV vaccine. Prev Med 2011; 55 suppl 1:S5-S11.

D'Souza G, Kluz N, Wentz A, Youngfellow RM et al. Oral human papillomavirus (HPV) infection among unvaccinated high risk young adults. Cancers 2014; 6:1691-1704.

El-Naggar K, Westra WH. p16 epxression as a surrogate marker for HPV-related oropharyngeal carcinoma: a guide for interpretative relevance and consistency. Head Neck 2011; 34:459-461.

Entiauspe L, Nunes E, Collares T, da Silveira MF et al. Comparison between two methods for molecular characterization of human papillomavirus. J bras Doencas Sex Transm 2013; 25:13-15.

Erickson BK, Alvarez RD, Huh WK. Human papillomavirus: what every provider should know. Am J Obstet Gynecol 2013; 208:169-175.

Evander M, Frazer IH, Payne E, Qi YM et al. Identification of the alpha 6 integrin as a candidate receptor for papillomaviruses. J Virol 1997; 71:2449-2456.

Fakhry C, Westra WH, Li S, Cmelak A et al. Improved survival of patients with human papillomavirus positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst 2008; 100:261–269.

Favre M, Ramoz N, Orth R. Human papillomaviruses: general features. Clin Dermatol 1997;

15:181-198.

Furniss CS, McClean MD, Smith JF, Bryan J et al. Human papillomavirus 6 seropositivity is associated with risk of head and neck squamous cell carcinoma, independent of tobacco and alcohol use. Ann Oncol 2009; 20:534-541.

Garland SM, Smith JS. Human papillomavirus vaccines. Drugs 2010; 70: 1079-1098.

Gavid M, Pillet S, Pozzetto B, Oriol M et al. Human papillomavirus and head and neck squamous cell carcinomas in the South-East of France: prevalence, viral expression, and prognostic implications. Acta Oto-Laryngologica 2013; 133: 538–543.

Gillison ML, Koch WM, Capone RB, Spafford M et al. Evidence for casual association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst 2000; 92:709-720.

Graham SV. Human papillomavirus: gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. Future Microbiol 2010; 5:1493-1506.

Gravitt PE, Peyton CL, Alessi TQ, Wheeler CM et al. Improved amplification of genital human papillomaviruses. J Clin Microbiol 2000; 38:357-361.

Haedicke J, Iftner T. Human papillomaviruses and cancer. Radiother Oncol 2013; 108:397–402.

Hille JJ, Markowitz S, Margolius KA, Isaacson C. Human papillomavirus and carcinoma of the esophagus. N Eng J Med 1985; 312:1707.

Hille JJ, Margolius KA, Markowitz S, Isaacson C. Human papillomavirus infection related to oesophageal carcinoma in black South Africans. A preliminary study. S Afr Med J 1986; 69:417-420.

Howie HL, Katzenellenbogen RA, Galloway DA. Papillomavirus E6 protein. Virology 2009; 384:324-334.

Howley PM, Lowy DR. Papillomaviruses. In: fields virology. Knipe DM, Howley PM (Eds). Lippincott Williams and Wilkins PA. USA 2007:2299-2354.

Husnjak K, Grce M, Magdi'c L, Paveli'c K. Comparison of five polymerase chain reaction methods for detection of human papillomavirus in cervical cell specimens. J Virol Methods 2000; 83:125-139.

Johansen J, Eriksen JG. Trends in cancer of the head and neck in the elderly in Denmark, 1980-2012. Acta Oncologica 2016; 55:13-18.

Joyce JG, Tung JS, Przysiecki CT, Cook JC et al. The L1 major capsid protein of human papillomavirus type 11 recombinant virus-like particles interacts with heparin and cell surface glycosaminoglycans of human keratinocytes. J Biol Chem 1999; 274:5810-5822.

Juckett G, Hartman-Adams H. Human papillomavirus: Clinical manifestation and prevention. Am Fam Phys 2010; 82:1209-1214.

Jung AC, Briolat J, Millon R, de Reynie's A et al. Biological and clinical relevance of transcriptionally active human papillomavirus (HPV) infection in oropharynx squamous cell carcinoma. Int J Cancer 2010; 126:1882-1894.

Kermani AI, Seifi SH, Dolatkhah R, Sakhinia E et al. Human papillomavirus in head and neck squamous cell cancer. Iran J Cancer Prev 2012; 5:21-26.

Kocjan BJ, Seme K, Poljak M. Comparison of the Abbott RealTime High Risk HPV test and INNO-LiPA HPV genotyping extra test for the detection of human papillomaviruses in formalin-fixed, paraffin-embedded cervical cancer specimens. J Virol Methods 2011; 175:117-119.

Kreimer AR, Clifford GM, Boyle P, Franceschi S. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. Cancer Epidemiol Biomarkers Prev. 2005; 14:467-475.

Krupar R, Hartl M, Wirsching K, Dietmaier W et al. Comparison of HPV prevalence in HNSCC patients with regard to regional and socioeconomic factors. Eur Arch Otorhinolaryngol 2014; 271: 1737-45.

Kumar R, Rai AK, Das D, Das R et al. Alcohol and tobacco increases risk of high risk HPV infection in head and neck cancer patients: study from the North-East region of India. PLoS One 2015; 10:e0140700.

Lajer CB, Von Buchwald C. The role of human papillomavirus in head and neck cancer. APMIS 2010; 118:510-519.

Langevin SM, Grandis JR, Taioli E. Female hormonal and reproductive factors and head and neck squamous cell carcinoma risk. Cancer Lett 2011; 310: 216–221.

Larque AB, Hakim S, Ordi J, Nadal A et al. High-risk human papillomavirus is transcriptionally active in a subset of sinonasal squamous cell carcinomas. Mod Pathol 2014; 27: 343–351.

Lassen P. The role of human papillomavirus in head and neck cancer and the impact on radiotherapy outcome. Radiother Oncol 2010; 95: 371-380.

Lassen P, Eriksen JG, Krogdahl A et al. The influence of HPV-associated p16-expression on accelerated fractionated radiotherapy in head and neck cancer: evaluation of the randomised DAHANGA 6&7 trial. Radiother Oncol 2011; 100:49-55.

Lazarczyk M, Cassonnet P, Pons C, Jacob Y et al. The EVER proteins as a natural barrier against papillomaviruses: a new insight into the pathogenesis of human papillomavirus infections. Microbial Mol Biol Rev 2009; 73:348-370.

Levin MJ, Moscicki AB, Song LY, Fenton T et al. Safety and immunogenicity of a quadravilent human papillomavirus (types 6, 11, 16 and 18) vaccine in HIV-infected children 7 to 12 years old. J Acquir Immune Defic Syndr 2010; 55:197-204.

Ljubojevic S, Skerlev M. HPV associated diseases. Clin Dermatol 2014; 32:227-234.

Longworth MS, Laimins LA. Pathogenesis of human papillomaviruses in differentiating epithelia. Microbiol Mol Biol Rev 2004; 68:362-372.

Mail and Guardian. Motsoaledi launches free HPV vaccine for school girls 2014.

Mannarini L, Kratochvil V, Calabrese L, Gomes-Silva L et al. Human papilloma virus (HPV) in head and neck region: review of literature. Acta Ortorhinolaryngol Ital 2009; 29:119-126.

Manos MM, Ting Y, Wright DK, Lewis J et al. The use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. Cancer Cells 1989; 7:209 –214

Matsha T, Erasmus R, Kafuko AB, Mugwanya D et al. Human papillomavirus associated with oesophageal cancer. J Clin Pathol 2002; 55:587-590.

McBride AA. The papillomavirus E2 proteins. Virology 2013; 445 (1-2):57-79.

McLaughlin-Drubin ME, Műnger K. The human papillomavirus E7 oncoprotein. Virology 2009; 384:335-344.

Miller DL, Puricelli MD, Stack MS. Virology and molecular pathogenesis of HPV (human papillomavirus)-associated oropharyngeal squamous cell carcinoma. Biochem J 2012; 443:339-353.

Mirghani H, Amen F, Moreau F, Guigay J et al. Human papilloma virus testing in oropharyngeal squamous cell carcinoma: what the clinicians should know. Oral Oncol 2014; 50:1-9.

Molano M, Van der Brule AJC, Posso H, Werderpass E et al. Low grade squamous cell intra-epithelial lesions and human papillomavirus infection in Colombian women. Br J Cancer 2002; 87:1417-1421.

Molijn A, Kleter B, Quint W, van Doom LJ. Molecular diagnosis of human papillomavirus (HPV) infections. J Clin Virol 2005; 32 Suppl 1:S43-S51.

Morbini P, Alberizzi P, Tinelli C, Paglino C et al. Identification of transcriptionally active HPV infection in formalin-fixed, paraffin-embedded biopsies of oropharyngeal carcinoma. Hum Pathol 2015; 46: 681-689.

Mork J, Lie AK, Glattre E, Hallmans G et al. Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. N Engl J Med 2001; 344:1125-1131.

Morris BJ. Cervical human papillomavirus screening by PCR: advantages of targeting the E6/E7 region. Clin Chem Lab Med 2005; 43:1171-1177.

Morshed K, Polz-Gruszka D, Szymański M, Polz-Dacewicz M. Human papillomavirus (HPV) structure, epidemiology and pathogenesis. Otolaryngol Pol 2014; 68:213-219.

Muńoz N, Castellsaguè X, de Gonzàlez AM, Gissmann L. Chapter 1: HPV in the etiology of human cancer. Vaccine 2006; 24 Suppl 3:S3/1-10.

Otero-Motta AP, Ordonez JL, Gonzales-Celador R, Rivas B et al. Prevalence of human papillomavirus genotypes in cytologic abnormalities from unvaccinated women living in north-western Spain. AMPIS 2011; 119:204-2015.

Pai SI, Westra WH. Molecular pathology of head and neck cancer: implications for diagnosis, prognosis and treatment. Annu Rev Pathol 2009; 4:49–70.

Pannone G, Santoro A, Papagerakis S, Lo-Muzio L et al. The role of human papillomavirus in the pathogenesis of head and neck squamous cell carcinoma: an overview. Infect Agent Cancer 2011; 6:4.

Panwar A, Batra R, Lydiatt WM, Ganti AK. Human papillomavirus positive oropharyngeal squamous cell carcinoma: a growing epidemic. Cancer Treat Rev 2014; 40:215-219.

Paquette C, Evans MF, Meer SS, Rajendran V et al. Evidence that alpha-9 papillomavirus infections are a major etiologic factor for oropharyngeal carcinoma in black South Africans. Head Neck Pathol 2013; 7:361-372.

Petrosky E, Bocchini JA Jr, Hariri S, Chesson H et al. Use of 9-valent human papillomavirus (HPV) vaccine: updated HPV vaccination recommendations of the advisory committee on the immunization practices. MMWR Morb Mortal Wkly Rep 2015; 64:300-304.

Piana A, Sotgiu G, Castiglia P, Pischedda S et al. Molecular methods for the detection of human papillomavirus infection: new insights into their role in diagnostics and epidemiological surveillance. Ital J Public Health 2009; 6:164-171.

Poljak M, Kocjan BJ, Ostrbenk A, Seme K. Commercially available molecular tests for human papillomaviruses (HPV): 2015 update. J Clin Virolo 2016; 76:S3-S13.

Printz C. FDA approves Gardasil 9 for more types of HPV. Cancer 2015; 121:1156-1157.

Psyrri A, DiMaio D. Human papillomavirus in cervical cancer and head-and-neck-cancer. Nat Clin Pract Oncol 2008; 5:24-31.

Rampias T, Sasaki C, Psyrri A. Molecular mechanisms of HPV induced carcinogenesis in head and neck. Oral Oncol 2014; 50:356-363.

Rautavara J, Kuuskosbi J, Syrjänen K, Greenman R. HPV genotypes and their prognostic significance in head and neck squamous cell carcinoma. J Clin Virol 2012; 53:116-120.

Robinson M, Sloan P, Shaw R. Refining the diagnosis of oropharyngeal squamous cell carcinoma using human papillomavirus testing. Oral Oncol 2010; 46:492-496.

Roman A, Munger K. The papillomavirus E7 proteins. Virology 2013; 445 (1-2):138–168.

Rusan M, Ovesen T. Worldwide prevalence of human papillomavirus in tonsillar squamous cell carcinoma and tumor-free tonsillar tissue. Ortolaryngol 2012; S2:001.

Ruttkay-Nedecky B, Jimenez-Jimenez AM, Nejdl L, Chudobova D et al. Relevance of infection with human papillomavirus: The role of the p53 tumor suppressor protein and E6/E7 zinc finger proteins Int J Oncol 2013; 43:1754-1762.

Saini R, Santhanam J, Othman NH, Saini D et al. Single-tube semi nested PCR assay for detecting human papillomavirus in clinical samples. Open Microbiol J 2009; 3:106-112.

Sambrook J, Russel DW. Molecular cloning: a laboratory manual (3rd edition). New York: Cold Springer Harbor Laboratory; 2001.

Santarelli A, Lo-Russo L, Bambini F, Campisi G, Lo-Muzio L. New perspectives in medical approach to therapy of head and neck squamous cell carcinoma. Minerva Stomatol 2009; 58:445-452.

Serrano B, Alemany L, Tous S, Bruni L et al. Potential impact of a nine-valent vaccine in human papillomavirus related to cervical disease. Infect Agent Cancer 2012; 7:38.

Sedaghat AR, Zhang Z, Begum S, Parlemo R et al. Prognostic significance of human papillomavirus in oropharyngeal cell carcinoma. Laryngoscope 2009; 119:1542-1549.

Seedat RY, Thukane M, Jansen AC, Rossouw I et al. HPV types causing juvenile recurrent laryngeal papillomatosis in South Africa. Int J Pediatr Otorhinolaryngol 2010; 74:255-259.

Smith EM, Rubenstein LM, Hoffman H, Haugen TH et al. Human papillomavirus, p16 and p53 expression associated with survival of head and neck cancer. Infect Agent Cancer 2010; 5:4.

Snijders PJF, Heideman DAM, Meijer CJLM. Methods for HPV detection in exfoliated cell and tissue specimens. AMPIS 2010; 118:520-528.

Snow AN, Laudadio J. Human papillomavirus detection in head and neck squamous cell carcinomas. Adv Anal Pathol 2010; 17:394-403.

Syrjänen K, Syrjänen S, Lamberg M, Pyrhönen S et al. Morphological and immunohistochemical evidence suggesting human papillomavirus (HPV) involvement in oral squamous cell carcinogenesis. Int J Oral Surg 1983; 12:418–424.

Syrjänen S. Human papillomavirus (HPV) in head and neck cancer. J Clin Virol 2005; 32 (suppl 1):59–66.

Syrjänen S. Current concepts on human papillomavirus infections in children. APMIS 2010a; 118 (6-7):494-509.

Syrjänen S. The role of human papillomavirus infection in head and neck cancers. Ann Oncol Suppl 7 2010b; 21: vii243–vii245.

Togawa K, Jaskiewiez k, Takahashi H, Meltz SJ et al. Human papillomavirus DNA sequences in esophagus squamous cell carcinoma. Gastroenterology 1994; 107:128-136.

Torrente MC, Rodrigo JP, Halgentz M Jr, Dikkers FG et al. Human papillomavirus infections in laryngeal cancer. Head Neck 2011; 33:581-586.

Torres M, Fraile L, Echevarria JM, Novoa BH et al. Human Papillomavirus (HPV) genotyping: Automation and application in routine laboratory testing. Open Virol 2012; 6:144-150.

Van Rensburg EJ, Van Heerden WF, Venter EH, Raubenheimer EJ. Detection of human papillomavirus DNA with in situ hybridisation in oral squamous carcinoma in a rural black population. S Afr Med J 1995; 85:894-896.

Van Rensburg EJ, Engelbrecht S, Van Heerden WF, Raubennheimer EJ et al. Human papillomavirus DNA in oral squamous cell carcinomas from Africa population sample. Anticancer Res 1996; 16:969-973.

Venceslau EM, Bezerra MM, Lopes ACM, Souza EV. HPV detection using primers MY09/MY11 and GP5+/GP6+ in patients with cytologic and/or colposcopic changes. J Bras Med Lab 2014; 50:280-285.

Wang XI, Thomas J, Zang S. Changing trends in human papillomavirus–associated head and neck squamous cell carcinoma. Ann Diagn Pathol 2012; 16:7–12.

Wang JW, Roden RB. L2, the minor capsid protein of papillomavirus. Virology 2013; 445:175-186.

Williamson AL, Jaskiesicz K, Gunning A. The detection of human papillomavirus in oesophageal lesions. Anticancer Res 1991; 11:263-265.

Winder DM, Ball SL, Vaughan K, Hanna N et al. Sensitive HPV detection in oropharyngeal cancers. BMC Cancer 2009; 9:440.

WHO International Agency for Research on Cancer (IARC). Human papillomaviruses. In: IARC Monographs on the evaluation of carcinogenic risks to humans. Lyon: WHO IARC 2007.

World Health Organization. Human papillomavirus vaccine: WHO position paper. WER 2014; 89:465-492.