Human Papillomavirus Integration and Methylation Events and Cervical Disease Progression Post-Vaccination

A thesis submitted for the Degree of Doctor of Philosophy at Cardiff University

By Rachel Louise Baldwin

Supervisors

Dr Sam Hibbitts

Dr Amanda Tristram

Professor Gavin Wilkinson

HPV Research Group

Cardiff University

2017

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed …………………………………………(candidate) Date …………………

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of, PhD

Signed ……………………………………(candidate) Date ……………………

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed …………………………………………(candidate) Date …………………

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed …………………………………………(candidate) Date …………………

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.

Signed …………………………………………(candidate) Date …………………

Acknowledgements

I would not be submitting this thesis if it wasn't for the unending support, advice and belief of Dr Sam Hibbitts, who believed in me when I didn't believe in myself and supported me at my best and my worst. Thank you also to Dr Amanda Tristram and Professor Gavin Wilkinson who have helped shape my research and this thesis. Thank you also to goes to Tenovus who funded this research.

In the HPV Group I have met many wonderful and helpful people; thank you to Tiffany, Áine, Rachel, Dean, Evelyne, Stefan, Sadie, Jon and Ned. A huge thank you is due to Jo Jones and Angharad Edwards, two wonderful friends and teachers who only laughed a little when I thought the Plant Room was where plants are grown!

Thank you to my family. To my parents; Mum, Dad, Alan and Carol, thank you for always being my biggest supporters. To my sisters Alana and Antonia, who inspire me to be happy. To my Gran, who sends the best care packages. To my new parents, Margaret and Stephen, thank you for making me part of your family. To Adele, Dini, Adam, Suzy and Claire, who have become my family; the others didn't have a choice but you guys did - thanks!

Finally, thank you to my extraordinary wife, Hayley, who keeps me focused on the bigger picture, inspires me to be the best I can be and repeatedly tells me to "get a grip." This thesis would not have been possible without your support (nagging) and encouragement (more nagging). In all seriousness, thank you for everything you have done for me over the past four years. This thesis is dedicated to you!

Collaborators

This thesis would not have been possible if it wasn't for the support, advice and assistance of numerous collaborators including Dr Sally Roberts (Birmingham University) who advised in the production of the raft culture, provided w12 Ser4B cells and also carried out the HPV E4 IHC.

Dr C Dawson from Birmingham University who carried out the SHH and GLI-I IHC and provided details of SHH expression in VIN cultures carried out in Birmingham.

Cardiff University's Central Biotech Services carried out the RNAseq presented in chapter 5. Thanks also go to and to Dr Kevin Ashelford and Dr Peter Giles from Wales Gene Park who assisted in the RNAseq library construction and advised in the analysis.

Dr Owen Weeks, an independent pathologist from Cardiff University took histology images presented in Chapter 5. Dr Weeks also assisted with the analysis and interpretation of pathology results.

Thank you to Professor Robert Hills from Cardiff University for his help and advice on the statistical analysis presented in this thesis.

Abbreviations

9v-HPV	9-valent HPV
APOT	Amplification of Papillomavirus Transcripts
BAK	Bcl-2 Homologous Antagonist/Killer
BIDD	Biomarkers and Imaging Discovery and Development Committee
BLAST	Basic Local Alignment Search Tool
bp	base pairs
cDNA	complementary DNA
CFS	common fragile sites
CIN	Cervical Intraepithelial Neoplasia
CNRQ	calibrated normalized relative quantities
CSW	Cervical Screening Wales
DAPI	4', 6'-diamidino-2-phenylindole
DIPS	Detection of Integrated Papillomavirus Sequences
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	deoxiribose nucleotide triphosphate
DSB	double stranded breaks
E6AP	E6 Adaptor Protein
E. coli	*Escherichia coli*
EGF	epithelial growth factor
EMBL-EBI	European Molecular Biology Laboratory – The European Bioinformatics Institute

ENA	European Nucleotide Archive
EZH2	*Enhancer of Zeste Homolog 2*
FISH	fluorescent in situ hybridization
FBS	foetal bovine serum
GLI-1	Zinc Finger Protein GLI1
GMEM	Glasgow Modified Eagle's Medium
HC2	Hybrid Capture 2
HEKn	human epithelial keratinocytes, neonates
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HI-BCD	Human Papillomavirus Integration as a Biomarker of Cervical Disease
HP/P	high positive/positive
HPRT	*Hypoxanthine Phosphoribosyltransferase*
HPV	Human Papillomavirus
HR-HPV	High-Risk Human Papillomavirus
hTERT	*Human Telomerase Reverse Transcriptase*
ICC	Invasive Cervical Carcinoma
IHC	immunohistochemistry
IL12A	Interleukin-12 Alpha
IPTG	isopropyl β-D-1-thiogalactopyranoside
ISH	*in-situ* hybridization
kb	kilo base
LB	Lysogeny Broth
LBC	liquid based cytology

LCR	long control region
LLETZ	large loop excision of the transformation zone
LR-HPV	low-risk Human Papillomavirus
miRNA	micro ribonucleic acid
mRNA	messenger ribonucleic acid
NCBI	National Institute of Biotechnology Information
NCR	non-coding region
NHS	National Health Service
ORF	open reading frame
PCR	polymerase chain reaction
qRT PCR	quantitative real-time polymerase chain reaction
RIN	ribonucleic acid integrity number
RNA	ribonucleic acid
RS-PCR	Restriction Site Polymerase Chain Reaction
SCC	Squamous Cell Carcinoma
SHH	Sonic Hedgehog Protein
SOC	super optimal broth with catabolite repression
TAE	tris-acetate-EDTA
TBP2	*TATA-Binding Protein 2*
URR	upstream regulatory region
VaIN	Vaginal Intraepithelial Neoplasia
VIN	Vulva Intraepithelial Neoplasia
VLP	virus like particle
X-gal	5-bromo-4-chloro-indolyl-β-D-galactopyranoside

List of Figures

List of Tables

Summary

Human papillomavirus infection is regulated by multiple factors including methylation, viral integration and aberrations in host and viral gene expression. In most patients, the infection is transient and cleared effectively by the host's immune system but in a minority of cases, the Human papillomavirus infection progresses to neoplasia and cancer if not detected and treated. Assays that detect the presence of Human Papillomavirus infection alone are not good prognostic markers of clinical outcome. Alternative clinical biomarkers that can measure other regulatory factors of Human Papillomavirus infection are required to more accurately predict patient outcome and help direct treatment options specifically to patients at risk of neoplasia and cancer progression.

This study aimed to examine several Human Papillomavirus regulatory factors to determine if they would be suitable as clinical biomarkers and explore further the link between molecular changes and associated pathology. This involved development of assays, application to samples obtained from different cohorts of women to ascertain prognostic validity and development of an *in vitro* system to model the *in vivo* pathology.

Initial work focused on investigating Human Papillomavirus 31, 33, 35 and 51 integration and methylation assays as prognostic biomarkers in young women attending their first routine cervical smear. Results indicated that Human Papillomavirus *E2* gene disruption and methylation were not common events and the assays investigated were not suitable biomarkers for predicting clinical outcome in

young women. The assays were then applied to clinical samples taken from patients with varying grades of cervical disease and high levels of viral methylation were present in women with high grade disease (Cervical Intraepithelial Neoplasia II+) and a correlation between methylation and HPV gene disruption was shown.

Novel *in vitro* organotypic raft cultures of cells from Vulval Intraepithelial Neoplasia and Vaginal Intraepithelial Neoplasia were employed to understand how molecular changes in viral methylation and gene expression correlated with observed pathology. The organotypic raft cultures showed diverse differentiation patterns. No correlation of pathology with viral integration status was detected. However, organotypic raft cultures with high-grade disease morphology all had a higher level of methylation and expression of regulatory genes p16, Ki-67 and Sonic Hedgehog in comparison cultures displaying a histology consistent with low-grade disease. p16 and Ki-67 are already being examined as part of cervical screening triage.

The findings presented in this thesis, highlight a need for further research into Human Papillomavirus infection and the molecular changes associated with Sonic Hedgehog gene expression and viral methylation as these show promise as prognostic biomarkers.

Contents

Chapter 1

Background

Human Papillomavirus (HPV) is a non-enveloped deoxyribonucleic acid (DNA) virus; approximately 8 kilo base (kb), which commonly infects epithelial cells. Most infections resolve spontaneously following the induction of an effective host immune response. Occasionally persistent infection with a High-Risk HPV (HR-HPV) type can lead to invasive disease. HPV causes almost half a million cancers each year worldwide, most of these in the cervix. Progress in understanding of the molecular biology of HPV has led to the development of a highly effective prophylactic vaccine that will reduce the incidence of anogenital cancer further. The optimal model of cervical screening in this vaccinated population still needs to be refined. Molecular biomarkers of disease presence and progression are a potential feature of future cervical screening programmes. More work is needed to define at which stage in anogenital disease these biomarkers can be used either discretely or independently.

1.1 Molecular biology of HPV

Viruses in the *Papillomaviridae* family have been detected in a wide variety of species and are host specific. 170 HPV types have been identified and are organised into genera, Alpha-, Beta-, Gamma-, Mu-, Nu-Papillomavirus primarily

based on the sequence of the *L1* gene (de Villiers et al., 2004). Alpha-HPV types are responsible for a range of benign warts (e.g. HPV 6 and 11) and 15 Alpha-HPV types have been classified as HR-HPV that have the potential to cause neoplastic changes (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, and 82) (Muñoz et al., 2003). This thesis will focus almost exclusively on HR-HPV types.

The Alpha-HPV types have circular ~8 kb double stranded DNA genomes that encode eight major proteins designated as either early (E) or late (L) according to when they are expressed in the viral lifecycle. The circular genome contains a non-coding regulatory region (NCR), also known as the upstream regulatory region (URR) or long-control region (LCR). The early proteins; E1, E2, E4, E5, E6 and E7, have regulatory functions and the late proteins, L1 and L2, are major and minor structural proteins respectively (Figure 1.1).
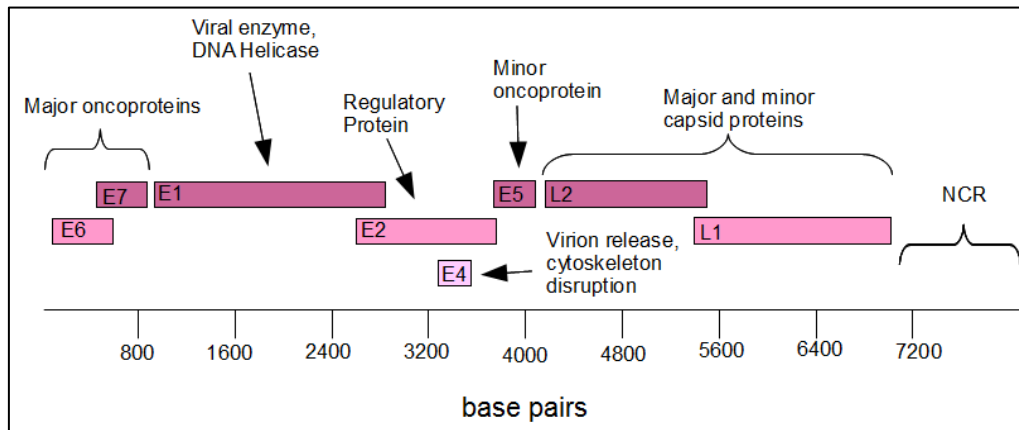
Figure 1.1. A schematic of the linearized HPV genome. The functions of the early and late viral proteins are highlighted and their location in the HPV genome. E = early protein; L = late protein; NCR = non-coding region.

1.1.1 Late gene expression: *L1* and *L2*

L1 and L2 are the major and minor capsid proteins that form the outer surface of the virus. New virus particles are made from 72 pentamers of L1 protein in a T = 7 icosahedral lattice (Conway and Meyers, 2009). The protein is folded with variable regions on the outer surface and the highly-conserved regions on the inside. L2 protein is located on the inner surface of the viral capsid and is much less abundant than L1. When L1 is produced in isolation it spontaneously forms an icosahedral lattice thus forming a virus like particle (VLP) (Joyce et al., 1999). VLPs consisting of pure L1 protein constitute the antigen in the HPV prophylactic vaccine that is proving

highly effective at reducing the incidence of HPV infection around the world in the cervical cancer prevention programmes (Kemp et al., 2011).

1.1.2 Functional Genes: *E1, E2* and *E4*

E1 encodes the only virally encoded enzyme, DNA helicase, which allows the HPV DNA to unwind and transcription of the viral genome to occur (Wilson et al., 2002).

E2 protein has many important functions in the HPV lifecycle, with four binding sites for E2 located in conserved regions of the NCR. Through interactions with these binding sites E2 can promote or repress viral gene expression. When *E2* is expressed in low quantities the protein only binds to the first binding site. This promotes transcription and *E6* and *E7* oncogenes are expressed. When high levels of E2 are present, the protein binds to all four promoters and transcription from the early promoter is suppressed (Figure 1.2) (Demeret et al., 1995, Thierry, 2009, Thierry and Yaniv, 1987).

E2 has a variety of additional functions that ensure the virus lifecycle is correctly maintained and prevents over-expression of the viral oncogenes (Blachon et al., 2005, Gammoh et al., 2009, McPhillips et al., 2005). This includes increasing the binding affinity of the E1 protein to the origin of replication.

The HPV E4 protein is produced from an E1^E4 spliced messenger ribonucleic acid (mRNA) (Doorbar et al., 2015). In productive infections, high levels of this protein are found and E4 expression can be used to establish an active HPV infection in which late HPV genes are expressed (Griffin et al., 2012). HR-HPV E1^E4 contains a leucine cluster (LLKLL) at the N-terminus region which is essential for interaction between E1^E4 and keratin (Wang et al., 2003). Keratin is a structural protein, which is part of the cellular cytoskeleton that contributes to the stability of the cell (Moll et al., 2008). This binding of E1^E4 to keratin causes destabilisation of the keratin network. Although not fully understood it is thought that this interaction between E1^E4 and keratin assists with virion release and subsequent spread of the virus (Doorbar et al., 2015).
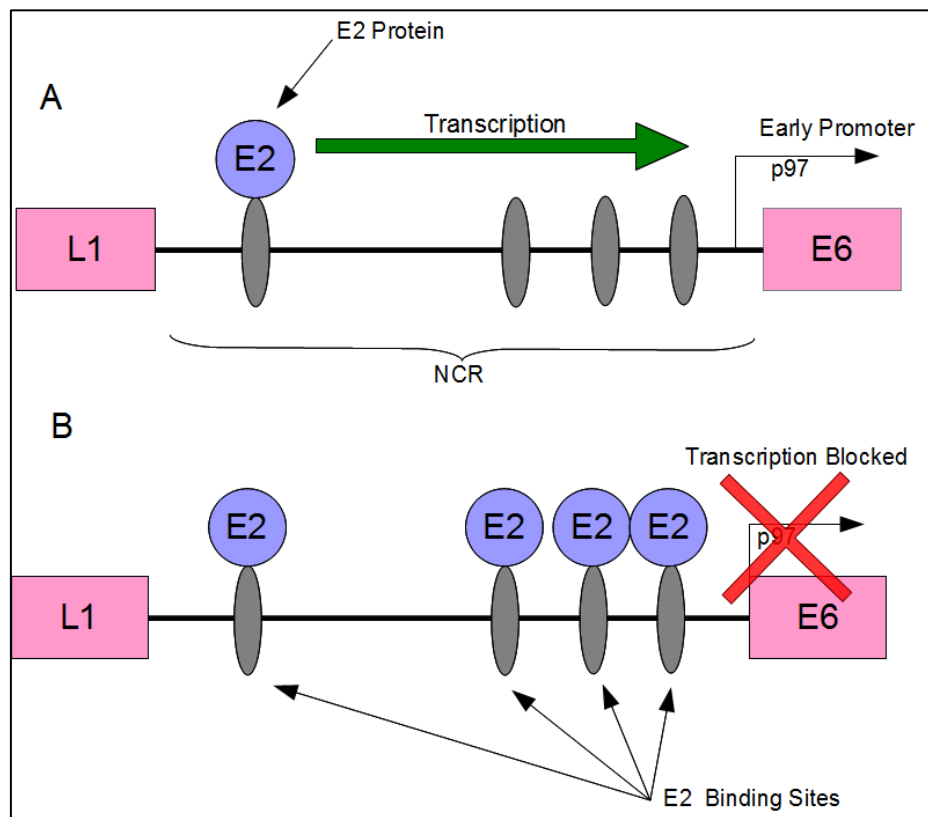
Figure 1.2. The *E2* negative feedback loop controlling HPV genome expression.
(A) When low levels of E2 are produced, transcription can occur and early genes can be expressed from the p97 early promoter. **(B)** When high levels of E2 are expressed E2 proteins bind to all four E2 binding sites which blocks transcription from the early promoter. This acts as a negative feedback loop to control unregulated expression of the HPV genome.

1.1.3 Viral Oncogenes: *E5, E6* and *E7*

The virus produces three proteins which contribute to the malignant transformation of the host cell: E5, E6 and E7. Changes in their levels of expression may lead to modifications to the normal cell cycle and subsequent DNA damage (Moody and Laimins, 2010). E6 and E7 proteins are responsible for oncogenic transformation and immortalisation of host cells (Hawley-Nelson et al., 1989, Münger et al., 1989). E6 protein contributes to cell transformation and immortalisation by coupling with and degrading the tumour suppressor protein p53, via an ubiquitin-proteasome pathway. p53 protein is an important factor involved in cell cycle repair and can subsequently induce apoptosis of damaged cells. The loss of p53 allows damaged DNA to accumulate without apoptosis (Scheffner et al., 1990, Howie et al., 2009). E6 also contributes to cell immortalisation by interacting and increasing activity of the human telomerase reverse transcriptase (hTERT) protein and inhibition of Bcl-2 homologous antagonist/killer (BAK) proteins (Liu et al., 2009, Thomas et al., 1999, Thomas and Banks, 1998).

The predominant oncogenic property of E7 is to bind and degrade pocket proteins, mainly pRB, which is involved in the regulation of the cell cycle. Disruption of this protein allows the cell to continue though the G1/S cell check point and contributes to the immortalisation of the infected cell (McLaughlin-Drubin and Munger, 2009a, Munger et al., 2001). E7 also has several secondary functions contributing to cell immortalization. These include decreasing levels of the Claspin protein leading to defective DNA repair and increasing levels of the gene *enhancer of zeste homologue 2* (*EZH2*) which inhibits apoptosis (Holland et al., 2008, Spardy et al., 2009) (Figure 1.3).
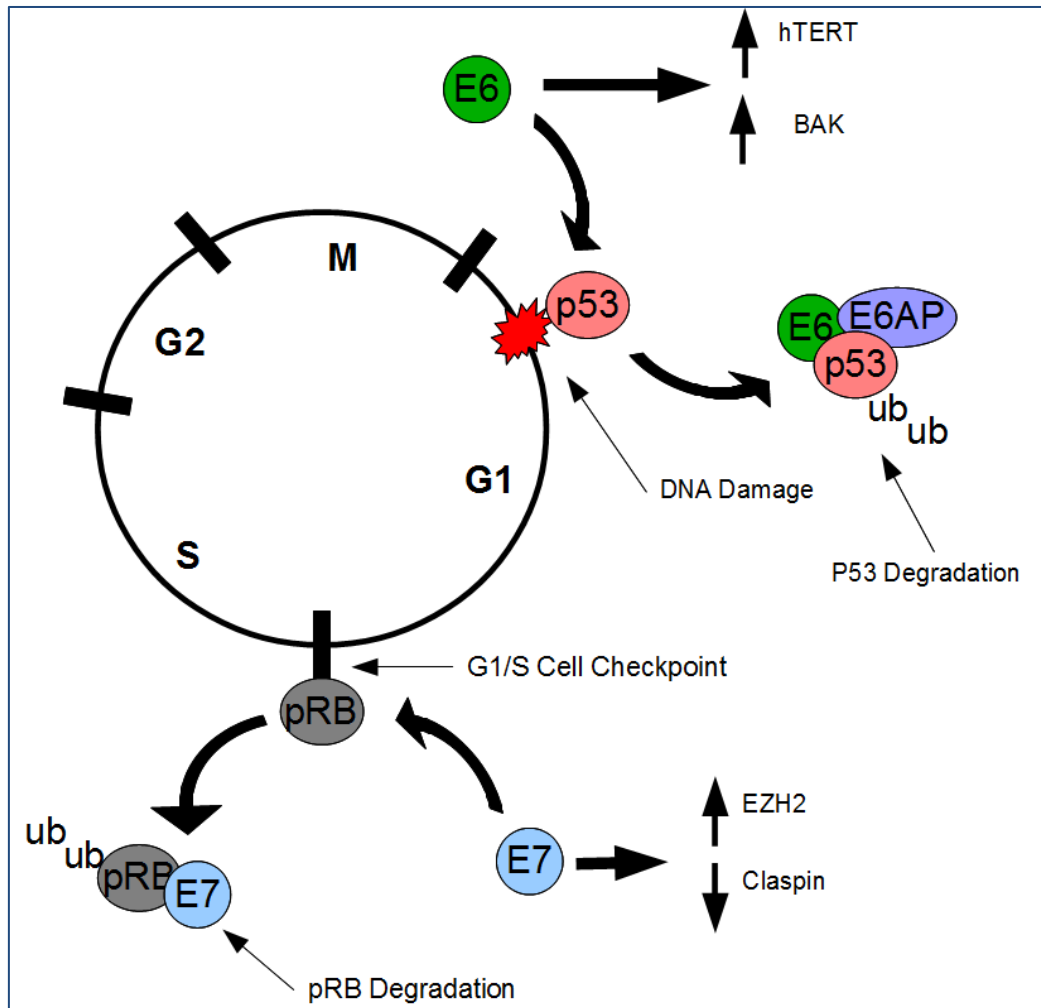
Figure 1.3. DNA damage and immortalization mechanisms employed by HPV E6 and E7 oncoproteins. E6 couples with E6 adaptor protein (E6AP) to degrade p53 tumour suppressor protein and increases levels of hTERT and BAK. E7 couples with and degrades pocket protein pRB. E7 also decreases levels of Claspin and increases EZH2 levels.

A large percentage of cervical cancers have been shown to have increased epithelial growth factor (EGF) receptor expression and research has shown that E5

expression contributes to this increase. There have been many studies to assess the mechanism by which E5 protein alters the normal activity of EGF receptors, the most widely accepted method being that E5 interacts with a 16 K subunit of the vacuolar ATPase (Müller et al., 2015), however the development of a 16 K antibody has cast this theory into question as the amount of E5 bound to 16 K appears to be too low to account for the increase in observed EGF reception expression (Suprynowicz et al., 2010). Regardless of method of action the E5 protein increases EGF-receptor expression which leads to an increase in DNA synthesis and maintains active cell division in the upper layers of the stratified epithelium (Fehrmann et al., 2003).

E5 also couples with E6 and E7 to promote transformation of the cell (DiMaio and Mattoon, 2001). It has been demonstrated that with prolonged oestrogen treatment, E5 caused oncogenic transformation in transgenic mice in the absence of E6 and E7. The same study also showed that mice expressing all three oncoproteins had only a marginal difference in number of tumours and tumour size compared with mice only expressing E6 and E7. These results suggest E5 is a minor oncoprotein compared with E6 and E7 (Maufort et al., 2010). Although E5 protein contributes to the oncogenic transformation of virus infected cells, this gene is often lost when the virus integrates into the host genome. It has therefore been suggested that E5 is a factor in cellular transformation rather than maintenance of the malignant phenotype (McLaughlin-Drubin and Munger, 2009b).

1.2 **Immunoregulation of HPV Infection**

In a transient infection, the virus remains below detectable levels from the host immune system, which allows the productive lifecycle to be complete before viral clearance occurs. HPV avoids the host immune system effectively by not infecting antigen presenting cells, not lysing cells, tightly controlling host gene expression and innate immune signalling and only producing new virus particles in the upper layers of the epithelium (Tindle, 2002, Westrich et al., 2017). It has been shown that HPV 16 E5 reduced MHC-I expression on the surface of the cells by accumulating them in the Golgi Apparatus (Campo et al., 2010). Women with multiple HPV infections were shown to be less likely to clear a HPV 16 infection; and it has been suggested this susceptibility may be due to failings in the host's immune system (Schmeink et al., 2013).

Although HPV is successful at evading the immune system the induction of the host immune system plays an important role in limiting spread and promoting viral clearance. Studies have shown that a localised increase in mononuclear cells, CD4+, CD8+ CD56+ and macrophages, and the expression of TH1 cytokines are signs of viral regression (Stanley, 2015). Also, HPV 16 clearance has been correlated with the presence of intraepithelial granzyme B+, CD8+ and CD 56+ cells (Westrich et al., 2017)

1.3 **Productive HPV lifecycle**

The HPV productive lifecycle leads to the production of new virus particles that can go on to infect other individuals. It is estimated that by the age of 50, 80% of women will have been infected with at least one type of HPV (CDC, 2013). However, the majority these infections are cleared by the host.

HPV transmission requires entry into basal epithelial cells and it does so via small micro abrasions in the skin or through the squamocolumnar junction that allows access to basal cells without micro abrasions. Most HPV induced cervical lesions occur in the transformation zone of the squamocolumnar epithelial junction. The transformation zone is a small region of differentiating squamous cells between the endocervix and ectocervix. The cells of the transformation zone are a single layer of differentiating cells, making them an ideal location for HPV entry to occur (Herfs et al., 2013, Herfs et al., 2012).

The productive HPV lifecycle leads to the release of new virus particles that can go on to infect new host cells. For a productive infection to take place the following stages must occur. These stages must all occur whilst evading the host's immune system to ensure the productive infection persists for as long as possible:

- Viral entry

- Replication of the viral genome

- Packaging and release of new virions

Although the receptor(s) involved in HPV have yet to be determined conclusively, there is evidence that glycosaminoglycans – more specifically heparin sulphate proteoglycans and syndecan-1 are involved in the uptake of HPV into host cells (Giroglou et al., 2001, Joyce et al., 1999, Patterson et al., 2005, Shafti-Keramat et al., 2003). Irrespective of receptor usage the virus is taken up by basal cells by means of endocytosis, this is a slow process taking up to 12 hours. The L1 protein on the cell surface attaches to the heparin sulphate proteoglycans which leads to a conformational change to the L2 protein. The L2 protein is then cleaved by furin on the cell surface (Aksoy et al., 2016). HPV 16 has been shown to enter the cells via clathrin-dependent, receptor-mediated endocytosis (Day et al., 2003), however it has been suggested that HPV 31 enters via caveola-mediated endocytosis (Bousarghin et al., 2003).

Once within the basal cells, the virus enters into the host nucleus via nucleus envelope breaks (Senapati et al., 2016). Once inside the nucleus of the cell the virus genome is maintained as an episome at low copy numbers, with the E2 protein regulating E6 and E7 transcription levels (as described in section 1.1.2). Two promoters control viral gene expression. The early promoter is located in the NCR

12

(P97 in HPV 16 and 31, P105 in HPV 18) and the late promoter is located in the E7 open reading frame (ORF) (P742 in HPV16 and 31) (Graham, 2010). Expression of *E6* and *E7* ensures that replication of viral DNA continues in terminally differentiated cells. In the upper epidermis, the structural proteins L1 and L2 are produced alongside E4 protein (Doorbar, 2005, Longworth and Laimins, 2004). E4 protein is important at this stage as it helps disrupt the cytokeratin structure of the host cell assisting virion release (Doorbar et al., 1991, Wang et al., 2003) from the epithelium. The virus is said to be cleared when HPV DNA can no longer be detected, however HPV may still be present but just at a level which cannot be detected by current DNA testing technology (Fernandes et al., 2013).

1.4 Transforming infection and neoplastic progression in HPV infected cells

Most individuals infected with HPV will naturally clear the infection, however a small percentage of individuals will not. This is due to the viruses' ability to evade the host immune system and down regulation of the host's innate immune signalling (Westrich et al., 2017). The failure of the immune system to clear the HPV virus leads to a persistent infection that can remain for many years; during the productive infection, the viral proteins E1, E2, E5, E6 and E7 are expressed. E4 is expressed further up in the epithelium and L1 and L2 are expressed in the upper layers of the epithelium and new viral particles are produced. Over time the infection can lead to

genetic instability and mutations within the viral genome, which can contribute to cancer.

The transition from a productive infection to a transforming infection is driven by the deregulation of the viral oncogenes. Prolonged infection leads to genetic instability; neoplastic changes and the virus lifecycle cannot be completed. This turns the infection from a productive infection into a transforming infection; oncoproteins E5, E6 and E7 are expressed throughout the epithelium and E4 is expressed in the upper-layers of the epithelium. Viral proteins E1, E2 and L1 and L2 are no longer expressed in the transformed epithelium and therefore the viral particles are no longer produced (Figure 1.4). This deregulation of HPV gene expression and neoplastic progression is ascribed to various factors including alterations in host gene expression, integration of the HPV DNA into the host genome or hyper- and/or hypo-methylation of the HPV genome. These mechanisms of viral genome deregulation and their by-products being considered as potential biomarker(s) for disease progression to either support or replace cervical cytology.
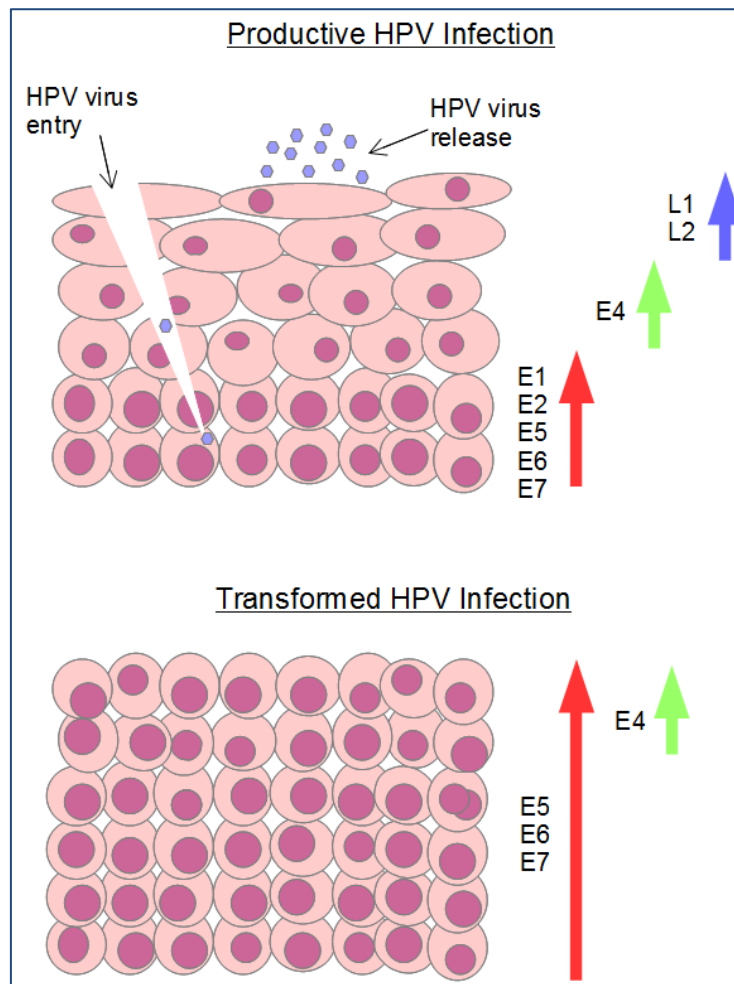
Figure 1.4. Representation of HPV gene expression in a productive and a transformed infection. (A) HPV gene expression in a productive infection. **(B)** HPV gene expression in a transformed infection.

1.4.1 Aberrant expression of host gene molecules in HPV infection

Aberrant expression of host genes, proteins and micro-ribonucleic acids (miRNA) is present in a transformed HPV infection. These include (but are not limited to) over expression of cell cycle regulators p16 and Ki-67, over expression of Hedgehog proteins, specifically Sonic Hedgehog protein (SHH), and suppression of tumour suppressor proteins p21 and p27. P16 is a tumour suppressor protein which inhibits cell cycle progression (Alcorta et al., 1996, Romagosa et al., 2011). In HPV infection p16 expression is increased in relation to an increase in cervical neoplasia (Benevolo et al., 2006, Romagosa et al., 2011). Ki-67 is a marker of cellular proliferation which in the normal epithelium is confined to the proliferating cells in the parabasal layer of the epithelium. In cervical intraepithelial neoplasia (CIN) Ki-67 expression is seen in the higher layers of the epithelium (Dwivedi et al., 2013, Sari Aslani et al., 2013). Both p16 and Ki-67 have been extensively studied as biomarkers for HPV infection and are now used as a clinical marker for HPV infection (Wentzensen et al., 2017, Ikenberg et al., 2013, Petry et al., 2011).

The Hedgehog proteins are key factors in embryo development, and inappropriate expression has been linked to cancer development. Hedgehog proteins are divided into three subgroups, Desert Hedgehog, Indian Hedgehog and SHH (Varjosalo and Taipale, 2008). SHH is the most studied of the three subgroups and aberrant expression has been shown to lead to several cancers including basal cell carcinoma, medulloblastoma and bladder cancer (Varjosalo and Taipale, 2008, Hahn et al., 1996, Gupta et al., 2010, Hanna and Shevde, 2016a, Hanna and Shevde, 2016b). SHH is aberrantly activated in vulval cancer and 8/10 solitary Vulval Intraepithelial Neoplasia (VIN) cases have shown over expression of SHH and its

downstream effector Zinc Finger Protein GLI-1 (GLI-1) (Dr C Dawson, personal communication). The expression of SHH is controlled by different factors of which GLI-1 is key to its regulation. GLI-1 expression is induced by Hedgehog ligands, and it provides a positive feedback and prolongs the effects of SHH (Ruiz i Altaba, 1999, Kasper et al., 2006).

miRNA is non-coding ribonucleic acid (RNA) which regulates the expression of mRNA. The over or under expression of miRNAs can influence the regulation of cell proliferation, apoptotic pathways and cell adhesion (Calin and Croce, 2006, Peng and Croce, 2016). In HPV infections it has been shown that HPV oncogenes E6 and E7 affect the miRNAs which regulate p53, leading to the degradation of p53 (Zheng and Wang, 2011).

1.4.2 Viral integration

Often in cervical cancer the HPV genome is integrated into the human genome with no remaining viral episomes. This results in the loss of several genes, including those that are essential for new virion production. HPV integration is a by-product of HPV infection rather than a step in the virus lifecycle; however it is a significant step in the process of viral transformation of the host cell (Pett and Coleman, 2007, McBride and Warburton, 2017). The virus integrates randomly into the host genome and this has been shown to occur in multiple chromosomes

(Wentzensen et al., 2004, Hu et al., 2015). Stages in the transition from episomal infection to integration are not well defined.

DNA damage and genomic instability are important prerequisites for HPV integration. In normal cells, various mechanisms are used to repair DNA damage and cell death is induced if the damage cannot be repaired. This is not always the case in HPV infected cells. HPV has been shown to induce DNA damage and disrupt normal cellular damage response mechanisms. This can lead to genomic instability and double stranded breaks (DSBs) forming. DSBs allow DNA recombination to occur via homologous recombination. High levels of DSB are seen in common fragile sites (CFS), and there DSB occur spontaneously or can be induced by environmental factors (Dmitrieva et al., 2011, Matovina et al., 2009). Research has shown that HPV DNA integrates into CFS at a higher frequency than in other locations (Gao et al., 2017, Dall et al., 2008). DNA plasmids also integrate into CFS at a higher frequency (Matzner et al., 2003) and the unstable nature of CFS make them an ideal location for HPV integration to occur.

When HPV integrates into the host genome it is common that the viral genome is only partially integrated. The *E6* and *E7* ORF are almost always fully intact; however other viral genes important in the production of new virus particles are frequently lost. The loss of the *E2* regulatory gene is a predominant finding in viral integrants (Pett, 2004, Cricca et al., 2009, Choo et al., 1987). Loss of *E2* through viral integration has been shown to increase the levels and stability of *E6*

and *E7* proteins in the cell, thus contributing to cell transformation (Choo et al., 1987, Jeon et al., 1995). Not all evidence agrees that integration leads to heightened levels of viral oncogenes. A study by Häfner and colleagues compared the level of viral transcripts against the physical state of the virus genome. Their results showed that the level of *E6* transcribed was similar in all samples with no marked increase in cells with only integrated HPV (Hafner et al., 2008).

1.4.3 Viral Genomic DNA Methylation

Methylation is a post-translational modification which is associated with cancer development. Methylation is also a normal consequence of aging with well characterised and reproducible changes occurring at specific sites of the human genome (Weidner and Wagner, 2014). Methylation occurs in cytosine located in a CpG dinucleotide where are methyl group is added to cytosines at the 5' position. The addition of the methyl group can alter chromatin conformation and DNA topology, leading to changes in gene expression (Stein, 2011, Deaton and Bird, 2011). DNA methylation is associated with a range of functions within the host that contribute to normal gene expression and foetal development. However methylation can also have an impact on malignant cells associated with the progression of cancers (Robertson, 2005).

Viral genomic DNA methylation in cancer cells has been shown prominently in HPV 16 infected cells with an increase in viral DNA methylation in the *E2*, *L1* and *L2* genes (Bryant et al., 2015, Bhattacharjee and Sengupta, 2006, Mirabello et al., 2012). Methylation of HPV DNA has also been shown to occur in cells infected with different HPV types such as HPV 18, 31 and 45, although this has not been extensively studied (Wentzensen et al., 2012). Several studies have shown a positive correlation between viral methylation and CIN that make it a potential biomarker for CIN presence or progression (Bhat et al., 2016, Chapman et al., 2016, Clarke et al., 2012, Shridhar et al., 2016). Following HPV vaccination, work is needed to identify where viral genomic DNA methylation testing would fit in a future screening programme and the cost benefit of introducing viral genomic DNA methylation as a screening test.

There are several methods of assessing DNA methylation. These can be broadly split into methods which detect global methylation within a sample, and can either profile whole genome methylation level or identify specific genes or regions which may have differential methylation expression levels (Kurdyukov and Bullock, 2016). Methods which can be employed to profile whole genome methylation patterns include HPLC-UV (Armstrong et al., 2011), LC-MS (Hu et al., 2013) and LINE-1 plus Pyrosequencing (Iwagami et al., 2012). These methods provide an overall picture of methylation in the sample without identifying exact regions of hyper-/hypo-methylation within the sample. These methods can be used to detect changes in the brain that occur during aging and learning (Gaglio et al., 2014) as well as looking at global patterns of methylation in disease such as cancer.

To directly identify genes or regions of the DNA which have been methylated bisulphide sequencing and array systems can be employed. Bisulphide sequencing involves firstly converting all unmethylation cytosines into uracil via treatment with sodium bisulphite, the methylated cytosine molecules are resistant to this change and are left unconverted. Therefore, when sequenced, unmethylated cytosine will be read as thyamine and methylated cytosines will be read as guanine (Figure 1.5). Bisulphite conversion is considered to be the gold standard in the identification of methylation changes (Patterson et al., 2011).

After bisulfide conversion, the DNA is sequenced to analyse the level of methylation at each CpG site. Different sequencing methods can be employed such as Sanger sequencing, Pyrosequencing and DNA microarrays. Sanger sequencing is the most traditional method however the results are not quantitative. The results of Sanger sequencing the DNA can be semi-quantitative by cloning the polymerase chain reaction (PCR) product into a bacterial population, plating out colonies and then sequencing individual colonies to assess the percentage methylation in each CpG site. Pyro sequencing is a fast method of DNA sequencing. The results achieved are quantitative giving a percentage methylated for each CpG site in the region of interest (Tost and Gut, 2007). DNA microarrays give a vast quantity of quantitative information that is not specific to the area of interest. DNA microarray analysis is also expensive particularly when working with a large number of samples (Deatherage et al., 2009).

With advances in next generation sequencing several commercial kits are now available to carry out whole genome bisulphide sequencing (such as the Illumina Whole Genome Bisulphide Sequencing WGBS) however these are still prohibitively expensive ($6000/samples) (Johnson et al., 2012b).
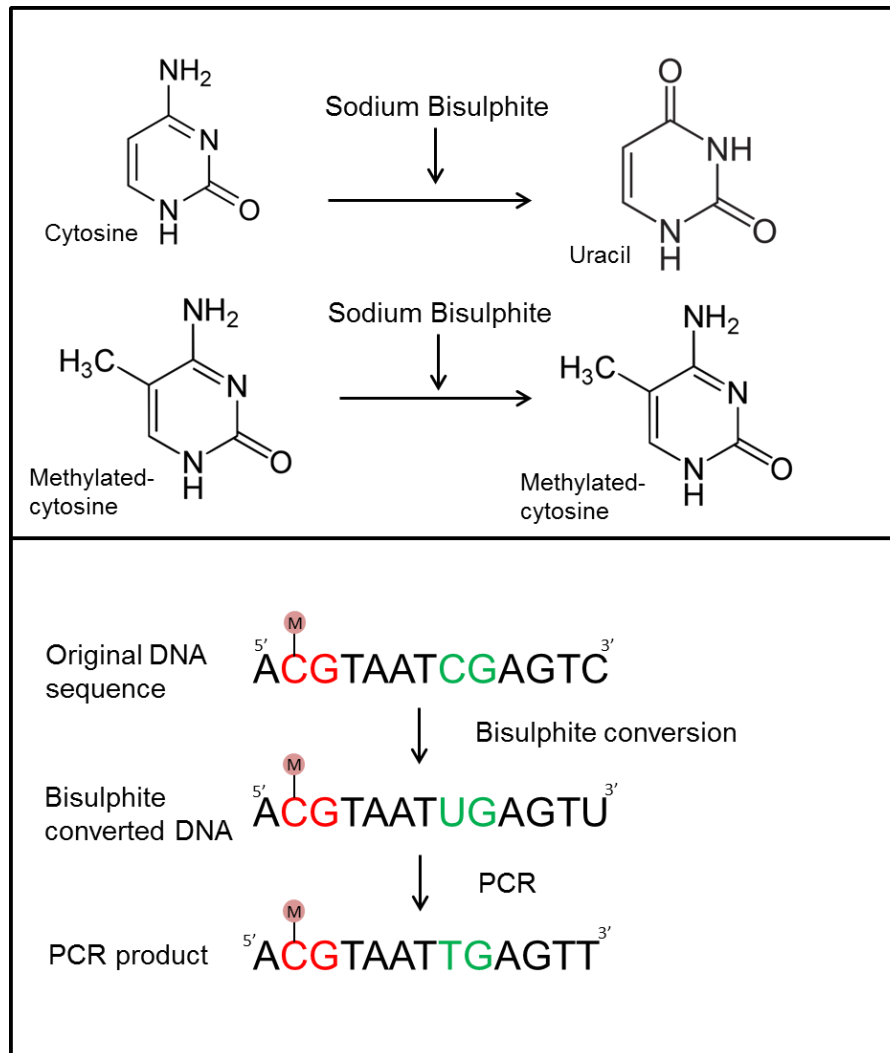
Figure 1.5. The process of bisulphite conversion of DNA with sodium bisulphite. A. Unmethylated cytosines are converted to uracil with the addition of sodium bisulphate. However methylated cytosine is resistant to bisulphide conversion and remains as methylated cytosine. B. In a sequence with methylation and unmethylated 'CG' sites bisulphide conversion converts unmethylated CG into UG (highlighted in green). Then after PCR UG is converted to TG. The methylated CG remains unchanged throughout (highlighted in red). This change can then be measured by Pyrosequencing, PCR with high resolution melt analysis or via methylation specific PCR.

Bisulphide conversion is also used to assess methylation in specific genes and regions of DNA. This can be does by Pyrosequencing, methylation specific PCR, PCR with high resolution melt or bead arrays. Pyrosequencing, methylation specific PCR and PCR with high resolution melt analysis all involve specific primers which amplify the region of interest and can be particularly useful for small cohort studies, where a small number of known points of interest have already been identified (Wojdacz et al., 2008). Bead arrays are similar to those used to identify genes of interest and multiple locations of interest can be tested using a single array (Bibikova et al., 2006).

1.5 Epidemiology of HPV in Women

80% of women will be infected with at least one type of HPV in their lifetime and only a small percentage of infections go on to cause any neoplastic changes (CDC, 2013). HPV is associated with neoplastic progression in the cervix, vulva and vagina as well as genital warts. Neoplastic changes can regress back to normal or can progress and lead to cancer. Factors such as age, number of sexual partners, pregnancy and smoking can all contribute to disease progression from a transient infection to neoplastic changes and disease. Low-risk HPV (LR-HPV) types such as HPV 6 and HPV 11 are responsible for 90% of genital warts (Pineros et al., 2013). The rest of this section will concentrate on HR-HPV infections and associated disease.

Intraepithelial neoplasia is spilt into three phases, I, II and III, which is characterised by increasing neoplastic changes to the cells. Individuals with intraepithelial neoplasia can regress back to normal naturally over time or can progress to a higher grade of neoplasia or cancer if left untreated (Sellors and Sankaranarayanan, 2003) (Figure 1.6).
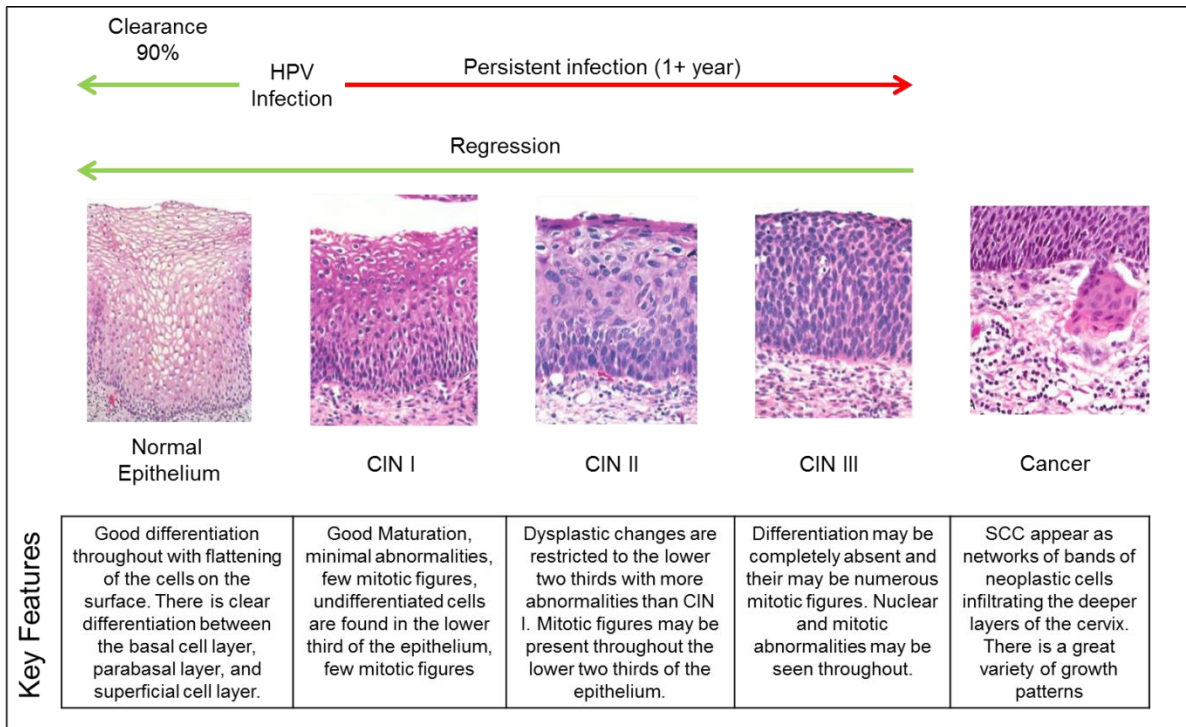
Figure 1.6. The progression of HPV infection to cancer and the key features of the cervical epithelium as seen using histological analysis. The red arrow represents disease progression. Around 90% of HPV infections will be cleared by the immune system of the host. The green arrows represent intraepithelial neoplasia regressing back to normal over time. Images of epithelium in the progression are from Robbins and Cotran Pathologic Basis of Disease.(Kumar et al., 2015).

Cervical cancer is the 4th most common female cancer worldwide (Figure 1.7) and the second most common cancer in women aged 15 to 44. Worldwide cervical cancer mortality is substantially lower than incidence with a ratio of mortality to incidence of 50.3% (Bruni et al., 2017). Less is known about HPVs role in other anogenital cancers. Vulva cancer represents 4% of all gynaecological cancers with 27,000 new cases in 2008, and vaginal cancer represents 2% of all gynaecological

cancers with 13,000 newly diagnosed cases in 2008 (de Martel et al., 2012). Both cervical and vaginal cancers occur more commonly in developing countries, however most vulvar cancers (60%) occur in more developed countries (Bruni et al., 2010).
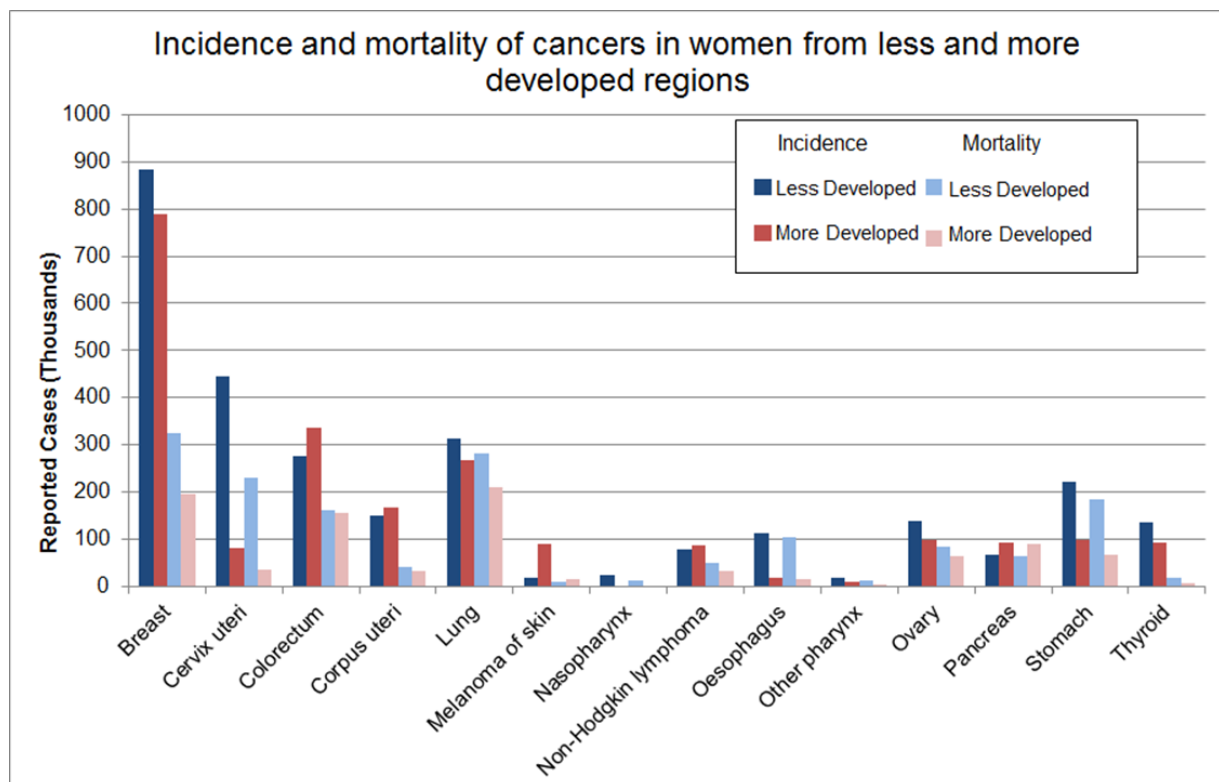
Figure 1.7. Cancer incidence and mortality in women in less developed and more developed regions. Incidence is shown in blue and mortality is shown in red. Figure adapted from Globocan 2012 (Ferlay et al., 2013).

1.6 Cervical cancer prevention models

Cervical screening has decreased the incidence of cervical cancer as precancerous lesions are treated before they progress to invasive carcinoma. Prevention strategies can be spilt into primary prevention with vaccination of young people and secondary prevention with cervical cytology, HPV DNA testing and future biomarker screening.

HPV prophylactic vaccinations are based on late proteins of the virus, cloned into vectors that spontaneously form VLPs in culture (Garland and Smith, 2010, Hagensee et al., 1993). The UK national HPV vaccine programme began in 2008 for girls aged 12-13 years and a catch-up program was introduced for older girls using the bivalent vaccine Cervarix™ (Jit et al., 2008). Initially two prophylactic vaccines were licensed against HPV; A quadrivalent vaccine, Gardasil® (Merck Sharp & Dohme), which protects against two HR-HPV types (HPV 16 and 18) and two LR-HPV types (HPV 6 and 11) and Cervarix™ (GlaxoSmithKline), a bivalent vaccine offering protection against HR-HPV types 16 and 18 (Keam and Harper, 2008). Approximately 70-76% cases of invasive cervical carcinoma (ICC) are caused by HPV types 16 and 18 (Li et al., 2011) and these vaccines are projected to offer significant protection against cervical cancer. Clinical trials have shown the bivalent and quadrivalent HPV vaccine offer partial cross-protection against other HR-HPV including types 31, 33 and 45 (Harari et al., 2016).

Recently a 9-valent HPV (9v-HPV) prophylactic vaccine (Gardasil 9, Merck and Co., Inc.) has been introduced which will offer prolonged protection against nine

HPV types including HR- HPV types 16,18, 31, 33, 45, 52 and 58 plus LR-HPV types 6 and 11 (Joura et al., 2015, Cuzick, 2015) and the level of protection offered against HPV 6, 11, 16 and 18 is non-inferior to the first generation HPV vaccine (Zhai and Tumban, 2016, Lopalco, 2017) This 9-valent vaccine have been approved by the Food and Drug Administration for females aged 9 – 26 years and males aged 9 – 15 years (Bonanni et al., 2010). With the addition of HPV 31, 34, 45, 52 and 58 this vaccine will potentially prevent almost 90% of cervical cancer cases (Zhai and Tumban, 2016).

Since the UK national cervical screening programme began testing women 20 to 64 years every 3-5 years in 1988, the number of cases of cervical cancer has significantly decreased (Levi et al., 2000, Michalas, 2000, Peto et al., 2004b). Screening costs around £150 million in England and prevents around 4,500 deaths (Peto et al., 2004b). This is a considerable cost to the National Health Service (NHS); however, this is offset by ICC treatment costs which equate to approximately £10,464 per person in their first year of treatment. Identifying individuals before they require invasive treatment has the dual benefit of increasing quality of life and providing a cost saving to the NHS (Brown et al., 2006).

There is an on-going debate about HPV testing as a suitable alternative to cytology. HPV DNA testing is more sensitive, less specific and has a higher negative predictive value than current cervical cytology (Waxman, 2008, Bulkmans et al., 2007). The sensitivity of HPV DNA testing appears to be equally high in a spectrum

of age ranges and geographical locations; this contrasts with cytology which varies in sensitivity (Cuzick et al., 2006). HPV testing was implemented in the UK in 2014, alongside cervical cytology with a view to reducing the number of ICC cases further (DH/NHS Finance, 2010).

1.7 Laboratory based research tools to investigate HPV infection

This section details the various molecular and cellular research tools available to investigate HPV infection.

1.7.1 Collection of clinical samples to investigate of HPV infection

Collection and preservation of clinical samples is important in cervical screening and residual material can be used for scientific research such as the development and optimisation of novel biomarkers. The use of cellular cytology as a screening method to predict neoplastic changes in the cervix was first described by Dr George N. Papanicolaou in 1941 (Michalas, 2000, Papanicolaou and Traut, 1997). This involved exfoliating cells from the surface of the cervix and transferring them to a microscope slide at the bedside. Slides were then processed and stained for microscopic examination to identify neoplastic changes. This enabled treatment of women with neoplastic changes and reduced risk of subsequent cervical cancer

development. Incorporating this into a systematic screening programme led to a reduction in cervical cancer associated mortality.

In recent years, there has been a move from the conventional cervical cytology to liquid based cytology (LBC). LBC involves storing cells in a protective media, transporting to a laboratory and transferring to a slide for automated analysis. LBC has a higher sensitivity in detecting cellular changes and residual cellular material can be used for molecular HPV testing (Bernstein et al., 2001, Zhu et al., 2007). LBC media such as Surepath^{TM} (Source BioScience, Nottingham, UK) and PreservCyt (Cytyc Corporation, Boxborough, Mass.) are commercial cellular preservatives used for collection of cervical smear tests. However, LBC media is designed for the preservation of cellular morphology, not for preservation of nucleic acid. LBC media has a detrimental effect on the nucleic acid preservation, degrading nucleic acid and cross-linking DNA due to the presence of formalin (BectonDickinson, 2017, Huijsmans et al., 2010). With the incorporation of HPV DNA testing into the cervical screening programme, the preservation of nucleic acid is as important as the preservation of cellular morphology structures within the cells. Cervical biopsies are used for the diagnosis of disease and can be cultured to generate cell lines for use in research.

Other types of material have been assessed for clinical HPV analysis including urine (Piyathilake et al., 2016), blood (Patel et al., 2011) and saliva (Turner

31

et al., 2011, Wimardhani et al., 2015), however these do not reliably detect HPV and are not used as part of screening.

1.7.2 Nucleic Acid Extraction and Purification

Proteinase K enzyme is used to break down cellular glycoproteins, DNase and RNase to release nucleic acid with limited denaturation (Roche, 2010, Ebeling et al., 1974). Cellular debris can subsequently be pelleted and purified nucleic acid in the aqueous phase can be collected. This is a crude method of extraction, but provides a high quantity of nucleic acid from clinical samples.

Many commercial nucleic acid extraction kits use Proteinase K to lyse cells, however the samples are then purified using a solid phase extraction (Janecka et al., 2015, Tan and Yiap, 2009, Bonin and Stanta, 2013). This involves binding nucleic acid to a solid membrane or beads. Centrifugal force is then used to wash the membrane or beads to remove contaminants including salt and cellular residue. The nucleic acid can then be eluted using a sterile buffer. A solid phase extraction reduces the chance of contamination that may occur in a normal proteinase K extraction.

1.7.3 Methods of HPV identification and typing

HPV identification and typing can be categorised into three broad categories; (i) nucleic acid hybridisation (ii) nucleic acid amplification, and (iii) signal amplification.

Southern blotting and *in-situ* hybridisation (ISH) both provide high quality information on the presence of HPV as well as the state of the viral genome. However, both methods are time consuming, suffer from low sensitivity and require a large quantity of purified DNA. Also, nucleic acid hybridisation requires well preserved DNA which can be difficult to achieve from clinical material which has been stored in a cellular preservative (Abreu et al., 2012, Villa and Denny, 2006).

The amplification of HPV DNA or RNA can identify the presence and genotype of HPV in a sample. DNA amplification identifies the presence of the viral genome and RNA amplification provides information of viral gene expression. RNA is sensitive to degradation and pure, intact RNA can be difficult to extract from clinical samples collected in LBC (Houseley and Tollervey, 2009). However, there are several commercially available assays for HPV mRNA available such as Aptima HPV Assay (Hologic, 2017) and the QuantiVirus[TM] HPV Cervical and Oral Tests (DiaCarta, 2017). Both of which have shown higher specificity and fewer false-positive results when compared to DNA based assays (Munson et al., 2014, Clad et al., 2011, Benevolo et al., 2011).

DNA is less sensitive to degradation and adequate purification can be achieved from clinical samples stored in LBC. Nucleic acid amplification can be achieved using consensus or type-specific HPV primers. Consensus primers target highly conserved regions of the HPV genome and can identify many HPV genotypes simultaneously. These include GP5+/6+ and MY09/MY11 consensus primers (Venceslau et al., 2014). Type specific primers are designed to only target a single HPV type; thus, the specific HPV type of infection can be identified. Many type-specific primers have been published and they can also be easily developed in-house. A range of commercial HPV nucleic acid amplification assays are available. These include INNO-LiPA® HPV test (Innogenetics, 2013), COBAS® (Roche Molecular Diagnostics, 2017) and PapilloCheck® (Greiner Bio-One, 2006, Vieira and Almeida, 2013). DNA-array technology can also be used to identify and type HPV from clinical samples. This involves amplification of a portion of the HPV genome, the PCR product is fluorescently labelled and hybridised to the DNAchip. The DNA chip can then be analysed and show the HPV types present in the sample. PapilloCheck® is a simple commercial assay which uses type-specific *E1* primers and DNA-array technology to simultaneously identify 24 different HPV types. HPV types present at a low copy number and multiple infections may not be identified if one genotype has a low copy number compared to other HPV types in the sample. A low viral copy number may represent a transient infection and not have a major impact on the clinical treatment of the patient (Depuydt et al., 2012).

HPV signal amplification involves hybridisation of HPV-DNA to labelled probes that produce a non-radioactive signal, which can then be assessed. These tests can differentiate between HR-HPV and LR-HPV infections but cannot differentiate between specific HPV types. Two commercially available signal amplification assays are Hybrid Capture 2 (HC2)(QIAGEN, 2017a) and Cervista (Hologic, 2010).

1.7.4 Detection of genetic changes in HPV infection

HPV gene expression is closely linked to the stratification of the epithelium and alters between transient and transforming infections, as described in section 1.3. HPV gene expression can be measured using various methods such as Northern Blotting, quantitative real-time polymerase chain reaction (qRT-PCR) and RNA sequencing. Northern Blotting is a classic technique however it does not provide sufficient information compared to qRT-PCR and RNA sequencing. qRT-PCR uses fluorescence to amplify and detect PCR targets in a single step and increasingly multiple targets can be detected in a single reaction. Further qRT-PCR can achieve greater accuracy than northern blotting, detecting small differences in gene expression between samples (Wong and Medrano, 2005). RNAseq is a relatively new method which provides in-depth quantitative detail of the whole transcriptome of the sample (Wang et al., 2009b).This involves parallel sequencing of either RNA or complementary DNA (cDNA) using next generation sequencing

technologies. These sequences can then be compared to a known reference genome to assess levels of transcript expression or *De novo*, in which a transcriptome can be assembled without a reference genome which may identify novel splice patterns (Surget-Groba & Montoya-Burgos, 2010)

Detecting HR-HPV infections identifies individuals at risk of precancerous lesions or who have already developed lesions. However many HR-HPV infections are transient and will be cleared naturally by the body's immune system (Fernandes et al., 2013). There is a need to develop molecular tests that can differentiate between transient infections and productive infections likely to progress to high-grade disease. Detecting genetic changes in HPV methylation of HPV DNA into the host genome may give an insight into which infections are transient and which are transforming (Clarke et al., 2012). Several methods of viral genomic DNA methylation analysis have been developed including bisulfide conversion, methylation sensitive restriction enzyme digestion and affinity purification of methylated DNA. As standard sequencing techniques cannot differentiate between cytosine and methylated cytosine an additional step is therefore necessary to identify methylated cytosines. Sodium bisulphite converts unmethylated cytosine into uracil, but not methylated cytosine (Clark et al., 1994). This is described in detail in section 1.4.3.

Integration is a random event in cervical neoplasia progression, which has been shown to occur frequently in CFS (described in section 1.3.1) (Thorland et al.,

2003). Various methods have been developed to identify viral integration events that may be useful in detecting HPV infections likely to progress to high grade CIN. Methods of detecting viral integration can be categorised as:

- Gene lost as a substitute marker of viral integration.
- Fusion transcripts that determine the viral integration event are transcriptionally active.
- Viral DNA integration events that may or may not be transcriptionally active.

As the *E2* gene is commonly disrupted in cervical cancer; the absence of the *E2* gene in the presence of HPV oncogenes E6 and/or E7 can be a potential marker for viral integration. This can be assessed by a tiling PCR which amplifies the E2 and E6 genes, assessing the absence (all or partial loss) of the *E2* gene in samples with detectable viral oncogenes. This test is less laborious than identification of individual integrants and is a good indicator of integration (Collins et al., 2009). This assay however is limited as *E2* gene presence could be a false positive indicator for 3 reasons: (1) Integrants and episomes can exist as a "mixed infection"; (2) Viral integration sites in the HPV genome can occur outside the E2 region; (3) The stage at which the *E2* gene is lost is undetermined and variable between patients (Boulet et al., 2009).

The *E2* tiling PCR uses type specific primers to amplify overlapping regions across the *E2* gene. When all reactions are positive, the *E2* gene is considered intact

in the test sample, however if one or more portions of the *E2* gene are missing then the sample is indicative of integration (Collins et al., 2009). *E6* viral oncogene relative to the *E2* gene is another substitute marker determined by qRT-PCR. The *E6* gene is typically maintained in viral integration and the *E2* gene is generally lost, therefore the ratio should decrease upon viral integration (Chang et al., 2013).

The identification of fusion transcripts can determine if the integration event is transcriptionally active. Methods involve the identification of the sequence and state of the RNA. Traditional methods of fusion transcript identification include RNA ISH and Amplification of Papillomavirus transcripts (APOT) (Klaes et al., 1999). However, technological advancements have increased the prominence of RNA sequencing as a method of identifying HPV-human fusion transcripts. There are several methods to identify viral integrants that may or may not be transcriptionally active, these include Southern Blotting (Villa and Denny, 2006), Restriction site Polymerase Chain Reaction (RS-PCR) (Sarkar et al., 1993), Fluorescent in situ-hybridisation (FISH) (Hopman et al., 2005) and Detection of Integrated of Papillomavirus Sequences (DIPS) (Luft et al., 2001) (Table 1.1).

Method of identification	Description	Reference
Southern Blotting	An enzyme that has a single cut site in the HPV genome digests samples. Samples are separated by electrophoresis and labelled used radioactive HPV DNA probes. This is the gold standard of HPV identification and genome status, however this methods is suffers from low sensitivity and a large quantity of purified full length DNA is required.	(Villa and Denny 2006)
RS PCR	Samples a digested and then DNA is amplified using HPV specific primers and primers that are specific to the restriction site. This is a time consuming method that can identify host/viral junctions.	(Sarkar et al. 1993)
FISH	Flourecent probes are used to identify integrated HPV in the host chromosomes. This method requires fresh cellular material and can identify the presence of HPV within the host chromosome but the exact host/viral junction cannot be identified.	(Hopman et al. 2005)
DIPS	Samples are digested and adapters are ligated to the digested DNA. HPV specific linear PCR is followed by a nested PCR with primers specific from the adapter and HPV. PCR products are sequenced to assess for integration fusion sites. This methods can identify the host viral junction site.	(Luft et al. 2001)

Table 1.1 A brief description of the different methods used to identify viral integrants which may or may not be transcriptionally active

1.7.5 Laboratory models for HPV study

Primary cells that have migrated from a piece of tissue grown under favourable conditions are a useful tool for the study of HPV infection. Primary cells have a limited life span, however with extended culture they can become immortalised, this may be due to chemical or viral stimulation, or a spontaneous process due to cellular stress during cell culture. Immortalised cells lines can be grown continually in the lab if suitable culture conditions are met and the cells are split at regular intervals. Over time cells with a growth advance are naturally selected for and cells can become different from the tissue of origin. This genetic drift is a significant problem in the development and culture of new models for disease research.

When developing novel cell lines, procedures need to be put in place to identify the genotype of the cells at the time of propagation and then cell lines can be tested to ensure that cells have not drifted away from the original patient phenotype (Gulliksen et al., 2012). If suitable controls are in place to check cells remain true to the original patient samples then cells lines can provide an invaluable tool in HPV research.

Another issue which affects cell line development is the misidentification of cell lines which can lead to questionable research results (Masters, 2002). This misidentification can be avoided with good laboratory practice and by limiting the number of laboratories the cells pass through from the origin. The production of in-house cell lines is an asset. Regardless, cell lines are an important cornerstone to the development of drug therapies and research into genetic changes driving disease progression. Some commonly used cell lines for HPV study are listed in Table 1.2.

Table 1.2 Description of origin and HPV genome status of immortalised cell lines that are commonly used in HPV research

Cell Name	Origin	HPV type/genome status	Reference
HeLa	Adenocarcinoma from a 31 year old Afro-Caribbean woman	10-50 copies of integrated HPV 18	Masters 2002; ATCC
Caski	Endometrial carcinoma from a 40 year old Caucasian woman	600 copies of integrated HPV 16 per cell plus some sequences related to HPV 18	ATCC; Pattillo *et al.* 1977
W12	A CIN I from a 22 year old woman	Approximately 100 copies of episomal HPV 16 per cells	Doorbar *et al.* 1990; Stanley *et al.* 1989
SiHa	A grade 2 squamous cell carcinoma from a 55 year old Asian woman	Integrated HPV 16 at 1-2 copies per cell	ATT; Friedl *et al.* 1970

One way of studying HPV biology in the lab is to grow cell lines in a monolayer culture. In a monolayer culture the cells are attached to a solid surface, such as a flask or a petri dish and then allowed to grow and colonise the surface. To assist cell growth, feeder cells such as irradiated J2 3t3 fibroblast cells (Stanley, 2002, Eiselleova et al., 2008) are used to support growth.

Cells grown in monolayer culture can provide information about how HPV infection influences cellular morphology and how cells respond to adverse conditions such as chemicals and radiation. However, as the HPV lifecycle is closely linked to the stratification of cells (as described in section 1.2) there are limitations to research carried out in monolayers. To overcome the limitations of cells grown in monolayer, organotypic raft culture was established as a model of the typical HPV life cycle. Some benefits of employing the raft culture include:

1. Cellular differentiation and inappropriate mitosis can be measured.
2. Changes in HPV gene expression linked with cellular stratification can be measured
3. Cell cycling markers can be assessed as to when they are expressed in a differentiated culture.

Cell lines are seeded on to a collagen matrix containing live J2 3t3 fibroblast cells. Cells are then grown to confluence and lifted onto a metal grid. Cells are fed from beneath with a liquid air interface for 13 days before being harvested and sent for pathological analysis (Anacker and Moody, 2012, Andrei et al., 2010, Delvenne et al., 2001). This model allows the entire viral life cycle to be produced and viral particles can be produced if the viral genome is intact (Anacker and Moody, 2012). This can be used to assess the cellular phenotype and levels of cellular differentiation, cell cycle markers and mitotic activity can also be measured. In normal skin, it would be expected that cells in the upper levels of the epithelium would stop all mitotic activity; however cells with a persistent HPV infection may still have mitotic activity high in the epithelium.

In normal morphology, cells differentiate from the basal cells at the bottom of the epithelium, to the elongated terminally differentiated cells at the surface. In low-grade disease, the viral particles have entered the basal cells and the viral lifecycle progresses normally during cell differentiation, with viral particle production and release. In high-grade disease, virus particles are not released and cells become less differentiated in the upper levels of the epithelium. Cell lines are seeded onto a collagen plug, placed upon a metal grid creating a liquid air interface with E media. After 14 days, the cells are stratified on top of the

collagen plug and the cells are fixed and sent for histological analysis (Figure 1.8).
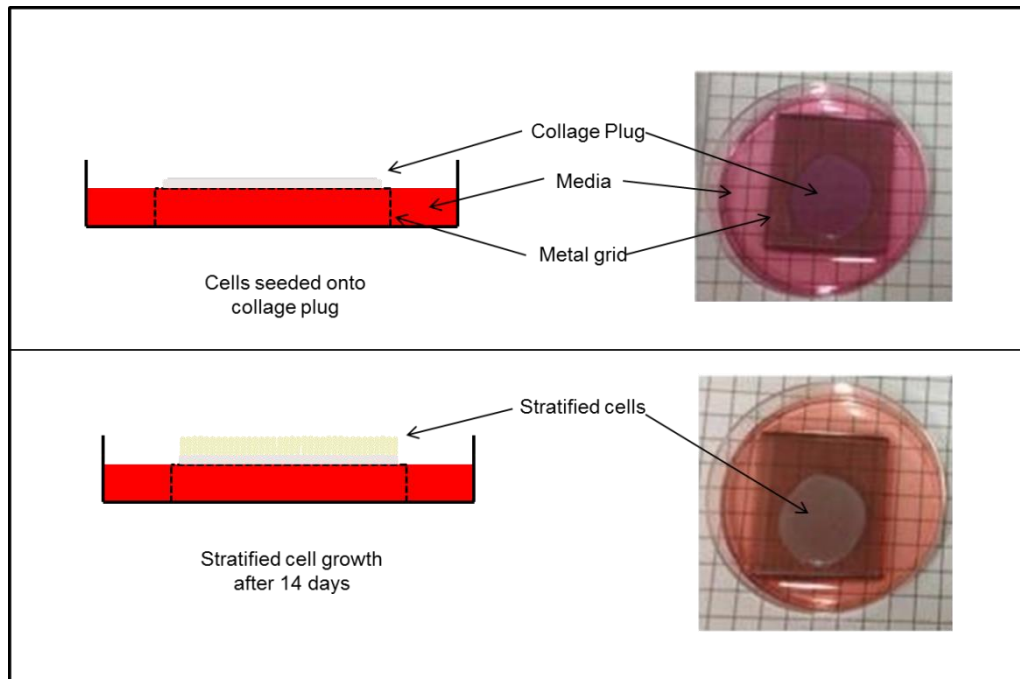
Figure 1.8. Process of cell growth in an organotypic raft culture. Cells were seeded onto collagen plugs and grown for 13 days. On the 14th day the collagen plugs (with the cells on top) were lifted onto a sterilised metal grid in a 10 cm petri dish using a metal spatula. E media was added below the culture until the top of the media touched the base of the collagen plug on top of the metal grid. Media was changed every day for 10 days until the cells had grown into a stratified layer, a white later can be seen on top of the collagen plug. At this point the cells were removed from the collagen and either placed into PBS for nuclear acid extraction or the collagen plug was lifted with the metal spatula and placed into formalin to fix the cells and allow downstream immunohistochemistry (IHC) analysis.

1.8 **Biomarker development**

Biomarkers are biological changes which can be measured to indicate the presence, progression or prognosis of disease. A suitable biomarker must show a specified and specific correlation between the biomarker and clinical outcome. The use of a biomarker must also show improved clinical outcome e.g. earlier treatment or more targeted treatment which will lead to a reduction of unnecessary treatment.

Biomarkers have already been employed in screening (BRCA1 in breast cancer), predicting response to therapy (KRAS mutation in colorectal cancer) and monitoring disease recurrence (AFP and LDH in germ cell tumours) (Henry and Hayes, 2012). Cancer Research UKs Biomarkers and Imaging Discovery and Development Committee (BIDD) has defined stages of biomarker research (Cancer Research UK, 2016) from initial basic research and rational to biomarker discovery and assay development and quantification (as shown in appendix 1). Biomarker discovery assay development is divided into three stages.

Stage 1: The development of accurate and reproducible assays. The distribution of the biomarker is defined in samples which are representative of the target patient population.

Stage 2: Biomarker distribution indicates a potential clinical utility. Standard operating procedures and performance of assays are defined. At this stage, the relationship between the biomarkers and clinical outcome is studied retrospectively.

Stage 3: Biomarkers in this stage have shown a correlation between distribution and clinical outcome and appropriate assay standards are met. At this stage, the assay should be fit for purpose.

If the potential biomarker fulfils the criteria at each stage the assay can then be quantified for clinical application. If a potential assay fails at any stage further basic research may be required or the research may be redirected.

Research in this thesis followed the Prognostic/Predictive Biomarker Roadmap for biomarker discovery, development and quantification. All results presented in this thesis will be considered within the framework set

up in the Prognostic/Predictive Biomarker Roadmap and next stages of biomarker discovery and development will be considered.

1.9 **Hypothesis**

This study aimed to examine several HPV regulatory factors to determine their suitability as clinical biomarkers. This involved development of assays, application to samples obtained from different cohorts of women to ascertain prognostic validity and development of an *in vitro* system to model the *in vivo* pathology. Table 1.3 outlines the hypothesis addressed by this PhD thesis.

Table 1.3 Hypothesis addressed by this PhD thesis

Hypothesis	
1	Novel HPV E2 tiling PCR assays specific for genotypes 31, 33, 35 and 51 will detect viral E2 gene disruption in DNA extracts from smears collected in SurePath™ LBC.
2	Disruption of the viral E2 (and/or E1 (HPV 16 only)) gene is (a) common in high-risk HPV infections (HPV 16, 31, 33, 35 and 51); (b) predominant in women with severe cytology and in women who go on to develop CIN/VaIN II+.
3	HPV hypermethylation is (a) Common in high-risk HPV 16 infections; and (b) Detectable at highest levels in women with severe cytology and in women who go on to develop CIN/VAIN II+.
4	Disruption of viral E2 gene coincides with hypermethylation of viral DNA in high-risk HPV 16 infection.
5	In vitro stratified epithelium models containing HPV integrated infections when compared to models of HPV episomal infection will display: (i) A less differentiated phenotype; (ii) Higher p16, Ki-67 and Hedgehog gene expression levels: (iii) Higher levels of viral methylation.

Chapter 2

Materials and Methods

2.1 Study Cohorts and Cell Lines

Two study cohorts were evaluated in this thesis: (1) The 'Baseline' cohort, which consisted of samples taken from young women attending their first cervical smear in Wales (Hibbitts et al., 2014); (2) The 'HPV Integration as a biomarker for cervical disease (HI-BCD)' cohort, representing a pilot study of potential biomarkers of disease in women attending colposcopy due to CIN. Study protocols for both studies are shown in Appendix 1, including inclusion and exclusion criteria for each research study respectively.

2.1.1 Study 1: Young women attending their first cervical smear

This study was conducted in April 2009 - July 2010 and involved taking pseudo-anonymous residual LBC (BD SurePath^{TM}, Source Bioscience) samples from women aged 20-22 attending their first cervical smear test. This was done in conjunction with Cervical Screening Wales

(CSW) who confirmed eligibility of samples and provided clinical data to link with the anonymous case ID of each sample. Participating cervical screening laboratories included Welsh laboratories in: Llandough, Singleton, Royal Gwent Hospital, Wrexham Maelor, Ysbyty Glan Clwyd, Withybush, Royal Glamorgan, Princess of Wales, Prince Charles and West Wales General, and Hereford, Chester and Shrewsbury laboratories which lie on the England-Wales border. Ethical approval was obtained from the Dyfed Powys Local Research Ethics Committee (08/WMW01/69).

2.1.1.1 Sample Collection, Storage and Processing

In total 14,128 LBC samples were collected and processed by CSW according to CSW guidelines (Cervical Screening Wales, 2005). These cases were managed according to standard guidelines and data collected from this study did not impact routine patient treatment or management. Samples that met all the study inclusion criteria were flagged and transported to the HPV Research laboratory in the University Hospital Wales, Cardiff.

Residual material from cervical smear samples were transferred from the SurePath™ pot to a 15 ml falcon tube. The falcons were spun for

10 minutes at 2,100 x g at 4°C and the supernatant was aspirated, leaving the cells pelleted at the bottom. These were then re-suspended in 2 ml of 10 mM Tris (pH 7.4) and 1 ml of the cell suspension was transferred into two 1.5 ml eppendorfs for each LBC sample 4 tubes were stored for further analysis: 1 x 1.5 ml tube containing the cell pellet, 1 x 1.5 ml tube containing 300 µl of cell suspension; and 2 x 0.5 ml tubes containing 100 µl of cell suspension each. Samples from this study were then used to evaluate the genome status of women attending their first routine cervical smear. Samples were included in subsequent analysis if they were HPV positive by type specific PCR for the HPV type under investigation (Figure 2.1).
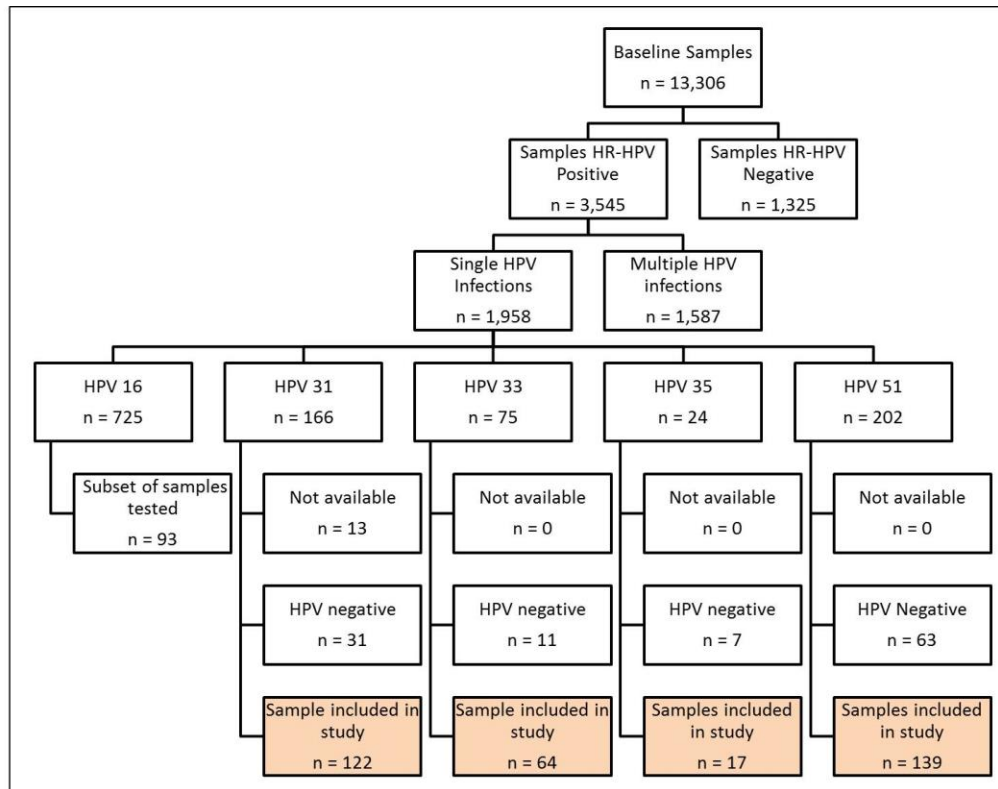
Figure 2.1 Flow diagram of samples included in a study of young women attending their first cervical smear. Samples included in this study had a single HPV infection and were HPV positive by type specific PCR (HPV 31, 33, 35 and 51 respectively).

2.1.2 Study 2: HI-BCD Study Cohort

The HI-BCD study ran from March 2011 to April 2014 and consisted of 200 women who attended colposcopy at a single site within the Cardiff and Vale Health Board. Eligible women were identified by Cervical

Screening Administration Department of CSW and were invited to donate a smear sample of convenience when attending colposcopy following an abnormal cervical smear result. An amendment to the study protocol allowed for the collection of smear samples from women who had cancer. Ethical approval was obtained from the Dyfed Powys Local Research Ethics Committee/National Institute for Social Care and Health Research Permission Co-ordinating Unit. The Cardiff and Vale R&D committee also approved this study.

2.1.2.1 Sample Collection, Storage and Processing

Cervical smear samples were collected at Llandough Hospital, Cardiff by colposcopy nurses and colposcopy gynaecologists. During the patients' visit a smear sample was taken with a standard cytology brush. This was then placed into 2 ml of RNA Protect® (QIAGEN Ltd, Manchester, UK) in a sterile 30 ml universal container. The tip of the brush was agitated in the media and then detached. Samples were electronically anonymised and then transported to the HPV Research Group, Cardiff University at room temperature, stored at -80°C. HPV type was assessed by Papillocheck® and by HPV 16 type specific *E6* PCR. Confirmed HPV 16 positive samples were used to assess viral integration, methylation and gene expression (Figure 2.2)
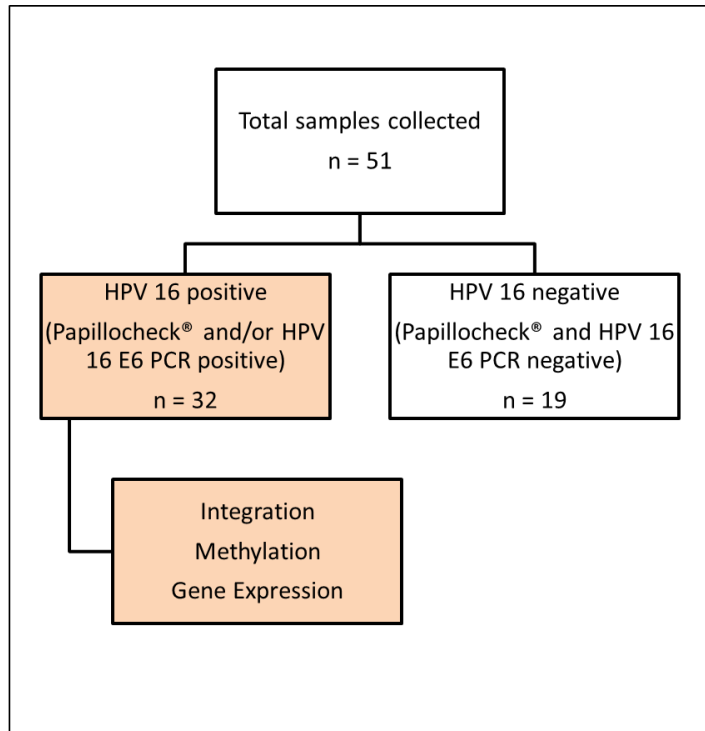
Figure 2.2. Flow diagram of samples included in HI-BCD study. Thirty two samples included in this study had a HPV 16 infection (single or multiple infections) as assessed by Papillocheck® and/or HPV 16 *E6* type specific PCR.

2.2 **Cell Lines**

Several cell lines were employed in this study as a model of HPV 16 disease. Cell lines used in this thesis along with their origin and growth conditions are detailed in Table 2.1.

Table 2.1 Origin, growth conditions and source/reference of cell lines used in this research. All cell lines used in this work were grown at $37^{o}C$ with 5% CO_2.

Cell Line Name	Origin	Growth Conditions	Source/reference
CU-VA-9A-16E	Homogeneous monoclonal adherent human epithelial short term cell lines originating from a 31-year-old with VaIN III (vaginal)	Grown on irradiated J2 3t3 fibroblast feeder cells in E media. Cells were split at 80% confluence and re-plated at a density of 8 x 10^4 cells/cm^2 with 2 x 10^6 irradiated J2 3t3 fibroblasts.	Onions, 2013
CU-VA-9H-16I	Homogeneous monoclonal adherent human epithelial short term cell lines originating from a 31-year-old with VaIN III (vaginal)		
CU-VU-8P-16E	Homogeneous monoclonal adherent human epithelial short term cell lines originating from46-year-old with VIN III (vulval)		
CU-VU-8M-16I	Homogeneous monoclonal adherent human epithelial short term cell lines originating from46-year-old with VIN III (vulval)		
W12 Ser4B	Immortalised adherent human epithelial cell line derived from 22-year old with CIN I (cervix)	Cells were grown on E Media. Cells were split at 80% confluence and inoculated at 8 x 10^4 cells/cm2.	Prof Nick Coleman, University of Cambridge
HEKn	Primary neonate epithelial keratinocytes isolated from neonatal foreskins.	Cells were inoculated at 2.5 x 10^3 cells/cm^2 in EpiLife® Medium with growth supplement and cells were split at 80% confluence and re-plated at 2.5 x 10^3 cells/cm^2.	Cascade Biologics, Invitrogen Cell Culture (Cat C0015C)
BU21HFK	Primary human foreskins were obtained from new born foreskin circumcision tissue.	Cells were inoculated at a density of 2 x 10^4 cells/cm2 and grown on irradiated J2 3t3 fibroblast feeder cells in E media.	Wilson et al, 2007
BU21HFK (HPV18)	Primary human foreskins were obtained from new born foreskin circumcision tissue. These cells were then transfected with wild type HPV 18 which was confirmed by Southern blot.	Cells were inoculated at a density of 2 x 10^4 cells/cm2 and grown on irradiated J2 3t3 fibroblast feeder cells in E media.	Wilson et al, 2007
J2 3t3 mouse fibroblasts	Immortalized mouse fibroblastic cell line used to supplement the growth of VaIN and VIN cell lines	J2 3t3 cell lines were inoculated at 1.2 x 10^4 cells/cm2 and grown using DMEM supplemented with 100x Penicillin/Streptomycin and 10x Foetal Calf serum.	Prof Nick Coleman, University of Cambridge

2.3 Molecular Cloning of Plasmid DNA

Reagents used in the molecular cloning of the plasmid DNA are listed with their supplier in Table 2.2.-

Table 2.2 Reagents used in the molecular cloning of the plasmid DNA and their supplier.

Reagents	Supplied by
Ampicillin	Sigma-Aldrich
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal)	
Isopropyl ß-D-1-thiogalactopyranoside (IPTG)	
lysogeny broth (LB) Agar	
LB Broth	
3-alpha competent *E. coli*	New England Biolabs
Super Optimal broth with Catabolite repression (SOC) Media	
Absolute Ethanol	Fisher
Glycerol	
QIAamp® DNA Mini and Blood Mini Kit	QIAGEN

2.3.1 HPV Plasmids

Information on positive control HPV plasmids obtained from Professor Dr Ethel-Michele de Villiers from the HPV Reference lab, Heidelberg, Germany and the plasmids that were kindly donated are detailed in Table 2.3

Table 2.3 Detail of Positive Control Plasmids used in this study

Plasmid	Vector	Cloning Site	Plasmid Concentration	Antibiotic selection	Supplied by
HPV 31	pT713	EcoR 1	1.05 mg/ml	Ampicillin	QIAGEN Inc. Gaithersburg, USA
HPV 33	pBR322	BGl I	10 ng/ul	Ampicillin	Dr M. Favre, Insitut Pasteur, Paris
HPV 35	pT713	BamHI°	1.12 mg/ml	Ampicillin	QIAGEN Inc. Gaithersburg, USA
HPV 51	pUC13	Hind III	unknown	Ampicillin and blue/white selection	Prof. Dr de Villers, Heidelberg, Germany on behalf of S. Silverstein

2.3.2 Agar Plate Production

Lysogeny broth (LB) Agar was produced by adding 250 ml of deionized water to 8.75 g of LB Agar powder. The LB agar was autoclaved and 500 µl of ampicillin was added to the agar solution. Plates were poured, set and inverted in the cold room until they were required. Plates

for HPV 51 cloning had 40 µl of 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-gal) and isopropyl β-D-1-thiogalactopyranoside (IPTG) added to the surface of each plate. These were then allowed to dry in the incubator at 37°C for 30 minutes before use. All plates were stored in the incubator at 37°C prior to *Escherichia coli* (*E. coli*) plating.

2.3.3 *E. coli* Transformation with Plasmid DNA

E. coli cells were thawed on ice and 50 µl was pipetted into two 1.5 ml eppendorf tubes. Plasmid DNA (2 µl) was added to one eppendorf and 2 µl of water was added to the negative control. The tubes were mixed and placed on ice for 30 minutes. The *E. coli* was then heat-shocked at 42°C for 30 seconds and subsequently placed on ice for a further 5 minutes. Room temperature super optimal broth with catabolite repression (SOC) media (950 µl) was added and placed in an incubator at 37°C for 60 minutes while shaking at approximately 0.35 x g. A serial dilution of *E. coli* cultures was performed and 50 µl of each dilution was plated onto an agarose gel and incubated for 16-18 hours at 37°C.

A single colony was picked from the transformed plate and placed into 2 ml of LB Broth and left overnight in a shaking incubator at 37°C

(approx. 0.35 x g). The following day a glycerol stock was produced from the culture and DNA was extracted from the culture.

To produce a glycerol stock 50 µl of the culture was diluted in 5 ml of LB broth and incubated at $37^{\circ}C$ with 0.5 x g shaking for 75 minutes. Then 800 µl of the growing *E. coli* was added to a 1.5 ml eppendorf with 200 µl of autoclaved glycerol. This was briefly mixed and stored at $-80^{\circ}C$ for future use.

2.4 Nucleic Acid Extraction from Clinical Samples and Cell Culture

The nucleic acid extracted in this study was assessed for quality and quantity using a Thermo Scientific Nanodrop® 1000 Spectrophotometer. As nucleic acids have absorbance maxima at 260 nm and proteins at 280 nm, the ratio between 260 nm and 280 nm is used as a measure of sample purity. For DNA an A260/A280 ratio of ~1.8 is considered pure and an A260/A280 ratio of ~2.0 is considered pure for RNA. A low 260/280 ratio may be indicative of residual reagent being carried over from the extraction protocol (Thermo Scientific, 2012).

2.4.1 Plasmid DNA Extraction

Plasmid DNA extraction was carried out using a QIAamp® DNA Mini and Blood Mini Kit. In brief, 1 ml of *E. coli* culture was centrifuged for 5 minutes at 4,600 x g. The supernatant was removed leaving the cell pelleted. DNA was extracted from the cell pellet using the bacteria protocol and the sample was eluted with AE buffer and stored at -20°C.

2.4.2 DNA Extraction from Baseline Study Cohort

Baseline samples were previously pelleted and stored at -80°C. C33a cell pellets and transformed *E. coli* containing the HPV type of interest were used as positive controls. Samples were re-suspended using 500 µl of Tris and 250 µl was used for DNA extraction. Sterile water was used as a negative extraction control.

To each sample 50 µl of Proteinase K was added and samples were placed in an incubator at 56°C with shaking for 16-18 hours. Samples were then placed into a dry heating bath at 80°C for 10 minutes

then transferred to a chilled rack and placed in the fridge ($4^{o}C$) for a further 10 minutes. Samples were then spun in a chilled microfuge at $4^{o}C$ for 10 minutes at 14,200 x g. The supernatant containing DNA was recovered and the pellet was discarded.

2.4.3 DNA and RNA Extraction from HI-BCD Study Cohort and cultured cells

All Prep DNA/RNA Mini Kit (QIAGEN Inc, Gaithersburg, USA) was used to extract DNA and RNA from both clinical samples in the HI-BCD cohort and cell lines grown in vitro.

Cells grown in vitro were trypsinised, resuspended in growth media and counted using a haemocytometer. For this extraction kit it was recommended that no more than 1×10^{7} cells were used per spin column. If the column was overloaded then it would lead to the co-purification of DNA and RNA. Both HI-BCD samples and cell lines DNA and RNA were then extracted using the protocol for 'Simultaneous Purification of Genomic DNA and total RNA from Animal Cells' according to manufacturer's instructions. In brief, RNA was eluted using 40 µl of RNase-free water and the DNA was eluted in two steps using 50 µl of EB

buffer in each step to total 100 µl of eluted DNA. DNA and RNA were then stored at -20°C until they were required for downstream analysis.

2.5 Identification of HPV Genes and Gene Expression by Molecular Analysis

The identification of HPV genes and gene expression in this thesis relied on PCR to amplify HPV genes of interest. PCR is a key molecular biology technique which involves cycles of denaturing DNA, annealing primers and the synthesis of a new DNA strand. The numerous cycles produce a large quantity of the specific DNA region of interest. This can then be sequenced or run on an agarose gel to visualise the size and quantity of DNA (Mullis et al., 1986, Bartlett and Stirling, 2003).

Reagents used in the molecular analysis are listed in Table 2.4. Table 2.5 shows the reagents and quantities used in each assay Mastermix and the PCR conditions are listed in Table 2.6. For BetaGlobin PCR, *E2* tiling PCR, *E1*/*E2* PCR, DIPS and methylation analysis the PCR Mastermix was produced in a PCR UV sterile hood. Mastermix was made for the appropriate number of samples plus one to account for pipetting loss and loaded into each well of the PCR plate and DNA was added to

the assigned location on the PCR plate. The plate was sealed using adhesive lids and placed in a thermocycler. Once the PCR cycle was complete the samples were run on a 2% agarose gel and viewed under UV using a transilluminator. All molecular tests presented in this thesis were carried out in triplicate.

Table 2.4 Reagents and their suppliers used in the identification of HPV genes and gene expression by molecular analysis

Reagents	Supplied by	Reagents	Supplied by
Zymo Ez DNA Methylation Bisulfite Conversion Kit	Cambridge Bioscience	10 x HotStar Taq Buffer with 15 nM MgCl$_2$	QIAGEN
Zymotaq Reagents		HotStar Taq	
Absolute Ethanol	Fisher	PyroMark Buffer Solution	
DECP treated water		PyroMark Gold Q96	
PCR grade water		LightCycler DNA Master SYBR Green I reagent kits	Roche Applied Diagnostics
TrisHCL		LightCycler reaction capillary tubes	
Streptavidin Sepharose Beads (30% w/v)	GE Healthcare Life Science	TAE (Tris-acetate-EDTA)	Sigma Aldridge
Molecular Grade Agarose	Geneflow	EDTA	
PCR Ladder (100kb and 1000kb)		Ethidium Bromide 10 mg/ml	
dNTPs	Invitorgen	Random Primers	
Taq			
MgCl$_2$ (50 mM)			
10 x PCR Buffer			
Superscript® III			
RNaseOUT™			
BSA	New England Biolabs		
Dnase 1 & Dnase 1 Reaction Buffer			
Sau3Al & Sau3Al Buffer			
T4 ligase and Ligase Buffer			

Table 2.5 (part 1) Mastermix used in the identification of HPV genes and gene expression assays

Assay	Reagent	Concentration	Volume (µl) n = 1
Beta-globin	DNA		5
	dNTPs	2 mM	2.5
	PCR Buffer	10 x	2.5
	MgCl2	50 mM	0.875
	Primer PCO3	5 µM	2.5
	primer PCO5	5 µM	2.5
	Taq		0.1
	H_2O		9.025
	Total Volume		25
E2/E1 tiling PCR	DNA		1.5
	dNTPs	2 mM	2.5
	PCR Buffer	10 x	2.5
	Forward Primer	5 µM	2.5
	Reverse Primer	5 µM	2.5
	HotStar Taq		0.125
	H_2O		13.375
	Total Volume		25
DIPS Adapter	AL1 Primer	100 µM	25
	AS Primer	100 µM	25
	TrisHCL	66 mM	50
	Total Volume		100
Sau3AI Digestion	DNA	1.2 µg	15
	Sau3AI	10 U	2.5
	Sau3AI Buffer	10 x	5
	BSA	100 x	0.5
	H_2O		27
	Total Volume		50
DIPS Ligation	DNA from Digestion		50
	Ligase Buffer	10 x	6
	T4 Ligase	400 U	1
	Sau3AI Adapter	0.25 µM	1.2
	H_2O		1.8
	Total Volume		60

Assay	Reagent	Concentration	Volume (µl) n = 1
DIPS Linear PCR	Ligation Product		3
	10x buffer with 15 mM $MgCl_2$	1 x	2.5
	dNTPs	2 mM	2.5
	Linear HPV PCR Primer	10 µM	0.5
	HotStar Taq	1 U	0.125
	H_2O		16.375
	Total Volume		25
DIPS Nested PCR	Linear PCR Product		2
	10 x buffer with 15 mM $MgCl_2$	1 x	2.5
	dNTPs	2 mM	2.5
	HotStar Taq	1 U	0.125
	Nested PCR Primer	10 µM	1
	AP1 Primer	10 µM	1
	H_2O		15.875
	Total Volume		25

Table 2.5 (part 2) Mastermix used in the identification of HPV genes and gene expression assays

Assay	Reagent	Concentration	Volume (µl) n = 1
Reverse Transcription Step 1	RNA	0.5 µM	5
	Random Primers	200 ng/µl	1
	dNTP	10 mM	1
	RNase free water		6
	Total Volume		13
Reverse Transcription step 2	RT Step 1		13
	First Strand buffer	5x	4
	DTT	0.1 M	1
	RNase™ OUT		1
	Superscript® III	200 U/µl	1
	Total Volume		20
TBP2 qRT-PCR	cDNA		2
	FS Mix		2
	Forward Primer	5 µM	2
	Reverse Primer	5 µM	2
	MgCl$_2$	25 mM	2.4
	H$_2$O		9.6
	Total Volume		20

Assay	Reagent	Concentration	Volume (µl) n = 1
E2 qRT-PCR	cDNA		2
	FS Mix		2
	Forward Primer	5 µM	2
	Reverse Primer	5 µM	2
	MgCl$_2$	25 mM	2
	H$_2$O		10
	Total Volume		20
HPV 16 E2 Methylation	DNA		2
	ZymoTaq Premix		25
	Forward Primer	5 µM	2
	Reverse Primer	5 µM	2
	MgCl$_2$	25 µM	2
	H$_2$O		17
	Total Volume		50
HPRT/E6 qRT-PCR	cDNA		2
	FS Mix		2
	Forward Primer	5 µM	2
	Reverse Primer	5 µM	2
	MgCl$_2$	25 mM	1.6
	H$_2$O		10.4
	Total Volume		20

Table 2.6 Thermocycler conditions used in the identification of HPV genes and gene expression by molecular analysis

Assay	Temperature (°C)	Time	Number of Cycles	Assay	Temperature (°C)	Time	Number of Cycles
Beta-globin	94	4 Minutes	1	HPV 16 *E2* Methylation	95	10 minutes	1
	94	30 seconds	40		95	30 seconds	40
	55	30 seconds			48	45 seconds	
	72	30 seconds			72	30 seconds	
	72	4 Minutes	1		72	10 minutes	1
E2 tiling PCR	95	15 minutes	1	Reverse Transcription	25	5 minutes	1
	94	30 seconds	40		50	50 minutes	1
	57	30 seconds			70	15 minutes	1
	72	30 seconds		qRT-PCR	95	10 minutes	1
	72	5 minutes	1		95	10 seconds	60
DIPS Digestion	37	2 hours	1		57	5 seconds	
	80	20 minutes	1		72	5 seconds	
DIPS Ligation	23	2 hours	1				
	65	10 minutes	1				
DIPS Linear PCR	95	15 minutes	1				
	94	30 Seconds	40				
	66	30 seconds					
	72	3 minutes					
	72	7 minutes	1				
DIPS Nested PCR	95	15 minutes	1				
	94	30 seconds	30				
	66	30 seconds					
	72	3 minutes					
	72	7 minutes	1				

2.5.1 PCR Primer Design for *E6, E7, E2* tiling PCR and DIPS Assays

Software and archives used for primer design is shown in Table 2.7

Table 2.7 Software and archives used in primer design and analysis.

Software/Archives	Supplied by
Artemis	Sanger Institute
BioEdit	Tom Hall, Ibis Biosciences
European Molecular Biology Laboratory – The European Bioinformatics Institute (EMBL-EBI)	European Nucleotide Archive (ENA)
NCBI Primer Blast	National Centre for Biotechnology Information (NCBI)
Net Primer	Premier Biosoft

Genome files of the HPV types of interest were obtained from the European Nucleotide Archive (ENA) (Hinxton, Cambridgeshire) and opened in Artemis (Sanger Institute) (Rutherford et al., 2000), this allowed for the genome to be annotated with restriction sites and the location of viral genes. A FASTA file was also downloaded and this was used to design the primers using National Institute of Biotechnology Information (NCBI) Primer Basic Local Alignment Search Tool (BLAST) (Ye et al., 2012). The NCBI Primer BLAST options were kept on the default setting

and the preferred primer location was selected. Appropriate primer information was saved and the primer sites were annotated on the Artimis file.

The primers that appeared to be suitable were input into Net Primer (PREMIER Biosoft International, Palo Alto, CA) to check for unwanted secondary structure formation. Primers, which formed secondary structures with a very low ΔG, were excluded. The suitable primers were also checked for unwanted secondary structure formation with the adapter primer as well as the corresponding linear/nested primer. Finally the primers were loaded onto BioEdit software (Ibis Bioscience) which checked the primer sequences against 13 of the most common HR-HPV types (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68) to ensure that primers were specific to the HPV type of interest.

2.5.2 Agarose Gel Electrophoresis of PCR Products

Agarose gel electrophoresis was used to visualise DNA amplification and provide a semi-quantitative analysis. The agarose powder was weighed out according to the required gel concentration. Tris-acetate-EDTA (TAE) buffer was added and placed into a 500 ml Duran

Bottle. This was placed into a standard microwave and heated for several minutes with occasional swirling to ensure the agarose and TAE were well mixed. Once the agarose was fully dissolved and the solution was clear and viscous the Duran was placed into cold water and allowed to chill to approximately 60°C at which point 3% ethidium bromide was added; the solution was mixed well and poured into the gel tray. Combs were added and the gel allowed to set for 15 to 20 minutes.

Once fully set the gel was placed into a suitably sized tank. 1% TAE buffer was added to the tank until the gel was covered with a thin layer of buffer. The PCR samples were then added to the gel with 20% Orange G and at least one DNA ladder (100 base pairs (bp) or 1 kb) was added per row of samples. Electrophoresis was carried out until the sample had almost reached the bottom of the gel. The gel was then transferred to the transilluminator and visualised.

2.5.3 *E2* and *E1/E2* tiling PCR

The *E2* PCR assay uses multiple PCRs spanning the *E2* gene (Collins et al., 2009) (Figure 2.3). This assay is a primary indicator of whether the virus has integrated into the host genome.
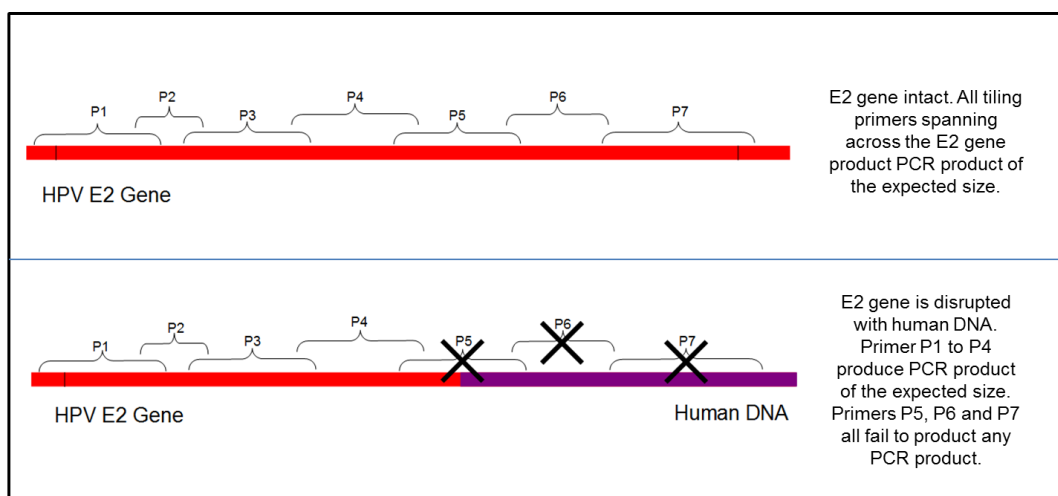
Figure 2.3. Principle of the *E2* PCR assay. P1-P7 represents the individual PCRs that span the *E2* gene of the HPV genome. (A) The genome is intact and all PCRs produced the expected product. (B) Integration of HPV into the human DNA. Here the *E2* gene is disrupted and three of the PCRs will fail, indicating integration.

To produce *E2*/*E1* tiling PCR Mastermix reagents were mixed as noted in Table 2.4. One master mix was made for each primer set (details can be found in Appendix 3). The Mastermix was aliquoted into 96 well PCR plates. DNA from clinical samples (2 µl) was added to the individual wells. Positive (type specific plasmid DNA) and negative controls (PCR grade water) was added into the remaining well. The PCR plate was sealed and placed in a thermocycler and *E2*/*E1* PCR protocol was carried out (as detailed in Table 2.5)

72

Once the thermocycler programme was complete five microliters of PCR product was then run on a 2% agarose gel. PCR products from clinical samples were compared to positive control to assess if the *E2*/*E1* gene was intact.

2.5.4 Detection of Integrated Papillomavirus Sequences PCR

The DIPS assay involves digesting clinical samples with restriction enzymes and ligating nucleotide adapters to the ends of the digested DNA. A HPV specific linear PCR is carried out followed by a nested PCR with primers specific for HPV and the adapter primer (Luft et al., 2001). PCR products are then run on an agarose gel and bands which deviate from the expected size can be sequenced to identify the sequence of the HPV-human viral junction (Figure 2.4).
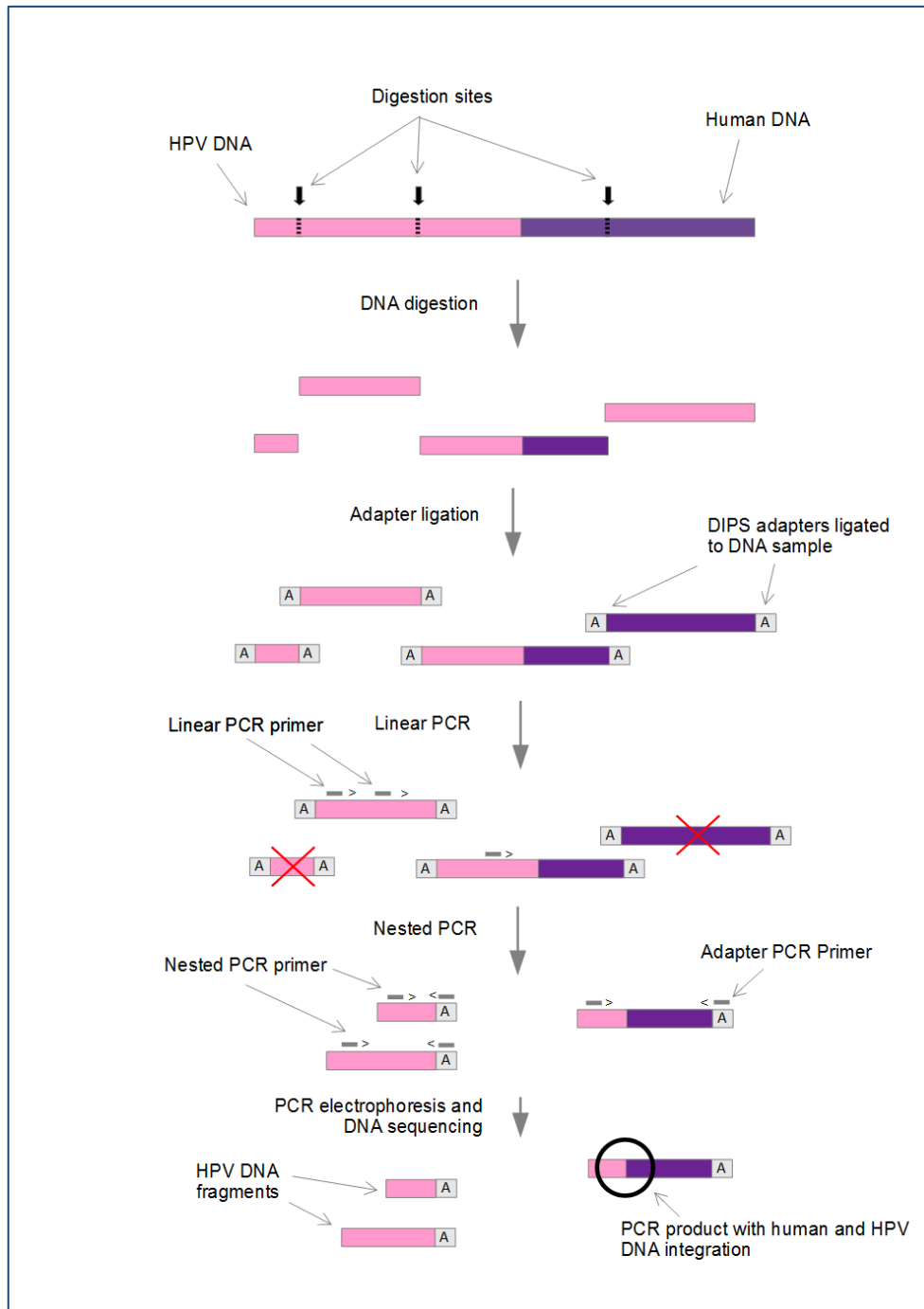
Figure 2.4 Principle of the DIPS assay. The sample is digested with digestion enzyme Sau3AI. DNA adapters are then ligated onto the DNA sample before a linear PCR is carried followed by a nested PCR with primers for HPV and the adapter sequence. The PCR products can then be visualised and sequenced.

To produce DIPS adapter's reagents were mixed as noted in Table 2.4. The mix was aliquoted into PCR tubes and placed in a thermocycler and gradually cooled from 90°C to 4°C overnight.

The genomic DNA was digested using Sau3AI enzyme by mixing reagents listed in table 2.4 and then incubated at 37°C for 2 hours followed by heat inactivation at 80°C for 20 minutes. The adapter was then ligated onto the digested PCR product and incubated as detailed in table 2.5. The linear PCR was then carried out using type specific primers, followed by a nested PCR using HPV 16 specific primers and AL1 primer which corresponds to the DNA sequence of the adapter.

PCR product (5 μl) was then run on a 2% agarose gel. If the sample produced dissimilar bands from those produced by the plasmid control the rest of the PCR product (20 μl) was run on a fresh 2% agarose gel and the band extracted from the gel for further sequencing.

2.5.5 PCR Gel Extraction

The PCR product was run on a gel and photographed using a low UV light intensity on the transilluminator. Individual wells were processed on to the table top UV transilluminator. Using a sterile scalpel, the bands of interest were cut from the gel and placed into individual tubes.

The DNA from each band was then purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare Life Sciences, Buckinghamshire, UK). In brief, 300 µl of Capture buffer type 3 was added to each gel slice and placed into a 60°C water bath for 15-30 minutes to allow the gel to dissolve. Once the gel was fully dissolved, if the solution was pink then ~10 µl of sodium acetate was added to bring the solution to a yellow/pale orange colour. The content of the eppendorf was placed into a GFX MicroSpin column in a 2 ml collection tube. This was incubated at room temperature for 1 minute and then spun at 16,000 x g for 30 seconds. The flow-through was discarded and the column was returned to the collection tube. The sample was then washed using 500 µl of wash buffer type 1. This step was carried out twice followed by a third spin (16,000 x g) to ensure the column was completely dry. The column was then transferred to a fresh 1.5 ml eppendorf and the sample was eluted using 25 µl of elution buffer type 6. This was incubated for 1 minute at room temperature and then spun at full speed for 1 minute. From the 25 µl of purified DNA, 5 µl was run on a new agarose gel to ensure the purification was successful and to estimate the concentration of DNA in

the sample. The remaining 20 µl was sent for sequencing, if the purification was successful.

2.5.6 DNA Sequencing and Analysis

PCR bands were sent to Source Bioscience (Nottingham, UK) and sequenced using Sanger sequencing. Sequencing results were then assessed using NCBI nucleotide BLAST to identify the location of virus-host integration site for the HPV genome and the host (Figure 2.5).
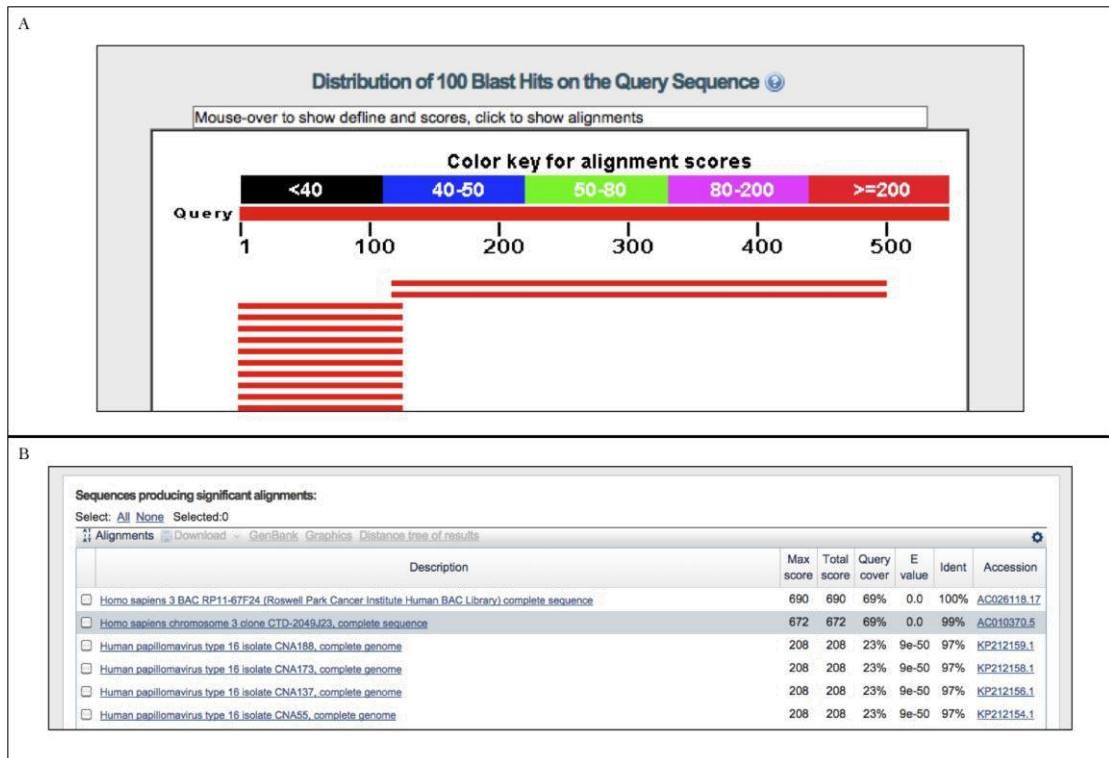
Figure 2.5 Example of NCBI nucleotide blast used to identify the location of virus-host integration site for the HPV genome and the host.

2.5.7 qRT PCR

A two-step qRT-PCR assay was used in this study to assess the level of mRNA from clinical samples. The first step involved converting mRNA into complementary DNA by a reverse transcriptase reaction. In this reaction the retroviral enzyme M-MLV reverse transcriptase enzyme is

used along with random primers to produce cDNA (Invitrogen, 2004). This is then used as a template for the second step.

The qRT-PCR reactions involves a PCR reaction like conventional PCR (section 2.5), however a fluorescent dye or probe is used to take a reading after each amplification cycle. This allows the amount of DNA produced after each PCR cycle to be measured and normalised against stable reference genes. In this work, SYBR Green I dye was used to quantify the mRNA expression. SYBR green I preferentially binds to double stranded DNA and in the qRT-PCR reaction the level of fluorescence increases relating to PCR production and subsequent fluorescence is measured. The level of expression of genes of interest is then reported compared to the expression levels of reference genes, in this thesis reference genes *TATA-binding protein 2* (*TBP2*) and *Hypoxanthine Phosphoribosyltransferase* (*HPRT*) were used.

2.5.7.1 Reverse Transcription

Firstly, DNase treatment was carried out to remove any DNA that has been carried over in the extraction procedure. RNaseZap (Sigma Aldrich) was used in the PCR hood to remove any RNase contamination

from the hood and equipment which could degrade the RNA of interest. Reagents detailed in Table 2.4 were mixed on ice, incubated at room temperature for 10 minutes and then DNaseI was inactivated with 25 mM EDTA. Finally, samples were heated at 65°C for 10 minutes.

To synthesise cDNA reagents in Table 2.4 (Reverse Transcription part 1) were utilised. This was heated at 65°C for 5 minutes before being placed on ice for at least 1 minute. Mastermix Reverse Transcription part 2 was then produced and 7 µl was added to each sample from Reverse Transcription part 1. Samples were then placed on a thermocycler as detailed in Table 2.5 and then diluted 1/10 with H_2O before being stored at -80°C.

2.5.7.2 qRT-PCR

qRT master mixes were prepared as detailed in table 2.4 and 18 µl of each Mastermix was placed in each capillary tube with 2 µl of cDNA. Glass capillaries were centrifuged at 750 x g for 5 seconds to transfer the PCR mix to the capillary tube. Glass capillaries were then placed in the light cycler carousel, which was set to the optimum PCR condition for the gene of interest (detailed in Table 2.5).

2.5.7.3 qRT-PCR analysis

qRT-PCR analysis was carried out using LightCycler® software and qBase^{PLUS} (Hellemans et al., 2007). LightCycler® software automatically calculated the Ct value (the number of cycles required for the fluorescent signal to exceed background) using second derivative maximum. LightCycler® software was also used to identify samples which failed to amplify the specific target; these samples were excluded from further analysis.

The data was then transferred into qBase^{PLUS} software which was used to calculate the ratio between the genes of interest (HPV *E2* and *E6*) compared to the reference genes (*HPRT* and *TBP2*). To reduce errors due to sample quality and quantity the PCR efficiency was calculated for each target by using a dilution series of CaSki cDNA (Bryant, 2012). This data was uploaded to qBase^{PLUS} and used to quantify PCR efficiency and compensate for sample variability (Hellemans et al., 2007). Expression levels were reported as calibrated normalised relative quantities (CNRQ) Expression levels were calculated relative to expression in the HPV 16 cell line CaSki. This allowed for the data to be comparable between studies.

2.5.8 DNA Methylation Study

HPV DNA methylation was assessed in this work by bisulphide conversion of sample DNA followed by PCR and Pyrosequencing. As described in Section 1.4.3. The bisulfide converted DNA is then used as the PCR template with primers designed specifically to not anneal to any cytosine bases which may have been converted to uracil. For the downstream Pyrosequencing reaction one of the primers used in this PCR reaction is biotinylated.

The PCR product is denatured; this allows the sequencing primer to bind to a single strand of DNA. To this is added DNA Polymerase, ATP sulfurylase, luciferase, apyrase and the substrates 5'phosphosulphate(APS) and luciferin. deoxiribose nucleotide triphosphates (dNTPs) are then added in sequential order; the DNA polymerase catalyses the addition of the dNTP to the sequencing primer/template. The incorporation of the dNTP is accompanied by a release of phosphate (PPi). ATP sulphurylase converts PPI into ATP in the presence of APS. This drives the luciferase –mediated conversion of luciferin to oxyluciferin which generated visible light. This intensity of the

light is proportional to the amount of ATP and therefore the amount of dNTPs incorporated. The addition of dNTPs is set to occur in sequential order and the level of signal with each dNTP addition can be measured. This can then be used to calculate the percentage of DNA methylation at each CpG site. To ensure there is no cross talk between nucleotide additions Apyrase degrades unincorporated dNTPs and ATP, when the degradation is complete the next nucleotide in sequence is added (QIAGEN, 2017b)

2.5.8.1 Bisulfide Conversion

All the reagents were prepared as specified in the kit protocol, in brief, CT conversion reagents were prepared by adding 750 µl of H_2O and 185 µl of M-Dilution buffer into a tube of CT conversion reagents and vortexed at a low speed until required. 7.5 µl of M-Dilution buffer was added to 500 ng of DNA and the volume was adjusted to 50 µl with PCR grade H_2O. Samples were mixed by repeat pipetting then incubated at $42^{o}C$ for 30 minutes on a hot block. Pre-prepared CT conversion reagents (97.5 µl) were added to each DNA sample and mixed. Samples were then incubated on the heating block at $50^{o}C$ for 12 – 16 hours covered with a black box to reduce the exposure to light. 12 – 16 hours later the samples were incubated on ice for 10 minutes. M-Binding Buffer (400 µl) was added to a Zymo-Spin column in a 2 ml collection tube. The sample were

then added to a spin column and mixed by inverting several times. The column and collection tube was then centrifuged at 14,200 x g for 30 seconds. 100 µl of M-Binding Buffer was added to each column and the sample was again spun at 14,200 x g for 30 seconds. M-Desulphonation Buffer (200 µl) was added to each sample and incubated at room temperature for 20 minutes. The samples were spun again at 14,200 x g for 30 seconds. Finally, 20 µl of M-Wash Buffer was added to the column, which was again spun at 14,200 x g for 30 seconds. 10 µl of M-Elution buffer was added to each tube and spun at 14,200 x g for 30 seconds to elute the sample. 90 µl of PCR grade H_2O was added to each sample to dilute to 1/10. These samples were then used in the pyro sequencing PCR or stored at -20^oC until required.

2.5.8.2 Pyro sequencing PCR

Pyro sequencing PCR reactions were carried out using ZymoTaq PCR reagents (Zymo Research Corporation, California, USA). Samples were diluted 1/100 by adding 90 µl to 10 µl of the bisulphide conversion product and 2 µl was added to the master mix in each well. PCR grade H_2O was added into a well as a negative control. The PCR was carried at the conditions shown in Table 2.5. Once complete 10 µl of PCR product was run on an agarose gel to ensure the PCR was successful.

2.5.8.3 Pyrosequencing Reaction

The workstation was set up according to the manufacturers' instructions with the wash buffer, 70% ethanol, water and denaturation buffer (QIAGEN, 2012). The details of the sample and the reaction were selected on the PyroMark CpG Software. Into each reaction well 1.75 µl of streptavidin sepharose bead suspension and PyroMark Binding Buffer (38.25 µl) were added. This was then placed onto a shaking hot plate at 10 x g at 22°C until required. The sequencing primer was diluted by adding 1.5 µl of 10 µM to 43.5 µl of PyroMark Annealing Buffer for each reaction. 45 µl of sequencing primer was placed the PSQ plate and the plate was securely placed on the correct position on the workstation.

The DNA and sepharose bead mix was taken off the hot plate and placed into the position on the workstation. The vacuum was dried and the vacuum prep tool captured the beads with the PCR product. The vacuum prep tool was then placed into the ethanol for 5 seconds, the denaturation solution for 5 seconds and the wash buffer for 5 seconds. All the fluid was aspirated and the vacuum switched off and the prep tool was lowered into the PSQ plate. The PSQ plate was then heated at 80°C for two minutes

and allowed to cool to room temperature. The PyroMark Gold Q96 reagents were prepared as specified in the kit and reagents were then placed into the correct location in the PSQ96 Reagent Cartridge. The PyroMark CpG Software calculated the volume of reagents required. The PSQ plate and cartridge were both placed into the PyroMark Q96 ID instrument and the run began.

2.5.8.4 Methylation Analysis

All samples included in this analysis were repeated in duplicate within a run to improve the reliability of the data. Analysis was performed using PyroMark CpG Software. Samples were categorised as either pass 'blue', check 'yellow' or fail 'red'. Samples which failed were repeated to ensure this was not a false-negative result and were excluded from the study after a second failed run. Samples which were categorised as 'check' were manually checked to assess the acceptability of the data, some of these had failed results at CpG sites CpG 3461 and CpG 3471, thus the first 6 CpG sites were included in the analysis and sites CpG 3461 and CpG 3471 were excluded.

2.6 **Growth Medium and Optimal Conditions for Cell Growth**

Information on cell lines used in this study is shown in Table 2.1 in section 2.2. All cells used in this project were incubated during growth at $37^{o}C$ with 5% CO_2. Reagents used for the growth of cells are listed in Table 2.8 and the media components are listed in Table 2.9.

Table 2.8. Reagents used in the growth of cells and their suppliers

Reagents	Supplied by
Hyclone Defined Foetal Bovine Serum	Fisher
Hydrochloric Acid	
Sodium Hydroxide	
H's F12	Life Technologies
Adenine	Sigma Aldrich
Cholera Toxin	
Dimethyl Sulfoxide (DMSO)	
Dulbecco's Modified Eagle's Medium (DMEM)	
Epidermal Growth Factor (EGF)	
EpiLife® Medium, with 60 µM calcium	
Foetal Bovine Serum (FBS)	
Glasgow's Modified Eagle's Medium (GMEM)	
Hydrocortisone	
Insulin	
L Glutamine	
Phosphate Buffered Saline (PBS)	
Penicillin/Streptomycin	
Supplement S7	
Transferrin	
Tri iodo-L-thyronine T3	

Table 2.9. Components of media used to culture cells

	E Media	HEK media	3T3 Media
Main media	DMEM 300 ml	HEK media	DMEM 450ml
Additional media	H's F12 160 ml		
Serum	Hyclone defined FBS (25 ml)		Foetal Calf serum 50 ml
Pen/Step	10 ml		5 ml
Hydrocortisone	0.5 ml		
Cholera toxin	0.5 ml		
L Glutamine	10 ml		
EGF	2.5ml per 500 ml		
Additional	100x cocktail 5 ml	HEK Supplement	

2.6.1 Vulval and Vaginal Cell Lines

The vulva and vaginal cell lines were cultured using E media plus 2 x 10^6 J2 3T3 and passaged when they reached ~80% confluence. The components for this E media are listed in Table 2.10 and 2.11. All

components were filter sterilized and EGF was only added for growing the cells in a monolayer.

Table 2.10. Components for 200 ml of E Media Cocktail

Component	Concentration	Volume
Adenine	0.18 M	20 ml
Insulin	5 mg/ml	20 ml
Transferrin	5 mg/ml	20 ml
T3	2×10^8 M	20 ml
PBS		120 ml

2.6.2 J2 3T3 Cells

J2 3T3 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) media made up as described in Table 2.8 and passaged every 3 days. After passage cells that were not re-plated were irradiated for use as feeder cells, losing the ability to replicate but still producing extracellular

matrix protein. Cells were irradiated for 22 minutes, absorbing 60 gray (Gy) radiation in a Caesium-137 irradiator.

2.6.3 Freezing and Thawing of Cell Lines

Cells were frozen down to preserve a stock of cells in case of infection and to preserve earlier passages of cell stocks in case of changes in composition through further passage. Cells were trypsanised and resuspended in 10 ml of normal growth media and centrifuged for 5 minutes at 160 x g and the supernatant removed. Cells were resuspended in (foetal bovine serum) FBS supplemented with 10% dimethyl sulfoxide (DMSO) to the desired concentration (from 5×10^5 cells/ml to 2×10^6 cells/ml). The resuspended cells were transferred to appropriately labelled cryovials in 1 ml aliquots. Vials were then placed in a Mr Frosty Freezing Container overnight, and placed in the -80^oC freezer. After 24 hours, cells were stored in liquid nitrogen and their location recorded for future use.

To thaw cell lines, cryovials were removed from the liquid nitrogen and defrosted. The cells were pipetted to ensure they were in single suspension and placed in the centre of a petri dish. If required, 1×10^6 J2 3T3 cells were added to the top of the defrosted cell suspension. Finally,

normal growth media was added. Cells were then incubated at $37^{\circ}C$ with 5% CO_2.

2.7 **Keratinocyte Differentiation Using Organotypic Raft Culture**

Organotypic raft culture (as described in section 1.7.5) was carried out to assess the differentiation level of cells and relative gene expression, methylation and integration status. In brief this involves seeding cells on a collagen matrix, with subsequent growth to confluence. Once cells reach confluence the collagen matrix is then lifted onto a metal grid above growth media (so the base of the collagen is touching the top of the media). The media is then replaced ever two days and the collagen matrix and cells are left untouched. After 13 days, the collagen matrixes along with the cells were lifted from the metal grid and were either placed into a cellular preservative or cells were removed from the collagen matrix and placed into saline for DNA/RNA extraction. Reagents used in organotypic raft culture are shown in Table 2.11.

Table 2.11. Reagents used in the organotypic raft culture and their suppliers.

Reagents	Supplied by
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Life Technologies
DMEM without sodium bicarbonate	Sigma Aldrich
Acetic Acid	
Sodium hydroxide	
Rat Tail Collagen Type 1	VWR

2.7.1 Production of Collagen Plugs

Reconstitution buffer (10x) was prepared by adding 2.2 g of sodium bicarbonate to 4.77 g of 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES). This was the dissolved in 100 ml of 0.05 M Sodium hydroxide. This was then filter sterilized, as was the DMEM and stored at - 2^oC until required. Once the plugs were ready to be produced J2 3T3 cells were harvested at 80% confluence and 1-2 x 10^6 cells were pelleted for each collagen plug. 0.3 ml of 10x DMEM and 0.3 ml of 10x reconstitution buffer was added to the cell pellet for each collagen gel. 2.4 ml of rat-tail

collagen at 4 mg/ml per collagen gel was then added using pipette tips that had previously been stored at -20°C and solution mixed by pipetting up and down. Sodium hydroxide was then added using a Pasteur pipette until the solution reached the correct pH range (pH 6 – 8).

Once the solution was prepared, 3 ml was pipetted slowly onto a 3 cm petri dish again using chilled pipette tips. The plugs were then stored in 10 cm petri dishes to ensure they were stable and transferred to an incubator at 37°C for 30 minutes. After 30 minutes 2 ml of complete E media was placed on top of the collagen plug. Collagen plugs must remain in the incubator for 1 day after it preparation and can be kept for up to 4 days before being used to grow the differentiated cells on. For optimum results the collagen gels were used within two days' storage.

2.7.2 Growing Differentiated Cells on Raft

Firstly, the stainless-steel metal grids were treated with chromic/sulphuric acid for an hour and rinsed overnight in tap water. After 24 hours they were rinsed with double distilled water for a further 3 – 5 hours and then autoclaved. Metal grids were then placed in a 10 cm petri dish using sterile forceps (autoclaved). Media was removed from the top of

the collagen gel and sides of the gel loosened from the edges of the petri dish using an autoclaved, sterile spatula. Carefully placing a spatula under the collagen gel, it was lifted away from the 3 cm petri dish. The collagen gel was then laid on to the metal grid ensuring no air bubbles formed. The liquid interface was then created by adding E media (without EGF) to the plate until it touched the bottom of the metal grid but did not touch the collagen, again this had to be done slowly to ensure no bubbles formed at the interface, this could lead to non-uniform growth. Rafts were then incubated at 37°C and the media changed every two days. On the thirteenth day the cells could be harvested.

2.7.3 Analysis of Organotypic Raft Culture

IHC results were quantified by grading the positive pixels in the selected raft cultures as a marker of the amount of stain the cells have taken in. Figure 2.3 shows the process by which the rafts were graded. The total number of pixels were counted and scored as either high-positive, positive, low positive or negative. IHC results were also quantified by measuring the depth of the culture that has taken in the IHC stains using ImageJ (Fiji) software (Jensen 2013; Varghese et al. 2014). The total depth of the culture was measured, and then the length from the base of the culture to the top of the stained area was measured and the

percentage was calculated. Each culture was measured three times and the average percentage of depth of IHC stain recorded. SHH and GLI-1 protein expression was also quantified using ImageJ. Each image had the colour adjusted to display the fluorescent stain as red and the proportion of the epithelium that stained red was measured. Results for SHH and GLI-1 expression were compared to the 4', 6'-diamidino-2-phenylindole (DAPI) nuclear control stain to produce a relative percentage stain compared to the control.

Ki-67

A

B

C

Pixel Count: 1793252
Percentage contibution of High Positive: 1.249
Percentage contibution of Positive: 15.2531
Percentage contibution of Low Positive: 36.7698
Percentage contibution of Negative: 46.7281
The score is Low Positive

p16

D

E

F

Pixel Count: 1652262
Percentage contibution of High Positive: 20.3985
Percentage contibution of Positive: 70.1483
Percentage contibution of Low Positive: 5.4881
Percentage contibution of Negative: 3.965
The score is Positive

Figure 2.6 IHC quantification using ImageJ. Images were loaded into the programme (A & D) and sections of the image containing the differentiated culture raft were selected (B & E) The total number of pixels were counted and scored as either high-positive, positive, low positive or negative (C & F). Each section was scored three times to ensure accurate results.

97

2.7.4 HPV IHC

IHC raft culture analysis was kindly performed by Dr Sally Roberts (HPV E4) and by Dr Christopher Dawson (SHH and GLI-1) at the University of Birmingham, table 2.11 shows the antibodies used in this IHC. p16 and Ki-67 IHC were performed by Contract Research Organisation Propath Laboratories (Hereford, UK).

Table 2.12 Details of Antibodies used in IHC including concentration used and details or supplier/reference

Protein	Antibody	Concentration	Supplier/Reference
SHH	ab53715	1:200	abcam
GLI1	sc20687	1:50	Santa Cruz
HPV E4 (primary)	1D11 mouse monoclonal	1:10	Roberts et al., J. Virol. January 2003 vol. 77 no. 1 673-684
HPV E4 (primary)	Rabbit anti-HPV18 E4 polyclonal r424	1:1000	Wilson et al., 2007 Virology 362:453
HPV E4 (secondary)	Alexa 488	1:500	Molecular Probes
HPV E4 (secondary)	Alexa 594	1:500	Molecular Probes

2.7.4.1 HPV 16 E4 IHC

In brief the raft sections were deparaffinised in Histoclear and low temperature antigen retrieval performed in 1 mM EDTA (pH 8.0), Tween 20 (0.1%) on a hot hotplate-stirrer at 65°C overnight (Watson et al., Carcinogenesis 2002 23:1791). Slides were blocked in a humidified chamber at room temperature in PBS containing 20% heat inactivated normal goat serum and 0.1% (w/v) bovine serum albumin (fraction V) for 90 min. The blocking buffer was removed and replaced with primary antibody solution prepared in blocking buffer and incubated at 4°C overnight. Primary antibodies used were 1D11, a mouse monoclonal E4 antibody (1:10 dilution) (Roberts et al., 2003) and a rabbit anti-HPV18 E4 polyclonal antibody r424 (1 in 1000 dilution) (Wilson et al., 2007). The slides were washed in PBS twice for 15 minutes and secondary antibodies applied (mouse and rabbit IgG – specific Alexa 488 and 594 antibodies (Molecular Probes; each diluted 1 in 500). Nuclei were counterstained with DAPI (4', 6'-diamidino-2-phenylindole), prior to mounting in ProFade (Molecular Probes Inc).

2.8 Statistical Analysis

Guidance on statistical analysis was kindly provided by Professor Robert Hills. Statistical analysis reported in this thesis was carried out using Microsoft Excel (Microsoft, Washington, USA) and IBM SPSS Statistics (IBM Corporation, New York, USA).

2.8.1 Chi-square Test of Independence

In this thesis categorical data was assessed using the Chi-square Test for Independence. This test assesses the association between two independent variables and demonstrates how likely the distribution of the data is due to chance. This is done by comparing the observed values 'O' to the expected 'E' (Laerd Statistics, 2016).

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

The significance of this Chi-square value was then determined by comparing the degrees of freedom in the data to the Chi-square value using a Chi-square distribution table, a result is considered statistically

significant if p<0.5. Chi-square test can also use Cramer's V as a measurement of the strength of the association between the two variables. Guidelines for interpreting the Cramer's V were established by Cohen (1988) (Table 2.13)

Table 2.13. The strength of association guidelines for Cramer's V as established by Cohen (1988).

Level of effect size	Cramer's V Value
Small	0.1
Moderate	0.3
Large	0.5

2.8.2 Spearmans Correlation coefficient

The Spearmans Correlation coefficient can be employed to measure both the strength of a correlation as well as the direction of the correlation. This test was used over the Pearsons Correlation Coefficient as Spearmans Correlation coefficient is less sensitive to data skewing by outliers (Mukaka, 2012). As sample sizes in this study were small this allowed for a robust statistical analysis without removing samples from the analysis.

In this equation, the total sample size = n and the two groups to be tested are separated into two groups, x and y. and S_x and S_y are the standard deviation of x and y respectively:

$$r = \frac{1}{n-1}\left(\frac{\sum_x \sum_y (x-\bar{x})(y-\bar{y})}{s_x s_y} \right)$$

The resulting value (r) is between -1 and +1, where a r = +1 is a perfect positive association of ranks, r = -1 is a perfect negative association of ranks and r = 0 is no correlation. Unlike the Pearsons correlation there is no set guidelines on how to determine the strength of the relationship. In this thesis the results will be interpreted using the following criteria (Table 2.14) (Hinkle et al., 2003):

Table 2.14. Interpretation of Spearman Correlation Coefficient reported in this thesis. Table adapted from Hinkle et al,.(2003).

Size of correlation	Interpretation
0.9 to 1.0 (-0.9 to -1.0)	Very high positive (negative) correlation
0.7 to 0.9 (-0.7 to -0.9)	High positive (negative) correlation
0.5 to 0.7 (-0.5 to -0.7)	Moderate positive (negative) correlation
0.3 to 0.5 (-0.3 to -0.5)	Low positive (negative) correlation
0.0 to 0.3 (-0.0 to -0.3)	Negligible correlation

2.8.3 Kruskal-Wallis H Test

The Kruskal-Wallis H test is a nonparametric test like a one-way ANOVA, however it can be used when the data fails to meet the assumptions of a one-way ANOVA (e.g. non-normally distributed, ordinal data). The Kruskal-Wallis H test was used in this study when the data was not normally distributed to determine if there are statistically significant differences between two or more groups of independent variables (Kruskal and Wallis, 1952, Laerd Statistics, 2015).

Firstly, the distributions of the data were compared for all groups by visual inspection of a box plot. If distributions were similarly shaped,

differences in the medians could be investigated. If the distribution of the independent variables were not similar then no inference about the median between groups could be made. In this case, the differences in distribution and mean ranks can still be assessed. The Kruskal-Wallis H test is reported as the value of the chi-squared distribution (X^2) with the degrees of freedom (k-1, where k is the number of groups).

In addition to the Kruskal-Wallis H test a pairwise comparison was carried out between individual groups to discover which groups were statistically different from each other. The pairwise comparison was interpreted using Dunn's (1964) procedure with a Bonferroi adjustment. In the pairwise comparison when you make multiple comparisons there is an increase in type 1 error, where a result is incorrectly reported as statistically significant, this error increases with each pairwise comparison carried out. The Bonferroni correction is used to compensate for this error, and the significance of the pairwise comparisons (p) will be reported as the "adjusted-significance".

Chapter 3

HPV genome status in first routine smear sample from women aged 20-22 years

Integration of episomal HPV DNA in to the host genome is not part of the normal virus replication cycle, and frequently results in loss of *E2* gene (McBride, 2013). *E2* plays a pivotal role in regulating HPV transcription, including limiting expression of the HPV oncogenes *E6* and *E7*. The disruption of the regulatory *E2* gene results in enhanced expression of *E6* and *E7*. Although integration is now known not to be an absolute pre-requisite for the progression to cervical cancer, HPV *E2* disruption could be used as a potential biomarker for predicting HPV malignant disease progression. *E2* disruption can be most readily assessed using the *E2* tiling PCR assay (discussed in sections 1.6.4).

DNA methylation is generally indicative of transcriptionally repressed chromatin. Studies performed in the HPV laboratory, Cardiff University and elsewhere have correlated the methylation state of the HPV genome with chromosomal integration (Bryant et al., 2014, Marongiu et al., 2014). Thus work was carried out to assess if HPV *E2* viral DNA methylation could be used as a potential biomarker for the prognosis of HPV related disease progression.

3.1 **DNA quality in clinical samples stored in SurePath^{TM} LBC media**

Reliable and reproducible purification of high quality nucleic acid is paramount in ensuring optimal results in HPV molecular assays. When working with clinical material it is therefore essential that the extraction and purification procedures be optimised. SurePath^{TM} LBC media is an ethanol based media containing formalin that preserves the morphology of the cellular and subcellular components within the cytoplasm and nucleus. This improves the sample quality for cytological examination; however, DNA damage and cross-linking can result following exposure to chemical agents, especially formalin. Formalin fixes cellular material for storage and subsequent histological analysis; however, it can cross-link DNA, making subsequent molecular analysis difficult. To overcome the potential degradation from storage in SurePath^{TM} five different extraction methods were assessed to determine the most effective method to extract DNA from clinical samples stored in Surepath^{TM} LBC media, and the merits of each is summarised in Table 3.1. DNA yields and purity was assessed by measuring absorbance at 260 and 280 nm (described in material and methods section 2.3) and semi-quantitative analysis of PCR product for the *BetaGlobin* gene.

Consistent with previous published findings from the laboratory, Protocol 1 resulted in relatively low yields of PCR products 0-2.750 µg DNA; (Hibbitts et al. 2014). Although Protocol 2 improved yields to 5 – 7.5 µg DNA, the amplification of target DNA by PCR was still not optimal (Figure 3.1). A Kruskal-Wallis H test was

3.1 **DNA quality in clinical samples stored in SurePath™ LBC media**

Reliable and reproducible purification of high quality nucleic acid is paramount in ensuring optimal results in HPV molecular assays. When working with clinical material it is therefore essential that the extraction and purification procedures be optimised. SurePath™ LBC media is an ethanol based media containing formalin that preserves the morphology of the cellular and subcellular components within the cytoplasm and nucleus. This improves the sample quality for cytological examination; however, DNA damage and cross-linking can result following exposure to chemical agents, especially formalin. Formalin fixes cellular material for storage and subsequent histological analysis; however, it can cross-link DNA, making subsequent molecular analysis difficult. To overcome the potential degradation from storage in SurePath™ five different extraction methods were assessed to determine the most effective method to extract DNA from clinical samples stored in Surepath™ LBC media, and the merits of each is summarised in Table 3.1. DNA yields and purity was assessed by measuring absorbance at 260 and 280 nm (described in material and methods section 2.3) and semi-quantitative analysis of PCR product for the *BetaGlobin* gene.

Consistent with previous published findings from the laboratory, Protocol 1 resulted in relatively low yields of PCR products 0-2.750 µg DNA; (Hibbitts et al. 2014). Although Protocol 2 improved yields to 5 – 7.5 µg DNA, the amplification of target DNA by PCR was still not optimal (Figure 3.1). A Kruskal-Wallis H test was

carried out to assess if there were differences in the DNA concentrations achieved for the two extraction methods. The mean ranks of the DNA concentration were statistically different between the two protocols $X^2(1)$ = 7.410, p = 0.006. There was no significant difference between protocol 1 and 2 when compared to the 260/280 ratio, $X^2(1)$ = 0.315, p = 0.575.

Three further extraction protocols were then assessed alongside Protocol 2 (Table 3.1). Protocol 2 obtained the highest-level DNA yield (28619.9 ng) in this cohort (Figure 3.2). A Kruskal-Wallis H test was carried out to assess if there were differences in the DNA concentrations achieved for the four extraction methods. The distributions of the DNA concentration were not similar for all groups, as assessed by visual inspection of the boxplot. The mean ranks of the DNA concentration were statistically different between the four groups $X^2(3)$ = 19.15, p > 0.001. A pairwise comparison revealed statistically significant differences in DNA concentration between Protocol 2 (mean rank = 28.00), Protocol 4 (mean rank = 6.00) (p < 0.001) and Protocol 5 (mean rank = 13.29) (p = 0.046). The mean ranks of DNA purity (as measured by 260/280) were statistically different between the four extraction methods tested, $X^2(3)$ = 18.916, p < 0.001. When a pairwise comparison was carried out the only significant difference in DNA purity was between Protocol 2 (mean rank = 5.00) and Protocol 5 (mean rank = 23.86), (p < 0.001). While product quality was improved by Protocol 3 (mean rank = 13.00), compared to Protocol 2 (mean rank = 5.00) this was not statistically significant (p = 0.413).

The enhanced Proteinase K extraction (protocol 2) was selected as the preferred method for this study as it was compatible with PCR amplification and significantly higher DNA yield compared to Protocols 1, 4 and 5. Protocol 2 resulted in the extraction of approximately 28691 ng of DNA from each sample, whereas Protocols 3, 4 and 5 yielded 2440 ng, 888 ng and 1779 ng DNA respectively. Although the samples extracted with Protocol 2 had reduced purity based upon the 260/280 ratio and the purity was only significantly reduced compared to Protocol 5 (which produced the lowest DNA yield). Subsequently the *BetaGlobin* PCR was applied to all samples extracted with a 95.3% success rate (n = 523/549).

Protocol 1

10 μl of proteinase K per 100 μl of clinical sample

Two-hour incubation period at 56°C

Protocol 2

20 μl of proteinase K per 100 μl of clinical material

Overnight incubation at 56°C

Band Number	Sample ID	Protocol 1		Protocol 2	
		DNA yield (ng)	260/280	DNA yield (ng)	260/280
1	3047	10915	0.67	33170	0.73
2	3416	13425	0.66	25407	0.72
3	4140	19327	0.75	22675	0.59
4	11659	12872	0.61	52495	0.85
5	13304	10647	0.60	17740	0.60
6	13420	13437	0.65	30297	0.69

Figure 3.1 Samples extracted by Protocol 1 and Protocol 2: Yield and purity of DNA and *BetaGlobin* PCR comparison (A) Comparison of protocol 1 and 2. **(B)** PCR product (110 bp) is highlighted with the yellow arrow and CaSki (positive) and water (negative) controls are labelled above. In samples extracted using protocol 1 *BetaGlobin* gene was detected in half of the samples. In samples extracted using protocol 2 all the clinical samples produced a weak PCR product. **(C)** DNA yield and purity results of the samples extracted using protocol 1 and 2 as assessed using Thermo Scientific Nanodrop® 1000 Spectrophotometer.

Table 3.1. Summary of five DNA extraction protocols evaluated in this study for optimum extraction of clinical samples stored in SurePath™ LBC

Protocol	Extraction Method	Overview	Procedure complexity	Time	Cost per 250ul sample (£)[1]	References
1	Preliminary Proteinase K	Proteinase K enzyme breaks down glycoproteins and releases nucleic acid wit limited denaturation. 10 µl of proteinase K per 100 µl of clinical material and a two hour incubation period at 56°C.	One step process. DNA is removed from the cellular debris at the end of the extraction.	2.5 - 3.0 hours	0.76	Roche 2015
2	Enhanced proteinase K	Proteinase K enzyme breaks down glycoproteins and releases nucleic acid wit limited denaturation. 20 µl of proteinase K per 100 µl of clinical material and a overnight incubation at 56°C.	One step process. DNA is removed from the cellular debris at the end of the extraction.	16+ hours	1.52	Roche 2015
3	Enhanced proteinase K +GET™ Clean DNA	Proteinase K enzyme breaks down glycoproteins and releases nucleic acid wit limited denaturation. 20 µl of proteinase K per 100 µl of clinical material and a overnight incubation at 56°C. The GET™ Clean DNA kit removes any left over salt, proteins, enzymes and short nucleic acid sequences (<100 base pairs).	Two step process. DNA is removed from the cellular debris and then is added to the spin column where the DNA is bound to the column and the impurities are washed away before the DNA is eluted.	17+ hours	3.94	Roche 2015 and G-Bioscience 2015
4	QIAamp DNA Blood Mini Kit	DNA is bound to a silica-based membrane while contaminants are washed through the membrane and discarded. The pure DNA is then detached from the membrane and eluted.	Four step process: cell lysing, binding to the column, washing and eluting the sample	1 - 2 hours	2.72	QIAGEN 2015
5	EZ1 Extraction	Fully automated protocol using silica coated magnetic beads. DNA is bound to the silica surface of the beads and then removed by magnets from the lysate, washed and eluted.	Fully automated process.	20 - 30 minutes	5.97	QIAGEN 2015
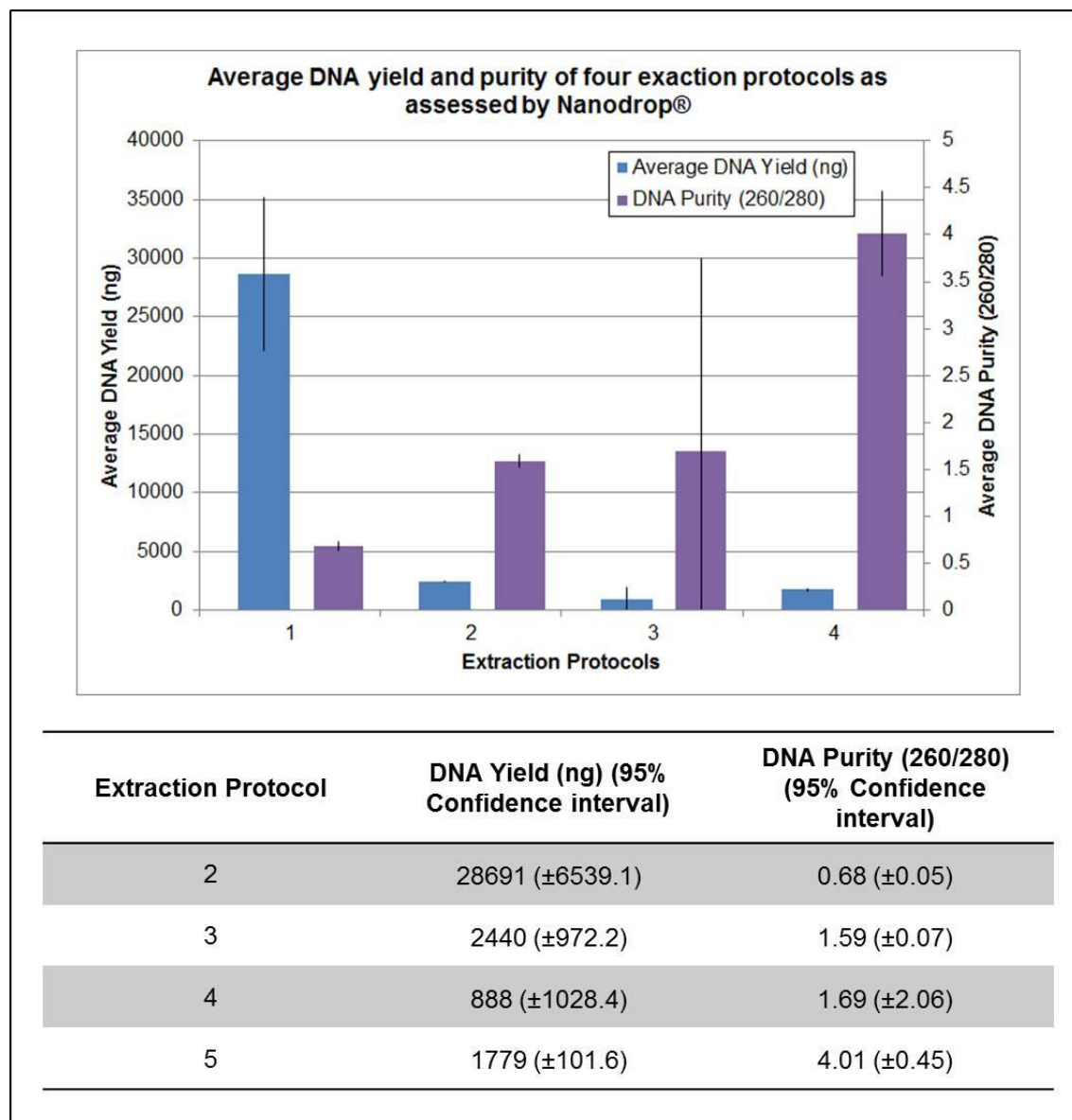
Figure 3.2. Average DNA yield and purity for four extraction procedures. Comparison of DNA yield (ng) and sample purity (as measured by 260/280 ratio) for samples using four extraction procedures to assess which procedure produces the highest quality and quantity of DNA for subsequent analysis. Analysis was carried out on 6 clinical samples and extraction and extraction and analysis was performed in triplicate. DNA yield and purity (260/280 ratio) as measured by Thermo Scientific Nanodrop® 1000 Spectrophotometer.

3.2   Identifying HPV *E2* gene in clinical samples

A HPV 16 *E2* tiling PCR assay is an established method to detect disruption of the *E2* by performing multiple PCRs across the HPV genome spanning *E2* (Collins et al., 2009). In this study, novel type-specific *E2* tiling PCR assays were developed for HPV 31, 33, 35 and 51. The assays were developed based on the consensus, type-specific, genome sequences extracted from the EMBL-EBI Database and primers were designed using NCBI Primer BLAST (Ye et al., 2012). A small PCR product was preferential (max 350 bp) due to the potential fragmented and cross-linked nature of the DNA stored in SurePath^{TM} LBC media. Suitable primer pairs were checked in Net Primer (PREMIER Biosoft International, Palo Alto, CA) for potential, unwanted, secondary structures. Primers were also compared to the genome of 12 of the most common HR-HPV types (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59) to ensure the primers were specific to only the HPV type of interest. Primer details can be found in Appendix 3.

In touchdown PCR, the initial annealing temperature is around 5-10°C higher than the calculated *T*m of the primers, it is then reduced every cycle until around 5°C below the calculated *T*m of the primers. Initially a touchdown PCR was used for the *E2* tiling PCR to reduce the time spent optimising assays as primer pairs with different optimal annealing temperatures could be run together (Hibbitts et al., 2014, Korbie and Mattick, 2008). The touchdown PCR resulted in faint bands with an average yield of 5.65 ng/µl DNA. The PCR bands remained faint even when the

cycle number was extended from 40 to 50 cycles. A standard PCR produced strong single bands with an average DNA concentration of 25 ng/µl of DNA, a greater than 4-fold increase from the touchdown PCR (Figure 3.3).

Assay sensitivity was established by means of a serial dilution from 50 ng to 5 x 10^{-7} ng of type specific plasmid DNA. All assays were sensitive to 0.05ng of plasmid DNA with some primer sets being sensitive to 5 x 10^{-3} ng of DNA (Table 3.2). *E2* tiling PCRs were successfully applied to 89.55% (n = 429) clinical samples that had been extracted using protocol 2 (Figure 3.4). Assays were considered successful based upon positive *E6*/*E7* PCR. These assays enable a quick and low cost assessment of the presence of the *E2* gene.
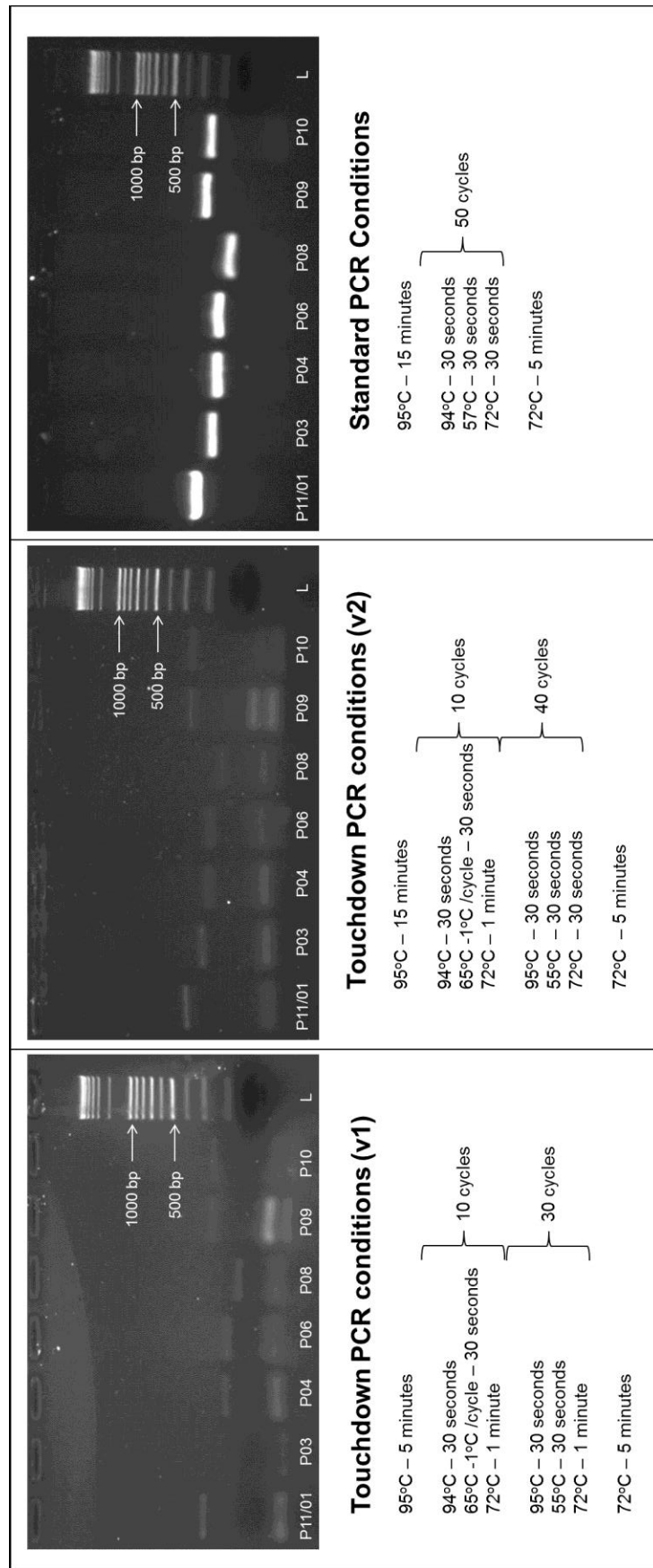
Figure 3.3. Optimisation of the *E2* tiling PCR assay PCR condition. Optimisation was carried out using plasmid control DNA (n = 3) and all three assays used HPV 51 *E2* PCR primers. Initially touchdown PCR *v1* and *v2* both produced faint bands when the PCR product was run on 2% agarose gel. Subsequent standard PCR conditions produced a large level of amplicons, therefore this was the methods employed for subsequent *E2/E1* PCR assays on clinical samples.

114

Table 3.2. Sensitivity of type specific *E2* tiling PCR assays. Sensitivity is reported the range of DNA concentrations (ng) that produced visible PCR product when run on a 2% agarose gel.

HPV 31		HPV 33		HPV 35		HPV 51	
Primers	Sensitivity (ng)	Primers	Sensitivity (ng)	Primers	Sensitivity (ng)	Primers	Sensitivity (ng)
31E2_F_11/31E2_R_01	$50 - 0.05$	33E2_F_02/3 3E2_R_02	$50 - 0.05$	35E2_F_01/3 5E2_R_01	$50 - 0.05$	51E2_F_01/5 1E2_R_06	$50 - 5 \times 10^{-3}$
31E2_F_03/3 1E2_R_03	$50 - 5 \times 10^{-3}$	33E2_F_03/3 3E2_R_03	$50 - 0.05$	35E2_F_02/3 5E2_R_02	$50 - 5 \times 10^{-3}$	51E2_F_03/5 1E2_R_03	$50 - 5 \times 10^{-3}$
31E2_F_04/3 1E2_R_04	$50 - 0.05$	33E2_F_04/3 3E2_R_04	$50 - 5 \times 10^{-3}$	35E2_F_03/3 5E2_R_03	$50 - 5 \times 10^{-3}$	51E2_F_04/5 1E2_R_04	$50 - 5 \times 10^{-3}$
31E2_F_06/3 1E2_R_06	$50 - 0.05$	33E2_F_05/3 3E2_R_05	$50 - 0.05$	35E2_F_04/3 5E2_R_04	$50 - 0.05$	51E2_F_05/5 1E2_R_05	$50 - 5 \times 10^{-3}$
31E2_F_08/3 1E2_R_08	$50 - 5 \times 10^{-3}$	33E2_F_06/3 3E2_R_06	$50 - 0.05$	35E2_F_05/3 5E2_R_10	$50 - 5 \times 10^{-3}$	51E2_F_07/5 1E2_R_07	$50 - 5 \times 10^{-4}$
31E2_F_09/3 1E2_R_09	$50 - 5 \times 10^{-3}$	33E2_F_09/3 3E2_R_07	$50 - 0.05$	35E2_F_11/35 E2_R_06	$50 - 5 \times 10^{-3}$	51E2_F_08/5 1E2_R_08	$50 - 0.05$
31E2_F_10/3 1E2_R_10	$50 - 0.05$	33E2_F_08/3 3E2_R_08	$50 - 5 \times 10^{-3}$	35E2_F_07/3 5E2_R_07	$50 - 5 \times 10^{-3}$	51E2_F_09/5 1E2_R_09	$50 - 0.05$
31E6_F_02/3 1E6_R_02	$50 - 5 \times 10^{-3}$	33E6_F_02/3 3E6_R_02	$50 - 5 \times 10^{-3}$	35E6_F_02/3 5E6_R_02	$50 - 5 \times 10^{-3}$	51E6_F_03/5 1E6_R_04	$50 - 5 \times 10^{-3}$

115
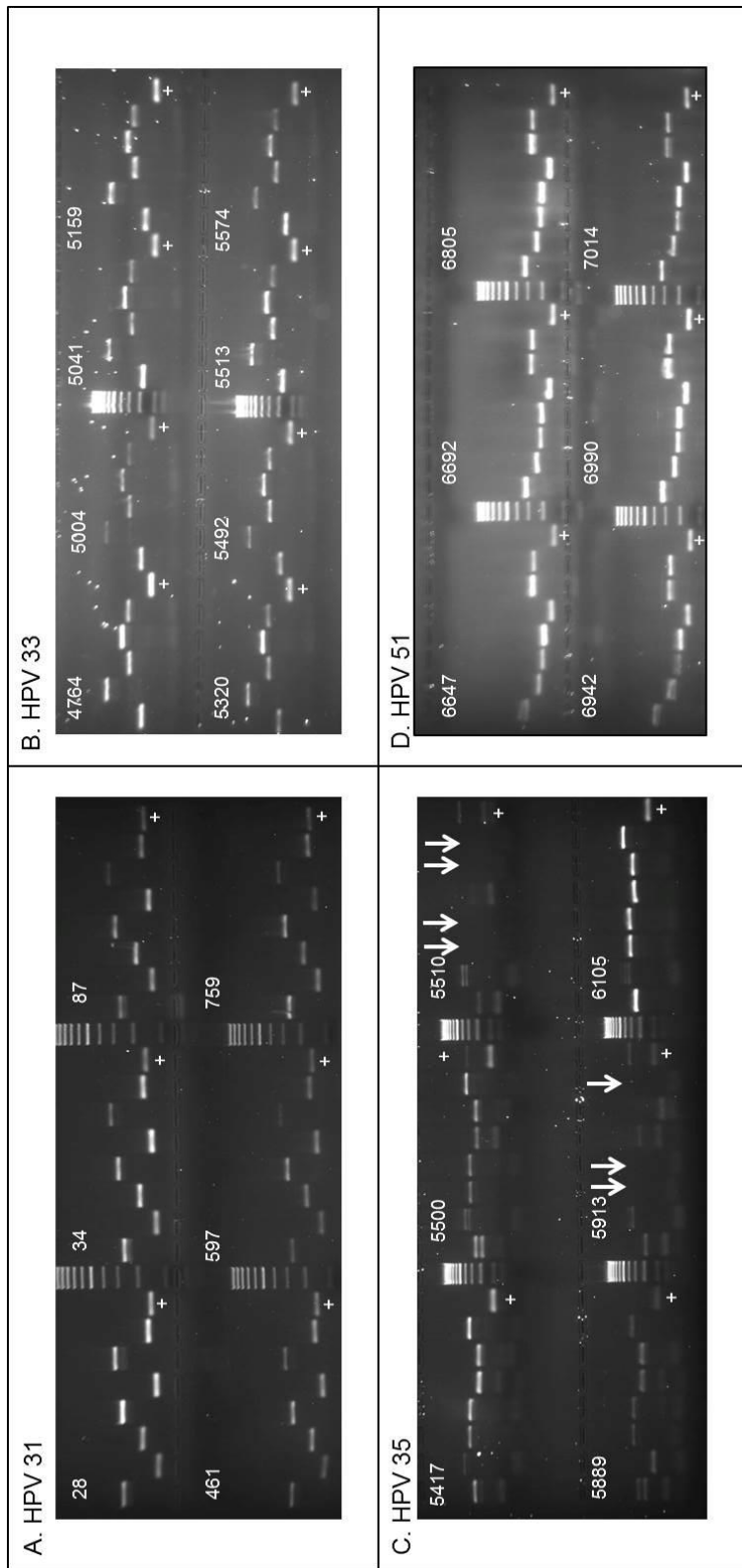
Figure 3.4 Examples of HPV _E2_ tiling PCR applied to a selection of clinical samples. Sample IDs are labelled above the band. The first 7 DNA bands represent the _E2_ tiling PCR primers). _E6_ is present in every sample (indicated with a '+'). Disruption is shown in samples 5510 and 5913 that are missing 4 and 3 amplicons respectively, primers that failed to amplify DNA are indicated with white arrows. _E2_ PCR was carried out three times on each clinical sample to ensure accurate reporting.

3.3 Disruption of HPV 16 *E2* in cervical smear samples

Starting with 13,306 samples in women aged 20 – 22 years old attending their first cervical smear, 4,870 were PCR-positive for HPV, of which 3,545 were attributed to HR-HPV types. This study aimed to answer hypothesis 2 by evaluating disruption of *E2* gene in a subset of less studied HPV types 31, 33, 35 and 51. Samples exhibiting single infections with HPV 31 (n = 122), 33 (n = 64), 35 (n = 17) or 51 (n = 139) were included in this analyses. This cohort included a higher prevalence of HPV 51 samples than expected, however analyses reported elsewhere has assessed that these are true HPV 51 infections (Bowden et al., 2013).

HPV 16 *E2* disruption has previously been assessed in this cohort but only assessing individuals with high-grade cytology results (Raybould, 2013). To produce a comparable assessment, a subset of HPV 16 samples was included in this analyses (n = 93) and were chosen randomly to mirror the prevalence of cytology grades shown in the total cohort (Figure 3.5). Samples that were either *E6*/*E7* negative by PCR were excluded from further analyses. *E6*/*E7* was chosen to evaluate sample eligibility as *E6*/*E7* are small genes which are present and expressed even in high-grade infections (Origoni et al., 2015).

	Negative cytology	Borderline cytology	Mild cytology	Moderate cytology	Severe cytology	Total
Total HPV 16 Cohort (%)	452 (62.4)	132 (18.2)	78 (10.8)	28 (3.9)	34 (4.7)	724
HPV 16 samples tested (%)	62 (66.6)	18 (19.4)	5 (5.4)	3 (3.2)	5 (5.4)	93
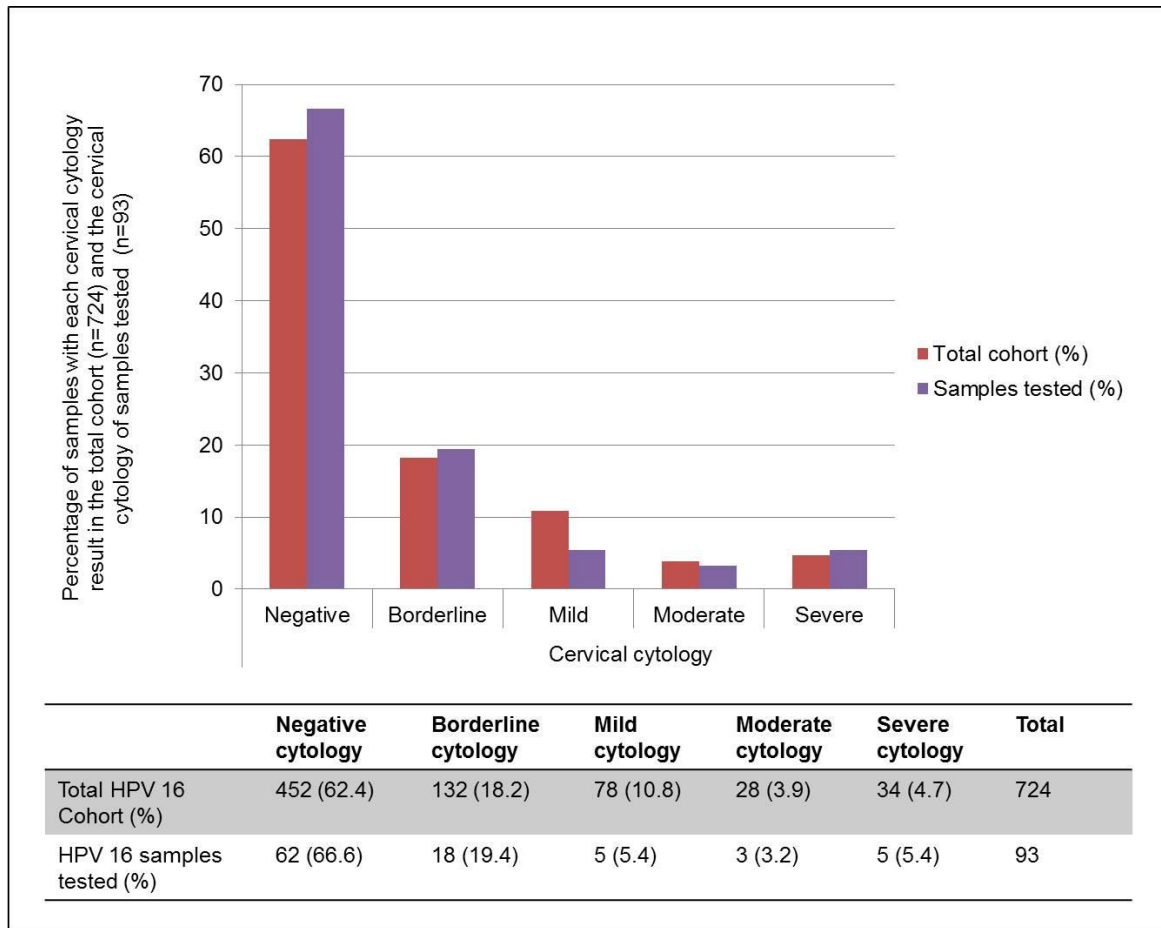
Figure 3.5. HPV 16 samples selection to reflect the proportion of cervical cytology in the total cohort. HPV 16 samples were separated by the cervical cytology and then samples were randomly selected to represent a similar proportion for each cervical cytology result (n = 93) compared to the total cohort (n = 724).

3.4 **Prevalence of *E2* disruption in a cohort of young unvaccinated women**

A systematic investigation of 412 clinical samples infected with a single HPV genotype identified *E2* gene was disrupted in 5.12% of samples. The levels of viral disruption ranged from 1.08% to 12.5% depending on HPV genotype (HPV 16 – 1.08% (n = 93), HPV31 – 8.26% (n = 121), HPV 33 – 6.25% (n = 64), HPV 35 – 12.5% (n = 16), HPV 51 – 5.12% (n = 116)) (Figure 3.6). There was no significant association between any specific HPV type and *E2* disruption (X^2 = 8.292, p = 0.081). All individuals with disrupted *E2* had negative or borderline smear results and subsequently none of these individuals attended colposcopy because of their first smear test.

The results suggested that *E2* was disrupted in a lower percentage of HPV 16 compared to the four high risk types not included in the vaccine (1.08% and 6.31% respectively) and when the results from the four non-vaccine types (HPV 31, 33, 35 and 51) were combined and compared with the results from HPV 16 the HPV 16 samples were significantly less likely to have disrupted *E2* gene in this cohort of young women (X^2 = 4.0533, p = 0.044). The hypothesis that the integrity of the *E2* locus (a biomarker of HPV genome integration) in HPV 16 infected samples would be equivalent to other HR-HPV types (HPV 31, 33, 35 and 51) was rejected.

	HPV 16 (n=93)	HPV 31 (n=121)	HPV 33 (n=64)	HPV 35 (n=16)	HPV 51 (n=116)	Total (n=410)
E2 Intact	92	111	60	14	112	389
E2 Intact (%)	98.92	91.74	93.75	87.50	96.55	94.88
E2 Disrupted	1	10	4	2	4	21
E2 Disrupted (%)	1.08	8.26	6.25	12.50	3.45	5.12
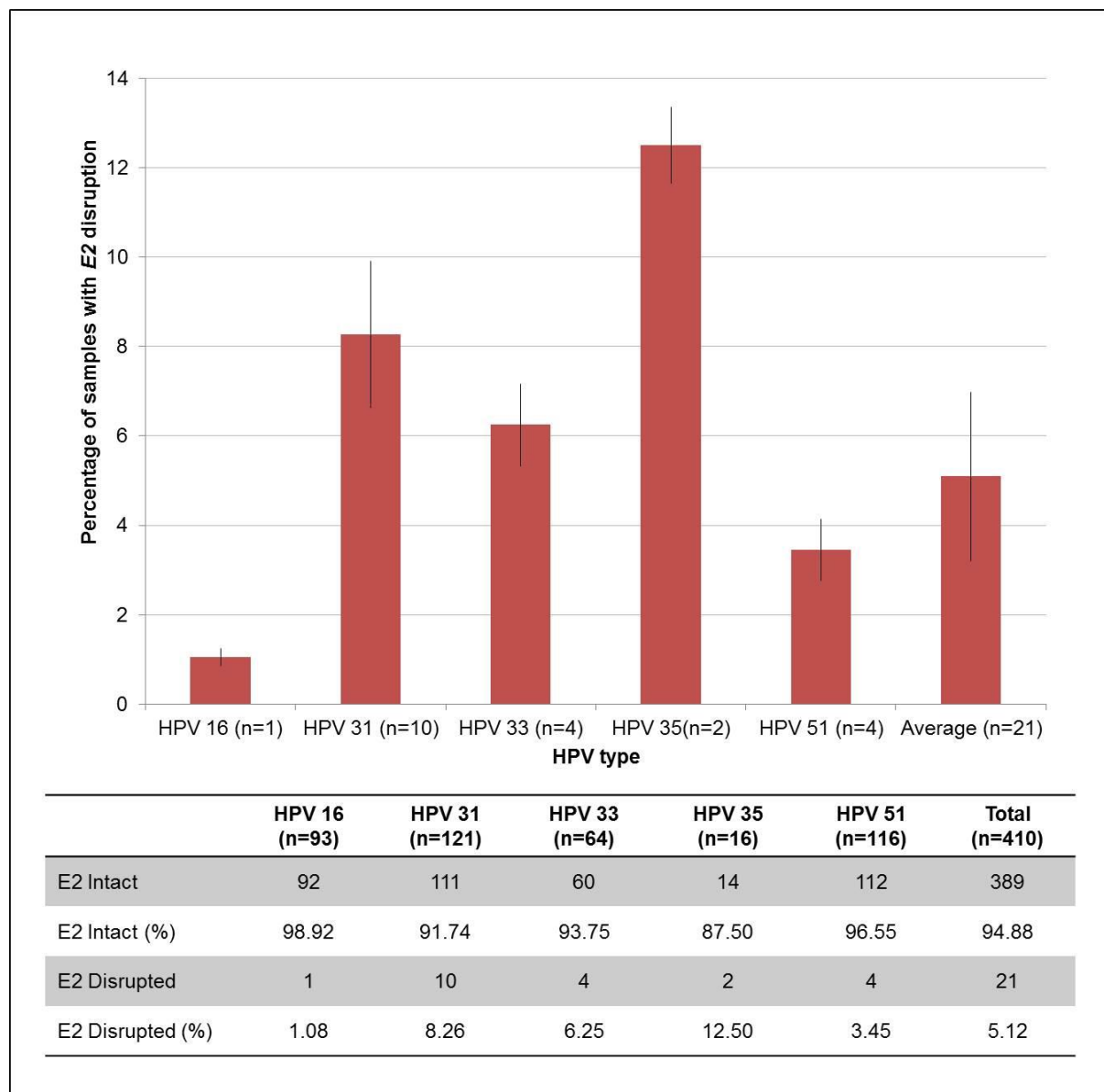
Figure. 3.6 *E2* gene disruption as detected by type specific *E2* tiling PCR for HPV 16, 31, 33, 35 and 51. (A) Table of the percentage of disruption for each HPV type **(B)** the percentage of samples with *E2* disruption (X^2 = 8.292, p = 0.081).

3.5 *E2* **status as a predictor of future HPV disease**

A further study was carried out to determining whether individuals infected with HPV with a disrupted *E2* were more likely to have developed CIN II+ in their first round of screening follow-up (≥ 36 months). Samples were included in this study if the individual had attended colposcopy in Wales from the date of their first smear until November 2013. Women who had not attended colposcopy and therefore had no reported histology, were excluded as it was not known if they had CIN II+. Samples were also excluded if they attended colposcopy and CIN was ungraded.

From this cohort, 89 individuals attended colposcopy (HPV 16 n = 22, HPV 31 n = 17, HPV 33 n = 20, HPV 51 n = 30). Only two samples in this cohort had *E2* disruption, one was HPV 31 positive and was diagnosed with CIN III and the other was HPV 16 positive and diagnosed with CIN I. The other 87 samples had an intact *E2* gene (no-CIN = 12, CIN I = 23, CIN II = 24, CIN III = 28). Chi-squared test revealed that there was no significant association between intact *E2* gene and follow up histology, $X^2(4) = 1.421$, p = 0.841.

The *E2* tiling PCR is not a suitable biomarker for HPV related disease progression in young women attending their first cervical smear. High-grade disease (CIN II+) does not correlate with *E2* disruption suggesting other viral/host factors are playing a role in disease progression. The *E2* gene was intact in 54 samples that

went on to have CIN II+ disease during the follow up period (≥ 36 months). Also 19 samples showed *E2 gene* disruption but to date have not attended colposcopy due to a suspicious smear result.

In this cohort of young women attending their first smear 5.12% had complete or partial *E2* disruption (n = 21/410). All samples with disrupted *E2* had either a borderline or negative cervical cytology, thus *E2* disruption, as measured by type specific tiling PCR alone, is not a suitable alternative for cervical cytology. This result may be due to the age of women in this cohort (20 – 22 years) as HPV related disease progression tends to result from persistent infection. (Molano, 2003, Bodily and Laimins, 2011)

3.6 HPV *E2* viral DNA methylation as a potential biomarker for HPV related disease

Methylation of cytosine located in CpG dinucleotides is known to have a major impact on gene expression (as described in Introduction 1.4.3). Methylation can have an impact on malignant cells associated with the progression of cancers. Viral genomic DNA methylation in cancer cells has been shown prominently in HPV 16-infected cells with an increase in viral genomic DNA methylation in the *E2*, *L1* and *L2* genes (Introduction section 1.3.2). Therefore HPV *E2* gene methylation was

investigated as a potential biomarker for disease presence in young women or an indicator of future disease progression.

HPV 16 samples (n = 93) were tested for *E2* methylation from those studied for *E2* disruption, however viral genomic DNA methylation level was not identified in 50% of samples. This result may have been due to the low copy number of HPV or due to sample disruption caused by cellular fixative. Samples that could not be sequenced were excluded from analyses. Therefore, further high-grade samples from the same cohort were sourced to increase the sample subset (Bryant *et al.* 2015). Samples in this study were chosen based on available cytology and histology results, leading to this cohort having a greater proportion of high grade disease compared to the entire cohort. The final cohort included 194 samples: No CIN = 21, CIN I = 32, CIN II = 52, CIN III = 88, adenocarcinoma = 1. All assays were carried out in triplicate and the average percentage of *E2* viral DNA methylation at each CpG site calculated for each sample.

When *E2* methylation was compared for samples with different histology results a trend can be seen with a decrease in viral DNA methylation from no-CIN to CIN III histology result, however the correlation between the average DNA methylation and histology was negligible, r_s = -0.203, p = 0.004 (Figure 3.7). Similarly, when the average DNA methylation at each CpG site is assessed a trend can be seen with decreased viral DNA methylation from no-CIN to CIN III, however the correlation between viral methylation and histology at all CpG sites was

negligible (Figure 3.8). From these results, no significant correlation can be drawn between histology and viral E2 methylation in this cohort.
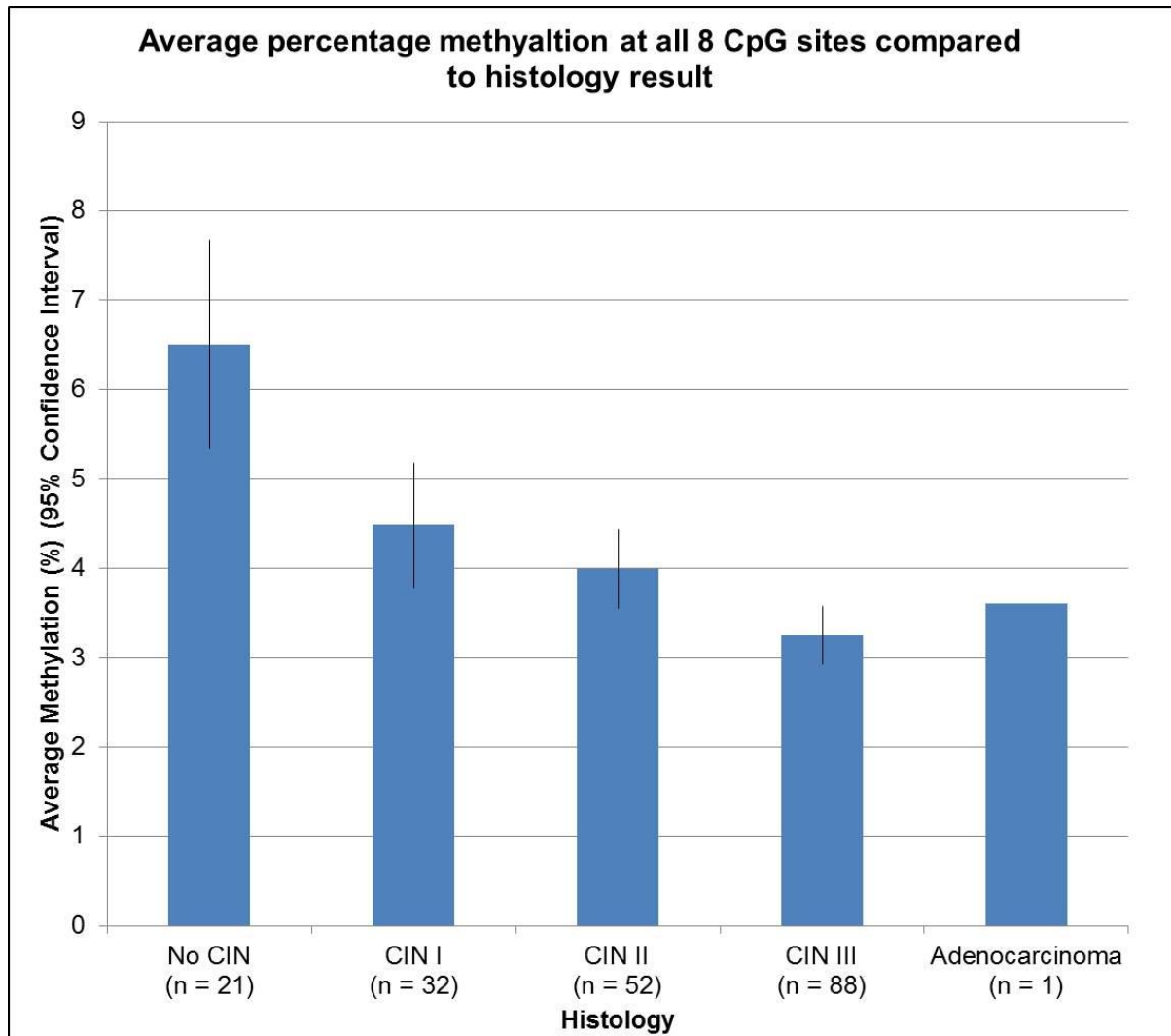
Figure 3.7. Average _E2_ viral DNA methylation values for samples with different Histology. Results show a negligible negative correlation between increasing cervical cytology results and the average viral DNA methylation r_s = -0.203, p = 0.004.

Histology	Number of Samples	Average methylation at each CpG site for all samples with the same histology (95% CI)								Average Methylation for all CpG sites
		3411	3414	3416	3432	3435	3447	3461	3472	
no-CIN	21	7.93 (±1.67)	1.70 (±0.46)	2.90 (±0.86)	9.00 (±1.40)	5.12 (±1.39)	6.94 (±1.13)	10.61 (±1.90)	7.83 (±1.56)	6.5 (±1.17)
CIN I	32	5.07 (±0.73)	1.09 (±0.28)	2.39 (±0.53)	7.64 (±1.02)	3.16 (±0.65)	5.82 (±1.06)	6.05 (±1.37)	4.59 (±1.17)	4.48 (±0.70)
CIN II	52	5.09 (±0.56)	1.01 (±0.30)	1.83 (±0.29)	5.92 (±0.59)	3.24 (±0.47)	4.38 (±0.82)	6.18 (±1.13)	6.19 (±0.77)	3.99 (±0.44)
CIN III	88	4.35 (±0.33)	0.45 (±0.11)	1.42 (±0.21)	5.12 (±0.42)	2.29 (±0.30)	3.99 (±0.45)	4.67 (±0.59)	4.5 (±0.58)	3.25 (±0.27)

Figure 3.8. Average *E2* viral DNA methylation at each CpG site for samples with varying cervical histology. (A) Table of viral DNA methylation results for each cervical cytology result at each CpG site. **(B)** Average *E2* viral DNA methylation at each CpG site (95% confidence interval is shown).

3.7 Discussion

This section will highlight the key findings from this chapter and the strength and limitations of this study. Each of the key findings will then be considered in greater detail compared to the available literature and the broader implications.

3.7.1 Key Findings

1. Novel *E2* primers were successfully applied to clinical samples stored in SurePath^{TM}.

2. Molecular tests to determine disruption and methylation of the HPV *E2* gene are not suitable biomarkers for CIN II+ in young women:

    a. The HPV *E2* gene was disrupted in 5.12% of samples tested.

    b. Viral methylation of the *E2* gene decreased with increased histology from no-CIN to CIN III, however this association was negligible.

3.7.2 Study Strengths

This work benefitted from a large cohort of samples (13,306 in total) with follow up histology data (up to 36 months). This allowed for samples with a single HPV infection to be used in further analysis. This resulted in a large cohort of samples with a single detectable HPV infection (n = 410) in this study. Along with the HPV type of interest a subset of HPV 16 samples were also included in this analysis. Previous research with this cohort had assessed viral integration in samples with high-grade cytology results (Raybould, 2013). To produce comparable results between all HPV types a subset of HPV 16 positive samples were included. The number of samples with each baseline cytology result (negative, borderline, mild, moderate and severe cytology) was chosen to represent a similar proportion of samples from the entire cohort. The individual samples within each cytology group were chosen at random to ensure that this cohort was a true representation of the total cohort.

This study benefitted from optimisation of extraction protocols as well as optimising *E2* tiling PCR assays. Extraction of DNA and assay optimisation was carried out specifically on samples generated from this study cohort which had undergone the exact same sampling storage and processing as the samples included in this study. This ensured that most samples were fit for analysis.

3.7.3 Study Limitations

This research aimed to investigate one potential measure of viral integration and methylation respectively, in four non-vaccine HPV types (HPV 31, 33, 35 and 51). Although biomarkers for common HPV types play an increasingly important role in cervical screening the introduction of the 9-valent HPV vaccine (Cuzick, 2015, Joura et al., 2015) may reduce the necessity for bookmarks specifically targeted against HPV types discussed in this thesis.

Work in this study does not consider samples with multiple HPV infections. This choice was made as in an infection with multiple HPV types it cannot be determined which infection/s are driving disease progression. Thus, any results would have been confounded by multiple infections. This reduced the number of samples available to this study.

It also could not be determined within this study if a sample had mixed infection consisting of both integrated and episomal HPV infections. If an individual had a mixed infection the integration event may be masked by the DNA present from viral episomes which would also be present, generating a potential false positive result and only viral integration events with purely integrated HPV infections would be accurately measured.

3.7.4 Novel *E2* primers were successfully applied to clinical samples stored in SurePath[TM].

This work has demonstrated that clinical samples stored in SurePath[TM] LBC media can be utilised in molecular assays to detect molecular changes in viral DNA. Four novel HPV *E2* tiling PCR assays have been developed which are both sensitive and specific to four highly prevalent HR-HPV types (HPV 31, 33, 35 and 51) which until recently women were not vaccinated against. This was the first study of HPV 31, 33, 35 and 51 *E2* disruption in young women attending their first cervical smear and the novel assays were successfully applied to 410 samples. Therefore, findings in this thesis support Hypothesis 1: Novel HPV *E2* tiling PCR assays specific for genotypes 31, 33, 35 and 51 will detect viral *E2* gene disruption in DNA extracts from smears collected in SurePath[TM] LBC media.

Two recent studies differed in the reported quality of nucleic acid after storage in SurePath[TM] LCB media. One study showed that 16% of samples were degraded and produced a false negative result after only three weeks storage in SurePath[TM] LBC media (Kubik et al., 2015) while another study by Gilbert *et al,* showed samples were viable up to 10 weeks of storage in SurePath[TM] LBC media at ambient room temperature (Gilbert et al., 2013). It has been shown that treating the clinical sample pre-extraction with a 1:1 alkaline solution and heating at 120°C for 20 minutes and improves the stability of the DNA while maintaining assay sensitivity (Tardif et al., 2016), in future work this may be an option to further improve DNA extraction and

assay sensitivity. All of these studies assess HPV presence and stability using the Roche Cobas 4800 assay (Roche Applied Science, Mannheim, Germany) which identifies a ~200 bp size fragment of the HPV *L1* gene (Rao et al., 2013). These studies do not address whether the HPV genome is intact in fragments > 200 bp. Work in this thesis restricted the size of DNA fragments amplified (< 352 bp) to account for potential DNA degradation.

3.7.5 Tests for disruption of viral E2 gene and viral methylation are not suitable biomarkers for CIN II+ in young women.

This study has shown that both *E2* tiling PCR and viral DNA methylation independently are unsuitable biomarkers in young women attending their first routine cervical smear test with as there was no distinct association with increased cervical histology (≥36 months), failing BIDD biomarker discovery stage 1. Following the prognostic and predictive biomarker roadmap further basic research is required to define potential biomarkers in young women. Previous work has shown HPV 16 *E2* disruption may be a suitable biomarker in women attending family planning clinics (Collins et al., 2009) and work by Bryant et al has suggested that methylation of HPV may be a biomarker for duration of HPV infection but the authors indicated a limited utility of methylation as a biomarker in young women (Bryant et al., 2015).

3.7.5.1 The HPV E2 gene was disrupted in 5.12% of samples tested.

A major aim of this work was to study HPV *E2* disruption as a potential biomarker in young women. In a population of young women (20 – 22 years old) attending their first routine cervical smear test with a single HPV infection (HPV 16, 31, 33, 35 or 51) HPV *E2* disruption was not a common event and was only detected in 5.12% of samples overall. The lowest level of HPV *E2* disruption was present in samples with HPV 16 infections. This result does not support Hypothesis 2(a): Disruption of the viral E2 (and/or E1 (HPV 16 only)) gene is (A) Common in HR-HPV infections (HPV 16, 31, 33, 35 and 51).

High levels of HPV *E2* gene disruption were shown by Collins et al, when *E2* disruption was assessed by *E2* tiling PCR in a cohort of young women (aged 15 to 19 years) who had actively sought medical advice and/or treatment at a family planning clinic (Collins et al., 2009). These samples were collected using an Ayre's spatula and placed in 10 ml of phosphate-buffered saline, stored at -80°C and then extracted using guanidinium thiocyanate acid. These women may have been exposed to behavioural risk factors for HR-HPV infection and disease such as; young age of sexual debut, multiple sexual partners, pregnancy and use of hormonal contraceptives (Brake and Lambert, 2005, Kjellberg et al., 2000, Johnson et al., 2012a). Another study by Ramanakumar et al invited college students (mean age 23, range 17 to 42) to participate in their study which assessed viral integration by the ratio of *E2*:*E6* gene expression (Ramanakumar et al., 2010). This study reported low

levels of viral gene disruption like those described in this thesis. Samples in this study were taken using Accelon Cervical Biosampler and extracted using QIAamp columns. This study does not report how samples were stored between collection and extraction.

Samples in both of the above studies and work reported in this thesis all varied in sample collection, storage and extraction procedure. This may have led to variation in downstream HPV gene analysis. If a biomarker for HPV disease presence or prognosis is to be added to future cervical screening procedures the collection, storage and extraction of cervical material must be assessed and a SOP for sample collection and molecular assays must be established (BIDD Prognostic/Predictive Biomarkers Roadmap - stage 2). This will ensure downstream that a statistically robust correlation between the biomarker and clinical outcome can be established which ensures that the clinical outcome of the patients is improved by the use of the biomarker.

Although HPV *E2* disruption does not appear to be a common event in young women (as reported in this thesis) several studies have shown that with increased cervical histology HPV *E2* disruption is a more common event (Graham and Herrington, 2000, Li et al., 2012, Sepetiene et al., 2011). In this study of young women there are few incidence of CIN II+ and no incidences of cancer, therefore predominant viral disruption in women with severe cytology and CIN II+ was not seen, rejecting Hypothesis 2 (b). This may be because this study only includes

women aged 20 – 22. The incidence of cervical cancer is strongly linked with age and in the UK the most common age of diagnosis is 25 – 29 year of age and more than 50% of cases are diagnosed in women under the age of 45 year of age (Cancer Research UK, 2017).

3.7.5.2 Viral methylation of the *E2* gene decreased with increased histology from no-CIN to CIN III, however this association was negligible.

In this study, a low level of viral methylation of the HPV *E2* gene was seen in samples (average methylation of less than 8%). With increasing cervical histology, there was a decrease in average viral methylation, although this correlation was negligible, therefore viral methylation failed to meet the criteria of BIDD biomarker discovery stage 1 and further basic research is required to define potential biomarkers in young women. This finding does not support Hypothesis 3: HPV hypermethylation is (a) Common in HR-HPV 16 infections; and (b) Detectable at highest levels in women with CIN II+.

In cancer progression both hypo- and hyper-methylation have been reported. A genome wide reduction in methylation has been shown to increase genetic instability which can contribute to cancer progression (Cadieux et al., 2006). Alternatively, hypermethylation of specific promoter genes have been shown to be a

critical hallmark of cancer progression (Paz et al., 2003, Kulis and Esteller, 2010). Studies into HPV 16 genome methylation have suggested that high levels of viral methylation in *L1*, *L2* and *E2* are associated with a risk of higher cervical histology (Mirabello et al., 2012, Yang, 2013).

Previous research into methylation of HPV DNA in young women has proposed that methylation of the HPV 16 *L1/L2* gene may be another potential biomarker for the duration of HPV disease, specifically methylation at CpG sites 5600 and 5609. However, like work presented in this thesis HPV 16 *L1/L2* methylation may only have limited value in young women. This research suggested that L1/L2 methylation may have clinical application in the triage of older women with HPV positive DNA (Bryant et al., 2015). Methylation of the HPV 16 *E2* gene (using the same assays as described in this thesis) has also been recommended for further study as a possible biomarker to help determine treatment pathways for VIN patients with topical drug treatments, cidofovir and imiquimod. This cohort of women had higher reported methylation than reported in this study, however the demographic of this cohort varied greatly from that included here, specifically women included in this study had VIN III and an average age of 47.2 years (range 20 – 81 years) (Jones, 2016). This shows that the methods of viral methylation reported here can be successfully applied and suggests that this young cohort of women is not a suitable cohort of women for application of this biomarker but does not rule out its application in other patient groups.

Chapter 4

Gene disruption, expression and methylation of HPV 16 in anogenital disease

Results in Chapter 3 indicated that neither disruption nor methylation of the HPV *E2* gene were reliable biomarkers for defining anogenital disease in young women attending their first cervical smear. Nevertheless, HPV methylation and integration into the host genome can be expected to have a major impact on the oncogenic process and could potentially have utility in predicting progression to high-grade cervical neoplasia in women with cervical disease. Previous work reported in this thesis utilised nucleic acid extracted from clinical samples which were stored in Surepath™ LBC media. To improve DNA/RNA quality (by reducing DNA/RNA degradation and sample impurities) in this chapter samples were investigated that had been collected in RNAprotect® media (QIAGEN Ltd, Manchester, UK) as described in section 2.1.2.

4.1 Clinical Cohort

Women attending a colposcopy clinic in University Hospital Llandough, Wales between March 2011 and April 2014 were invited to donate a cervical smear sample for investigation in this study. The study protocol originally aimed to obtain 15

samples from women with a mild, moderate and severe initial colposcopy impression. The study target population was extended to include samples with an initial cancerous colposcopy impression. Samples were obtained by taking an additional cervical smear with a standard cytology brush which was then transported to the HPV Research Group, Cardiff University. A detailed study protocol is included in Appendix 2.

Samples were collected from 51 patients with mild (n = 11), moderate (n = 5), severe (n = 16) and cancerous (n = 19) initial colposcopy impression. Additional clinical information was made available and included any treatment for anogenital disease and cervical histology results (Table 4.1). Results were then categorised based on cervical histology result where available (90.2%) as this is the gold standard for assessing the presence and degree of cervical disease. In samples with histology reported between two categories (e.g. CIN I/CIN II or CIN II/CIN III) samples were categorised by the highest histology grade. Of the five samples categorised as 'no reported histology' three had no histology data reported and two had ungraded CIN. As these two samples could not be correctly categorised they were designated as not reported and were not included in any subsequent analysis. On average, women with cancer were older than women with negative histology or CIN and had the largest range of patient ages (18 to 84 years of age, average age 49.8 years).

Table 4.1. Summary of clinical cohort categorised by sample histology. The average age of patients at the time of sampling has been shown along with the age range of all individuals within each category. Colposcopy impression refers to the initial categorisation of patients when the smear sample was taken. HR-HPV infections were identified using Papillocheck® and divided into HPV 16 single infections (HPV 16 SI), multiple infections with HPV 16 plus 1 or more other HR-HPV types (HPV 16 MI), other single HPV infections, excluding HPV 16 (HR-HPV SI) and samples with multiple HPV types, excluding HPV 16 (HR-HPV MI).

	Histology					
	Not Reported	Negative	CIN I	CIN II	CIN III	Cancer
Number of samples	5	6	1	4	18	17
Average age (years) (Range)	34.4 (22 – 53)	36.7 (21 – 59)	36 (-)	27 (23 – 32)	28.8 (21 – 32)	49.8 (18 – 84)
Colposcopy Impression						
Mild	2	6	-	2	1	-
Moderate	2	-	1	1	1	-
Severe	-	-	-	-	15	1
Cancer	1	-	-	1	1	16
Treatment						
Not reported	-	-	-	-	-	-
Biopsy	5	6	1	2	7	5
LLETZ	-	-	-	2	11	2
Chemotherapy	-	-	-	-	-	1
Hysterectomy	-	-	-	-	-	9
HR-HPV Infection						
Negative	1	3	-	-	1	5
HPV16 SI	-	-	-	-	5	6
HPV 16 MI	-	1	-	1	4	3
Other HR-HPV SI	-	1	-	-	-	3
Other HR-HPV MI	4	1	1	3	8	0

4.2 **HPV typing**

HPV typing was performed using commercial DNA-array PapilloCheck$^®$ (as described in section 2.1.2). A single HPV type was detected in 15 samples, while 26 samples were infected with two or more HPV types, accounting for 98 HPV infections in total. Samples with cancer accounted for most single infections (60%) and HPV 16 was the most prevalent single infection (73.3%). The most prevalent HPV types in samples with multiple infections were HPV 33 (12.0%), HPV 16 (10.8%) and HPV 39 (10.8%). Samples with cancer had on average 0.88 HPV infections per sample; however negative, CIN I, CIN II and CIN III samples contained on average 2.55 HPV types per sample. A negligible negative correlation is present between the severity of histology and the average number of HPV infections per sample (r_s = -0.115, p = 0.531).

Ten samples tested negative for HPV using the Papillocheck® assay (Not reported n = 1, negative n = 3, CIN III n = 1, Cancer n = 5). To verify this result, 'negative' samples were tested a second time using type specific primers for HPV 16. This showed that 6 samples (negative n = 2, Cancer n = 4) were positive for HPV 16 DNA. These samples were included in subsequent analysis of HPV 16

biomarkers but not included in this HPV typing analysis (Figures 4.1), as these samples may or may not contain more HR-HPV types.

	Negative	HPV 16	HPV 18	HPV 31	HPV 33	HPV 35	HPV 39	HPV 44/55	HPV 45	HPV 51	HPV 52	HPV 53	HPV 56	HPV 58	HPV 59	HPV 66	HPV 68	HPV 82
Not Reported (2)	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Negative (7)	3	1	0	0	0	1	1	0	0	1	0	0	1	0	2	1	0	0
CIN I (4)	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
CIN II (16)	0	1	0	0	2	1	1	1	1	1	2	1	1	0	1	1	1	1
CIN III (47)	1	9	3	2	6	4	6	0	1	2	1	0	0	7	3	2	1	0
Cancer (15)	5	9	1	0	0	0	0	0	0	0	1	0	0	0	3	0	1	0
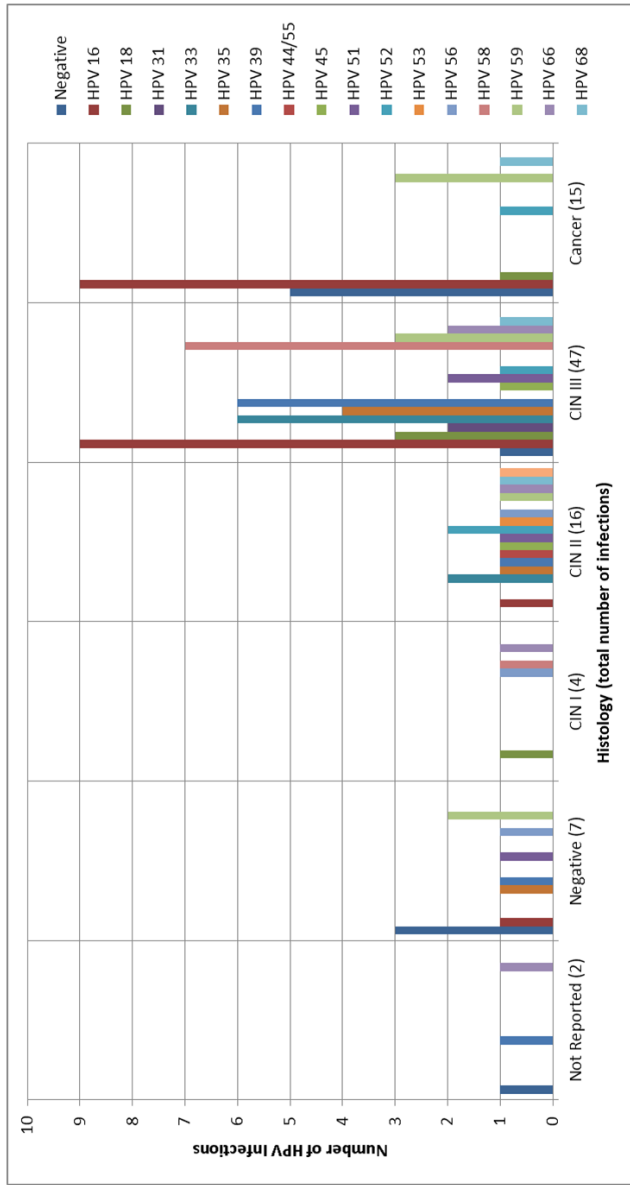
Figure 4.1 HR-HPV infections as reported by Papillocheck DNA Chip. The total number of HPV infections for each histology grade is show along with the number of samples with no HPV detected.

4.3 HPV 16 Genome Disruption

HPV genome disruption was used in this study as a substitute marker for HPV integration. Samples were considered to have a disrupted HPV genome if the *E6* gene was detectable but the *E1* and/or *E2* genes could not be detected by PCR analysis. This model is indicative of viral integration although is not a definitive biomarker (described in section 1.3.1). HPV genome disruption was assessed in HPV 16 positive samples as determined by Papillocheck® and/or type specific *E6* PCR (n = 32). The average age of samples tested for genome disruption was 40 years old (18 to 84 years) and had varying histology (not reported n = 1, no CIN n = 4, CIN II n = 3, CIN III n = 10 and Cancer n = 14). HPV gene disruption was seen in 12 samples (37.5%), with CIN II (n = 2), CIN III (n = 4) and cancer (n = 6) (Figure 4.2), although there was not a significant difference in level of integration compared to disease grade ($X^2(1) = 0.545$, p = 0.46).

There was very little difference in the average age of samples with and without viral integration, 40.1 years and 39.9 years respectively. There was a difference in the prevalence of viral integration when comparing samples with single (63.63% integration) and multiple HPV infections (45.45% integration) however this difference is not statistically significant ($X^2 = 0.64$, p < 0.1).

Viral gene disruption was present in 66.7% of samples with CIN II (n = 2), 40% of samples with CIN III (n = 4) and 42.9% of samples with cancer (n = 6), (Figure 4.2). When patient treatment was compared, 50% of women with viral genome disruption had a radical hysterectomy as part of their treatment compared 10% of women with no genome disruption (Table 4.2).

The *E1* region was at least partially disrupted in all samples with genome disruption and four samples had no *E1* gene detectable by PCR (HI-BCD-5, HI-BCD-8, HI-BCD-31 and HI-BCD-66). HI-BCD-5, HI-BCD-8 and HI-BCD-66 had previously tested positive for HPV 16 by Papillocheck®, indicating the *E1* gene is present, but may be fragmented and/or only partially present. HI-BCD-31 was the only sample to have no detectable *E1* or *E2*. This sample was also HPV 16 negative when assessed by Papillocheck®. The HPV 16 *E2* gene was disrupted in only one other sample, HI-BCD-29, which had CIN II and received large loop excision of the transformation zone (LLETZ) treatment. Both samples with *E2* disruption (HI-BCD-31 and HI-BCD-29) came from young women, aged 26 and 27 years respectively with mild and mild/moderate colposcopy impression. Both women were diagnosed with low-grade HPV disease (CIN II and CIN I/II) and treated with LLETZ. All samples from women with moderate, severe and cancer colposcopy impression had intact *E2* (Table 4.2).
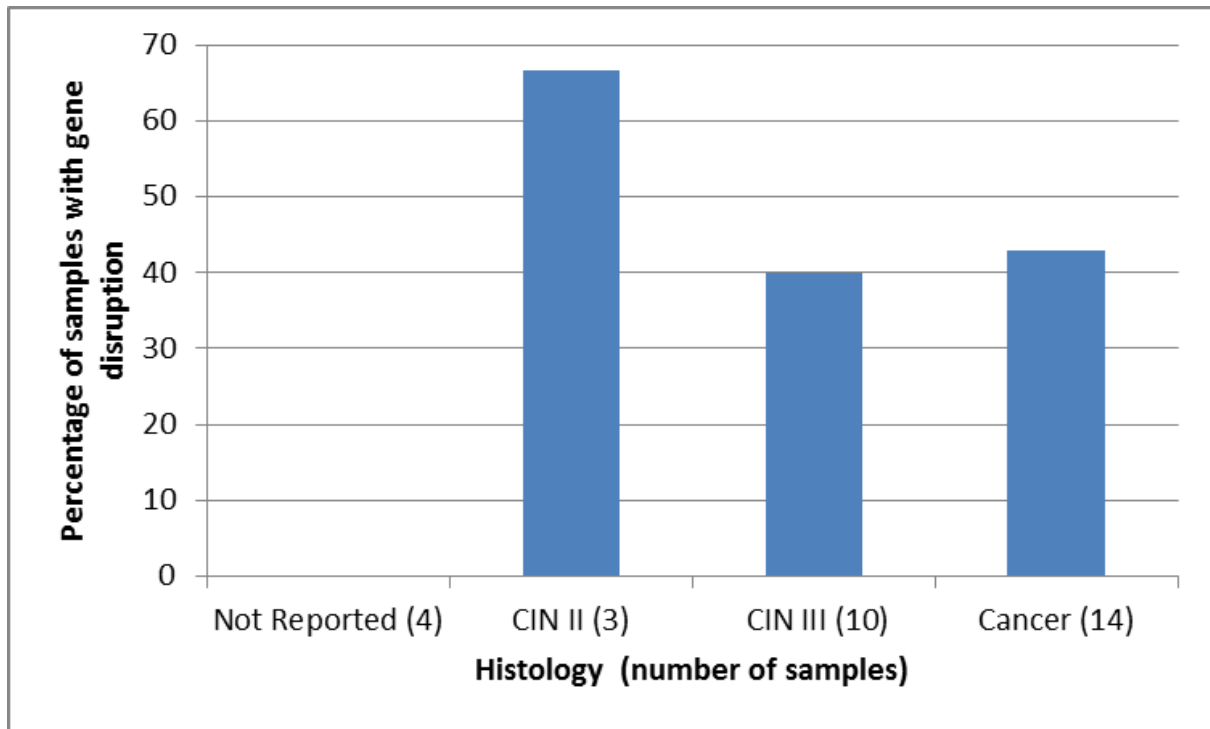
Figure 4.2 HPV 16 viral disruptions in samples with no reported disease, CIN II, CIN III and Cancer. Samples were considered disrupted if one or more of full-length HPV 16 *E1*/*E2* primers failed to amplify the target DNA with amplification of the HPV E6 gene ($X^2(1) = 0.545$, p = 0.46).

Table 4.2 Summary of samples with viral genome disruption. All samples had the *E6* gene intact and one or more of the *E1* and *E2* primers did not produce any amplicons. Assays that produced amplicons are noted with a 1 and assays that failed to produce any product are noted with 0.

Study ID	Colposcopy Impression	Age	HPV Type/s	Histology	Treatment	E1 Part1	E1 Part2	E2	E6
HIBCD-29	Mild to moderate	27	16, 33, 59, 56, 51, 68	CIN 1/2	LLETZ	0	1	0	1
HIBCD-31	Mild	26	52, 66	CIN 2	LLETZ	0	0	0	1
HIBCD-5	Severe	35	16	CIN 2/3	LLETZ	0	0	1	1
HIBCD-20	Moderate	28	16	CIN 3	LLETZ	0	1	1	1
HIBCD-8	Severe	29	16, 18, 31, 59, 66	CIN 3	LLETZ	0	0	1	1
HIBCD-46	Cancer	31	16	CIN 3	cone biopsy	0	1	1	1
HIBCD-58	Cancer	43	16, 18	adenosquamous carcinoma (FIGO 1B1)	radical hysterectomy	1	0	1	1
HIBCD-53	Cancer	70	16	adenosquamous carcinoma (FIGO 1B1)	radical hysterectomy	0	1	1	1
HIBCD-52	Cancer	45	16	adenosquamous carcinoma (FIGO 1B1)	radical hysterectomy	0	1	1	1
HIBCD-50	Cancer	34	16	SCC (FIGO 1B1)	radical hysterectomy	0	1	1	1
HIBCD-49	Cancer	59	16	SCC (FIGO 1B1)	radical hysterectomy	0	1	1	1
HIBCD-66	Cancer	55	16, 52	SCC (FIGO 1B1)	radical hysterectomy	0	0	1	1

HPV genome disruption was further investigated by using the DIPS assay to identify virus/host integration sites. In DIPS analysis samples are digested and adapters are ligated to the digested DNA. HPV specific linear PCR is followed by a nested PCR with primers specific from the adapter and HPV. PCR products are then sequenced to identify viral/host fusion sites (described in Methods section 2.4.3). DIPS analysis was carried out on 32 samples. HPV 16 only transcripts were detected in 13 samples and viral-host fusion sites were identified in 5 samples with CIN I (n = 1), CIN III (n = 3) and Cancer (n = 1). Findings from the DIPS analysis have been summarised in Table 4.3.

Table 4.3 HI-BCD Samples with viral integration identified by DIPS. Samples HI-BCD-50 showed integration at two sites in the HPV and human genome. Human locations noted with a * were integrated into a known common fragile site of the human genome.

Sample ID	HPV Type/s	Colposcopy Impression	Histology	HPV Location	Human Location	Accession Number
HIBCD-19	16, 56, 58, 66	Moderate	CIN I	L1 (4479 bp)	9q21.3	NT_008470.19
HIBCD-8	16, 18, 31, 59, 66	Severe	CIN III	E1 (2431 bp)	11p12	NT_0009237.18
HIBCD-9	16, 18	Severe	CIN III	E1 (2877 bp)	6q15*	NT_007299.13
HIBCD-46	16	Cancer	CIN III	E1 (2583 bp)	3q25.33*	NC_018914.2
HIBCD-50	16	Cancer	Invasive SCC (Figo 1b1)	E1 (2611 bp)	9p21.2*	AL451137.8
				E2 (3200 bp)	3q26.32	AC110992

HI-BCD-19 came from a 36-year-old woman who had a moderate colposcopy impression and histological analysis diagnosed CIN I. This sample had multiple HPV infections identified (HPV 16, 56, 58 and 66). The integration event was detected in the *L1* region of the HPV genome and at locus 9q21.3 in the human genome. As would be expected, *E1*/*E2* PCR did not detect this integration event.

HI-BCD-8 and HI-BCD-9 both had a severe colposcopy impression and both were subsequently diagnosed with CIN III. HI-BCD-8 came from a 29-year-old woman with multiple HPV infections (HPV 16, 18, 31, 59 and 66). Viral disruption was detected in the *E1* region of the HPV genome and integrated into chromosome 11 at locus 11p12. HI-BCD-9 also came from a 29-year-old women with two HPV infections identified (HPV 16 and 18). This sample had viral genome disruption identified in the *E1* region of HPV genome and at locus 6q15 of the human genome. This is a known common fragile site in the DNA, FRA6G. HI-BCD-8 had no *E1* detected by the *E1*/*E2* PCR however HI-BCD-9 had full length *E1* and *E2* present. These results suggest that HI-BCD-8 has a fully integrated HPV 16 infection and HI-BCD-9 is a mixed infection with both integrated and episomal HPV DNA present which is masking the genome disruption in the *E1*/*E2* PCR assay.

Sample HI-BCD-46 had initial cancerous colposcopy impression however the patient was the patient was diagnosed with CIN III by histological analysis. HI-BCD-

46 came from a 31-year-old woman with a single HPV 16 infection. Viral disruption was detected by the *E1*/*E2* PCR assays and DIPS confirmed disruption in the *E1* region of HPV and integrated into chromosome 3, locus 3q25.33. This is the location of common fragile site FRA3D and encodes for Interleukin-12A (IL12A), a subunit of Interleukin 12.

Individual HI-BCD-50 had initial cancerous colposcopy impression and was subsequently diagnosed with invasive squamous cell carcinoma (SCC) (Figo Stage 1b1). This sample came from a 34-year-old woman with a single HPV 16 infection. Two viral-host fusion sites were identified in this sample, the first was between the *E1* gene and chromosome 9 within a common fragile site (9p21.3) and between HPV *E2* gene and chromosome 3 (3q26.32). Histological analysis identified invasive SCC (FIGO stage 1b1) in HI-BCD-50. In the *E1*/*E2* PCR analysis the *E1* genome was disrupted however full length *E2* was detected, suggesting that sample HI-BCD-50 has a mixed infection of integrated and episomal HPV forms.

4.4 Viral genomic DNA methylation

In chapter 3, a moderate negative correlation between increasing cytology grade and viral methylation was observed (r_s = -0.203, p = 0.004) however this work did not support the use of methylation as a biomarker of HPV disease in young

women, this may have been due to a high level of transient HPV infections in this young population. This subsequent work aimed to assess viral methylation as a biomarker for HPV disease in a population attending a colposcopy clinic, including a mix of ages and cervical disease.

Analysis of HPV 16 DNA methylation was successfully carried out in 68% (n = 22/32) of HPV 16 positive samples and in 40% (n = 4/10) of samples that were HPV 16 negative by Papillocheck[®]. The average age women included in the methylation analysis was 39.54 years (18 to 80 years) and when comparing the age of patients at the time of sampling to the average level of viral methylation there is a trend for increased viral methylation with an increase in age, however this was not significant (r_s = 0.413, p = 0.56) (Figure 4.3).
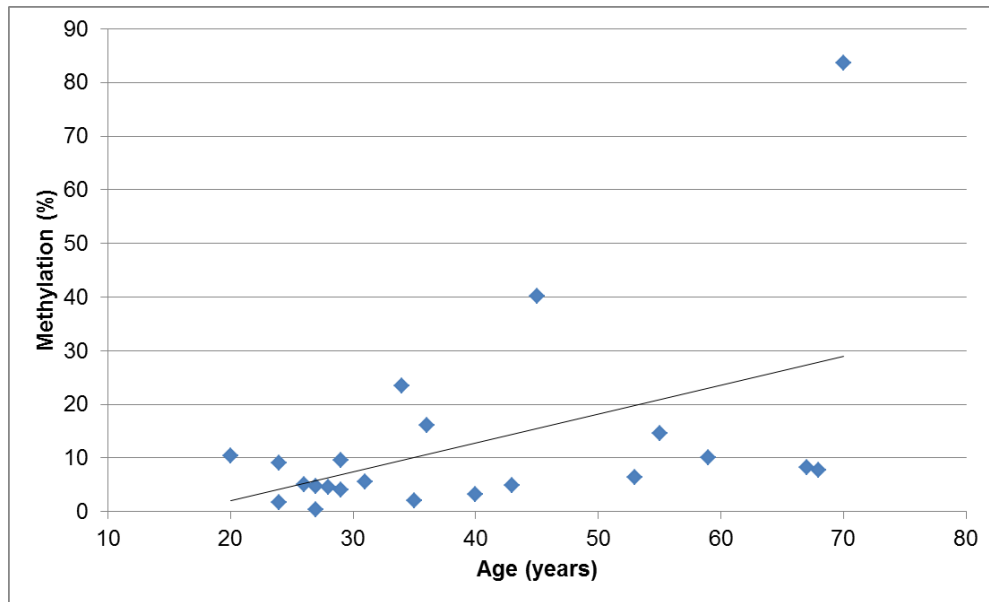
Figure 4.3. Age of participants compared to the average viral methylation across all 8 CpG points analysed. The blue dots each represent a patient sample and the black line is the linear trend of the data. (r_s = 0.413, p = 0.56)

The average viral methylation detected in samples with no CIN (6.22%, n = 3), CIN II (4.67%, n = 1), CIN III (6.41%, n = 10) and cancer (21.60%, n = 9) histology showed a positive correlation between the level of viral methylation and disease grade, although this did not reach significance (r_s = 0.408, p = 0.06) (Figure 4.4). When Figure 4.4 is considered there are two samples with higher levels of viral methylation compared to the cohort, HI-BCD-52 and HI-BCD-53. If these outlier samples are removed from the analysis, the average viral methylation in the cancerous colposcopy impression is reduced to 10.10% and the correlation between disease grade and viral methylation is reduced (r_s = 0.283, p = 0.22).

Samples with high methylation HI-BCD-52 and HI-BCD-53 both have a single HPV 16 infection that was disrupted in the *E1* region as identified by *E1*/*E2* PCR and in both samples viral-host fusion transcripts were not identified by DIPS. Both women had adenocarcinoma identified by histological analysis and went on to have radical hysterectomies.
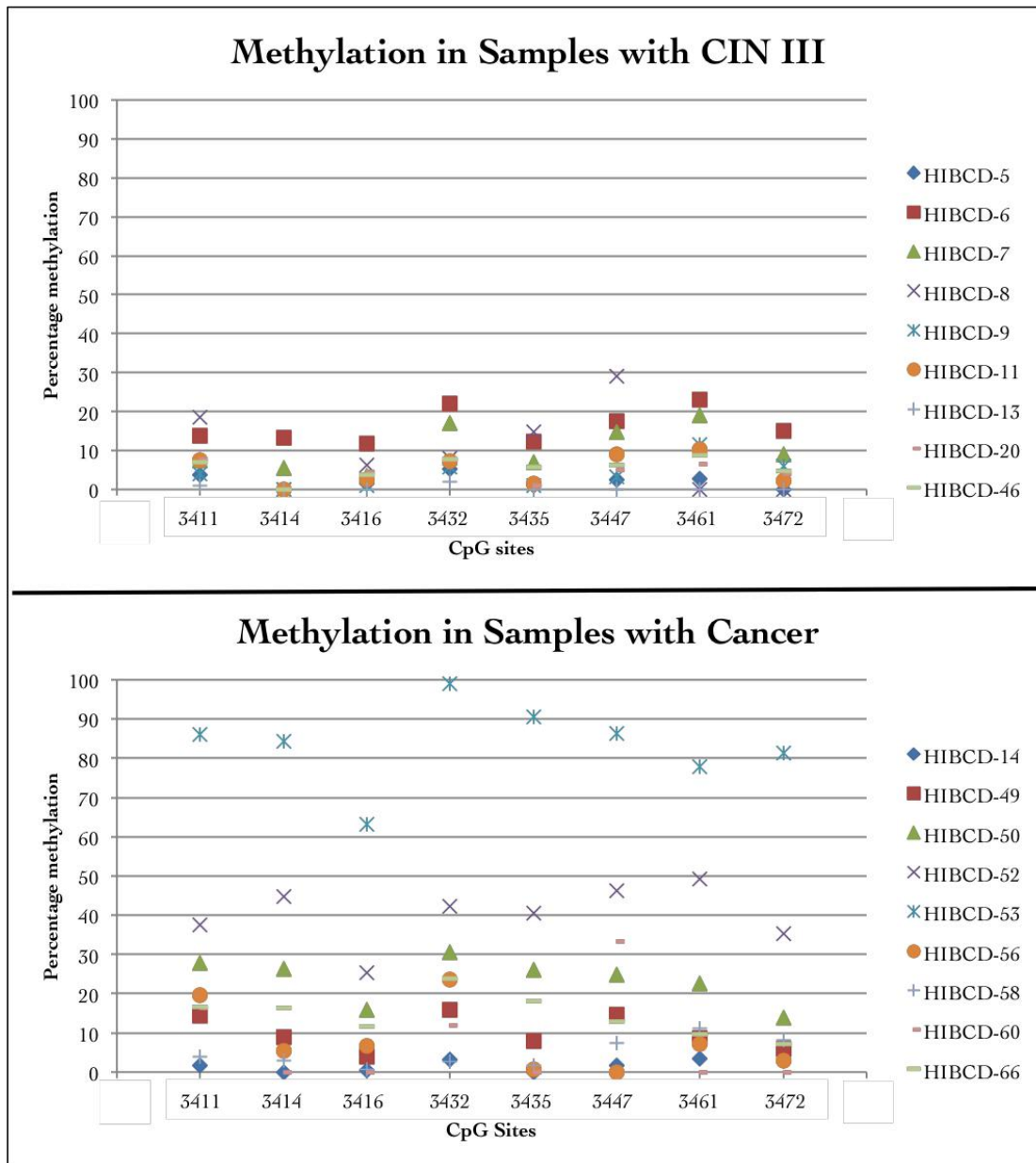
Figure 4.4 Comparison of viral DNA methylation levels in samples with (A) CIN III histology and (B) Cancer histology There was a positive correlation between the level of viral methylation and disease grade, although this did not reach significance ($r_s = 0.408$, $p = 0.06$).

When viral methylation levels were compared between samples with disrupted genome vs. undisrupted genome there was no significant variation in the levels of viral methylation ($X^2(1)$ = 0.085, p = 0.770) (Figure 4.5). However when individual CpG sites were considered there was a significant variation in levels of viral methylation compared to gene disruption at three sites; CpG 3416, CpG 3435 and CpG 3447 (CpG 3416 $X^2(1)$ = 4.281, p = 0.039; CpG 3435 ($X^2(1)$ = 7.113, p = 0.008); CpG 344 ($X^2(1)$ = 4.020, p = 0.045).
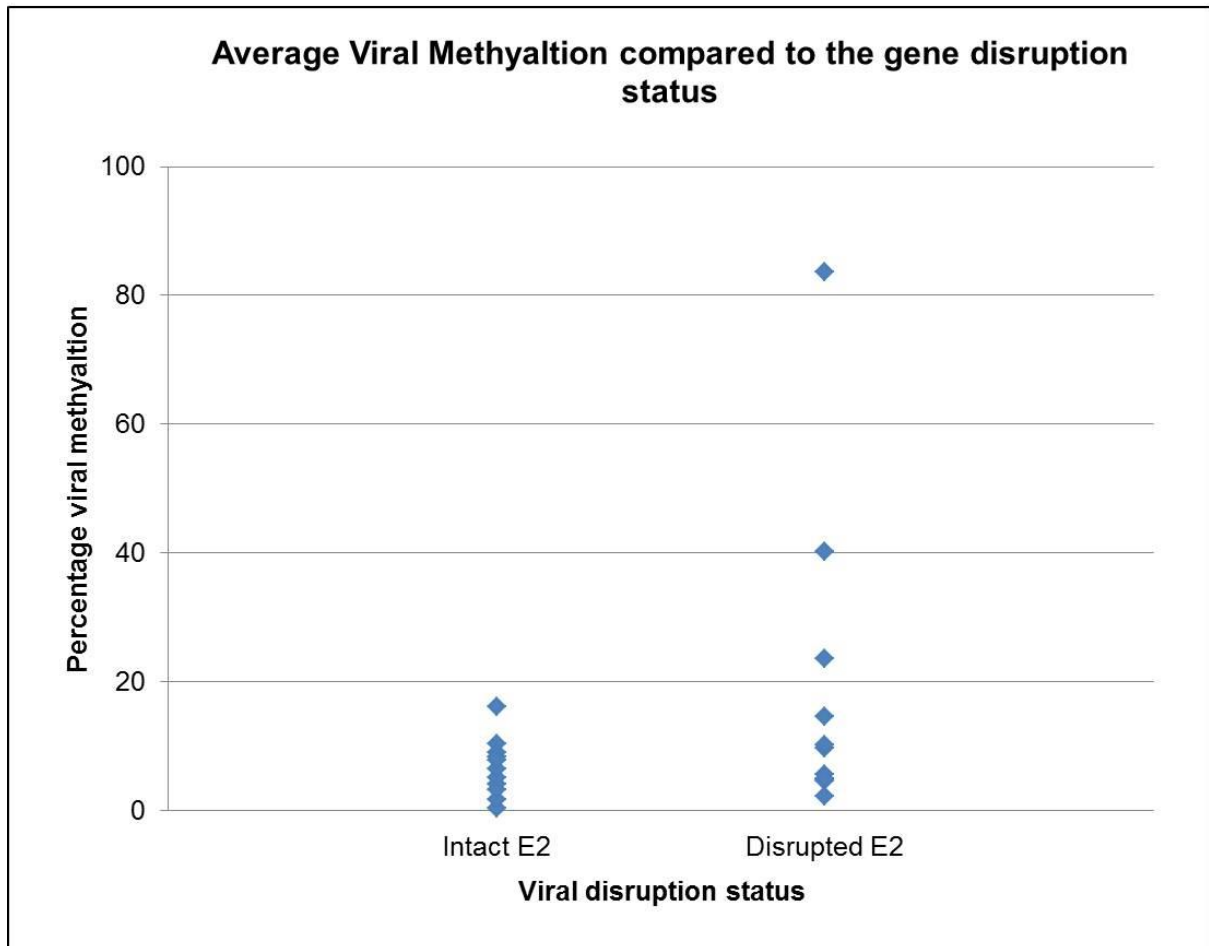
Figure 4.5 viral methylation levels were compared between samples with disrupted genome and intact _E2_ genome. ($X^2(1) = 0.085$, p = 0.770).

Across all 8 CpG sites, samples with a single HPV 16 infection had lower methylation compared to samples with multiple infections; however this result was not statistically significant ($X^2(2)= 1.11$, p = 0.53) (Figure 4.6).

	CpG Sites								
HPV infection	3411	3414	3416	3432	3435	3447	3461	3472	Average methylation
Single infection	14.91	9.89	7.96	16.28	10.74	14.83	15.31	10.24	12.52
Multiple infection	16.49	13.15	9.77	18.70	14.39	17.74	16.76	12.32	14.91

Figure 4.6. Trends in viral DNA methylation in samples with a single HPV16 infection compared to samples with HPV 16 plus one or more other HR-HPV types. At all CpG sites samples with multiple HPV types (shown in red) had a higher level of viral DNA methylation compared to samples with only HPV 16 infection detected (shown in blue) ($X^2(2) = 1.11$, p = 0.53).

4.5 Viral RNA expression

In Chapter 3, it was not possible to assess viral RNA expression in samples from young women attending their first cervical smear due to sample storage in SurePath™ LBC Media, which meant viable viral RNA was not available for analysis. In this study, however, storage conditions were designed to ensure that RNA could be obtained. HPV *E2* and *E6* gene expression was assessed using qRT-PCR and normalised to human housekeeping genes *HPRT* and *TBP2* (described in Methods section 2.5). When *E2* and *E6* expression levels are equivalent to zero their expression is comparable to housekeeping genes. When viral gene expression levels are quantified above or below zero, this indicates HPV gene expression is greater or lower than that of the house keeping genes (Tables 4.4 and 4.5).

In the 32 samples analysed, 28% of samples failed to amplify any gene target. In addition, 25% of samples (n = 8/32) amplified the *E2* gene but failed to amplify *E6*. Most samples (87%, n = 7/8) with detectable *E2* but no *E6* gene expression had cancerous colposcopy impression and 50% of samples had genome disruption in the *E1* gene. Notably, samples HI-BCD-52 and HI-BCD-53 had high levels of viral DNA methylation and both amplified HPV 16 *E2* gene but failed to amplify *E6* (Table 4.5). Subsequent analysis described in this chapter includes samples which successfully amplified reference genes *HPRT* and *TBP2* as well as both HPV genes; this reduced the sample size for analysis down to 14 samples. These 14 samples included two samples with no CIN detected, two samples with CIN II, six samples with CIN III and four samples diagnosed with Cancer.

Table 4.4 Summary of samples tested for HPV 16 viral gene expression with no CIN to CIN III. CNRQs of *E2* and *E6* gene expression are shown and these values are normalised to two reference genes *HPRT* and *TBP2*.

Sample	Infection/s	Age	Cytology	Histology*	Disruption	E2	E6	Methylation
HIBCD-21	16, 35, 39, 59	22	mild	no CIN	0	-2.07 (±0.61)	-4.74 (±0.3)	ND
HIBCD-24	ND	40	mild	no CIN	0	-0.06 (±0.18)	-0.84 (±0.07)	3.25
HIBCD-26	ND	21	mild	no CIN	0	-2.32 (±2.14)	ND	ND
HIBCD-29	16, 33, 59, 56, 51, 68	27	mild	CIN I/II	1	1.03 (±0.11)	0.63 (±0.09)	4.67
HIBCD-10	33, 35, 39, 45, 52	32	moderate	CIN II	0	-3.26 (±0.29)	-3.09 (±0.34)	ND
HIBCD-20	16	28	moderate	CIN III	1	ND	ND	4.50
HIBCD-5	16	35	severe	CIN II/III	1	1.35 (±0.06)	0.92 (±0.05)	2.09
HIBCD-7	16, 35	20	severe	CIN III	0	0.97 (±0.04)	0.81 (±0.04)	10.38
HIBCD-9	16, 18	29	severe	CIN III	0	0.76 (±0.04)	0.58 (±0.12)	4.07
HIBCD-11	16, 52	26	severe	CIN III	0	0.92 (±0.43)	0.10 (±0.33)	5.06
HIBCD-13	16	27	severe	CIN III	0	2.18 (±0.13)	1.41 (±0.05)	0.37
HIBCD-46	16	31	cancer	CIN III	1	1.17 (±0.11)	0.74 (±0.04)	5.50

Table 4.5 Summary of samples tested for HPV 16 viral gene expression with Cancer. CNRQs of *E2* and *E6* gene expression are shown and these values are normalised to two reference genes *HPRT* and *TBP2*.

Sample	Infection/s	Age	Cytology	Histology*	Disruption	E2	E6	Methylation
HIBCD-14	16, 52	24	severe	SCC	0	1.89 (±0.11)	1.13 (±0.11)	1.69
HIBCD-49	16	59	cancer	SCC 1B1	1	1.61 (±0.11)	0.91 (±0.11)	10.05
HIBCD-52	16	45	cancer	Adenocarcinoma grade 1	1	1.32 (±0.08)	ND	40.17
HIBCD-53	16	70	cancer	Adenocarcinoma 1B1, moderate differentiation	1	1.7 (±0.09)	ND	83.57
HIBCD-54	ND	54	cancer	SCC 1B1, moderate differentiation	0	-2.4 (±1.8)	ND	ND
HIBCD-55	16	84	cancer	Complex squamous neoplastic proliferative process, papillary in-situ squamous carcinoma,	0	ND	0.67 (±0.05)	ND
HIBCD-56	16	67	cancer	CIN3/SCC	0	1.55 (±0.05)	0.8 (±0.03)	8.30
HIBCD-58	16, 18	43	cancer	Adenocarcinoma 1B1, moderate differentiation	1	0.43 (±0.2)	ND	4.83
HIBCD-60	ND	68	cancer	Poorly differentiated basaloid carcinoma with neuroendocrine differentiation	0	0.66 (±0.27)	ND	7.71
HIBCD-64	ND	52	cancer	SCC 3b	0	-3.29 (±0.39)	ND	ND
HIBCD-65	ND	18	cancer	SCC 1B2	0	-4.49(±0.12)	-2.39(±0.07)	ND
HIBCD-66	16, 52	55	cancer	SCC 1B1	1	-0.86(±0.28)	ND	14.55

The gene expression results were compared to histology and low $E2$ and $E6$ gene expression is seen in samples with no CIN ($E2$ = -1.06, $E6$ = -2.79) and CIN II ($E2$ = -1.11, $E6$ = -1.23). This highest $E2$ and $E6$ levels are present in samples with CIN III ($E2$ = 1.22, $E6$ = 0.76) and Cancer ($E2$ = 0.14, $E6$ = 0.11). A trend can be seen with increased $E2$ and $E6$ gene expression in CIN III and cancer samples compared to no CIN and CIN II, however these results were not significant ($E2$ expression, $X^2(3)$ = 4.07, p = 0.254, $E6$ expression, $X^2(3)$ = 5.75, p = 0.13) (Figure 4.7).
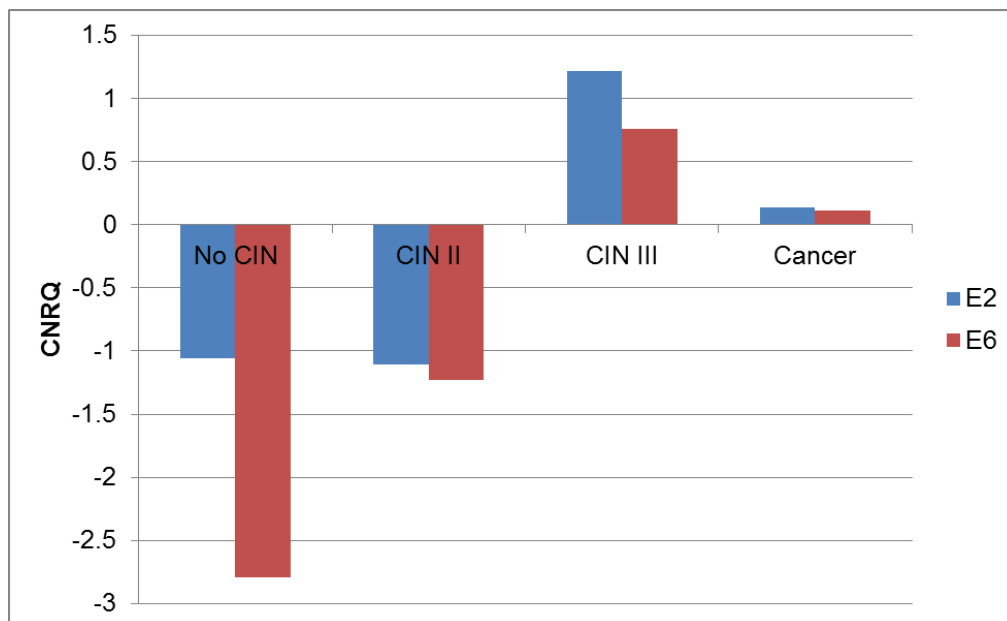
Figure 4.7 Viral *E2* and *E6* gene expression in samples with varying histological outcomes (n = 15). The CNRQs of $E2$ and $E6$ gene expression are shown. Values are normalised to reference genes *HPRT* and *TBP2*. ($E2$ expression, $X^2(3)$ = 4.07, p = 0.254, $E6$ expression, $X^2(3)$ = 5.75, p = 0.13)

4.6 Discussion

This section will highlight the key findings from this chapter and the strength and limitations of this study. Each of the key findings will then be considered in greater details compared to the available literature.

4.6.1 Key Findings

The key findings from work presented in this chapter:

1. Samples with cancer accounted for most single infections and HPV 16 was the most prevalent single infection.
2. Viral disruption was present in 37.5% of samples; however, there was no significant disruption between different histology.
3. There was a non-significant trend of increased viral methylation with increase in patient age and cervical histology.
4. Overall there was no significant association between viral methyaltiuon and genome disruption; however there was a significant difference between methylation levels and gene disruption at three sites; CpG 3416, CpG 3435 and CpG 3447.
5. There was no significant association between HPV 16 *E2* and *E6* gene expression compared to cervical histology.

4.6.2 Strengths

This study was designed to obtain DNA and RNA which were of a suitable quality for analysis. In the research presented in Chapter 3 of this thesis RNA could not be obtained for analysis, these samples were stored in SurePath^{TM} which lead to nucleic acid degradation. To eliminate this problem samples in this study were stored in RNAprotect$^{®}$ (Raybould, 2013). RNAprotect$^{®}$ stabilises the RNA within the sample which can then be stored at room temperature during storage and transport before DNA and RNA extraction can be performed (QIAGEN, 2017c). This made it possible for HPV 16 *E2* and *E6* expression to be measured by qRT-PCR.

In this study samples were collected from women with different cervical cytology and subsequent cervical histology which produced a broad range of disease grades to be included in this study. Women included in this study had a wide age range so viral biomarkers studied in chapter three could be studied in a cross-section of women with different ages.

4.6.3 Limitations

This pilot study originally intended to collect 45 samples (15 samples with mild dyskaryosis, 15 with moderate dyskaryosis and 15 with severe dyskaryosis) however a low number of women with mild and moderate dyskaryosis were recruited onto this study (mild dyskaryosis – n = 11, moderate dyskaryosis – n = 5). The decision was taken to extend the study to include women with cervical cancer. The final cohort contained a small number of samples (n = 51) and consequently it was not possible to draw statistically significant conclusions from trends seen in the data.

The small sample size also resulted in the inclusion of samples with multiple HPV infections in the molecular analysis of HPV 16 integration, methylation and gene expression. As these samples were obtained via cervical smear test, multiple sites of HPV infection may be sampled in each smear test, only one of which is driving neoplastic progression. It therefore would be advantageous in the assessment of potential biomarkers to only assess samples with a single HPV infection in order that the infection driving disease progression can be identified. Samples with multiple infections were excluded from analysis in chapter 3 as it cannot be determined which HPV type is driving cervical disease progression in samples with multiple HPV infections. Similarly, there are several unknown variables such as comorbidities, use of hormonal contraceptives, previous pregnancies etc. This inclusion of multiple HPV infections and unknown comorbidities must be considered when interpreting the results presented in this chapter.

4.6.4 Samples with cancer accounted for most single infections and HPV 16 was the most prevalent single infection.

This pilot data showed a greater number of multiple HPV infections in samples with CIN I, CIN II and CIN III when compared to samples with cancer. In progression from CIN I to cancer, samples with multiple HPV infections may be less prevalent due to clonal selection and outgrowth of a single, dominant transformed cell with one HPV infection (Van Tine et al., 2004). This theory is in agreement with other studies which demonstrate that women with CIN III+ are less likely to have multiple HPV infections (Lau et al., 2015, Tjalma et al., 2015).

Recent work by Depuydt et al (2016) has described a method of determining the active HPV type driving disease in a sample with multiple HPV infections. This method used qRT-PCR to calculate which HPV-induced processes were ongoing. In samples with CIN I the results were similar regardless of single or multiple infections. However, in samples from CIN III+ a HPV type driving disease could clearly be identified. This model should be employed in further studies with multiple HPV infections to identify which HPV type is driving disease (Depuydt et al., 2016)

4.6.5 Viral disruption was present in 37.5% of samples; however there was no significant disruption between different histology.

In this study, HPV gene disruption was identified in 37.5% of samples with gene disruption in 50% of cancer cases (n = 7/14). This result was lower than expected with other studies reporting viral integration in around 70% of cervical cancer cases (Pett and Coleman, 2007, The Cancer Genome Atlas Research, 2017). Work reported elsewhere has shown a positive correlation between severity of histology and viral integration (Hudelist et al., 2004, Li et al., 2008, Manawapat et al., 2012). The lack of significant conclusions reported in this thesis may in part be due to the small sample size. Increasing the size of clinical cohort in future work may allow for significant associations to be made between disease grade and integration.

Interestingly in this work HPV *E1* gene was most commonly disrupted site in the viral genome and the *E2* gene was identified in all but two HPV 16 samples. As described earlier (Introduction section 1.3.1) integration of HPV is a random event which has been shown to occur at a higher frequency in or near common fragile sites (Thorland et al., 2003, Jiang et al., 2015). Three of the six fusion transcripts identified by DIPS, in samples HI-BCD-9, HI-BCD-46 and HI-BCD-50, were integrated into known common fragile sites, FRA6G, FRA3D and FRA9C respectively. Also, HI-BCD-46 was inserted into the host genome at locus 3q25.33 that encodes for Interleukin-12A (IL-12A), a subunit of Interleukin 12 that acts on natural killer and T

cells (NCBI, 2017). IL-12A has been considered as a potential marker of cervical and vulval cancer (Chang et al., 2015, Zhang et al., 2014).

The low level of gene disruption and integration reported here may have been caused by a mixed infection of episomal and integrated HPV, either within the same lesion or multiple lesions on the cervix. In the work by The Cancer Genome Atlas Research Group (2017) they assessed viral integration by RNA-Seq Whole Transcription Sequencing which can detect viral integration events in samples which have mixed infections of episomal and integrated HPV (The Cancer Genome Atlas Research, 2017).

The APOT assay or RNA-Seq could be utilised to overcome the limitations presented by the HPV 16 *E2* tiling PCR assay and identify actively transcribed HPV-human transcripts in samples with mixed HPV infections, however APOT requires high-quality RNA to provide suitable results (Klaes et al., 1999). Previous work by Raybould (2013) showed that due to RNA degradation APOT could not be carried out on these clinical samples. It was suggested that this may be due to natural degradation of RNA in cells at the surface of the cervix which are sampled during a cervical smear test (Raybould, 2013). APOT and RNA-Seq may be more suitable for application in RNA taken from biopsy tissue specifically preserved and extracted to ensure high quality and quantity of RNA can be obtained. Future research should consider carrying out RNA-Seq to determine if viral integration events are present even in a background of episomal RNA transcripts.

4.6.6 There is a trend observed with increased viral methylation and increased patient age and cervical histology and there was a significant difference between methylation levels and gene disruption at CpG sites CpG 3416, CpG 3435 and CpG 3447

This work has shown a trend of increased viral methylation with an increase in age of the patients. Changes in host methylation has been shown to occur as people age (Jung and Pfeifer, 2015) and this change in global and local methylation may in part contribute to the methylation levels observed here. Methylation has also been linked with other comorbidities including obesity and smoking which can increase the incidence of age-associated DNA methylation (Noreen et al., 2014). Information on potential co-morbidities was not available for samples in this cohort and these may be contributing to the increase in age-associated methylation seen in this study.

In this study, the highest levels of viral methylation were present in samples with cancer, compared to samples with no CIN, CIN II or CIN III. This agrees with the suggested association from Chapter 3 of this thesis and elsewhere (Mirabello et al., 2012) that viral methylation may be a suitable biomarker in an older population of women. Apart from HPV methylation, the methylation of human genes has also been studied as a potential biomarker for HPV disease progression. Methylation of

human genes *p16*, *MGMT*, *FHIT*, *CADM1* and *MAL* have all been suggested as potential biomarkers for HPV driven cervical disease (Bierkens et al., 2013, Virmani et al., 2001, van Baars et al., 2016)

This work has also demonstrated a significant difference between viral methylation levels with gene disruption at CpG 3416, CpG 3435 and CpG 3447. However overall there was no significant association between viral disruption and viral methylation when all CpG sites were considered. This result suggests that there is a potential clinical utility for methylation as a biomarker for HPV integration, although viral integration is not a prerequisite for disease presence and progression. Further study into the relationship between viral methylation and clinical outcome is required to assess methylation as a biomarker in BIDD assay development stage 2. This result partially supports Hypothesis 4; Disruption of viral E2 gene coincides with hypermethylation of viral DNA in HR-HPV 16 infections.

4.6.7 There was no significant association between HPV 16 E2 and E6 expression compared to cervical histology.

In this study, there was an increase in both HPV 16 *E2* and *E6* gene expression with an increase in cervical histology. These results did not reach significance, however that may in part be due to the small sample size (n = 14). A recent study by

Cerasuolo et al (2017) has shown that HPV 16 *E6* gene expression were consistently high in cervical cancer with low expression of *E2* in cancer and CIN (Cerasuolo et al., 2017). This study used only a single reference gene, *GAPDH*, to standardise the results of the qRT-PCR analysis. MIQE Guidelines suggest that normalization against a single reference gene is not an acceptable form of normalization and the optimal number and choice of reference gene must be reported (Bustin et al., 2009). In this thesis reference genes were decided based on work carried out by geometric averaging of multiple internal control genes for a given set of tissues, based on high stability and expression levels in epidermal keratinocytes *HPRT* and *TBP-2* were chosen as reference genes (Vandesompele et al., 2002, Allen et al., 2008).

Chapter 5

A model of HPV16 neoplasia - gene expression, viral genomic DNA methylation and integration in differentiated epithelium

A major aim of the study was to characterise molecular changes induced by HPV 16 neoplasia to identify useful biomarkers of disease progression. To facilitate this process, an *in vitro* system was sought to establish a model of *in vivo* pathology. To this end, a system to culture stratified epithelium (organotypic raft cultures) was established which could be seeded with HPV-transformed cells lines. This technology was used to assess the regulation of virus gene expression, cell proliferation and histology of HPV-transformed cells in stratified cultures. This work was carried out with to establish the association between histology of HPV-transformed cell lines and molecular markers of disease. Molecular profiles could then be applied as part of future anogenital screening to predict disease presence and progression.

5.1 Rational for cells and organotypic raft culture

Cell lines used in this study were derived at Cardiff University via ring cloning in heterogeneous populations from patients with high grade intraepithelial neoplasia. (Bryant, Onions, *et al.* 2014). There is a direct causal relationship between HPV 16

infection and VIN/Vaginal Intraepithelial Neoplasia (VaIN), the precursor to vulval/vaginal cancer. Cell lines CU-VU-8M-16I and CU-VU-8P-16E originated from a biopsy taken from a 46-year-old women undergoing her first treatment for VIN grade III, while cell lines CU-VA-9A-16E and CU-VA-9H-16I came from a 31-year-old woman with VaIN III (Onions 2013). As a positive control cell line W12 Ser4B was used to represent a high-grade cervical disease, this is a well-defined cell line with integrated HPV 16 genome (Doorbar *et al.* 1990). Human epithelial keratinocytes, neonate (HEKn) primary cells were used to represent a HPV negative culture. Information on the origin of cell lines used and culture conditions can be found in Methods section 2.6.

Organotypic raft culture of epithelial cells provides an efficient system that recapitulated the full differentiation of primary human keratinocytes, thereby supporting the complete HPV lifecycle through to the production of virus particles. (Andrei *et al.* 2010; Delvenne *et al.* 2001). Differentiation is the process by which cells become more specialised and in a 'normal' epithelium the cells in the upper layers become elongated and unable to carry out mitosis. They provide a barrier for the differentiating cells in the basal layer and are slowly lost over time as they are replaced with new terminally differentiated cells. However in cases where differentiation is lost the cells continue to mitosis in the upper layers of the epithelium, and do not become elongated (Thiery et al., 2009).

In this work, the organotypic raft culture was used to produce a stratified epithelium to emulate the natural process of differentiation. Histology was defined as "in keeping with low-grade intraepithelial neoplasia" or "in keeping with high-grade intraepithelial neoplasia" by an independent pathologist from UHW, Cardiff, Dr Owen Weeks.

When seeded on to raft culture w12 Ser4B cells had little differentiation and show features such as keratin pearls, irregularities in the cell nucleus and apoptotic debris. HEKn cells showed differentiation through the stratified epithelium and cells detaching from the top of the epithelium as would be expected in a non-transformed epithelium. These phenotypes match the anticipated outcome with w12 Ser4B showing histology consistent with high-grade intraepithelial neoplasia and HEKn cells displaying a differentiated phenotype as would be expected in a non-immortalised cell line.

Cell lines CU-VA-9A-16E and CU-VU-8M-16I both exhibited abnormal morphology like that of a low-grade intraepithelial neoplasia (Figure 5.1). CU-VA-9A-16E (Figure 5.2 (A)) shows minimal differentiation and significant desquamation of the cells. Numerous apoptotic figures are also present along with prominent formation of keratin pearls. CU-VU-8M-16E has formed a prominent band of parakeritosis on the surface, suggesting this model is more like that of a condyloma (Figure 5.1 (D)). Parakeritosis on the surface is normally present in true mucosal membranes such as the vagina and the oral cavity (Ferey *et al.*1985). The presence

170

of the parakeritosis in cells derived from the vulva is unusual and would be expected in cells such as CU-VA-9A-16E or CU-VA-9H-16I that originated in the vagina.

Cell lines CU-VA-9H-16I and CU-VU-8P-16E showed morphology in keeping with high-grade intraepithelial neoplasia. CU-VA-9H-16I shows a thick epithelium, which lacks maturation, however there is a slight flattening of cells on the surface of the raft (Figure 5.1 (B)). The cells are pleomorphic, including multinucleate cells and numerous apoptotic bodies. An occasional cell in mitosis was detected high up the epithelium. This morphology is very like that of the positive control W12 Ser4B. CU-VU-8P-16E lacks differentiation and has similar morphology to CU-VA-9H-16I. There are numerous apoptotic bodies and pleomorphic cells with morphology consistent with a high-grade lesion (Figure 5.1 (C)).
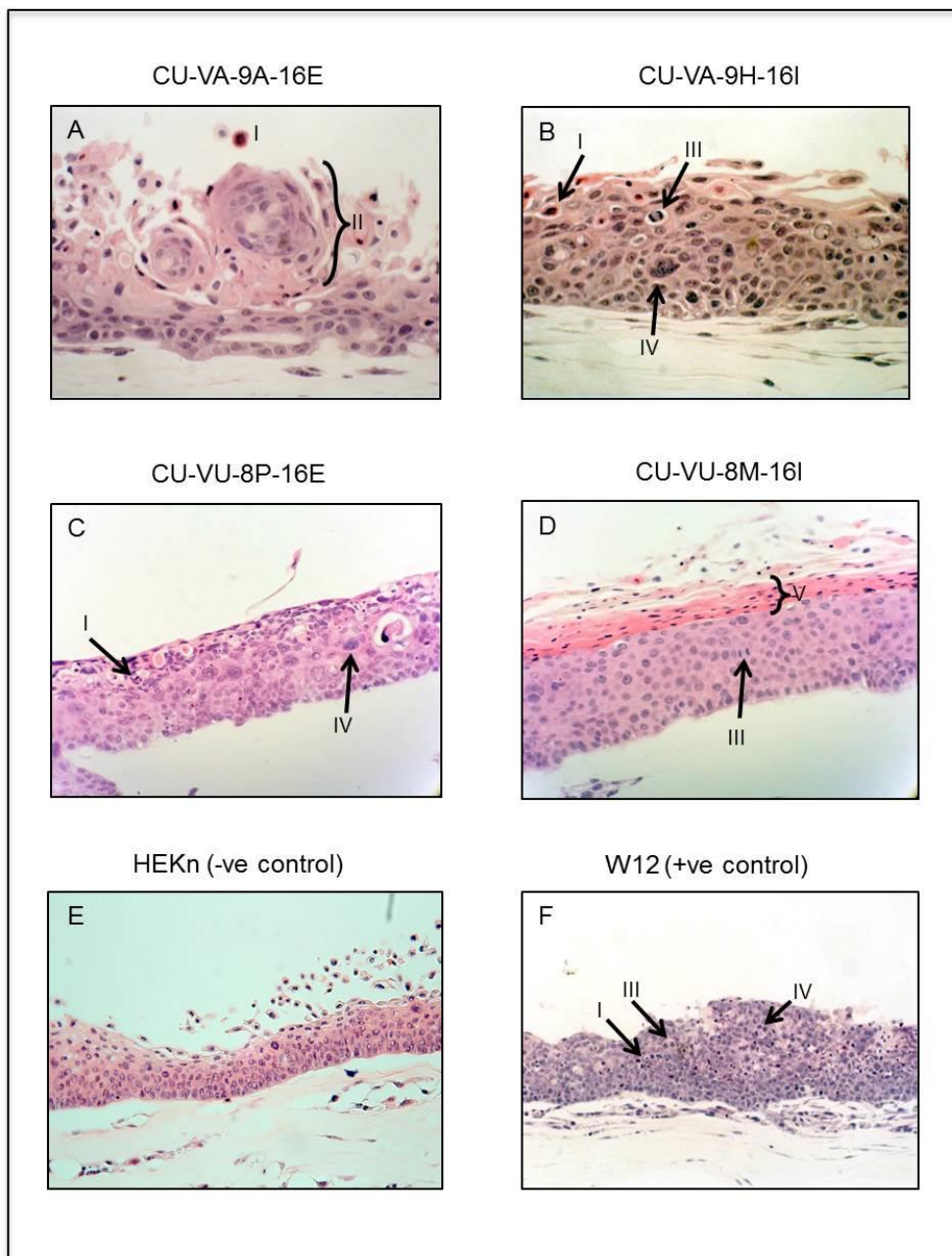
Figure 5.1 H&E staining of raft sections (A) CU-VA-9A-16E **(B)** CU-VA-9H-16I **(C)** CU-VU-8P-16E **(D)** CU-VU-8M-16I **(E)** W12 Ser4B **(F)** HEKn. Features: apoptotic bodies (I); keratin pearls (II); multinucleated cells (IV); mitosis in the high epithelium (III); irregularities in the cell nucleus (IV); parakeritosis on the surface (V); apoptotic debris (VI). Images and pathology assessment carried out by independent pathologist Dr Owen Weeks.

This is the first time that HPV 16 positive cell lines derived from VIN and VaIN lesions have been grown in an organotypic raft culture. HPV genome status (episomal or integrated) defined in monolayer did not predict morphology of cells when grown in a stratified epithelium. Cell lines that originated from a single patient sample (CU-VA-9A-16E / CU-VA-9H-16I and CU-VU-8M-16I / CU-VU-8P-16E) both showed markedly different phenotypes. This work has established a suitable system for further analysis for the regulation of HPV and human genes to establish if cell lines with morphology in-keeping with a low-grade lesion have a distinct molecular profile compared to cell lines with morphology in-keeping with a high-grade lesion.

5.2 Gene expression in organotypic raft cultured cells

The HPV genome encodes eight proteins that are differentially expressed at various stages of the viral replication cycle (Introduction section 1.1). In a productive infection, the expression of viral proteins is tightly regulated in parallel with the host cell differentiation to enable efficient release of high titre virus. However, in a transforming viral infection, the natural progression of viral transcriptional programme is disrupted, viral capsid proteins are not synthesised and the expression of viral oncogenes enhanced (Graham, 2010). HPV can also lead to changes in host cellular proteins involved in tumour suppression and cellular proliferation, leading to a

productive infection developing into a transforming infection. Changes in the host cell cycle and host cell proliferation can be used as surrogate markers for HPV infection and transformation (Litjens *et al.* 2013).

In this study, expression of human proteins p16 and Ki-67 were used as a substitute marker of HPV gene expression. Further it has been suggested that the *Hedgehog* genes are inappropriately expressed during HPV infection (Samarzija and Beard, 2012). Therefore SHH and GLI1 expression was investigated to assess if expression was altered in HPV 16 positive VIN and VaIN cells when cultured in stratified epithelium. The viral lifecycle was also assessed using the viral E4 protein as a marker of late gene expression to determine if the whole viral lifecycle is being completed. Histology was quantified using ImageJ, IHC Profiler and by measuring the proportion of the culture stained (Methods section 2.8.3)

Images of IHC analysis for p16 and Ki-67 expression in organotypic raft cultures is shown in Figures 5.2 and 5.3. IHC results were quantified by grading the positive pixels in the selected raft cultures as a marker of the amount of stain the cells have taken; data is shown in Table 5.1 and Figure 5.4. IHC results were also quantified by measuring the depth of the culture that has taken in the IHC stains and the results of this are shown in Figure 5.5.

Host tumour suppressor protein p16 is instrumental in regulating the cell cycle and is up regulated by HPV *E7* protein. In 'normal' epithelium, p16 expression is high only in the basal layer where cells are actively cycling. However, with increasing epithelial neoplasia, p16 expression expands to the upper layers of the epithelium. The positive control, w12 Ser4B had high p16 expression (86.1% high positive/positive pixels (HP/P pixels)) throughout the culture typical of a high-grade epithelial neoplasia (Figure 5.2 (A)). The negative control, HEKn had p16 expressed in under half of the cultures and HP/P pixels were 65.3% lower than those expressed in w12 Ser4B epithelium (Figure 5.2 (C)).

CU-VA-9H-16I and CU-VU-8P-16E both show similar p16 expression levels to w12 Ser4B, with high levels of HP/P pixels throughout the culture (69.8% and 71.2% respectively). CU-VU-M8-16I has p16 expressed in the lower 63% of the epithelium with 32.2% HP/P pixels but no p16 expression in the parakeratin layer on the surface of the culture. CU-VA-9A-16E has the lowest expression of p16 and is only expressed at low levels (7.3% HP/P pixels) in the lower quarter of the epithelium (22%) (Figure 5.3)

Ki-67 is a marker of cell proliferation, in 'normal' epithelial cultures Ki-67 should be expressed in the basal layer of the epithelium as these cells are actively proliferating. Positive control w12 Ser4B has Ki-67 expression in the lower 87% of the culture (16.3% HP/P pixels) indicating that cellular proliferation is occurring beyond the basal layer of the epithelium (Figure 5.2 (B)). The negative control, HEKn

had low levels of Ki-67 expression (11.8% HP/P pixels) confined to basal layer of the culture in keeping with normal epithelial cultures (lower 34% of the culture) (Figure 5.2 (D)).

Raft cultures CU-VA-9H-16I, CU-VU-M8-16I and CU-VU-8P-16E all showed similar levels of Ki-67 when the percentage spread of stain in the culture was measured (76%, 69% and 60% respectively) however when the number of positive pixels is compared only CU-VA-9H-16I has a level of HP/P pixels comparative to w12 Ser4B. CU-VU-M8-16I and CU-VU-8P-16E both have a similar percentage of HP/P pixels, measured at 4.9% and 6.8% respectively. Culture CU-VA-9A-16E has the lowest level of Ki-67 expression compared to all other cultures including negative control HEKn, (25% of culture, 1.2% HP/P pixels) (Figure 5.3). Full depth p16 staining is present in CU-VA-9H-16I, CU-VU-8P-16E and w12 Ser4B with histology "in keeping with high-grade morphology". Cultures with histology 'in keeping with low-grade morphology' CU-VA-9A-16E, CU-VU-M8-16I and HEKn had lower p16 expression than samples with histology "in keeping with high-grade morphology" (22%, 63% and 47% respectively). There was a significant difference in both p16 ($X^2(1)$ = 4.511, p = 0.034) and Ki-67 ($X^2(1)$ = 12.816, p >0.001) stain depth when cell lines with histology "in keeping with high-grade morphology" and cell lines with histology "in keeping with high-grade morphology" were compared. CU-VA-9A-16E, which originated from the same biopsy as CU-VA-9H-16I, had the lowest level of p16 and Ki-67 staining of all samples tested.
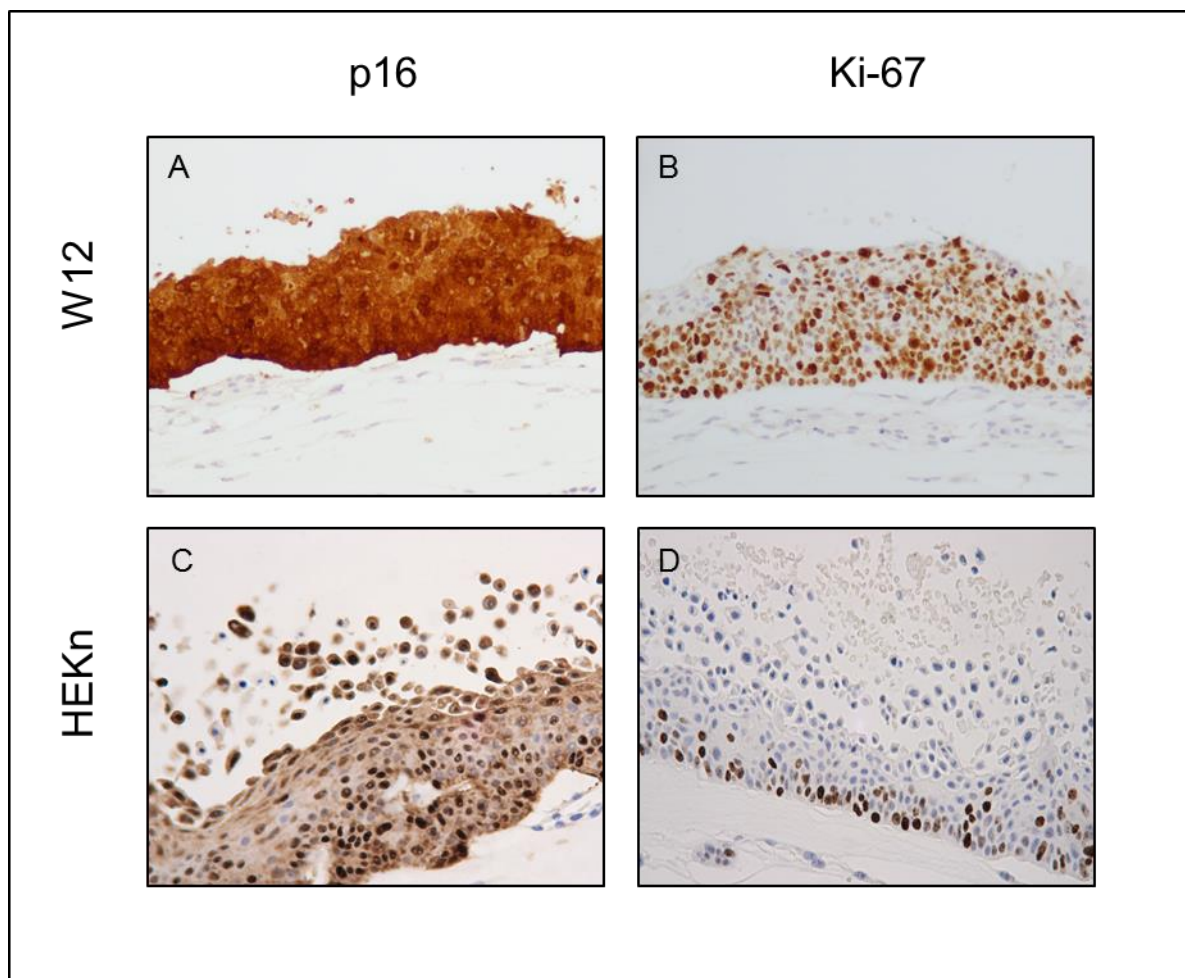
Figure 5.2 p16 and Ki-67 IHC analysis of W12 Ser4B (positive control) and HEKn (negative control) raft cultures. W12 Ser4B has a high level of full thickness p16 and Ki-67 expression. HEKn has p16 expression throughout the culture, although Ki-67 expression is restricted to the basal layer of the culture.
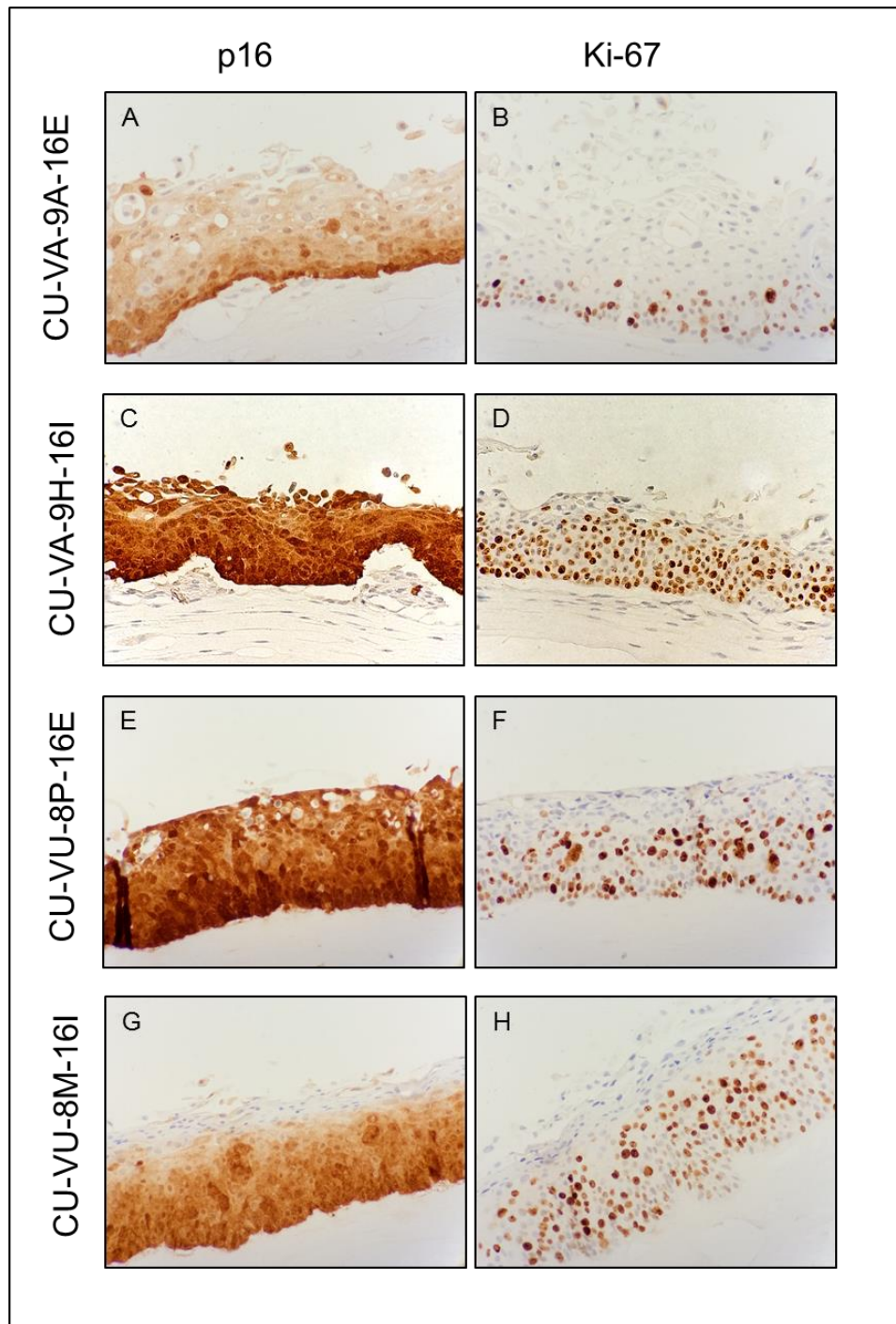
Figure 5.3 p16 and Ki-67 IHC analysis of vulval and vaginal organotypic raft cultures. Lowest expression is seen in cultures CU-VA-9A-16E and both cultures CU-VA-9H-16I and CU-VU-8P-16E have high expression of p16 and Ki-67. All cultures have lower expression of Ki-67 compared to p16.

Table 5.2 p16 and Ki-67 IHC quantification of pixel staining using ImageJ. Each organotypic raft culture section was scored three times and a percentage of pixels with each result reported with the 95% confidence interval.

Cell Lines	p16 IHC Profile Percentage (95% Confidence Interval)				Ki-67 IHC Profile Percentage (95% Confidence Interval)			
	High Positive	Positive	Low Positive	Negative	High Positive	Positive	Low Positive	Negative
CU-VA-9A-16E	0.01 (±0.01)	7.35 (±0.60)	35.07 (±0.10)	57.57 (±0.50)	0.16 (±0.02)	0.99 (±0.20)	5.08 (±0.84)	93.77 (±1.06)
CU-VA-9H-16I	16.17 (±1.51)	53.65 (±3.98)	18.61 (±0.71)	11.56 (±5.94)	4.18 (±2.23)	10.04 (±4.29)	23.49 (±3.47)	62.29 (±9.92)
CU-VU-8M-16I	0.31 (±0.01)	31.94 (±0.98)	35.55 (±1.05)	32.21 (±2.04)	0.38 (±0.03)	4.51 (±0.31)	14.25 (±0.84)	80.86 (±1.18)
CU-VU-8P-16E	20.78 (±0.91)	50.40 (±2.05)	19.60 (±0.61)	9.21 (±3.53)	1.35 (±0.09)	5.49 (±0.35)	15.66 (±0.96)	77.50 (±1.39)
W12 Ser4B	19.37 (±1.32)	66.76 (±4.53)	5.04 (±0.12)	8.83 (±5.90)	1.23 (±0.07)	15.06 (±0.76)	35.59 (±1.65)	48.12 (±2.48)
HEKn	10.94 (±0.98)	18.60 (±1.05)	31.80 (±3.02)	37.22 (±2.62)	2.56 (±0.98)	9.27 (±4.33)	21.22 (±4.64)	65.27 (±8.12)
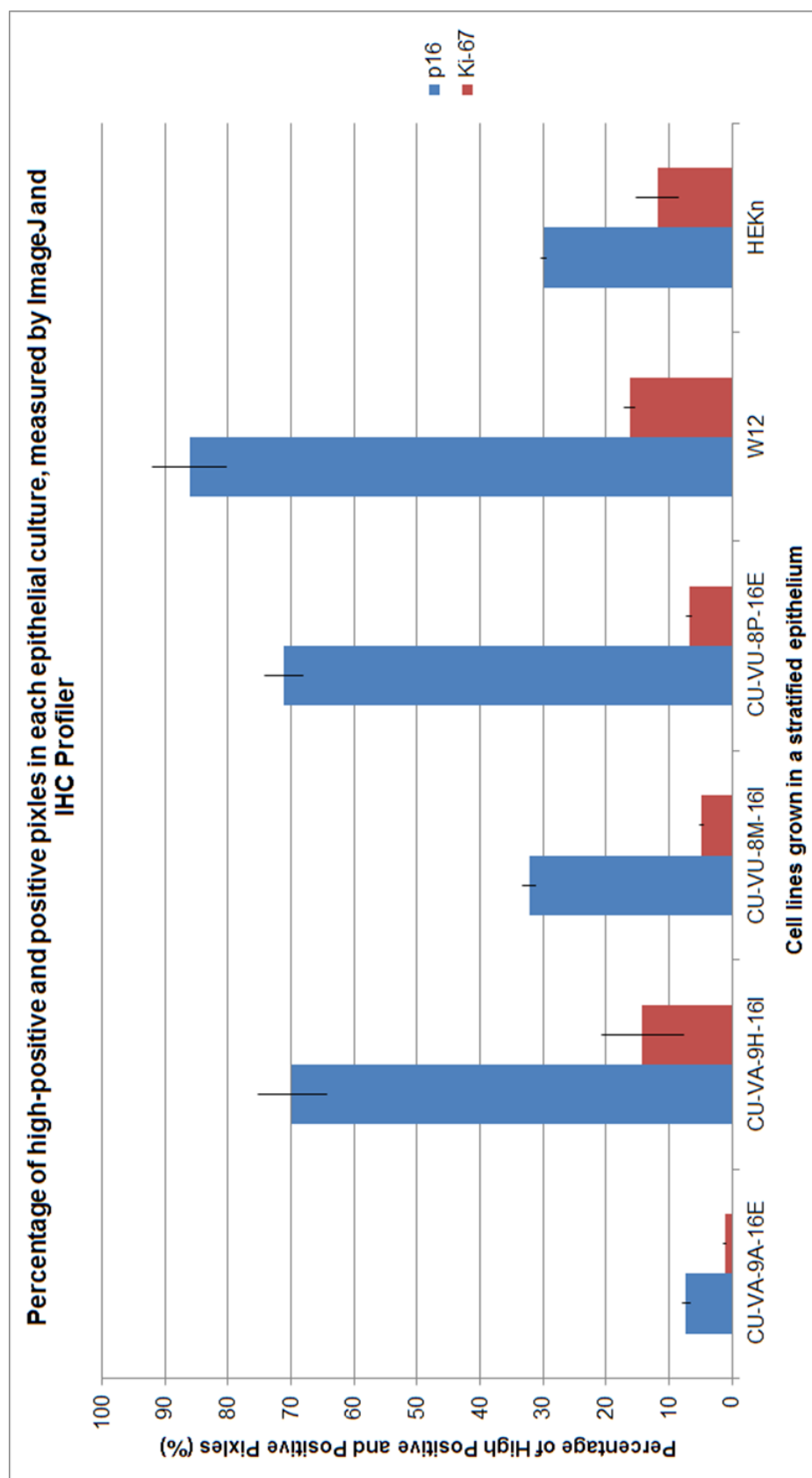
Figure 5.4 Percentage of HP/P pixels in each of the epithelial cultures as measured by ImageJ and IHC Profiler. p16 results are shown in blue and Ki-67 results are shown in red.

180

	Percentage of epithelium stained with p16 antibody	Percentage of epithelium stained with Ki-67 antibody
CU-VA-9A-16E	22 (±3.64)	25 (±15.22)
CU-VA-9H-16I	100 (±0)	76 (±1.96)
CU-VU-8M-16I	100 (±0)	60 (±4.28)
CU-VU-8P-16E	63 (±3.65)	69 (±2.61)
W12 Ser4B	100 (±0)	87 (±24.83)
HEKn	47 (±5.35)	34 (±14.24)
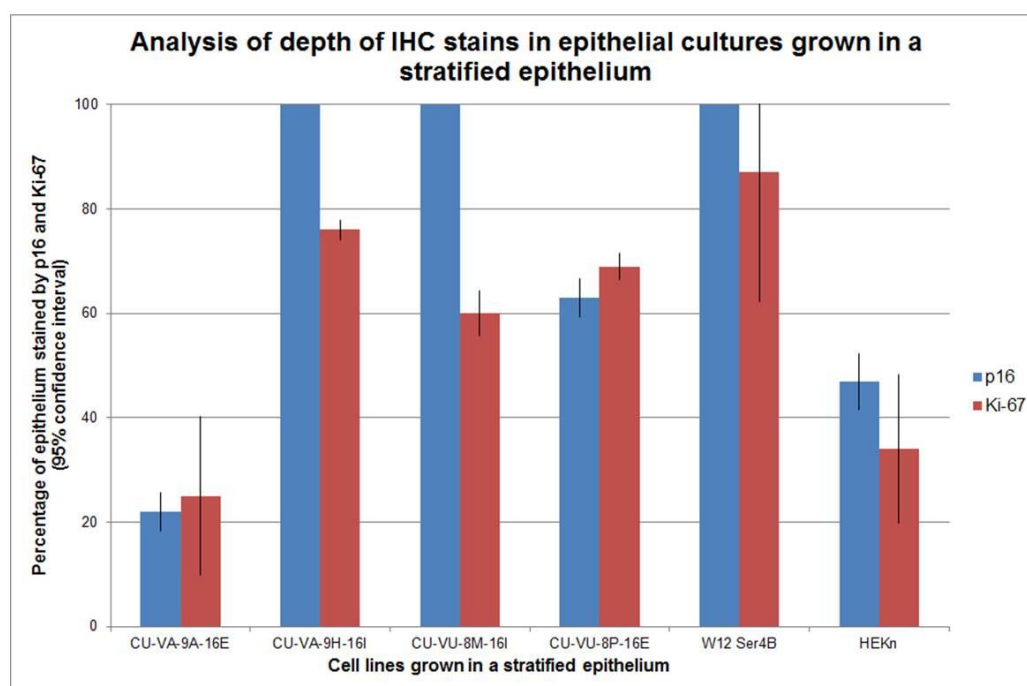
Figure 5.5 Analysis of the depth of stain in each epithelial culture. The total depth of the cultures and the depth of the stained culture were measured in triplicate and the percentage depth of stain was calculated. There was a significant difference in both p16 ($X^2(1) = 4.511$, p = 0.034). and Ki-67 ($X^2(1)= 12.816$, p > 0.001). stain depth when cell lines with histology "in keeping with high-grade morphology" and cell lines with histology "in keeping with high-grade morphology" were compared.

5.2.1 SHH Gene Expression

Cell lines in this study were stained for SHH (Figure 5.6 and 5.8) and GLI1 expression (Figures 5.7 and 5.8). Epithelial cultures derived from human foreskin keratinocytes, BU21HFK and BU21HFK (HPV18), were used as the negative and positive controls respectively. BU21HFK (HPV 18) had the highest SHH expression (176.8%) and BU21HFK has low SHH expression (19.63%) (Figure 5.9). Cell lines CU-VA-9H-16I and CU-VU-M8-16I both had similarly high levels of SHH expression (142.9% and 119.5% respectively). The lowest level of SHH expression was present in cell line CU-VA-9A-16E (19.63%), cultures with morphology in-keeping with high-grade disease. CU-VA-9H-16I and CU-VU-P8-16E both have high levels of p16 (100% and 100%) and Ki-67 (76% and 60%), however culture CU-VU-M8-16I which shows morphology in-keeping with low-grade disease also showed similarly high levels of SHH. There was a significant variation in SHH expression when cell lines with histology "in keeping with high-grade morphology" and cell lines with histology "in keeping with high-grade morphology" were compared (X^2(3) = 13.205, p = 0.004).

GLI1 expression is lower than SHH expression in all cultures and increased expression is not associated with increasing neoplastic morphology (X^2(3) = 0.105, p = 0.991) (Figure 5.9). Cells that have originated from the same clinical samples

appear to have similar GLI1 expression regardless of differing morphology (vaginal sample: CU-VA-9A-16E: 42.88%, CU-VA-9H-16I: 37.17%, Vulval sample: CU-VU-M8-16I: 62.64%, CU-VU-P8-16E: 72.17%). There results suggest that changes to SHH expression may be modulated by HPV, but GLI1 is not. The raft culture we have employed here will facilitate the further study of SHH and GLI1 expression in HPV infected cells.
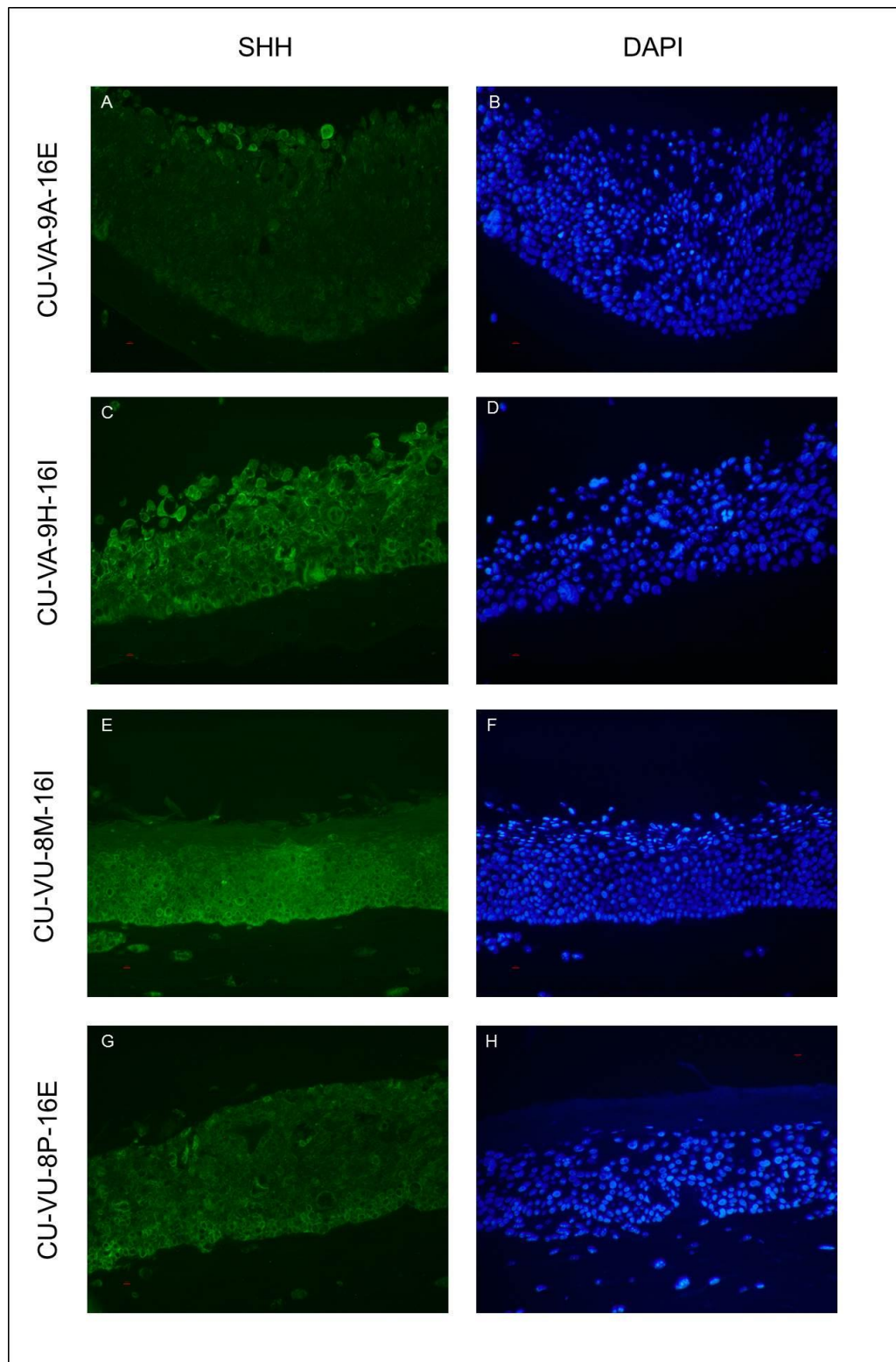
Figure 5.6 VIN and VaIN epithelial culture immunofluorescent staining for SHH protein. Cultures CU-VA-9H-16I, CU-VU-8P-16E and CU-VU-8M-16I all show high levels of SHH expression throughout the culture (142.92%, 143.19% and 119.49% respectively) DAPI is shown as the control staining.
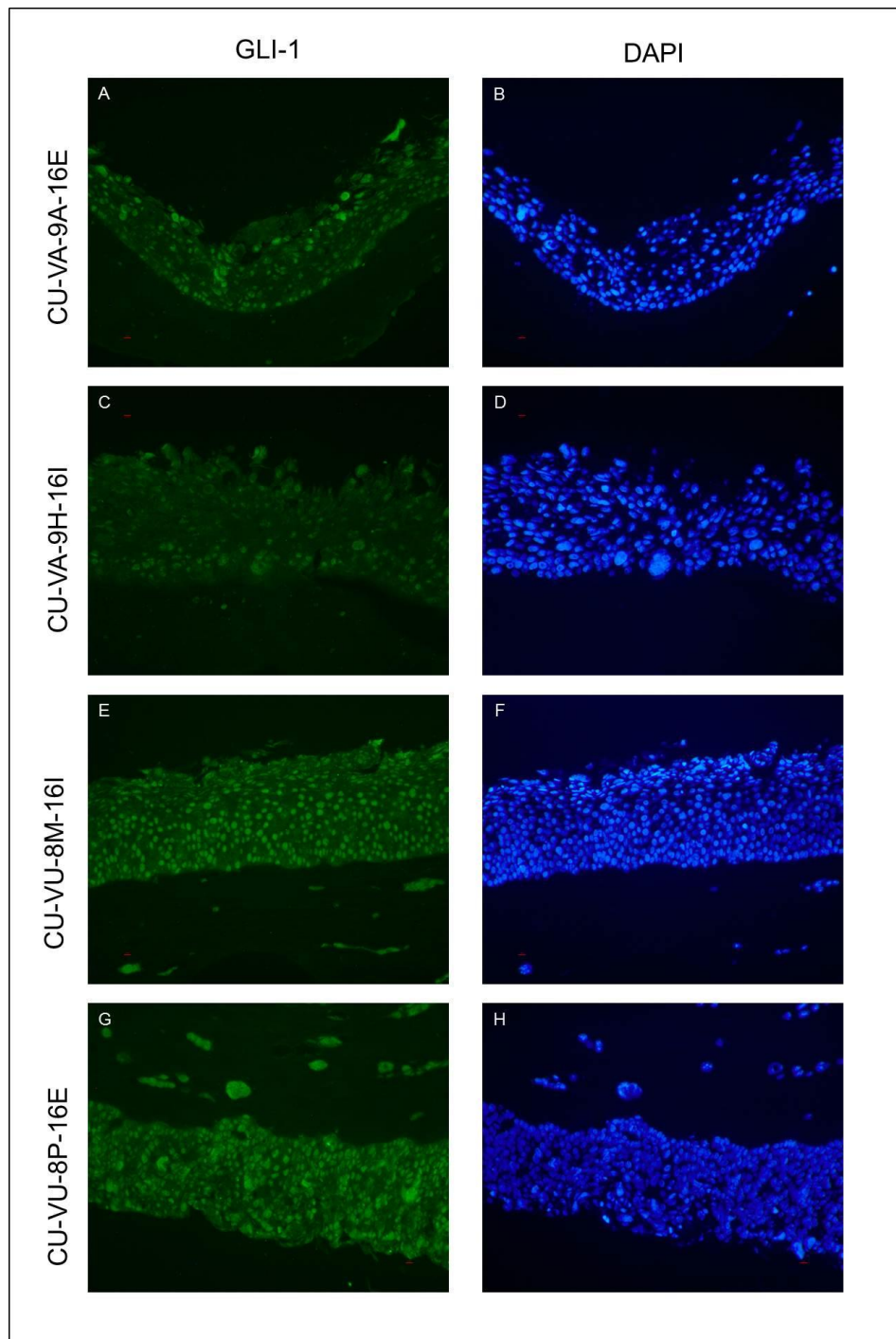
Figure 5.7 Immunofluorescent staining for GLI1 protein in VIN and VaIN epithelial cultures. CU-VU-8P-16E has the highest GLI1 expression is present in (72.17%) however there is less variation in expression compared to SHH expression. DAPI is shown as the control staining.
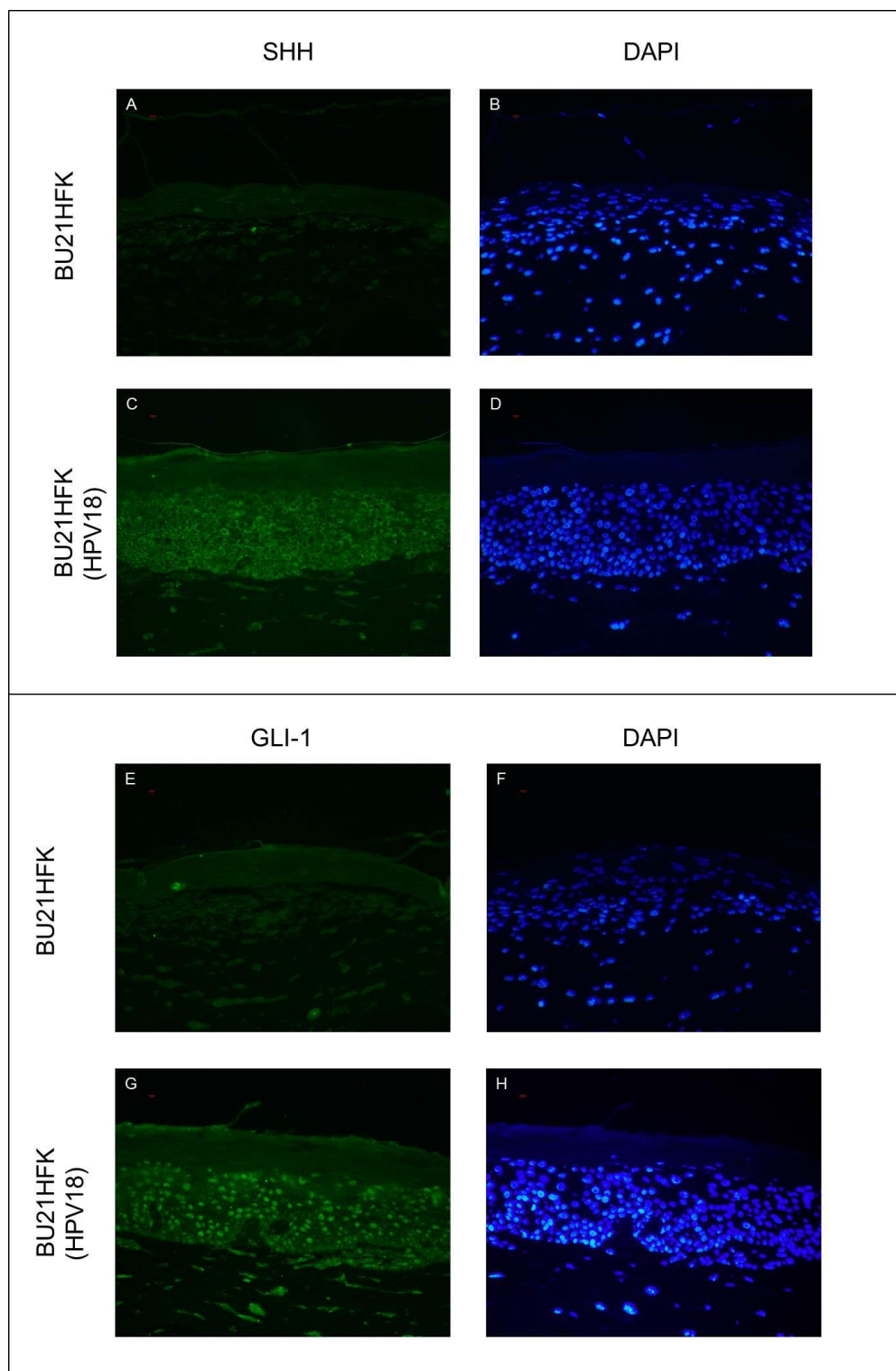
Figure 5.8 Immunofluorescent staining for GLI1 protein in epithelial cultures BU21HFK and BU21HFK (HPV 18).

A

Cell Lines	Percentage of section stained for GL1-1 vs DAPI	Percentage of section stained for SHh vs DAPI
CU-VA-9A-16E	42.88	19.63
CU-VA-9H-16I	37.17	142.92
CU-VU- 8M-16I	62.64	119.49
CU-VU- 8P-16E	72.17	143.19
BU21HFK	47.76	31.39
BU21HFK (HPV18)	53.53	176.78

B
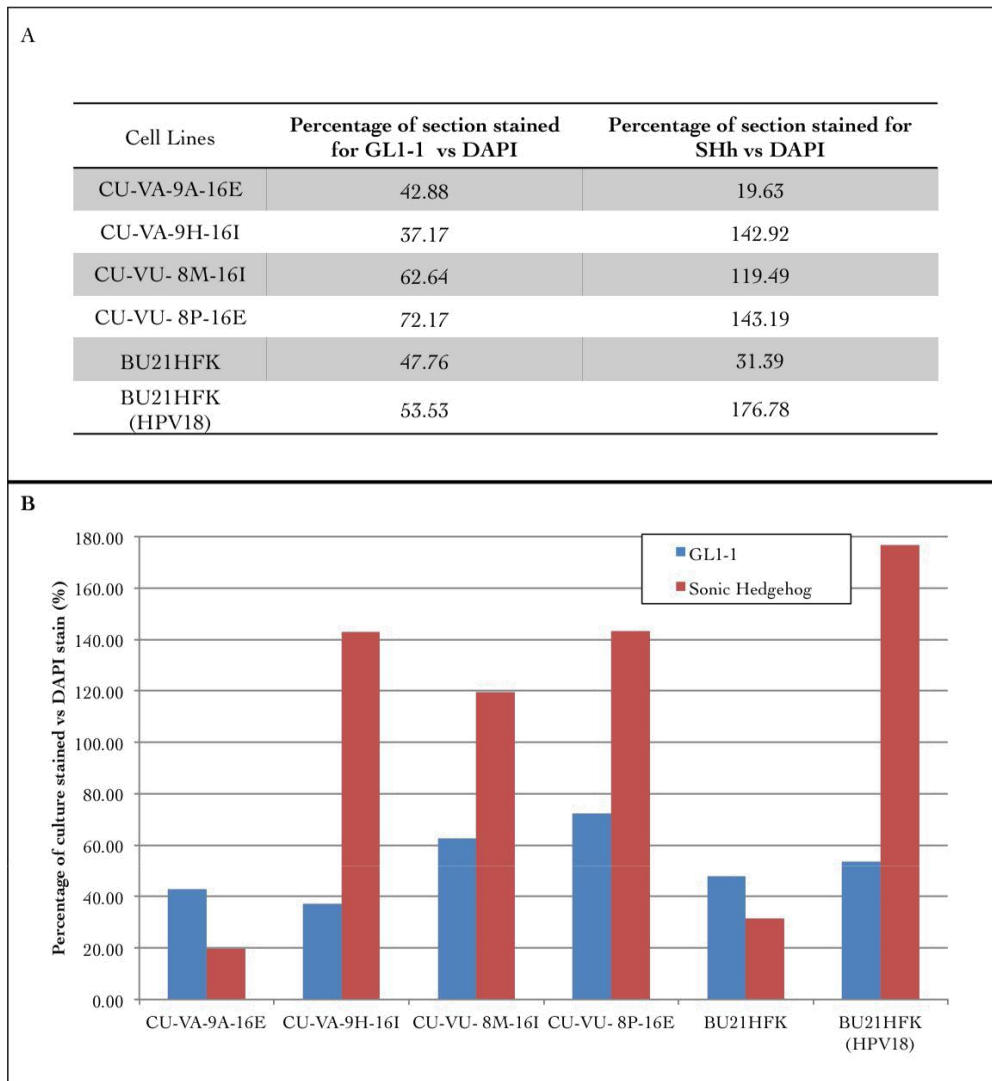
Figure 5.9 Analysis of SHH and GLI1 expression compared to the expression of DAPI staining within the cultures. There was a significant difference in SHH gene expression ($X^2(3) = 13.205$, p = 0.004), but not with GLI1 expression ($X^2(3) = 0.105$, p = 0.991) when cell lines with histology "in keeping with high-grade morphology" and cell lines with histology "in keeping with high-grade morphology" were compared.

5.2.2 HPV late gene expression

E4 protein is expressed late in the virus lifecycle and facilitates the release of newly formed virus capsids from upper layers of the epithelium [Introduction section 1.1] (Doorbar *et al.* 1997). Detection of the E4 protein in the upper epithelium is a substitute marker for late gene expression and indicative of a complete virus lifecycle. E4 expression was only detected in sample CU-VA-9A-16E while all other samples tested were negative. The E4 expression in CU-VA-9A-16E is in the mid to upper suprabasal layer and is primarily cytoplasmic (Figure 5.10). CU-VA-9A-16E had morphology 'in keeping with low-grade disease' and had the lowest levels of p16, Ki-67 and SHH when compared to other cultures.
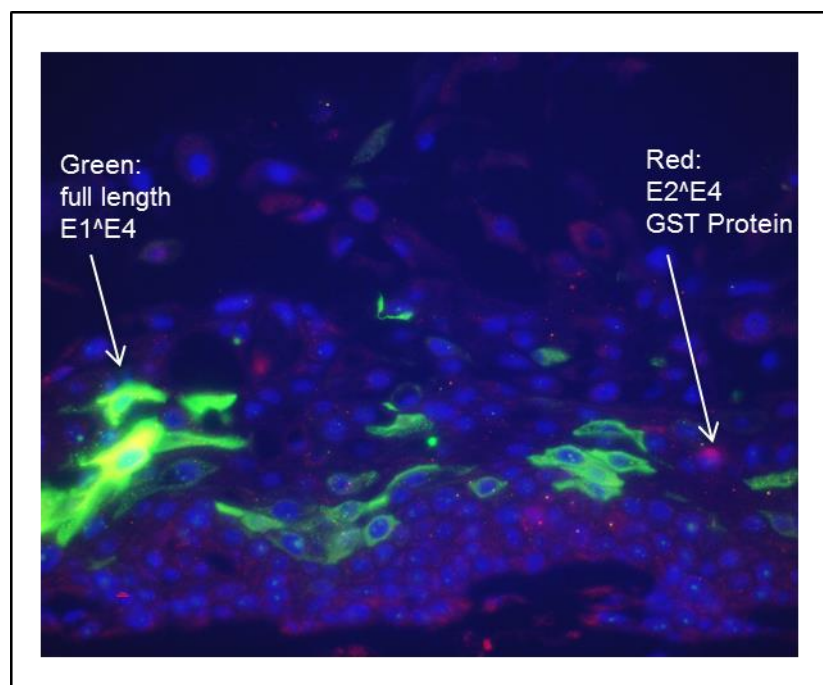
Figure 5.10 Example of E4 staining in CU-VA-9A-16E.

5.3 **HPV Genomic DNA Methylation**

Work in Chapter 4 showed that HPV hypermethylation is common in HR-HPV 16 infections and the highest level of hypermethylation was present in women who went on to be diagnosed with CIN/ II+. This increase in hypermethylation was also associated with samples which had a disrupted HPV genome. Cell lines used in the organotypic raft culture had varying HPV genome status (integrated or episomal) and this work aimed to evaluate if cultures with viral gene disruption had a higher level of hypermethylation of the viral DNA.

HPV genomic DNA methylation analysis show cultures CU-VA-9H-16I and CU-VU-8M-16I have hypermethylation in the *E2* region of the HPV genome (49.8% and 78.05% average methylation respectively). In previous analyses both of these cell lines have reported viral integration into the human genome (Bryant et al., 2014). Samples with previously defined episomal HPV infections, CU-VA-9A-16E and CU-VU-8P-16E, had low level of viral genomic DNA methylation (3.0% and 12.7% average methylation respectively). This result was not significant ($X^2(3)$ = -1.209, p = 0.313) however this may, in part, be due to the small sample size. Figure 6.8 show the average level of methylation in eight CpGs sites of the *E2* gene of HPV16 positive rafts.

	CpG 3411	CpG 3414	CpG 3416	CpG 3432	CpG 3435	CpG 3447	CpG 3461	CpG 3472
CU-VA-9A-16E	3.75	0.00	1.75	3.88	0.00	3.20	7.54	4.04
CU-VA-9H-16I	70.59	50.70	33.48	68.98	50.84	52.62	36.19	35.15
CU-VU-8P-16E*	21.45	0.00	7.63	21.56	16.70	8.69		
CU-VU-8M-16I	92.74	59.66	59.76	96.13	85.31	84.56	78.97	67.26
W12 Ser4B	4.39	0.00	1.08	2.30	0.87	0.99	0.00	1.26
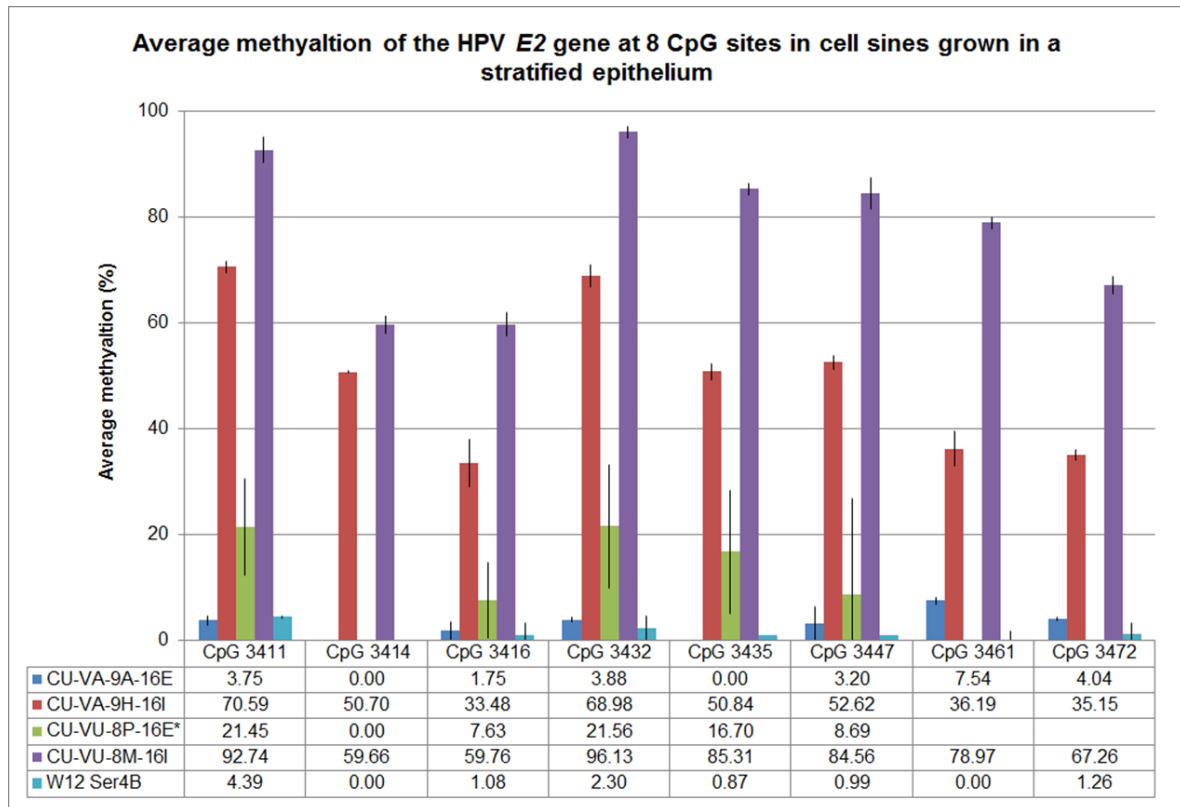
Figure 5.11 Average *E2* viral genomic DNA methylation at 8 CpG sites for cell lines grown in a stratified epithelium. There was no significant variation in viral methylation levels. when cell lines with histology "in keeping with high-grade morphology" and cell lines with histology "in keeping with high-grade morphology" were compared ($X^2(1) = 1.582$, p = 0.208).

5.4   HPV Whole transcriptome sequencing and viral integration

HPV whole transcriptome sequencing evaluates all the RNA present in the cells at a single point in time, allowing for the evaluation of HPV and human gene

expression. Whole transcriptome has previously been carried out on cell lines CU-VU-8M-16I and CU-VU-8P-16E. These cell lines were grown in monolayer and sequenced using SOLiD RNA (Bryant et al., 2014). Here, whole transcriptome sequencing was employed to validate HPV gene expression and to confirm actively transcribed viral integration events in the epithelial culture.

5.4.1 Quality of samples

Prior to whole transcriptome analysis each sample was tested in duplicate for RNA quality and quantity. Quantity was assessed using Thermo Scientific Nanodrop® 1000 Spectrophotometer and RNA integrity was assessed using Agilent 2100 Bioanalyzer (carried out by Central Biotechnology Services, Cardiff University). Samples CU-VA-9A-16E, CU-VA-9H-16I and CU-VU-8P-16E achieved a suitable RNA integrity number (RIN) and quantity of RNA for whole transcriptome analysis. However, sample CU-VU-8M-16I was excluded from further analysis as the RNA was at a level that could not be detected by either analysis (Table 5.2). To further improve results for samples CU-VA-9A-16E, CU-VA-9H-16I and CU-VU-8P-16E the RNA was concentrated using a RNA Clean and Concentrator-5 (Zymo Research Corporation, California, USA).

Samples CU-VA-9A-16E, CU-VA-9H-16I and CU-VU-8P-16E underwent transcriptome sequencing and RNA libraries were constructed by Dr P Giles and Dr K Ashelford, Cardiff University. The percentage of reads that mapped successfully to the target DNA was low for all samples but sample CU-VU-8P-16E produced a very low percentage of mapped reads, 22.6%. Therefore only results for HPV gene expression and viral integration are presented for organotypic raft cultures CU-VA-9A-16E and CU-VA-9H-16I (61.2% and 60.9% mapping percentage respectively).

Table 5.2 Quality and quantity of RNA used for whole transcriptome analysis Concentration and RNA integrity of VIN and VaIN organotypic raft cultures

VIN/VaIN Cell Lines	RNA Concentration		RIN	
	Sample 1	Sample 2	Sample 1	Sample 2
CU-VA-9A-16E	11	26	8.9	9.2
CU-VA-9H-16I	26	22	9.5	9.3
CU-VU-8M-16I	0	0	NA	NA
CU-VU-8P-16E	16	20	9.3	9.6

5.4.2 Validation of gene expression

Whole transcriptome sequencing was used to validate the HPV expression levels in the organotypic raft culture. HPV gene expression between samples CU-VA-9A-16E and CU-VA-9H-16I were assessed and transcript reads were mapped to the HPV 16 genome (NCBI, accession no: NC_001526.2). In IGV, the reference HPV genome was used to define transcript locations; transcript reads that map to the location on the HPV 16 genome are represented by gray lines (Figure 5.12). This data confirms findings in section 5.2.2 which showed that cell line CU-VA-9A-16E was the only cell line which expressed HPV late genes. In the gene expression analysis, we can see cell line CU-VA-9A-16E (Figure 5.12 (A)) produces transcripts for all viral genes including late genes *E4*, *L1* and *L2*. RNA expression of culture CU-VA-9H-16I which has high grade morphology, a high level of *E6*/*E7* expression but low level late HPV genes.
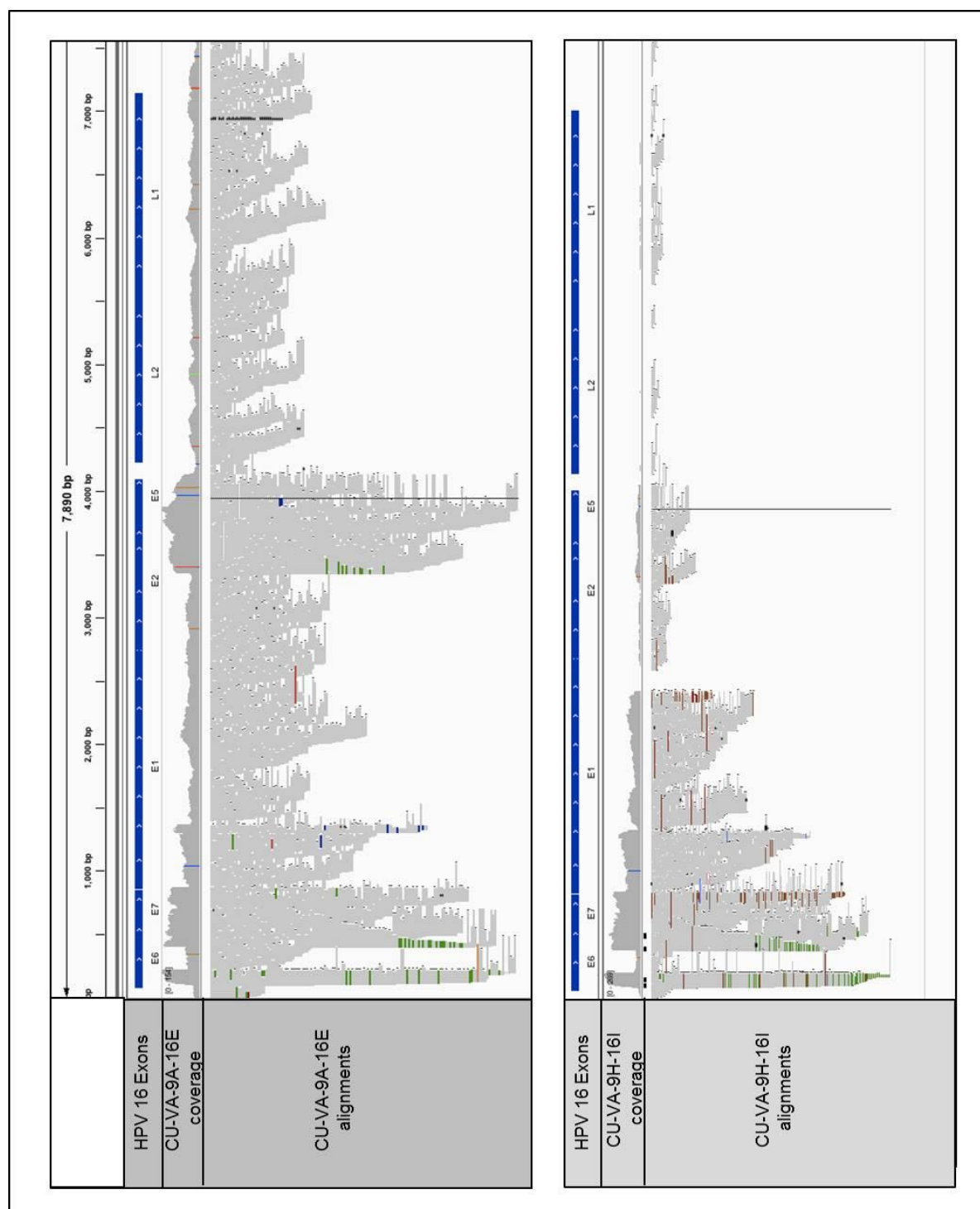
Figure 5.12 Viral gene expression of VaIN organotypic raft cultures visualised in IGV (A) CU-VA-9A-16E and (B) CU-VA-9H-16I.

5.4.3 Validation of viral integration

Viral integration events were validated by the whole transcriptome sequencing and actively transcribed human-viral integration event were visualised using circos plots (Krzywinski *et al.*, 2009). Grey lines represent paired end reads where one end maps to the HPV genome and the other end links to the location that it is associated with in the human genome. Previous work has shown cell lines CU-VA-9A-16E had no identifiable viral integration transcripts either by DIPS or APOT and the E2 gene is intact, cell lines CU-VA-9H-16I has integration defined by DIPS into 11p15.3 and 22q12.3 and by APOT integration into 5q11.2 and 1p36.13 (Onions, 2013). Circos plot of CU-VA-9A-16E shows HPV/human fusions transcripts into 5 different human chromosomes, however these lines are all light grey and may not represent an actively transcribed integration event (Figure 5.13). Circos plot for CU-VA-9H-16I shows a very strong integration site from the HPV genome into the human chromosome 11, represented by a black band, and less transcribed integration events into chromosome 1 and 17 (Figure 5.14). This result suggested that the integration into chromosome 11 may be a true, transcribed integration event and this integration site had previously been reported in DIPS analysis.
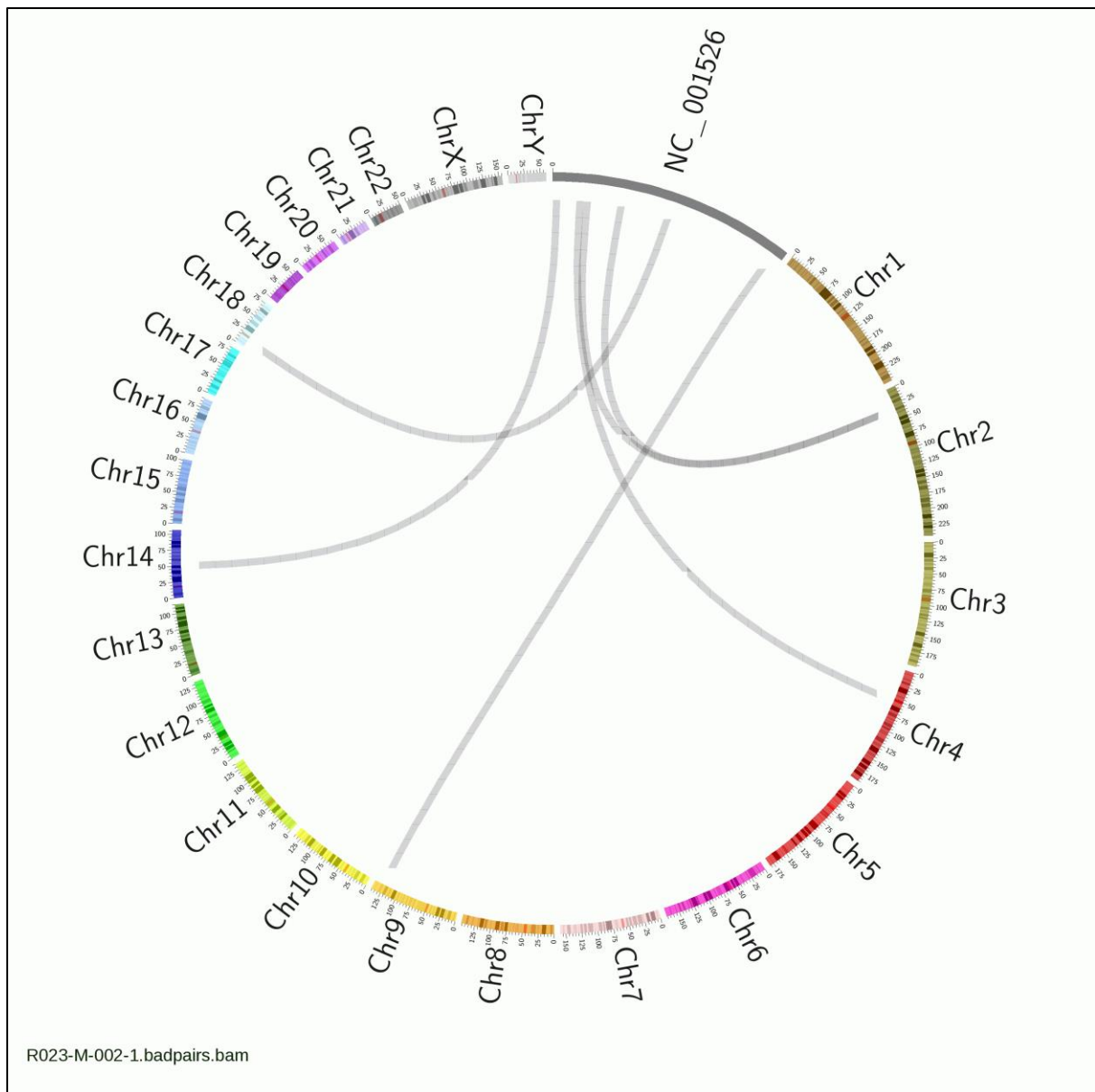
Figure 5.13 Circos plot of CU-VA-9A-16E which shows HPV/human fusions transcripts into 5 different human chromorsomes (2, 4, 9,14 and 17).
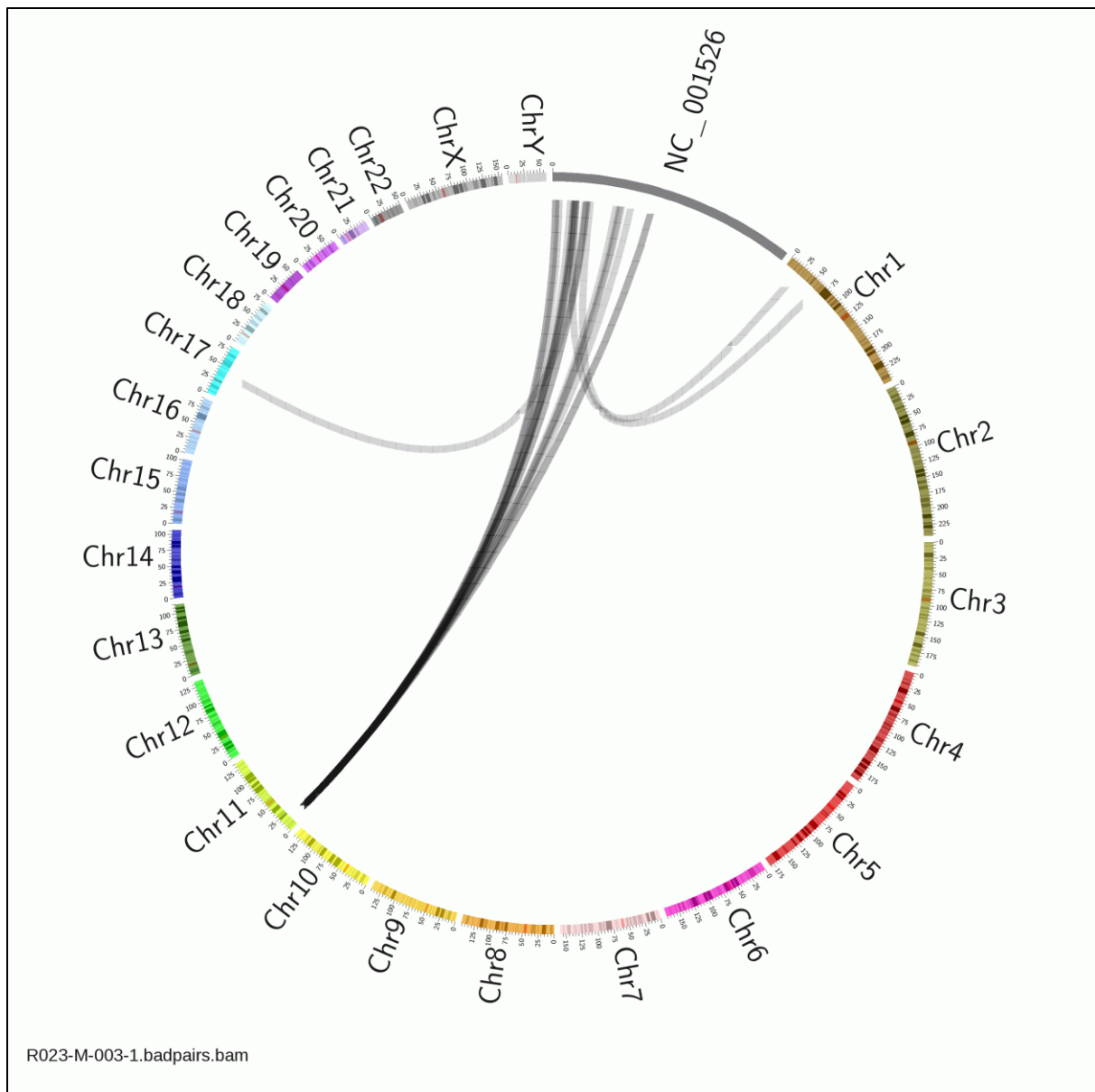
Figure 5.14 Circos plot for CU-VA-9H-16I shows a very strong integration site from the HPV genome into the human chromosome 11, represented by a black band, and less transcribed integration events into chromosome 1 and 17

5.5 Discussion

This section will highlight the key findings from this chapter and the strength and limitations of this study. Each of the key findings will then be considered in greater detail compared to the available literature and the broader implications.

5.5.1 Key Findings

1. This is the first reported organotypic raft culture of VIN and VaIN.

2. HPV status defined in monolayer did not predict histology of cells when grown in a stratified epithelium.

3. Samples with morphology in keeping with high grade disease had higher expression of p16 and SHH compared to samples with morphology of samples in keeping with low grade disease.

4. Late gene expression E4 was only detected in CU-VA-9A-16E which had morphology in keeping with low grade disease.

5. The highest levels of viral methylation were present in CU-VA-9H-16I and CU-VU-8M-16I which both had previously defined viral integration.

6. Whole transcriptome sequencing showed CU-VA-9A-16E expressed late genes compared to CU-VA-9H-16I which had high levels of *E6*, *E7* and *E1*

and showed HPV-human transcripts in both CU-VA-9A-16E and CU-VA-9H-16I.

5.5.2 Strengths

This is the first VIN and VaIN organotypic raft culture to be described. The use of the organotypic raft culture allowed for histological analysis as well as the molecular analysis of cultures with a mixed population of differentiated cells. All work in this section was carried out in triplicate to ensure the reproducibility of the histological and molecular results.

All cellular material was fresh frozen and subsequently high quality DNA and RNA were obtained for molecular analysis, allowing for whole transcription sequencing to be carried out on three of the four raft cultures.

5.5.3 Limitations

The interpretation of the morphology of the organotypic raft cultures is a subjective analysis, however to overcome any technical variation each raft was

produced in triplicate and the same pathologist interpreted the cultures throughout to ensure the same criteria and standard of interpretation. The pathologist was blinded to the nature of the HPV infection in the raft to ensure an unbiased analysis.

There is a possibility that artefacts may have been introduced into the system due to the production and handling of the raft cultures. This was countered using positive and negative controls (w12 Ser4B and HEKn). This allowed for the comparison of the raft cultures to a highly neoplastic control as well as a control with no neoplastic properties. This highlighted any artefacts that the model introduced when samples were analysed.

5.5.4 This is the first reported organotypic raft culture of VIN and VaIN.

Historically HPV research was dampened by the lack of a recognised model for the HPV lifecycle and production of HPV in the lab. This is due to the fact that HPV infection is species specific and there is little cross-species infection. Additionally unlike bacteria, viruses require a host cell in order to be grown – making reproduction in a laboratory difficult (Campo, 2002). Animal models have been pivotal in assessing the original identification of HPV as a transmissible virus (Rowson and Mahy, 1967, Rous and Beard, 1934) and also provided a model for disease progression and the interplay of host genes and papillomavirus. Although

200

animal models are still key to our understanding of HPV driven disease (Borzacchiello et al., 2009) the development of cellular models has increased our understanding of HPV related disease and are useful for testing therapeutic treatments against HPV.

An organotypic raft culture facilitates the differentiation of cells consistent with a typical anogenital lesion, therefore produces a more robust model of HPV driven neoplasia than is possible with cells grown in a monolayer (Andrei et al., 2010, Anacker and Moody, 2012, Frattini et al., 1996). This model allows for the study of genetic and epigenetic changes due to the process of cellular differentiation and stratification. In the raft culture model HPV can be propagated which allows the study of the viral lifecycle from infection to virion production (Wang et al., 2009a, Jagu et al., 2013, Chow, 2015). New advances in 3D cellular *in vitro* cellular cultures include scaffold free stromal tissue resembling extracellular matrix, (De Gregorio et al., 2017, He et al., 2016).

5.5.5 HPV status defined in monolayer did not predict histology of cells when grown in a stratified epithelium.

Cell lines in this study with previously defined viral integration showed morphology in keeping with both a low-grade and high-grade lesion, demonstrating

that cellular differentiation is not solely driven by HPV genome status. This is consistent with claims elsewhere that whilst HPV integration is common in transformation, it is not a necessary step in HPV infected cells (Wentzensen and von Knebel, 2007, Senapati et al., 2016). The level of viral integration has been shown to vary between HPV types with integration being at a lower rate than in cancers driven by HPV 16 compared to samples with HPV 18 (Badaracco et al., 2002, Woodman et al., 2003). Further work should be carried out to assess if HPV status in monolayer can predict the histology of cells in a stratified epithelium for different HPV types.

This work has shown distinct morphological changes with a marked difference seen in cultures originating from a single biopsy. The phenotype of integrated organotypic raft models, CU-VA-9H-16I and CU-CU-8M-16I, exhibit both high and low levels of differentiation respectively. Similarly, organotypic raft models with episomal DNA, CU-VU-8P-16E and CU-VA-9A-16E, displayed both high and low levels of differentiation. Therefore, the phenotype of the cell lines was not influenced by the integration status of the cell lines.

5.5.6 Samples with morphology in keeping with high grade disease had higher expression of p16 and SHH compared to samples with morphology of samples in keeping with low grade disease.

Organotypic raft cultures with integrated HPV 16 infections showed both low and high levels of cellular differentiation and varied levels of p16, Ki-67 and SHH protein expression, rejecting Hypothesis 5(A) and 5(B): *In vitro* stratified epithelium models containing HPV integrated infections when compared to models of HPV episomal infection will display (A) a less differentiated phenotype and (B) higher p16, Ki-67 and SHH gene expression levels.

In this research there was a significant difference in both p16 ($X^2(1)$ = 4.511, p = 0.034) and Ki-67 ($X^2(1)$ = 12.816, p > 0.001) stain depth when cell lines with histology "in keeping with high-grade morphology" and cell lines with histology "in keeping with low-grade morphology" were compared. p16 and Ki-67 are both potential biomarkers for HPV infection and as such have been introduced into cervical screening protocols. Previous work has suggested that Ki-67 independently is a poor discrete indicator of HPV disease (Agoff et al., 2003). However work elsewhere has suggested that Ki-67 together with p16, p53 or both may be a useful biomarker for use in cervical screening (Kotzev et al., 2012).

The expression of SHH in all VIN/VaIN cultures in this study is consistent with the cell lines being transformed. This work has shown an increase in expression of SHH in cell lines with histology "in keeping with high-grade morphology" when compared to cell lines with histology "in keeping with high-grade morphology" ($X^2(3)$ = 13.205, p = 0.004), thus suggesting that elevated SHH expression may be a potential prognostic/predictive biomarker for HPV related disease worthy of

developing an accurate and reproducible assay to measure SHH in a clinical cohort (BIDD Assay Development – stage 1). CU-VU-8M-16I with low-grade morphology also exhibited a high level of SHH expression (119.49%). However, GLI-1 expression was variable and (37.17% to 72.17%) and no association could be seen to either HPV integration status or culture morphology. Although in this study GLI-1 was not associated with HPV status or morphology, it has been shown that GLI-1 and other GLI proteins have an association with HPV disease outcome in head and neck cancer, although studies have disagreed whether this is a positive or negative outcome (Yan et al., 2011, Enzenhofer et al., 2016).

5.5.7 Late gene expression of HPV 16 E4 was only detected in CU-VA-9A-16E which had morphology in keeping with low grade disease.

Although cell lines CU-VA-9A-16E and CU-VU-8M-16I both show low-grade morphology only the CU-VA-9A-16E raft culture shows late HPV gene expression based on HPV16 *E4* expression. This is the only cell line that is completing the full viral lifecycle leading to HPV virus particle production, this data indicates that with the low p16, Ki-67 and SHH expression, CU-VA-9A-16E is representative of a transient HPV infection, which would be naturally cleared by the immune system of the host. This would be a successful model for use in future work to better understand the host and viral biology in a normal transient infection and comparison

204

with the other line from the same host may reveal the changes that take place to move it from transient to progressive.

It has been suggested that HPV 16 E4 may be another potential biomarker for disease as HPV E4 has been detected in low- and high-grade CIN but not in HPV negative lesions (Yajid et al., 2017), contrary to work presented in this thesis which only detected HPV E4 protein in CU-VA-9A-16E which had low grade morphology. It has also been suggested that HPV E4 and p16 detection could be used as a dual biomarker which could differentiate between CIN I/CIN II and CIN II+ as E4 staining is only present in low grade lesions (van Baars et al., 2015). This is similar to with work presented here for both p16 and E4 gene expression and with the use of a pan-HPV E4 antibody (an antibody against E4 in 15 HR-HPV types) may be a suitable biomarker for triage of vaccinated women.

5.5.8 The highest levels of viral methylation were present in CU-VA-9H-16I and CU-VU-8M-16I which both had previously defined viral integration.

Higher levels of viral genomic HPV 16 DNA methylation were present in samples with previously defined HPV integration, agreeing with Hypothesis 5(C): *In vitro* stratified epithelium models containing HPV integrated infections when

Integrated cell line CU-VU-8M-16I also displayed low grade morphology, however it had several molecular factors, (including high SHH expression, hyper methylation) consistent with a high-grade HPV infection. The range of organotypic models presented in this thesis represents a range of naturally prevalent HPV infections that may occur *in vivo*:

i.   A transient infection;

ii. A high-grade infection;

iii. A low-grade infection with molecular characteristics of a high-grade infection, perhaps representative of the transition from a low-grade to a high-grade infection.

When the cell lines are divided by the observed phenotype high p16 and SHH expression is shown in cultures with morphology characteristic of high-grade neoplasia, CU-VA-9H-16I and CU-VU-8P-16E. Cultures with morphology in-keeping with low-grade neoplasia, CU-CU-8M-16I and CU-VA-9A-16E, had lower levels of p16 and SHH expression. Interestingly culture CU-VU-8M-16I had the highest level of viral genomic DNA hypermethylation, no HPV late gene expression and SHH expression consistent with cultures CU-VA-9H-16I and CU-VU-8P-16E. Therefore, culture CU-VU-8M-16I may represent the intermediate state between low-grade and high-grade neoplasia compared to culture CU-VA-9A-16E which has all the features

of a low-grade transient infection including late gene expression, hypomethylation of viral genomic DNA and low p16, Ki-67 and SHH expression.

5.5.9 Whole transcriptome sequencing showed CU-VA-9A-16E expressed late genes compared to CU-VA-9H-16I which had high levels of E6, E7 and E1 and showed HPV-human transcripts in both CU-VA-9A-16E and CU-VA-9H-16I.

Whole transcriptome sequencing analysis was carried out to confirm integration events and gene expression levels as reported in monolayer cultures (Onions, 2016). The results confirmed that cell line CU-VA-9A-16E transcribes both early and late HPV genes when grown in an organotypic raft culture. There were several points of integration shown by whole transcriptome sequencing; however these were infrequent events that are believed to not be transcriptionally active. It was also confirmed that cell lines CU-VA-9H-16I showed a transcriptionally active integration event into chromosome 11 and late HPV gene expression was reduced in this culture compared to CU-VA-9A-16E. From these results we confirmed the gene expression and transcriptionally active integration events reported in the monolayer equivalent cultures. To further analyse which integration events are transcriptionally active in future work it may be of interest to look at the integration events reported by whole transcriptome sequencing using digital PCR to quantify these integration events and assess the true level of viral integration events (Sedlak et al., 2014).

Chapter 6

General Discussion

This thesis reports several key findings:

1. Disruption and methylation of viral *E2* gene are not suitable biomarkers for detecting progression to CIN II+ in young women.

2. In population with a wide age range, viral methylation was positively correlated with disease grade.

3. Models of HPV positive organotypic raft cultures from VIN and VaIN derived cell lines identified that:

    I. The HPV genome status in monolayer cell cultures did not predict subsequent histology of cells grown in a stratified epithelium.

    II. Samples with morphology in-keeping with high grade morphology have high expression of p16 and SHH genes compared to morphology in-keeping with low grade disease.

HPV integration and methylation were assessed in a young population and a mixed age population to assess whether they were associated with increased cervical histology. This work showed a random prevalence of low level of viral integration in both populations studied, which was not associated with higher grade disease. It has previously been reported that integration frequency is highest for HPV

types 16, 18 and 45 (Vinokurova et al., 2008), which is in opposition to data presented in chapter three, which detected the lowest level of viral integration in HPV 16 positive samples, however this was not statistically significant. The work by Vinokurova et al (2008) was carried out in a mixed age population of women whereas work presented in chapter three focused on young women (20 – 22 years of age), which may account for the disparity in findings. Previous studies of viral integration in young women have found a disagreement in the prevalence of viral integration (Collins et al., 2009, Ramanakumar et al., 2010).These difference may have been a result of discrepancies in how the clinical samples were obtained and managed before molecular testing. Findings in this study support that HPV integration is a rare event in young women and further basic research is required to assess potential biomarkers for HPV related disease in this demographic. It would also be advantageous to consider behavioural and environmental factors that could be assessed in combination with molecular changes, e.g. use of hormonal contraceptives, pregnancy or urinary tract infections (Nelson et al, 2009, Winer et al, 2003). These factors may highlight a subset of women for whom molecular biomarkers could be successfully applied to predict HPV disease.

Viral integration events are commonly seen in or near CFS of the human genome. Work in this thesis showed 50% of viral integration sites were in or near known CFS's (6q15, 3q25.33 and 9p21.2). Recently an updated review of viral integration sites has been published which reported that over 1,500 integration sites have been published which were statistically more likely to be near CpG regions,

fragile sites and transcriptionally active regions (Bodelon et al., 2016). It has been suggested here and elsewhere that integration of HPV into the human genome may be related to regions of genomic instability, and could suggest that HPV may be an intentional mutagen of human DNA which contributes to cancer formation by promoting mutagenesis (Akagi et al., 2014).

Methylation analysis in young women did not show any association between viral methylation of HPV E2 and high grade disease in young women. However, work described in Chapter 4 and 5 of this thesis both supported HPV methylation as a potential marker of high grade disease. Viral methylation has been considered as a potential biomarker of HPV disease. In this work, viral methylation has shown to be associated with an increase in HPV driven disease potentially suggesting that viral methylation should be further studied as a HPV biomarker, this suggestion supports work presented elsewhere that has shown HPV 16 *L1*/*L2* viral methylation is a potential biomarker for disease duration and that *E2* viral methylation has potential as a biomarker for predicting response to Cidofovir in the treatment of VIN (Jones, 2016, Bryant et al., 2015). Work in this thesis showed that three CpG sites (CpG 3416, CpG 3435 and CpG 3447) in the E2 region of HPV 16 were significantly hyper methylated in samples with viral genome disruption. As discussed in section 1.4.2 viral integration is a common event in cervical cancer but not a prerequisite for disease. Further examination of the relationship between viral integration, methylation and disease progression may highlight a few individual CpG sites in the

human and/or HPV genome, such as CpG 3416, CpG 3435 and CpG 3447, which may have clinical utility.

Considering the Prognostic and Predictive Biomarker Road map (Lioumi and Newell, 2010, Cancer Research UK, 2016) work presented in this thesis has reported the development and potential clinical utility of accurate and reproducible assays to measure viral methylation as a biomarker for HPV integration. The next stage in biomarker development is to refine the molecular assays, assays for HPV driven anogenital neoplasia must be refined to facilitate high assay performance in samples taken during cervical screening (BIDD Assay Development – Stage 2).

This thesis considers several further biomarkers for HPV disease of which P16, Ki-67 and SHH showed the greatest potential as biomarkers for high-grade HPV disease as all of these showed significant variation between cell lines with histology "in keeping with high-grade morphology" and cell lines with histology "in keeping with low-grade morphology". A dual stain assay has been evaluated in as a part of cervical screening and triage and has been shown to have comparable specificity to standard cervical screening (97.2% vs 95.4%, p = 0.15) with increased sensitivity (86.7% vs 68.5%, p<0.001) (Ikenberg et al., 2013, Petry et al., 2011). Also as the p16/Ki-67 dual stain is not HPV type specific but a marker of cellular changes this should be considered as a potential biomarker in vaccinated women. This would ensure monitoring of disease presence even regardless of the HPV type driving

disease, compared to HPV DNA testing which may only assess a small range of HPV types.

SHH protein expression was also shown to be unregulated in organotypic raft cultures with morphology in-keeping with a high-grade infection. This work on SHH and GLI-1 protein expression presented in this thesis was carried out in collaboration with Dr Chris Dawson from The University of Birmingham. Work from Dr Dawson group has shown that the deregulation of Hedgehog protein is activated by Epstein-Barr virus through the induction of SHH (Port et al., 2013) and has reported SHH is aberrantly activated in vulval cancer and 8/10 solitary VIN cases have shown over expression of SHH and its downstream effector GLI-1 (Dr Dawson, Personal Communication). It has also been shown that the inhibition of GLI-1 in cancer cell lines reduced the 'stemness' of cervical cancer cells, leading to a less-cancerous phenotype (Vishnoi et al., 2016). This work suggested there was an interaction between HPV E6 and GLI-1 which may be a useful relationship to exploit in the search for novel therapeutic approaches against HPV driven cancer. Work in this thesis also promotes further development of SHH expression as a potential prognostic or predictive biomarker for HPV disease. Future work is required to develop clinical assays for these biomarkers and to define the biomarker distribution in specimens from the target population (BIDD Assay Development – Stage 1).

Apart from biomarkers for the prediction of disease presence or progression, biomarkers can also be used to predict the success of treatment of HPV infections or

risk of disease recurrence. Recent work in oropharyngeal SCC have shown that E6 and E7 serum antibody level was significantly higher in patients with recurring disease (Spector et al., 2016). This biomarker could be implemented after oropharyngeal SCC treatment to monitor likelihood of disease recurrence to enhance patient management strategies.

Work presented in this thesis must be considered within the broader context of the HPV vaccination programme. Cervical screening programmes must evolve to maintain their effectiveness in a vaccinated population, when fewer abnormalities will be prevalent and cytologists have the potential to become deskilled without appropriate new measures in place to counteract this. New molecular screening methods for HPV presence, which have a higher sensitivity than traditional cytological methods, must avoid over-treatment of women with HPV associated lesions to allow time for viral regression to occur naturally, wherever possible. Work in this thesis supports the observed natural history of HPV infection that can change over time with differences between transient and persistent infections detected in young and older women. It has been demonstrated that CIN lesions in young women are more likely to regress compared to older women (Moore et al., 2007, Fuchs et al., 2007). An optimal strategy for future cervical screening programmes may be to combine different detection methods to ensure that the most appropriate treatment is delivered dependent on the stage of infection. It may also be beneficial to have age specific treatment protocols to ensure that the most efficient method of screening is taken place depending on the women's age. On the 1st of September 2013 CSW

changed the age criteria for screening women in Wales. The new protocols include screening women every three years from 25 to 50 years of age, then screening every 5 years from 50 to 64 years. More recently CSW has also changed the screening protocol; from April 2nd 2017 primary HPV testing has been introduced into the cervical screening programme in Wales (CSW, 2017). This is an initial pilot of the programme which is due to be rolled out across Wales by 2018/2019. Testing for HPV is carried out as a primary test and only if HPV is detected then cervical cytology is assessed. Current UK National Screening Committee have assessed the introduction of primary HPV screening and concluded that replacing HPV cytology with primary HPV DNA testing is likely to be a cost effective move (UK NSC, 2016).

Several other challenges are faced in routine cervical screening programmes following the introduction of HPV vaccination. These include a potential reduction in women attending for call and recall smear testing, changes in the vaccination used and understanding of the long-term effectiveness of the HPV vaccination.

Globally, implications must also be considered for women in low resource settings. Introduction and uptake of the HPV vaccine is a significant step in disease prevention but optimal strategies for screening in this population must be considered. Several studies have tried to consider novel biomarkers and screening protocols, which includes the use of trained sniffer dogs to detect cancerous changes (Guerrero-Flores et al., 2017), self-collected samples (Bansil et al., 2014) & how to enhance uptake rates of the vaccine in the overall population.

This thesis has described the first VIN and VaIN raft cultures and as expected has highlighted increased p16, Ki-67 and SHH expression as potential biomarkers for HPV disease. Future work should assess the effect of p16, Ki-67 and SHH gene expression on cellular morphology by assessing if knocking out p16, Ki-67 and/or SHH genes effect the growth of cultures CU-VA-9H-16I and CU-VU-8P-16E. These cultures can also be employed in future HPV drug discovery and evaluation to increase understanding of drug uptake, changes to human and HPV gene expression and changes in phenotype of the culture after exposure to drugs and/or radiation. Organotypic raft cultures have already been employed to assess compounds against Alphaherpesviruses. In this study antiviral drugs were added directly to lesions in the raft culture and the changes to cellular histology with various compounds were measured. The antiviral effect of the various compounds were then measured using plaque assays and qRT-PCR (Andrei, et al. 2005).

Chapter 7

Conclusion

HPV anogenital disease progression is driven by a variety of contributing factors. This work examined HPV methylation, HPV gene disruption and aberrations in host and viral gene expression and their suitability as biomarkers for HPV disease. This study has shown that HPV genome disruption and HPV DNA methylation are not suitable biomarkers for young women. However, this study showed a correlation between HPV methylation at three CpG sites and HPV gene disruption. More research is required to define if HPV methylation is a suitable biomarker which shows improved clinical outcome by its use in the cervical screening programme or to predict response to treatments.

This thesis is the first to report the growth of VIN and VaIN cell lines in an Organotypic Raft Culture model which facilitated the growth of cells in a stratified epithelium. These models can also be utilised to further understand the natural history of VIN and VaIN by identifying phenotypic changes with related molecular changes. These Organotypic Raft Cultures can also be employed as a model for testing of potential topical treatments for VIN and VaIN.

VIN and VaIN cell lines showed diverse differentiation patterns which were not associated with viral integration status as defined in monolayer. This work

supports that integration is unlikely to be an appropriate biomarker in predicting disease progression. The increased expression of p16, Ki-67 and SHH in cell lines with high-grade histology supports the further investigation of these biomarkers in cervical screening as a potential prognostic biomarkers

Chapter 8

References

ABREU, A. L. P., SOUZA, R. P., GIMENES, F. & CONSOLARO, M. E. L. 2012. A review of methods for detect human Papillomavirusinfection. *Virology Journal,* 9**,** 262.

AGOFF, S. N., LIN, P., MORIHARA, J., MAO, C., KIVIAT, N. B. & KOUTSKY, L. A. 2003. p16(INK4a) expression correlates with degree of cervical neoplasia: a comparison with Ki-67 expression and detection of high-risk HPV types. *Mod Pathol,* 16**,** 665-73.

AKAGI, K., LI, J., BROUTIAN, T. R., PADILLA-NASH, H., XIAO, W., JIANG, B., ROCCO, J. W., TEKNOS, T. N., KUMAR, B., WANGSA, D., HE, D., RIED, T., SYMER, D. E. & GILLISON, M. L. 2014. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res,* 24**,** 185-99.

AKSOY, P., GOTTSCHALK, E. Y. & MENESES, P. I. 2016. HPV entry into cells. *Mutation Research/Reviews in Mutation Research.* 772: 13-22

ALCORTA, D. A., XIONG, Y., PHELPS, D., HANNON, G., BEACH, D. & BARRETT, J. C. 1996. Involvement of the cyclin-dependent kinase inhibitor p16 (INK4a)

in replicative senescence of normal human fibroblasts. *Proceedings of the National Academy of Sciences,* 93**,** 13742-13747.

ALLEN, D., WINTERS, E., KENNA, P. F., HUMPHRIES, P. & FARRAR, G. J. 2008. Reference gene selection for real-time rtPCR in human epidermal keratinocytes. *J Dermatol Sci,* 49**,** 217-25.

ANACKER, D. & MOODY, C. 2012. Generation of organotypic raft cultures from primary human keratinocytes. *J Vis Exp.*

ANDREI, G., VAN DEN OORD, J., FITEN, P., OPDENAKKER, G., DE WOLF-PEETERS, C., DE CLERCQ, E. & SNOECK, R. 2005. Organotypic epithelial raft cultures as a model for evaluating compounds against alphaherpesviruses. *Antimicrobial Agents and Chemotherapy,* 49**,** 4671-4680.

ANDREI, G., DURAFFOUR, S., VAN DEN OORD, J. & SNOECK, R. 2010. Epithelial raft cultures for investigations of virus growth, pathogenesis and efficacy of antiviral agents. *Antiviral Res,* 85**,** 431-49.

ARBYN, M., PARASKEVAIDIS, E., MARTIN-HIRSCH, P., PRENDIVILLE, W. & DILLNER, J. 2005. Clinical utility of HPV-DNA detection: triage of minor cervical lesions, follow-up of women treated for high-grade CIN: an update of pooled evidence. *Gynecol Oncol,* 99**,** S7-11.

ARMSTRONG, K. M., BERMINGHAM, E. N., BASSETT, S. A., TRELOAR, B. P., ROY, N. C. & BARNETT, M. P. 2011. Global DNA methylation measurement by HPLC using low amounts of DNA. *Biotechnol J,* 6**,** 113-7.

ATCC. 2016. *SiHa (ATCC HTB-35)* [Online]. https://www.lgcstandards-atcc.org: LGC Standards. Available: https://www.lgcstandards-atcc.org/Products/Cells_and_Microorganisms/By_Tissue/Cervix/HTB-35.aspx?geo_country=gb [Accessed 14/02/2017 2017].

BADARACCO, G., VENUTI, A., SEDATI, A. & MARCANTE, M. L. 2002. HPV16 and HPV18 in genital tumors: Significantly different levels of viral integration and correlation to tumor invasiveness. *J Med Virol,* 67**,** 574-82.

BANSIL, P., WITTET, S., LIM, J. L., WINKLER, J. L., PAUL, P. & JERONIMO, J. 2014. Acceptability of self-collection sampling for HPV-DNA testing in low-resource settings: a mixed methods approach. *BMC Public Health,* 14**,** 596.

BARTLETT, J. M. S. & STIRLING, D. 2003. A Short History of the Polymerase Chain Reaction. *In:* BARTLETT, J. M. S. & STIRLING, D. (eds.) *PCR Protocols.* Totowa, NJ: Humana Press.

BECTON DICKINSON. 2017. *BD SUREPATH™ PAP TEST* [Online]. http://www.bd.com/tripath/physicians/surepath.asp. [Accessed 12/02/2017 2017].

BENEVOLO, M., MOTTOLESE, M., MARANDINO, F., VOCATURO, G., SINDICO, R., PIPERNO, G., MARIANI, L., SPERDUTI, I., CANALINI, P., DONNORSO, R. P. & VOCATURO, A. 2006. Immunohistochemical expression of p16(INK4a) is predictive of HR-HPV infection in cervical low-grade lesions. *Mod Pathol,* 19**,** 384-91.

BENEVOLO, M., VOCATURO, A., CARACENI, D., FRENCH, D., ROSINI, S., ZAPPACOSTA, R., TERRENATO, I., CICCOCIOPPO, L., FREGA, A. & ROSSI, P. G. 2011. Sensitivity, Specificity, and Clinical Value of Human Papillomavirus (HPV) E6/E7 mRNA Assay as a Triage Test for Cervical Cytology and HPV DNA Test. *Journal of Clinical Microbiology,* 49**,** 2643-2650.

BERNSTEIN, S. J., SANCHEZ-RAMOS, L. & NDUBISI, B. 2001. Liquid-based cervical cytologic smear study and conventional Papanicolaou smears: a metaanalysis of prospective studies comparing cytologic diagnosis and sample adequacy. *Am J Obstet Gynecol,* 185**,** 308-17.

BHAT, S., KABEKKODU, S. P., NORONHA, A. & SATYAMOORTHY, K. 2016. Biological implications and therapeutic significance of DNA methylation regulated genes in cervical cancer. *Biochimie,* 121**,** 298-311.

BHATTACHARJEE, B. & SENGUPTA, S. 2006. CpG methylation of HPV 16 LCR at E2 binding site proximal to P97 is associated with cervical cancer in presence of intact E2. *Virology,* 354**,** 280-5.

BIBIKOVA, M., LIN, Z., ZHOU, L., CHUDIN, E., GARCIA, E. W., WU, B., DOUCET, D., THOMAS, N. J., WANG, Y., VOLLMER, E., GOLDMANN, T., SEIFART, C., JIANG, W., BARKER, D. L., CHEE, M. S., FLOROS, J. & FAN, J.-B. 2006. High-throughput DNA methylation profiling using universal bead arrays. *Genome Research,* 16**,** 383-393.

BIERKENS, M., HESSELINK, A. T., MEIJER, C. J. L. M., HEIDEMAN, D. A. M., WISMAN, G. B. A., VAN DER ZEE, A. G. J., SNIJDERS, P. J. F. & STEENBERGEN, R. D. M. 2013. CADM1 and MAL promoter methylation levels in hrHPV-positive cervical scrapes increase proportional to degree and duration of underlying cervical disease. *International Journal of Cancer,* 133**,** 1293-1299.

BLACHON, S., BELLANGER, S., DEMERET, C. & THIERRY, F. 2005. Nucleo-cytoplasmic shuttling of high risk human Papillomavirus E2 proteins induces apoptosis. *J Biol Chem,* 280**,** 36088-98.

BODELON, C., UNTEREINER, M. E., MACHIELA, M. J., VINOKUROVA, S. & WENTZENSEN, N. 2016. Genomic characterization of viral integration sites in HPV-related cancers. *Int J Cancer,* 139**,** 2001-11.

BODILY, J. & LAIMINS, L. A. 2011. Persistence of human papillomavirus infection: keys to malignant progression. *Trends Microbiol,* 19**,** 33-9.

BONANNI, P., COHET, C., KJAER, S. K., LATHAM, N. B., LAMBERT, P. H., REISINGER, K. & HAUPT, R. M. 2010. A summary of the post-licensure surveillance initiatives for GARDASIL/SILGARD. *Vaccine,* 28**,** 4719-30.

BONIN, S. & STANTA, G. 2013. Nucleic acid extraction methods from fixed and paraffin-embedded tissues in cancer diagnostics. *Expert Rev Mol Diagn,* 13**,** 271-82.

BORZACCHIELLO, G., ROPERTO, F., NASIR, L. & CAMPO, M. S. 2009. Human papillomavirus research: Do we still need animal models? *International Journal of Cancer,* 125**,** 739-740.

BOSCH, F. X., LORINCZ, A., MUNOZ, N., MEIJER, C. J. & SHAH, K. V. 2002. The causal relation between human papillomavirus and cervical cancer. *J Clin Pathol,* 55**,** 244-65.

BOULET, G. A., BENOY, I. H., DEPUYDT, C. E., HORVATH, C. A., AERTS, M., HENS, N., VEREECKEN, A. J. & BOGERS, J. J. 2009. Human papillomavirus 16 load and E2/E6 ratio in HPV16-positive women: biomarkers for cervical intraepithelial neoplasia >or=2 in a liquid-based cytology setting? *Cancer Epidemiol Biomarkers Prev,* 18**,** 2992-9.

BOUSARGHIN, L., TOUZÉ, A., SIZARET, P.-Y. & COURSAGET, P. 2003. Human Papillomavirus Types 16, 31, and 58 Use Different Endocytosis Pathways To Enter Cells. *Journal of Virology,* 77**,** 3846-3850.

BOWDEN, S., HOUGHTON, R., EDWARDS, A., FIANDER, A. & HIBBITTS, S. Is HPV 51 a candidate for type-replacement following vaccination? SpROGs 2013 Conference, 2013 Nottingham, United Kingdom. BJOG An International Journal of Obstetrics and Gynaecology.

BRAKE, T. & LAMBERT, P. F. 2005. Estrogen contributes to the onset, persistence, and malignant progression of cervical cancer in a human papillomavirus-transgenic mouse model. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 2490-2495.

BROWN, R. E., BREUGELMANS, J. G., THEODORATOU, D. & BENARD, S. 2006. Costs of detection and treatment of cervical cancer, cervical dysplasia and genital warts in the UK. *Curr Med Res Opin,* 22**,** 663-70.

BRUNI, L., BARRIONUEVO-ROSAS, L., ALBERO, G., SERRANO, B., MENA, M., GOMEZ, D., MUNOZ, J., BOSCH , F. X. & DE SANJOSÉ , S. 2017. Human Papillomavirus and Related Diseases in the World. Summery Report. ICO Information Centre on HPV and Cancer (HPV Information Centre). Avaliable at: www.hpvcentre.net/statistics/reports/XWX.pdf [Accessed: 13/06/2017]

BRUNI, L., DIAZ, M., CASTELLSAGUE, X., FERRER, E., BOSCH, F. X. & DE SANJOSE, S. 2010. Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. *J Infect Dis,* 202**,** 1789-99.

BRYANT, D. 2012. *Methylation of Human Papillomavirus DNA: biological significance and clinical utility.* Doctor of Philosophy, Cardiff University.

BRYANT, D., HIBBITTS, S., ALMONTE, M., TRISTRAM, A., FIANDER, A. & POWELL, N. 2015. Human papillomavirus type 16 L1/L2 DNA methylation shows weak association with cervical disease grade in young women. *J Clin Virol,* 66**,** 66-71.

BRYANT, D., ONIONS, T., RAYBOULD, R., FLYNN, A., TRISTRAM, A., MEYRICK, S., GILES, P., ASHELFORD, K., HIBBITTS, S., FIANDER, A. & POWELL, N. 2014. mRNA sequencing of novel cell lines from human papillomavirus type-16 related vulval intraepithelial neoplasia: consequences of expression of HPV16 E4 and E5. *J Med Virol,* 86**,** 1534-41.

BULKMANS, N. W. J., BERKHOF, J., ROZENDAAL, L., VAN KEMENADE, F. J., BOEKE, A. J. P., BULK, S., VOORHORST, F. J., VERHEIJEN, R. H. M., VAN GRONINGEN, K., BOON, M. E., RUITINGA, W., VAN BALLEGOOIJEN, M., SNIJDERS, P. J. F. & MEIJER, C. 2007. Human papillomavirus DNA testing for the detection of cervical intraepithelial neoplasia grade 3 and cancer: 5-year follow-up of a randomised controlled implementation trial. *The Lancet,* 370**,** 1764-1772.

BUSTIN, S. A., BENES, V., GARSON, J. A., HELLEMANS, J., HUGGETT, J., KUBISTA, M., MUELLER, R., NOLAN, T., PFAFFL, M. W., SHIPLEY, G. L., VANDESOMPELE, J. & WITTWER, C. T. 2009. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry,* 55**,** 611-622.

CADIEUX, B., CHING, T. T., VANDENBERG, S. R. & COSTELLO, J. F. 2006. Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation. *Cancer Res,* 66**,** 8469-76.

CALIN, G. A. & CROCE, C. M. 2006. MicroRNA signatures in human cancers. *Nat Rev Cancer,* 6**,** 857-66.

CAMPO, M. S. 2002. Animal models of papillomavirus pathogenesis. *Virus Res,* 89**,** 249-61.

CAMPO, M. S., GRAHAM, S. V., CORTESE, M. S., ASHRAFI, G. H., ARAIBI, E. H., DORNAN, E. S., MINERS, K., NUNES, C. & MAN, S. 2010. HPV-16 E5 down-regulates expression of surface HLA class I and reduces recognition by CD8 T cells. *Virology,* 407**,** 137-42.

CANCER RESEARCH UK. 2016. *Prognostic and Predictive Biomarker Roadmap* [Online]. Cancer Research UK. Available: http://www.cancerresearchuk.org/prod_consump/groups/cr_common/@fre/@fun/documents/generalcontent/cr_027486.pdf [Accessed 13/02/2017 2017].

CANCER RESEARCH UK. 2017. *Cervical cancer incidence statistics* [Online]. Cancer Research UK. Available: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/cervical-cancer/incidence [Accessed 14/04/2017 2017].

CDC. 2013. *Basic Information About HPV and Cancer* [Online]. https://www.cdc.gov/cancer/hpv/basic_info/: Center for Disease Control. Available: https://www.cdc.gov/cancer/hpv/basic_info/ [Accessed 20th September 2015 2015].

CERASUOLO, A., ANNUNZIATA, C., TORTORA, M., STARITA, N., STELLATO, G., GREGGI, S., MAGLIONE, M. G., IONNA, F., LOSITO, S., BOTTI, G., BUONAGURO, L., BUONAGURO, F. M. & TORNESELLO, M. L. 2017. Comparative analysis of HPV16 gene expression profiles in cervical and in oropharyngeal squamous cell carcinoma. *Oncotarget,* 8**,** 34070-34081.

CERVICAL SCREENING WALES 2005. National Service Framework for the Cervical Screening Programme in Wales Public Health Wales.

CERVICAL SCREENING WALES. 2007. KC53/61/65 Statistical Report 2008/2009. Avaliable at: www.cervicalscreeningwales.wales.nhs.uk/opendoc/225753 [Accessed 19/04/2017]

CERVICAL SCREENING WALES 2017 Key messages about the cervical screening programme. Avaliable: http://www.cervicalscreeningwales.wales.nhs.uk/key-messages-about-the-screening-program [Accessed 14/04/2017]

CHANG, L., HE, X., YU, G. & WU, Y. 2013. Effectiveness of HPV 16 viral load and the E2/E6 ratio for the prediction of cervical cancer risk among Chinese women. *J Med Virol,* 85**,** 646-54.

CHANG, S.-W., XU, G.-Q. & FAN, Y.-L. 2015. Association of interleukin-12 gene polymorphisms with cancer susceptibility: a meta-analysis. *International Journal of Clinical and Experimental Medicine,* 8**,** 5317-5322.

CHAPMAN, V. L., TERRANOVA, R., MOGGS, J. G., KIMBER, I. & DEARMAN, R. J. 2016. Evaluation of 5-methylcytosine and 5-hydroxymethylcytosine as potential biomarkers for characterisation of chemical allergens. *Toxicology,* 340**,** 17-26.

CHENG, S., SCHMIDT-GRIMMINGER, D. C., MURANT, T., BROKER, T. R. & CHOW, L. T. 1995. Differentiation-dependent up-regulation of the human papillomavirus E7 gene reactivates cellular DNA replication in suprabasal differentiated keratinocytes. *Genes Dev,* 9**,** 2335-49.

CHOO, K.-B., PAN, C.-C. & HAN, S.-H. 1987. Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: Preferential deletion of the E2 gene and invariable retention of the long control region and the E6/E7 open reading frames. *Virology,* 161**,** 259-261.

CHOW, L. T. 2015. Model systems to study the life cycle of human papillomaviruses and HPV-associated cancers. *Virologica Sinica,* 30**,** 92-100.

CLAD, A., REUSCHENBACH, M., WEINSCHENK, J., GROTE, R., RAHMSDORF, J. & FREUDENBERG, N. 2011. Performance of the Aptima High-Risk Human

Papillomavirus mRNA Assay in a Referral Population in Comparison with Hybrid Capture 2 and Cytology. *Journal of Clinical Microbiology,* 49**,** 1071-1076.

CLARK, S. J., HARRISON, J., PAUL, C. L. & FROMMER, M. 1994. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res,* 22**,** 2990-7.

CLARKE, M. A., WENTZENSEN, N., MIRABELLO, L., GHOSH, A., WACHOLDER, S., HARARI, A., LORINCZ, A., SCHIFFMAN, M. & BURK, R. D. 2012. Human papillomavirus DNA methylation as a potential biomarker for cervical cancer. *Cancer Epidemiol Biomarkers Prev,* 21**,** 2125-37.

COHEN, J. 1988. *Statistical power analysis for the behavioral sciences,* New York, NY, Psychology Press.

COLLINS, S. I., CONSTANDINOU-WILLIAMS, C., WEN, K., YOUNG, L. S., ROBERTS, S., MURRAY, P. G. & WOODMAN, C. B. 2009. Disruption of the E2 gene is a common and early event in the natural history of cervical human papillomavirus infection: a longitudinal cohort study. *Cancer Res,* 69**,** 3828-32.

CONWAY, M. J. & MEYERS, C. 2009. Replication and Assembly of Human Papillomaviruses. *Journal of Dental Research,* 88**,** 307-317.

CRICCA, M., VENTUROLI, S., LEO, E., COSTA, S., MUSIANI, M. & ZERBINI, M. 2009. Disruption of HPV 16 E1 and E2 genes in precancerous cervical lesions. *Journal of Virological Methods,* 158**,** 180-183.

CUZICK, J. 2015. Gardasil 9 joins the fight against cervix cancer. *Expert Rev Vaccines,* 14**,** 1047-9.

CUZICK, J., CLAVEL, C., PETRY, K. U., MEIJER, C. J., HOYER, H., RATNAM, S., SZAREWSKI, A., BIREMBAUT, P., KULASINGAM, S., SASIENI, P. & IFTNER, T. 2006. Overview of the European and North American studies on HPV testing in primary cervical cancer screening. *Int J Cancer,* 119**,** 1095-101.

CUZICK, J., SZAREWSKI, A., CUBIE, H., HULMAN, G., KITCHENER, H., LUESLEY, D., MCGOOGAN, E., MENON, U., TERRY, G., EDWARDS, R., BROOKS, C., DESAI, M., GIE, C., HO, L., JACOBS, I., PICKLES, C. & SASIENI, P. 2003. Management of women who test positive for high-risk types of human papillomavirus: the HART study. *Lancet,* 362**,** 1871-6.

DALL, K. L., SCARPINI, C. G., ROBERTS, I., WINDER, D. M., STANLEY, M. A., MURALIDHAR, B., HERDMAN, M. T., PETT, M. R. & COLEMAN, N. 2008. Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions. *Cancer Res,* 68**,** 8249-59.

DAVIES, P., ARBYN, M., DILLNER, J., KITCHENER, H. C., MEIJER, C. J., RONCO, G. & HAKAMA, M. 2006. A report on the current status of European research on the use of human papillomavirus testing for primary cervical cancer screening. *Int J Cancer,* 118**,** 791-6.

DAY, P. M., LOWY, D. R. & SCHILLER, J. T. 2003. Papillomaviruses infect cells via a clathrin-dependent pathway. *Virology,* 307**,** 1-11.

DE GREGORIO, V., IMPARATO, G., URCIUOLO, F., TORNESELLO, M. L., ANNUNZIATA, C., BUONAGURO, F. M. & NETTI, P. A. 2017. An Engineered Cell-Instructive Stroma for the Fabrication of a Novel Full Thickness Human Cervix Equivalent In Vitro. *Adv Healthc Mater*, 6(11)

DE MARTEL, C., FERLAY, J., FRANCESCHI, S., VIGNAT, J., BRAY, F., FORMAN, D. & PLUMMER, M. 2012. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *The Lancet Oncology,* 13**,** 607-615.

DE VILLIERS, E. M., FAUQUET, C., BROKER, T. R., BERNARD, H. U. & ZUR HAUSEN, H. 2004. Classification of papillomaviruses. *Virology,* 324**,** 17-27.

DEATHERAGE, D. E., POTTER, D., YAN, P. S., HUANG, T. H. & LIN, S. 2009. Methylation analysis by microarray. *Methods Mol Biol,* 556**,** 117-39.

DEATON, A. M. & BIRD, A. 2011. CpG islands and the regulation of transcription. *Genes Dev,* 25**,** 1010-22.

DELVENNE, P., HUBERT, P., JACOBS, N., GIANNINI, S. L., HAVARD, L., RENARD, I., SABOULARD, D. & BONIVER, J. 2001. The organotypic culture of HPV-transformed keratinocytes: an effective in vitro model for the development of new immunotherapeutic approaches for mucosal (pre)neoplastic lesions. *Vaccine,* 19**,** 2557-64.

DEMERET, C., LE MOAL, M., YANIV, M. & THIERRY, F. 1995. Control of HPV 18 DNA replication by cellular and viral transcription factors. *Nucleic Acids Research,* 23**,** 4777-4784.

DEPUYDT, C. E., CRIEL, A. M., BENOY, I. H., ARBYN, M., VEREECKEN, A. J. & BOGERS, J. J. 2012. Changes in type-specific human papillomavirus load predict progression to cervical cancer. *J Cell Mol Med,* 16**,** 3096-104.

DEPUYDT, C. E., THYS, S., BEERT, J., JONCKHEERE, J., SALEMBIER, G. & BOGERS, J. J. 2016. Linear viral load increase of a single HPV-type in women with multiple HPV infections predicts progression to cervical cancer. *International Journal of Cancer,* 139**,** 2021-2032.

DH/NHS FINANCE, P. A. O. 2010. The Operating Framework for the NHS in England 2011/12. National Health Service.

DIMAIO, D. & MATTOON, D. 2001. Mechanisms of cell transformation by papillomavirus E5 proteins. *Oncogene,* 20**,** 7866-73.

DMITRIEVA, N. I., CUI, K., KITCHAEV, D. A., ZHAO, K. & BURG, M. B. 2011. DNA double-strand breaks induced by high NaCl occur predominantly in gene deserts. *Proc Natl Acad Sci U S A,* 108**,** 20796-801.

DOORBAR, J. 2005. The papillomavirus life cycle. *J Clin Virol,* 32 Suppl 1**,** S7-15.

DOORBAR, J., EGAWA, N., GRIFFIN, H., KRANJEC, C. & MURAKAMI, I. 2015. Human papillomavirus molecular biology and disease association. *Rev Med Virol,* 25 Suppl 1**,** 2-23.

DOORBAR, J., ELY, S., STERLING, J., MCLEAN, C. & CRAWFORD, L. 1991. Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. *Nature,* 352**,** 824-7.

DOORBAR, J., PARTON, A., HARTLEY, K., BANKS, L., CROOK, T., STANLEY, M. & CRAWFORD, L. 1990. Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line. *Virology,* 178**,** 254-62.

DUNN, O. J. 1964. Multiple Comparisons Using Rank Sums. *Technometrics,* 6**,** 241-252.

DWIVEDI, N., CHANDRA, S., KASHYAP, B., RAJ, V. & AGARWAL, A. 2013. Suprabasal expression of Ki-67 as a marker for the severity of oral epithelial dysplasia and oral squamous cell carcinoma. *Contemporary Clinical Dentistry,* 4**,** 7-12.

EBELING, W., HENNRICH, N., KLOCKOW, M., METZ, H., ORTH, H. D. & LANG, H. 1974. Proteinase K from Tritirachium album Limber. *Eur J Biochem,* 47**,** 91-7.

EISELLEOVA, L., PETERKOVA, I., NERADIL, J., SLANINOVA, I., HAMPL, A. & DVORAK, P. 2008. Comparative study of mouse and human feeder cells for human embryonic stem cells. *Int J Dev Biol,* 52**,** 353-63.

ENZENHOFER, E., PARZEFALL, T., HAYMERLE, G., SCHNEIDER, S., KADLETZ, L., HEIDUSCHKA, G., PAMMER, J., OBERNDORFER, F., WRBA, F., LOADER, B., GRASL, M. C., PERISANIDIS, C. & EROVIC, B. M. 2016. Impact of Sonic Hedgehog Pathway Expression on Outcome in HPV Negative Head and Neck Carcinoma Patients after Surgery and Adjuvant Radiotherapy. *PLoS One,* 11**,** e0167665.

FEHRMANN, F., KLUMPP, D. J. & LAIMINS, L. A. 2003. Human Papillomavirus Type 31 E5 Protein Supports Cell Cycle Progression and Activates Late Viral Functions upon Epithelial Differentiation. *Journal of Virology,* 77**,** 2819-2831.

FERLAY, J., SOERJOMATARAM, I., ERVIK, M., DIKSHIT, R., ESER, S., MATHERS, C., REBELO, M., PARKIN, D. M., FORMAN, D. & BRAY, F. 2013. *Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]* [Online]. Lyon, France: International Agency for Research on Cancer: GLOBOCAN 2012 v1.0. [Accessed 10/05/2017 2017].

FERNANDES, J., ARAÚJO, J. & FERNANDES, T. 2013. Biology and natural history of Human Papillomavirus infection. *Open Access Journal of Clinical Trials,* 2013:5**,** 1-12.

FRATTINI, M. G., LIM, H. B. & LAIMINS, L. A. 1996. In vitro synthesis of oncogenic human papillomaviruses requires episomal genomes for differentiation-dependent late expression. *Proceedings of the National Academy of Sciences,* 93**,** 3062-3067.

FUCHS, K., WEITZEN, S., WU, L., PHIPPS, M. G. & BOARDMAN, L. A. 2007. Management of cervical intraepithelial neoplasia 2 in adolescent and young women. *J Pediatr Adolesc Gynecol,* 20**,** 269-74.

GAGLIO, D., CAPITANO, F., MASTRODONATO, A., MINICOCCI, E., DEIANA, C., FRAGAPANE, P., CAMILLONI, G. & MELE, A. 2014. Learning induced epigenetic modifications in the ventral striatum are necessary for long-term memory. *Behav Brain Res,* 265**,** 61-8.

GAMMOH, N., ISAACSON, E., TOMAIC, V., JACKSON, D. J., DOORBAR, J. & BANKS, L. 2009. Inhibition of HPV-16 E7 oncogenic activity by HPV-16 E2. *Oncogene,* 28**,** 2299-304.

GAO, G., JOHNSON, S. H., VASMATZIS, G., PAULEY, C. E., TOMBERS, N. M., KASPERBAUER, J. L. & SMITH, D. I. 2017. Common fragile sites (CFS) and extremely large CFS genes are targets for human papillomavirus integrations and chromosome rearrangements in oropharyngeal squamous cell carcinoma. *Genes Chromosomes Cancer,* 56**,** 59-74.

GARLAND, S. M. & SMITH, J. S. 2010. Human papillomavirus vaccines: current status and future prospects. *Drugs,* 70**,** 1079-98.

GILBERT, L., OATES, E. & RATNAM, S. 2013. Stability of cervical specimens in SurePath medium for human papillomavirus testing with the Roche cobas 4800 assay. *J Clin Microbiol,* 51**,** 3412-4.

GIROGLOU, T., FLORIN, L., SCHAFER, F., STREECK, R. E. & SAPP, M. 2001. Human papillomavirus infection requires cell surface heparan sulfate. *J Virol,* 75**,** 1565-70.

GRAHAM, D. A. & HERRINGTON, C. S. 2000. HPV-16 E2 gene disruption and sequence variation in CIN 3 lesions and invasive squamous cell carcinomas of the cervix: relation to numerical chromosome abnormalities. *Molecular Pathology,* 53**,** 201-206.

GRAHAM, S. V. 2010. Human papillomavirus: gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. *Future microbiology,* 5**,** 1493-1506.

GREINER BIO-ONE. 2006. *PapilloCheck HPV-Screening* [Online]. Greiner Bio-One. Available: https://shop.gbo.com/en/row/products/diagnostics/ocheck/papillocheck/ [Accessed 20/01/2017 2017].

GRIFFIN, H., WU, Z., MARNANE, R., DEWAR, V., MOLIJN, A., QUINT, W., VAN HOOF, C., STRUYF, F., COLAU, B., JENKINS, D. & DOORBAR, J. 2012. E4 Antibodies Facilitate Detection and Type-Assignment of Active HPV Infection in Cervical Disease. *PLoS ONE,* 7**,** e49974.

GUERRERO-FLORES, H., APRESA-GARCIA, T., GARAY-VILLAR, O., SANCHEZ-PEREZ, A., FLORES-VILLEGAS, D., BANDERA-CALDERON, A., GARCIA-PALACIOS, R., ROJAS-SANCHEZ, T., ROMERO-MORELOS, P., SANCHEZ-ALBOR, V., MATA, O., ARANA-CONEJO, V., BADILLO-ROMERO, J., TANIGUCHI, K., MARRERO-RODRIGUEZ, D., MENDOZA-RODRIGUEZ, M., RODRIGUEZ-ESQUIVEL, M., HUERTA-PADILLA, V., MARTINEZ-CASTILLO, A., HERNANDEZ-GALLARDO, I., LOPEZ-ROMERO, R., BANDALA, C., ROSALES-GUEVARA, J. & SALCEDO, M. 2017. A non-invasive tool for detecting cervical cancer odor by trained scent dogs. *BMC Cancer,* 17**,** 79.

GULLIKSEN, A., KEEGAN, H., MARTIN, C., O'LEARY, J., SOLLI, L. A., FALANG, I. M., GRØNN, P., KARLGÅRD, A., MIELNIK, M. M., JOHANSEN, I.-R., TOFTEBERG, T. R., BAIER, T., GRANSEE, R., DRESE, K., HANSEN-HAGGE, T., RIEGGER, L., KOLTAY, P., ZENGERLE, R., KARLSEN, F., AUSEN, D. & FURUBERG, L. 2012. Towards a "Sample-In, Answer-Out" Point-of-Care Platform for Nucleic Acid Extraction and Amplification: Using an HPV E6/E7 mRNA Model System. *Journal of Oncology,* 2012**,** 12.

GUPTA, S., TAKEBE, N. & LORUSSO, P. 2010. Targeting the Hedgehog pathway in cancer. *Therapeutic Advances in Medical Oncology,* 2**,** 237-250.

HAFNER, N., DRIESCH, C., GAJDA, M., JANSEN, L., KIRCHMAYR, R., RUNNEBAUM, I. B. & DURST, M. 2008. Integration of the HPV16 genome

does not invariably result in high levels of viral oncogene transcripts. *Oncogene,* 27**,** 1610-7.

HAGENSEE, M. E., YAEGASHI, N. & GALLOWAY, D. A. 1993. Self-assembly of human papillomavirus type 1 capsids by expression of the L1 protein alone or by coexpression of the L1 and L2 capsid proteins. *Journal of Virology,* 67**,** 315-322.

HAHN, H., WICKING, C., ZAPHIROPOULOUS, P. G., GAILANI, M. R., SHANLEY, S., CHIDAMBARAM, A., VORECHOVSKY, I., HOLMBERG, E., UNDEN, A. B., GILLIES, S., NEGUS, K., SMYTH, I., PRESSMAN, C., LEFFELL, D. J., GERRARD, B., GOLDSTEIN, A. M., DEAN, M., TOFTGARD, R., CHENEVIX-TRENCH, G., WAINWRIGHT, B. & BALE, A. E. 1996. Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome. *Cell,* 85**,** 841-51.

HANNA, A. & SHEVDE, L. A. 2016a. Erratum to: 'Hedgehog signaling: modulation of cancer properties and tumor microenvironment'. *Molecular Cancer,* 15**,** 35.

HANNA, A. & SHEVDE, L. A. 2016b. Hedgehog signaling: modulation of cancer properies and tumor mircroenvironment. *Molecular Cancer,* 15**,** 24.

HARARI, A., CHEN, Z., RODRÍGUEZ, A. C., HILDESHEIM, A., PORRAS, C., HERRERO, R., WACHOLDER, S., PANAGIOTOU, O. A., BEFANO, B.,

BURK, R. D. & SCHIFFMAN, M. 2016. Cross-protection of the Bivalent Human Papillomavirus (HPV) Vaccine Against Variants of Genetically Related High-Risk HPV Infections. *The Journal of Infectious Diseases,* 213**,** 939-947.

HARPER, D., FRANCO, E., WHEELER, C., FERRIS, D., JENKINS, D., SCHUIND, A., ZAHAF, T., INNIS, B., NAUD, P., DE CARVALHO, N., ROTELI-MARTINS, C., TEIXEIRA, J., BLATTER, M., KORN, A., QUINT, W. & DUBIN, G. 2004. Efficacy of a bivalent L1 virus-like particle in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. *Lancet,* 364**,** 1757-1765.

HAWLEY-NELSON, P., VOUSDEN, K. H., HUBBERT, N. L., LOWY, D. R. & SCHILLER, J. T. 1989. HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *The EMBO Journal,* 8**,** 3905-3910.

HE, B., CHEN, G. & ZENG, Y. 2016. Three-dimensional cell culture models for investigating human viruses. *Virologica Sinica,* 31**,** 363-379.

HELLEMANS, J., MORTIER, G., DE PAEPE, A., SPELEMAN, F. & VANDESOMPELE, J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology,* 8**,** R19-R19.

HENRY, N. L & HAYES, D. F. 2012. Cancer Biomarkers. *Molecular Oncology*, 6 140-146

HERFS, M., VARGAS, S. O., YAMAMOTO, Y., HOWITT, B. E., NUCCI, M. R., HORNICK, J. L., MCKEON, F. D., XIAN, W. & CRUM, C. P. 2013. A novel blueprint for 'top down' differentiation defines the cervical squamocolumnar junction during development, reproductive life, and neoplasia. *J Pathol,* 229**,** 460-8.

HERFS, M., YAMAMOTO, Y., LAURY, A., WANG, X., NUCCI, M. R., MCLAUGHLIN-DRUBIN, M. E., MÜNGER, K., FELDMAN, S., MCKEON, F. D., XIAN, W. & CRUM, C. P. 2012. A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer. *Proceedings of the National Academy of Sciences,* 109**,** 10516-10521.

HIBBITTS, S., RIECK, G. C., HART, K., POWELL, N. G., BEUKENHOLDT, R., DALLIMORE, N., MCREA, J., HAUKE, A., TRISTRAM, A. & FIANDER, A. N. 2006. Human papillomavirus infection: an anonymous prevalence study in South Wales, UK. *Br J Cancer,* 95**,** 226-32.

HIBBITTS, S., TRISTRAM, A., BEER, H., MCREA, J., ROSE, B., HAUKE, A., NUTTALL, D., DALLIMORE, N., NEWCOMBE, R. G. & FIANDER, A. 2014. UK population based study to predict impact of HPV vaccination. *J Clin Virol,* 59**,** 109-14.

HINKLE, D. E., WIERSMA, W. & JURS, S. G. 2003. Applied statistics for the behavioral sciences. Boston, Mass, Houghton Mifflin.

HOLLAND, D., HOPPE-SEYLER, K., SCHULLER, B., LOHREY, C., MAROLDT, J., DURST, M. & HOPPE-SEYLER, F. 2008. Activation of the enhancer of zeste homologue 2 gene by the human papillomavirus E7 oncoprotein. *Cancer Res,* 68**,** 9964-72.

HOLOGIC. 2010. *Introducting Cervicta HPV* [Online]. thinprep.com: Hologic. Available: http://www.thinprep.com/hcp/cervista_hpv.html [Accessed 12/05/2017 2017].

HOPMAN, A. H., KAMPS, M. A., SMEDTS, F., SPEEL, E. J., HERRINGTON, C. S. & RAMAEKERS, F. C. 2005. HPV in situ hybridization: impact of different protocols on the detection of integrated HPV. *Int J Cancer,* 115**,** 419-28.

HOUSELEY, J. & TOLLERVEY, D. 2009. The Many Pathways of RNA Degradation. *Cell,* 136**,** 763-776.

HOWIE, H. L., KATZENELLENBOGEN, R. A. & GALLOWAY, D. A. 2009. Papillomavirus E6 proteins. *Virology,* 384**,** 324-34.

HU, C. W., LEE, H., CHEN, J. L., LI, Y. J. & CHAO, M. R. 2013. Optimization of global DNA methylation measurement by LC-MS/MS and its application in lung cancer patients. *Anal Bioanal Chem,* 405**,** 8859-69.

HU, Z., ZHU, D., WANG, W., LI, W., JIA, W., ZENG, X., DING, W., YU, L., WANG, X., WANG, L., SHEN, H., ZHANG, C., LIU, H., LIU, X., ZHAO, Y., FANG, X., LI, S., CHEN, W., TANG, T., FU, A., WANG, Z., CHEN, G., GAO, Q., LI, S., XI, L., WANG, C., LIAO, S., MA, X., WU, P., LI, K., WANG, S., ZHOU, J., WANG, J., XU, X., WANG, H. & MA, D. 2015. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat Genet,* 47**,** 158-163.

HUDELIST, G., MANAVI, M., PISCHINGER, K. I. D., WATKINS-RIEDEL, T., SINGER, C. F., KUBISTA, E. & CZERWENKA, K. F. 2004. Physical state and expression of HPV DNA in benign and dysplastic cervical tissue: different levels of viral integration are correlated with lesion grade. *Gynecologic oncology,* 92**,** 873-880.

HUIJSMANS, C. J., DAMEN, J., VAN DER LINDEN, J. C., SAVELKOUL, P. H. & HERMANS, M. H. 2010. Comparative analysis of four methods to extract DNA from paraffin-embedded tissues: effect on downstream molecular applications. *BMC Res Notes,* 3**,** 239.

IKENBERG, H., BERGERON, C., SCHMIDT, D., GRIESSER, H., ALAMEDA, F., ANGELONI, C., BOGERS, J., DACHEZ, R., DENTON, K., HARIRI, J., KELLER, T., VON KNEBEL DOEBERITZ, M., NEUMANN, H. H., PUIG-TINTORE, L. M., SIDERI, M., REHM, S. & RIDDER, R. 2013. Screening for cervical cancer precursors with p16/Ki-67 dual-stained cytology: results of the PALMS study. *J Natl Cancer Inst,* 105**,** 1550-7.

INNOGENETICS 2013. Human Papillomavirus INNO-LiPA HPV Genotyping Extra II Based on SPF10. Fujirebio.com. Avaliable at: https://www.fujirebio-europe.com/products-services/product-browser/inno-lipar-hpv-genotyping-extra-ii [Accessed: 15/01/2017]

INVITROGEN 2004. SuperScript III Reverse Transcriptase. *In:* INVITROGEN (ed.). Thermo Fisher Scientific, Avaliable at: https://www.thermofisher.com/order/catalog/product/18080093 [Accessed 15/01/2017]

IWAGAMI, S., BABA, Y., WATANABE, M., SHIGAKI, H., MIYAKE, K., IDA, S., NAGAI, Y., ISHIMOTO, T., IWATSUKI, M., SAKAMOTO, Y., MIYAMOTO, Y. & BABA, H. 2012. Pyrosequencing assay to measure LINE-1 methylation level in esophageal squamous cell carcinoma. *Ann Surg Oncol,* 19**,** 2726-32.

JACOBS, M. V., SNIJDERS, P. J., VAN DEN BRULE, A. J., HELMERHORST, T. J., MEIJER, C. J. & WALBOOMERS, J. M. 1997. A general primer

GP5+/GP6(+)-mediated PCR-enzyme immunoassay method for rapid detection of 14 high-risk and 6 low-risk human papillomavirus genotypes in cervical scrapings. *J Clin Microbiol,* 35**,** 791-5.

JAGU, S., KWAK, K., SCHILLER, J. T., LOWY, D. R., KLEANTHOUS, H., KALNIN, K., WANG, C., WANG, H. K., CHOW, L. T., HUH, W. K., JAGANATHAN, K. S., CHIVUKULA, S. V. & RODEN, R. B. 2013. Phylogenetic considerations in designing a broadly protective multimeric L2 vaccine. *J Virol,* 87**,** 6127-36.

JANECKA, A., ADAMCZYK, A. & GASINSKA, A. 2015. Comparison of eight commercially available kits for DNA extraction from formalin-fixed paraffin-embedded tissues. *Anal Biochem,* 476**,** 8-10.

JEON, S., ALLEN-HOFFMANN, B. L. & LAMBERT, P. F. 1995. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J Virol,* 69**,** 2989-97.

JIANG, Y., LUCAS, I. & LE BEAU, M. M. 2015. Common Chromosomal Fragile Sites and Cancer. *In:* ROWLEY, J. D., LE BEAU, M. M. & RABBITTS, T. H. (eds.) *Chromosomal Translocations and Genome Rearrangements in Cancer.* Cham: Springer International Publishing.

JIT, M., CHOI, Y. H. & EDMUNDS, W. J. 2008. Economic evaluation of human papillomavirus vaccination in the United Kingdom. *BMJ,* 337**,** a769.

JOHNSON, A. M., MERCER, C. H., BEDDOWS, S., DE SILVA, N., DESAI, S., HOWELL-JONES, R., CARDER, C., SONNENBERG, P., FENTON, K. A., LOWNDES, C. & SOLDAN, K. 2012a. Epidemiology of, and behavioural risk factors for, sexually transmitted human papillomavirus infection in men and women in Britain. *Sexually Transmitted Infections,* 88**,** 212-217.

JOHNSON, M. D., MUELLER, M., GAME, L. & AITMAN, T. J. 2012b. Single nucleotide analysis of cytosine methylation by whole-genome shotgun bisulfite sequencing. *Curr Protoc Mol Biol,* Chapter 21**,** Unit21.23.

JONES, J., POWELL, N. G., TRISTRAM, A., FIANDER, A. N. & HIBBITTS, S. 2009. Comparison of the PapilloCheck DNA micro-array Human Papillomavirus detection assay with Hybrid Capture II and PCR-enzyme immunoassay using the GP5/6+ primer set. *J Clin Virol,* 45**,** 100-4.

JONES, M. H., SINGER, A. & JENKINS, D. 1996. The mildly abnormal cervical smear: patient anxiety and choice of management. *J R Soc Med,* 89**,** 257-60.

JONES, S. 2016. *HPV biology in VIN: viral biomarkers to predict response to treatment.* Doctor of Philosophy PhD, Cardiff University.

JOURA, E. A., GIULIANO, A. R., IVERSEN, O. E., BOUCHARD, C., MAO, C., MEHLSEN, J., MOREIRA, E. D., JR., NGAN, Y., PETERSEN, L. K.,

LAZCANO-PONCE, E., PITISUTTITHUM, P., RESTREPO, J. A., STUART, G., WOELBER, L., YANG, Y. C., CUZICK, J., GARLAND, S. M., HUH, W., KJAER, S. K., BAUTISTA, O. M., CHAN, I. S., CHEN, J., GESSER, R., MOELLER, E., RITTER, M., VUOCOLO, S., LUXEMBOURG, A. & BROAD SPECTRUM, H. P. V. V. S. 2015. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. *N Engl J Med,* 372**,** 711-23.

JOYCE, J. G., TUNG, J. S., PRZYSIECKI, C. T., COOK, J. C., LEHMAN, E. D., SANDS, J. A., JANSEN, K. U. & KELLER, P. M. 1999. The L1 major capsid protein of human papillomavirus type 11 recombinant virus-like particles interacts with heparin and cell-surface glycosaminoglycans on human keratinocytes. *J Biol Chem,* 274**,** 5810-22.

JUNG, M. & PFEIFER, G. P. 2015. Aging and DNA methylation. *BMC Biol,* 13**,** 7.

KASPER, M., REGL, G., FRISCHAUF, A. M. & ABERGER, F. 2006. GLI transcription factors: mediators of oncogenic Hedgehog signalling. *Eur J Cancer,* 42**,** 437-45.

KEAM, S. J. & HARPER, D. M. 2008. Human papillomavirus types 16 and 18 vaccine (recombinant, AS04 adjuvanted, adsorbed) [Cervarix]. *Drugs,* 68**,** 359-72.

KEMP, T. J., HILDESHEIM, A., SAFAEIAN, M., DAUNER, J. G., PAN, Y., PORRAS, C., SCHILLER, J. T., LOWY, D. R., HERRERO, R. & PINTO, L. A. 2011. HPV16/18 L1 VLP vaccine induces cross-neutralizing antibodies that may mediate cross-protection. *Vaccine,* 29**,** 2011-4.

KJELLBERG, L., HALLMANS, G., AHREN, A. M., JOHANSSON, R., BERGMAN, F., WADELL, G., ANGSTROM, T. & DILLNER, J. 2000. Smoking, diet, pregnancy and oral contraceptive use as risk factors for cervical intra-epithelial neoplasia in relation to human papillomavirus infection. *Br J Cancer,* 82**,** 1332-8.

KLAES, R., WOERNER, S. M., RIDDER, R., WENTZENSEN, N., DUERST, M., SCHNEIDER, A., LOTZ, B., MELSHEIMER, P. & VON KNEBEL DOEBERITZ, M. 1999. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. *Cancer Res,* 59**,** 6132-6.

KORBIE, D. J. & MATTICK, J. S. 2008. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat Protoc,* 3**,** 1452-6.

KOTZEV, A. I., KAMENOVA, M. A. & TCHERVENIAKOV, A. P. 2012. Immunohistochemical Expression of Ki-67, PCNA, pRb, p16, p53, Bcl-2 and Bax in Esophageal Adenocarcinoma and Barrett¡¯s Associated Dysplasia. *Journal of Cancer Therapy,* Vol.03No.06**,** 9.

KOUTSKY, L. 1997. Epidemiology of genital human papillomavirus infection. *Am J Med,* 102**,** 3-8.

KRUSKAL, W. H. & WALLIS, W. A. 1952. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association,* 47**,** 583-621.

KUBIK, M. J., PERMENTER, T. & SAREMIAN, J. 2015. Specimen Age Stability for Human Papilloma Virus DNA Testing Using BD SurePath. *Lab Med,* 46**,** 51-4; quiz e13.

KULASINGAM, S. L., KIM, J. J., LAWRENCE, W. F., MANDELBLATT, J. S., MYERS, E. R., SCHIFFMAN, M., SOLOMON, D. & GOLDIE, S. J. 2006. Cost-effectiveness analysis based on the atypical squamous cells of undetermined significance/low-grade squamous intraepithelial lesion Triage Study (ALTS). *J Natl Cancer Inst,* 98**,** 92-100.

KULIS, M. & ESTELLER, M. 2010. DNA methylation and cancer. *Adv Genet,* 70**,** 27-56.

KUMAR, V., ABBAS, A. K., ASTER, J. C. & PERKINS, J. A. 2015. *Robbins and Cotran pathologic basis of disease.*

KURDYUKOV, S. & BULLOCK, M. 2016. DNA Methylation Analysis: Choosing the Right Method. *Biology,* 5**,** 3.

LAERD STATISTICS. 2015. *Kruskal-Wallis H test using SPSS Statistics.* [Online]. Laerd Statistics. Available: https://statistics.laerd.com/ [Accessed 13/03/2017 2017].

LAERD STATISTICS. 2016. *Chi-square test for independence using SPSS Statistics.* [Online]. Laerd Statistics. Available: https://statistics.laerd.com/ [Accessed 12/04/2017 2017].

LAU, Y. M., CHEUNG, T. H., YEO, W., MO, F., YU, M. Y., LEE, K. M., HO, W. C. S., YEUNG, A. C. M., LAW, P. T. Y. & CHAN, P. K. S. 2015. Prognostic Implication of Human Papillomavirus Types and Species in Cervical Cancer Patients Undergoing Primary Treatment. *PLOS ONE,* 10**,** e0122557.

LEVI, F., LUCCHINI, F., NEGRI, E., FRANCESCHI, S. & LA VECCHIA, C. 2000. Cervical cancer mortality in young women in Europe: patterns and trends. *Eur J Cancer,* 36**,** 2266-71.

LI, K., JIN, X., FANG, Y., WANG, C., GONG, M., CHEN, P., LIU, J., DENG, D. & AI, J. 2012. Correlation between physical status of human papilloma virus and cervical carcinogenesis. *Journal of Huazhong University of Science and Technology [Medical Sciences],* 32**,** 97-102.

LI, N., FRANCESCHI, S., HOWELL-JONES, R., SNIJDERS, P. J. & CLIFFORD, G. M. 2011. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. *Int J Cancer,* 128**,** 927-35.

LI, W., WANG, W., SI, M., HAN, L., GAO, Q., LUO, A., LI, Y., LU, Y., WANG, S. & MA, D. 2008. The physical state of HPV16 infection and its clinical significance in cancer precursor lesion and cervical carcinoma. *Journal of cancer research and clinical oncology,* 134**,** 1355-1361.

LIOUMI, M. & NEWELL, D. 2010. CR-UK biomarker roadmaps. *Clinical Cancer Research,* 16**,** B33-B33.

LIU, X., DAKIC, A., ZHANG, Y., DAI, Y., CHEN, R. & SCHLEGEL, R. 2009. HPV E6 protein interacts physically and functionally with the cellular telomerase complex. *Proc Natl Acad Sci U S A,* 106**,** 18780-5.

LONGWORTH, M. S. & LAIMINS, L. A. 2004. Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiol Mol Biol Rev,* 68**,** 362-72.

LOPALCO, P. L. 2017. Spotlight on the 9-valent HPV vaccine. *Drug Des Devel Ther,* 11**,** 35-44.

LUFT, F., KLAES, R., NEES, M., DURST, M., HEILMANN, V., MELSHEIMER, P. & VON KNEBEL DOEBERITZ, M. 2001. Detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells. *Int J Cancer,* 92**,** 9-17.

MANAWAPAT, A., STUBENRAUCH, F., RUSS, R., MUNK, C., KJAER, S. K. & IFTNER, T. 2012. Physical state and viral load as predictive biomarkersfor persistence and progression of HPV16-positive cervical lesions: results from a population based long-term prospective cohort study. *Am J Cancer Res,* 2**,** 192-203.

MARONGIU, L., GODI, A., PARRY, J. V. & BEDDOWS, S. 2014. Human Papillomavirus 16, 18, 31 and 45 viral load, integration and methylation status stratified by cervical disease stage. *BMC Cancer,* 14**,** 384.

MASTERS, J. R. 2002. HeLa cells 50 years on: the good, the bad and the ugly. *Nat Rev Cancer,* 2**,** 315-9.

MATOVINA, M., SABOL, I., GRUBISIC, G., GASPEROV, N. M. & GRCE, M. 2009. Identification of human papillomavirus type 16 integration sites in high-grade precancerous cervical lesions. *Gynecol Oncol,* 113**,** 120-7.

MATZNER, I., SAVELYEVA, L. & SCHWAB, M. 2003. Preferential integration of a transfected marker gene into spontaneously expressed fragile sites of a breast cancer cell line. *Cancer letters,* 189**,** 207-219.

MAUFORT, J. P., SHAI, A., PITOT, H. C. & LAMBERT, P. F. 2010. A role for HPV16 E5 in cervical carcinogenesis. *Cancer Res,* 70**,** 2924-31.

MAYRAND, M. H., DUARTE-FRANCO, E., COUTLEE, F., RODRIGUES, I., WALTER, S. D., RATNAM, S. & FRANCO, E. L. 2006. Randomized controlled trial of human papillomavirus testing versus Pap cytology in the primary screening for cervical cancer precursors: design, methods and preliminary accrual results of the Canadian cervical cancer screening trial (CCCaST). *Int J Cancer,* 119**,** 615-23.

MCBRIDE, A. A. 2013. The Papillomavirus E2 Proteins. *Virology,* 445**,** 57-79.

MCBRIDE, A. A. & WARBURTON, A. 2017. The role of integration in oncogenic progression of HPV-associated cancers. *PLoS Pathogens,* 13**,** e1006211.

MCLAUGHLIN-DRUBIN, M. E. & MUNGER, K. 2009a. The human papillomavirus E7 oncoprotein. *Virology,* 384**,** 335-44.

MCLAUGHLIN-DRUBIN, M. E. & MUNGER, K. 2009b. Oncogenic activities of human papillomaviruses. *Virus Res,* 143**,** 195-208.

MCPHILLIPS, M. G., OZATO, K. & MCBRIDE, A. A. 2005. Interaction of bovine papillomavirus E2 protein with Brd4 stabilizes its association with chromatin. *J Virol,* 79**,** 8920-32.

MICHALAS, S. P. 2000. The Pap test: George N. Papanicolaou (1883-1962). A screening test for the prevention of cancer of uterine cervix. *Eur J Obstet Gynecol Reprod Biol,* 90**,** 135-8.

MIRABELLO, L., SUN, C., GHOSH, A., RODRIGUEZ, A. C., SCHIFFMAN, M., WENTZENSEN, N., HILDESHEIM, A., HERRERO, R., WACHOLDER, S., LORINCZ, A. & BURK, R. D. 2012. Methylation of human papillomavirus type 16 genome and risk of cervical precancer in a Costa Rican population. *J Natl Cancer Inst,* 104**,** 556-65.

MOLANO, M. 2003. Determinants of Clearance of Human Papillomavirus Infections in Colombian Women with Normal Cytology: A Population-based, 5-Year Follow-up Study. *American Journal of Epidemiology,* 158**,** 486-494.

MOLL, R., DIVO, M. & LANGBEIN, L. 2008. The human keratins: biology and pathology. *Histochem Cell Biol,* 129**,** 705-33.

MOODY, C. A. & LAIMINS, L. A. 2010. Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer,* 10**,** 550-60.

MOORE, K., COFER, A., ELLIOT, L., LANNEAU, G., WALKER, J. & GOLD, M. A. 2007. Adolescent cervical dysplasia: histologic evaluation, treatment, and outcomes. *Am J Obstet Gynecol,* 197**,** 141.e1-6.

MOSS, S., GRAY, A., LEGOOD, R., VESSEY, M., PATNICK, J. & KITCHENER, H. 2006. Effect of testing for human papillomavirus as a triage during screening for cervical cancer: observational before and after study. *Bmj,* 332**,** 83-5.

MUKAKA, M. M. 2012. A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi,* 24**,** 69-71.

MÜLLER, M., PRESCOTT, E. L., WASSON, C. W. & MACDONALD, A. 2015. Human papillomavirus E5 oncoprotein: function and potential target for antiviral therapeutics. *Future Virology,* 10**,** 27-39.

MULLIS, K., FALOONA, F., SCHARF, S., SAIKI, R., HORN, G. & ERLICH, H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol,* 51 Pt 1**,** 263-73.

MUNGER, K., BASILE, J. R., DUENSING, S., EICHTEN, A., GONZALEZ, S. L., GRACE, M. & ZACNY, V. L. 2001. Biological activities and molecular targets of the human papillomavirus E7 oncoprotein. *Oncogene,* 20**,** 7888-98.

MÜNGER, K., PHELPS, W. C., BUBB, V., HOWLEY, P. M. & SCHLEGEL, R. 1989. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *Journal of Virology,* 63**,** 4417-4421.

MUNOZ, N. 2000. Human papillomavirus and cancer: the epidemiological evidence. *J Clin Virol,* 19**,** 1-5.

MUÑOZ , N., BOSCH , F. X., DE SANJOSÉ , S., HERRERO , R., CASTELLSAGUÉ , X., SHAH , K. V., SNIJDERS , P. J. F. & MEIJER , C. J. L. M. 2003. Epidemiologic Classification of Human Papillomavirus Types Associated with Cervical Cancer. *New England Journal of Medicine,* 348**,** 518-527.

MUNSON, E., KROEGER, L., BALZER, S., AMRHEIN, R., MUNSON, K. L., NAPIERALA, M., HUDSPETH, R. & DILLON, P. J. 2014. Comparison of commercial hybridization and automated transcription-mediated amplification modalities for detection of high-risk human papillomavirus nucleic acid. *J Clin Microbiol,* 52**,** 331-4.

NCBI. 2017. *IL12A interleukin 12A [Homo sapiens (human)]* [Online]. Available: https://www.ncbi.nlm.nih.gov/gene/3592 [Accessed 28.05.2017 2017].

NIELSON, C., HARRIS, R., FLORES, R., ABRAHAMSEN, M., PAPENFUSS, M., DUNNE, E., MARKOWITZ, L. & GIULIANO, A. 2009. Multiple-Type Human Papillomavirus Infection in Male Anogenital Sites: Prevalence and Associated Factors. Cancer Epidemiology Biomarkers & Prevention, 18, 1077-1083.

NOREEN, F., RÖÖSLI, M., GAJ, P., PIETRZAK, J., WEIS, S., URFER, P., REGULA, J., SCHÄR, P. & TRUNINGER, K. 2014. Modulation of Age- and Cancer-Associated DNA Methylation Change in the Healthy Colon by Aspirin and Lifestyle. *JNCI: Journal of the National Cancer Institute,* 106**,** dju161-dju161.

ONIONS, T. 2016. *Validation and use of a model system to investigate topical treatment of vulval intraepithelial neoplasia.* Doctor of Philosophy, Cardiff Univerisity.

ORIGONI, M., CRISTOFORONI, P., CARMINATI, G., STEFANI, C., COSTA, S., SANDRI, M. T., MARIANI, L. & PRETI, M. 2015. E6/E7 mRNA testing for human papilloma virus-induced high-grade cervical intraepithelial disease (CIN2/CIN3): a promising perspective. *ecancermedicalscience,* 9**,** 533.

PAAVONEN, J., JENKINS, D., BOSCH, F. X., NAUD, P., SALMERON, J., WHEELER, C. M., CHOW, S. N., APTER, D. L., KITCHENER, H. C., CASTELLSAGUE, X., DE CARVALHO, N. S., SKINNER, S. R., HARPER, D. M., HEDRICK, J. A., JAISAMRARN, U., LIMSON, G. A., DIONNE, M., QUINT, W., SPIESSENS, B., PEETERS, P., STRUYF, F., WIETING, S. L.,

LEHTINEN, M. O. & DUBIN, G. 2007. Efficacy of a prophylactic adjuvanted bivalent L1 virus-like-particle vaccine against infection with human papillomavirus types 16 and 18 in young women: an interim analysis of a phase III double-blind, randomised controlled trial. *Lancet,* 369**,** 2161-70.

PAPANICOLAOU, G. N. & TRAUT, H. F. 1997. The diagnostic value of vaginal smears in carcinoma of the uterus. 1941. *Arch Pathol Lab Med,* 121**,** 211-24.

PATEL, D. A., UNGER, E. R., WALLINE, H., OPIPARI, A. W., LEE, D. R., FLOWERS, L. C. & RUFFIN, M. T. 2011. Lack of HPV 16 and 18 Detection in Serum of Colposcopy Clinic Patients. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology,* 50**,** 342-344.

PATTERSON, K., MOLLOY, L., QU, W. & CLARK, S. 2011. DNA Methylation: Bisulphite Modification and Analysis. *Journal of Visualized Experiments : JoVE***,** 3170.

PATTERSON, N. A., SMITH, J. L. & OZBUN, M. A. 2005. Human papillomavirus type 31b infection of human keratinocytes does not require heparan sulfate. *J Virol,* 79**,** 6838-47.

PAZ, M. F., FRAGA, M. F., AVILA, S., GUO, M., POLLAN, M., HERMAN, J. G. & ESTELLER, M. 2003. A systematic profile of DNA methylation in human cancer cell lines. *Cancer Res,* 63**,** 1114-21.

PEDERSEN, K., SORBYE, S. W., KRISTIANSEN, I. S. & BURGER, E. A. 2017. Using novel biomarkers to triage young adult women with minor cervical lesions: a cost-effectiveness analysis. *BJOG,* 124**,** 474-484.

PENG, Y. & CROCE, C. M. 2016. The role of MicroRNAs in human cancer. *Signal Transduction And Targeted Therapy,* 1**,** 15004.

PETO, J., GILHAM, C., DEACON, J., TAYLOR, C., EVANS, C., BINNS, W., HAYWOOD, M., ELANKO, N., COLEMAN, D., YULE, R. & DESAI, M. 2004a. Cervical HPV infection and neoplasia in a large population-based prospective study: the Manchester cohort. *Br J Cancer,* 91**,** 942-53.

PETO, J., GILHAM, C., FLETCHER, O. & MATTHEWS, F. E. 2004b. The cervical cancer epidemic that screening has prevented in the UK. *The Lancet,* 364**,** 249-256.

PETO, J., GILHAM, C., FLETCHER, O. & MATTHEWS, F. E. 2004c. The cervical cancer epidemic that screening has prevented in the UK. *Lancet,* 364**,** 249-56.

PETRY, K. U., SCHMIDT, D., SCHERBRING, S., LUYTEN, A., REINECKE-LÜTHGE, A., BERGERON, C., KOMMOSS, F., LÖNING, T., ORDI, J., REGAUER, S. & RIDDER, R. 2011. Triaging Pap cytology negative, HPV

positive cervical cancer screening results with p16/Ki-67 Dual-stained cytology. *Gynecologic Oncology,* 121**,** 505-509.

PETT, M. & COLEMAN, N. 2007. Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? *J Pathol,* 212**,** 356-67.

PETT, M. R. 2004. Acquisition of High-Level Chromosomal Instability Is Associated with Integration of Human Papillomavirus Type 16 in Cervical Keratinocytes. *Cancer Research,* 64**,** 1359-1368.

PINEROS, M., HERNANDEZ-SUAREZ, G., ORJUELA, L., VARGAS, J. C. & PEREZ, G. 2013. HPV knowledge and impact of genital warts on self esteem and sexual life in Colombian patients. *BMC Public Health,* 13**,** 272.

PIYATHILAKE, C. J., BADIGA, S., CHAMBERS, M. M., BRILL, I. K., MATTHEWS, R. & PARTRIDGE, E. E. 2016. Accuracy of urinary human papillomavirus testing for the presence of cervical human papillomaviruses and higher grades of cervical intraepithelial neoplasia. *Cancer,* 122**,** 2836-44.

PORT, R. J., PINHEIRO-MAIA, S., HU, C., ARRAND, J. R., WEI, W., YOUNG, L. S. & DAWSON, C. W. 2013. Epstein-Barr virus induction of the Hedgehog signalling pathway imposes a stem cell phenotype on human epithelial cells. *J Pathol,* 231**,** 367-77.

QIAGEN 2012. PyroMark Q24 Vacuum Workstation quick-start guide. *In:* QIAGEN (ed.) 12/2012 ed.: QIAGEN. Avaliable at: https://www.qiagen.com/us/resources/download.aspx?id=41eed6af-3e3d-4085-ae1b-3607e80539ea&lang=en [Accessed: 4/5/2017]

QIAGEN. 2017a. *Principle of Hybrid Capture 2 (HC2) Technology* [Online]. qiagen.com. Available: https://www.qiagen.com/us/resources/knowledge-area/hpv-na/principle%20of%20hc2%20technology/ [Accessed 16/05/2017 2017].

QIAGEN. 2017b. *Pyrosequencing Technology and Platform Overview* [Online]. QIAGEN. Available: https://www.qiagen.com/at/resources/technologies/pyrosequencing-resource-center/technology-overview/ [Accessed 19/05/2017 2017].

QIAGEN. 2017c. *RNAprotect Cell Reagent* [Online]. QIAGEN.com. Available: https://www.qiagen.com/gb/shop/sample-technologies/rna/rnaprotect-cell-reagent/ [Accessed 14/02/2017 2017].

RAMANAKUMAR, A. V., GONCALVES, O., RICHARDSON, H., TELLIER, P., FERENCZY, A., COUTLEE, F. & FRANCO, E. L. 2010. Human papillomavirus (HPV) types 16, 18, 31, 45 DNA loads and HPV-16 integration in persistent and transient infections in young women. *BMC Infect Dis,* 10**,** 326.

RAO, A., YOUNG, S., ERLICH, H., BOYLE, S., KREVOLIN, M., SUN, R., APPLE, R. & BEHRENS, C. 2013. Development and Characterization of the cobas Human Papillomavirus Test. *Journal of Clinical Microbiology,* 51**,** 1478-1484.

RAYBOULD, R. 2013. *Human Papillomavirus integration: The mechanism(s) behind the high-risk associated with this event and cervical disease progression.* Doctor of Philosophy, Cardiff University.

ROBERTS, S., HILLMAN, M. L., KNIGHT, G. L. & GALLIMORE, P. H. 2003. The ND10 component promyelocytic leukemia protein relocates to human papillomavirus type 1 E4 intranuclear inclusion bodies in cultured keratinocytes and in warts. *J Virol,* 77**,** 673-84.

ROBERTSON, K. D. 2005. DNA methylation and human disease. *Nat Rev Genet,* 6**,** 597-610.

ROCHE 2010. Proteinase K, recombinant, PCR Grade. *In:* DIAGNOSTICS, R. (ed.) 4 ed. Sigmaaldrich.com: Roche. Avaliable at: http://www.sigmaaldrich.com/catalog/product/roche/rprotksolro?lang=en&region=GB [Accessed: 26/4/2017]

ROCHE MOLECULAR DIAGNOSTICS. 2017. *cobas HPV* [Online]. Roche Molecular Systems Inc. Available: https://molecular.roche.com/assays/cobas-hpv-test/ [Accessed 16/05/2017 2017].

ROMAGOSA, C., SIMONETTI, S., LOPEZ-VICENTE, L., MAZO, A., LLEONART, M. E., CASTELLVI, J. & RAMON Y CAJAL, S. 2011. p16Ink4a overexpression in cancer: a tumor suppressor gene associated with senescence and high-grade tumors. *Oncogene,* 30**,** 2087-2097.

ROUS, P. & BEARD, J. W. 1934. A virus-induced mammalian growth with the characters of a tumor (the shope rabbit papilloma) : I. The growth on implantation within favorable hosts. *J Exp Med,* 60**,** 701-22.

ROWSON, K. E. & MAHY, B. W. 1967. Human papova (wart) virus. *Bacteriological Reviews,* 31**,** 110-131.

RUIZ I ALTABA, A. 1999. Gli proteins and Hedgehog signaling: development and cancer. *Trends Genet,* 15**,** 418-25.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics (Oxford, England),* 16**,** 944-945.

SAMARZIJA, I. & BEARD, P. 2012. Hedgehog pathway regulators influence cervical cancer cell proliferation, survival and migration. *Biochem Biophys Res Commun,* 425**,** 64-9.

SARI ASLANI, F., SAFAEI, A., POURJABALI, M. & MOMTAHAN, M. 2013. Evaluation of Ki67, p16 and CK17 Markers in Differentiating Cervical Intraepithelial Neoplasia and Benign Lesions. *Iranian Journal of Medical Sciences,* 38**,** 15-21.

SARKAR, G., TURNER, R. T. & BOLANDER, M. E. 1993. Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. *Genome Research,* 2**,** 318-322.

SCHEFFNER, M., WERNESS, B. A., HUIBREGTSE, J. M., LEVINE, A. J. & HOWLEY, P. M. 1990. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell,* 63**,** 1129-36.

SCHMEINK, C. E., MASSUGER, L. F., LENSELINK, C. H., QUINT, W. G., WITTE, B. I., BERKHOF, J., MELCHERS, W. J. & BEKKERS, R. L. 2013. Prospective follow-up of 2,065 young unscreened women to study human papillomavirus incidence and clearance. *Int J Cancer,* 133**,** 172-81.

SEDLAK, R. H., COOK, L., HUANG, M.-L., MAGARET, A., ZERR, D. M., BOECKH, M. & JEROME, K. R. 2014. Identification of Chromosomally Integrated Human Herpesvirus 6 by Droplet Digital PCR. *Clinical chemistry,* 60**,** 765-772.

SELLORS, J. W. & SANKARANARAYANAN, R. (eds.) 2003. *Colposcopy and Treatment of Cervical Intraepithelial Neoplasia: A Beginners' Manual* France: International Agency for Research on Cancer.

SENAPATI, R., SENAPATI, N. N. & DWIBEDI, B. 2016. Molecular mechanisms of HPV mediated neoplastic progression. *Infect Agent Cancer,* 11**,** 59.

SEPETIENE, A., GUDLEVICIENĖ, Z., BUMBULIENE, Z., DRASUTIENE, G. & DIDZIAPETRIENE, J. 2011. State of HPV16 integration in Lithuanian women with cervical neoplasia. *Central European Journal of Medicine,* 6**,** 205-212.

SHAFTI-KERAMAT, S., HANDISURYA, A., KRIEHUBER, E., MENEGUZZI, G., SLUPETZKY, K. & KIRNBAUER, R. 2003. Different Heparan Sulfate Proteoglycans Serve as Cellular Receptors for Human Papillomaviruses. *Journal of Virology,* 77**,** 13125-13135.

SHRIDHAR, K., WALIA, G. K., AGGARWAL, A., GULATI, S., GEETHA, A. V., PRABHAKARAN, D., DHILLON, P. K. & RAJARAMAN, P. 2016. DNA methylation markers for oral pre-cancer progression: A critical review. *Oral Oncol,* 53**,** 1-9.

SOLOMON, D., SCHIFFMAN, M. & TARONE, R. 2001. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. *J Natl Cancer Inst,* 93**,** 293-9.

SPARDY, N., COVELLA, K., CHA, E., HOSKINS, E. E., WELLS, S. I., DUENSING, A. & DUENSING, S. 2009. Human papillomavirus 16 E7 oncoprotein attenuates DNA damage checkpoint control by increasing the proteolytic turnover of claspin. *Cancer Res,* 69**,** 7022-9.

SPECTOR, M. E., SACCO, A. G., BELLILE, E. L., TAYLOR, J. M. G., JONES, T. R., SUN, K., BROWN, W. C., BIRKELAND, A. C., BRADFORD, C. R., WOLF, G. T., PRINCE, M. E., MOYER, J. S., MALLOY, K. M., SWIECICKI, P., EISBRUCH, A., MCHUGH, J. B., CHEPEHA, D. B., ROZEK, L. S. & WORDEN, F. P. 2016. E6 and E7 antibody levels are potential biomarkers of recurrence in patients with advanced stage human papillomavirus positive oropharyngeal squamous cell carcinoma. *Clinical Cancer Research*, 23 (11): 2723 - 2729

STANLEY, M. 2015. Immunology of HPV Infection. *Current Obstetrics and Gynecology Reports,* 4**,** 195-200.

STANLEY, M. A. 2002. Culture of Human Cervical Epithelial Cells. *Culture of Epithelial Cells.* John Wiley & Sons, Inc.

STANLEY, M. A., BROWNE, H. M., APPLEBY, M. & MINSON, A. C. 1989a. Properties of a non-tumorigenic human cervical keratinocyte cell line. *International Journal of Cancer,* 43**,** 672-676.

STANLEY, M. A., BROWNE, H. M., APPLEBY, M. & MINSON, A. C. 1989b. Properties of a non-tumorigenic human cervical keratinocyte cell line. *Int J Cancer,* 43**,** 672-6.

STEIN, R. A. 2011. DNA methylation profiling: a promising tool and a long road ahead for clinical applications. *Int J Clin Pract,* 65**,** 1212-3.

STUNKEL, W. & BERNARD, H. U. 1999. The chromatin structure of the long control region of human papillomavirus type 16 represses viral oncoprotein expression. *J Virol,* 73**,** 1918-30.

SUPRYNOWICZ, F. A., KRAWCZYK, E., HEBERT, J. D., SUDARSHAN, S. R., SIMIC, V., KAMONJOH, C. M. & SCHLEGEL, R. 2010. The human papillomavirus type 16 E5 oncoprotein inhibits epidermal growth factor trafficking independently of endosome acidification. *J Virol,* 84**,** 10619-29.

SURGET-GROBA, Y. & MONTOYA-BURGOS, J. 2010. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Research,* 20**,** 1432-1440.

TAN, S. C. & YIAP, B. C. 2009. DNA, RNA, and Protein Extraction: The Past and The Present. *Journal of Biomedicine and Biotechnology,* 2009.

TAN, S. H., LEONG, L. E., WALKER, P. A. & BERNARD, H. U. 1994. The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. *J Virol,* 68**,** 6411-20.

TARDIF, K. D., PYNE, M. T., MALMBERG, E., LUNT, T. C. & SCHLABERG, R. 2016. Cervical Cytology Specimen Stability in Surepath Preservative and Analytical Sensitivity for HPV Testing with the cobas and Hybrid Capture 2 Tests. *PLoS One,* 11**,** e0149611.

THE CANCER GENOME ATLAS RESEARCH, N. 2017. Integrated genomic and molecular characterization of cervical cancer. *Nature,* 543**,** 378-384.

THERMO SCIENTIFIC 2012. Interpretation of Nucleic Acid 260/280 Ratios. Rev 1/2012 ed.: Thermo Scientific. Avaliable at: https://tools.thermofisher.com/content/sfs/brochures/T123-NanoDrop-Lite-Interpretation-of-Nucleic-Acid-260-280-Ratios.pdf [Accessed: 13/01/2017]

THIERRY, F. 2009. Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. *Virology,* 384**,** 375-9.

THIERRY, F. & YANIV, M. 1987. The BPV1-E2 trans-acting protein can be either an activator or a repressor of the HPV18 regulatory region. *The EMBO Journal,* 6**,** 3391-3397.

THIERY, J. P., ACLOQUE, H., HUANG, R. Y. J. & NIETO, M. A. 2009. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell,* 139**,** 871-890.

THOMAS, J. T., HUBERT, W. G., RUESCH, M. N. & LAIMINS, L. A. 1999. Human papillomavirus type 31 oncoproteins E6 and E7 are required for the maintenance of episomes during the viral life cycle in normal human keratinocytes. *Proc Natl Acad Sci U S A,* 96**,** 8449-54.

THOMAS, M. & BANKS, L. 1998. Inhibition of Bak-induced apoptosis by HPV-18 E6. *Oncogene,* 17**,** 2943-54.

THORLAND, E. C., MYERS, S. L., GOSTOUT, B. S. & SMITH, D. I. 2003. Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. *Oncogene,* 22**,** 1225-37.

TINDLE, R. W. 2002. Immune evasion in human papillomavirus-associated cervical cancer. *Nat Rev Cancer,* 2**,** 59-64.

TJALMA, W. A. A., TRINH, X. B., ROSENLUND, M., MAKAR, A. P., KRIDELKA, F., ROSILLON, D., VAN DAM, P. A., COLLAS DE SOUZA, S., HOLL, K., SIMON, P. & JENKINS, D. 2015. A cross-sectional, multicentre, epidemiological study on human papillomavirus (HPV) type distribution in adult women diagnosed with invasive cervical cancer in Belgium. *Facts, Views & Vision in ObGyn,* 7**,** 101-108.

TOST, J. & GUT, I. G. 2007. DNA methylation analysis by pyrosequencing. *Nat Protoc,* 2**,** 2265-75.

TURNER, D. O., WILLIAMS-COCKS, S. J., BULLEN, R., CATMULL, J., FALK, J., MARTIN, D., MAUER, J., BARBER, A. E., WANG, R. C., GERSTENBERGER, S. L. & KINGSLEY, K. 2011. High-risk human papillomavirus (HPV) screening and detection in healthy patient saliva samples: a pilot study. *BMC Oral Health,* 11**,** 28.

UK NATIONAL SCREENING COMMITTEE. 2016 UK NSC HPV reccomendation. Avaliable: https://legacyscreening.phe.org.uk/policydb_download.php?doc=594 [Accessed: 20/03/2017]

VAN BAARS, R., GRIFFIN, H., WU, Z., SONEJI, Y. J., VAN DE SANDT, M., ARORA, R., VAN DER MAREL, J., TER HARMSEL, B., JACH, R., OKON, K., HURAS, H., JENKINS, D., QUINT, W. & DOORBAR, J. 2015. Investigating Diagnostic Problems of CIN 1 and 2 Associated with High-Risk HPV by Combining the Novel Molecular Biomarker PanHPV E4 with P16(ink4a). *The American journal of surgical pathology,* 39**,** 1518-1528.

VAN BAARS, R., VAN DER MAREL, J., SNIJDERS, P. J. F., RODRIQUEZ-MANFREDI, A., TER HARMSEL, B., VAN DEN MUNCKHOF, H. A. M., ORDI, J., DEL PINO, M., VAN DE SANDT, M. M., WENTZENSEN, N., MEIJER, C. J. L. M. & QUINT, W. G. V. 2016. CADM1 and MAL methylation status in cervical scrapes is representative of the most severe underlying lesion in women with multiple cervical biopsies. *International Journal of Cancer,* 138**,** 463-471.

VAN TINE, B. A., KAPPES, J. C., BANERJEE, N. S., KNOPS, J., LAI, L., STEENBERGEN, R. D., MEIJER, C. L., SNIJDERS, P. J., CHATIS, P., BROKER, T. R., MOEN, P. T., JR. & CHOW, L. T. 2004. Clonal selection for transcriptionally active viral oncogenes during progression to cancer. *J Virol,* 78**,** 11172-86.

VANDESOMPELE, J., DE PRETER, K., PATTYN, F., POPPE, B., VAN ROY, N., DE PAEPE, A. & SPELEMAN, F. 2002. Accurate normalization of real-time

quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol,* 3**,** Research0034.

VARJOSALO, M. & TAIPALE, J. 2008. Hedgehog: functions and mechanisms. *Genes Dev,* 22**,** 2454-72.

VENCESLAU, E. M., BEZERRA, M. M., LOPES, A. C. M., SOUZA, É. V., ONOFRE, A. S. C., MELO, C. M. D., JERALDO, V. D. L. S. & ONOFRE, F. B. D. M. 2014. HPV detection using primers MY09/MY11 and GP5+/GP6+ in patients with cytologic and/or colposcopic changes. *Jornal Brasileiro de Patologia e Medicina Laboratorial,* 50**,** 280-285.

VIEIRA, L. & ALMEIDA, A. 2013. The cytology and DNA detection by the PapilloCheck(®) test in the diagnosis of human papillomavirus infection. *European Journal of Microbiology & Immunology,* 3**,** 61-67.

VILLA, L. L., COSTA, R. L., PETTA, C. A., ANDRADE, R. P., AULT, K. A., GIULIANO, A. R., WHEELER, C. M., KOUTSKY, L. A., MALM, C., LEHTINEN, M., SKJELDESTAD, F. E., OLSSON, S. E., STEINWALL, M., BROWN, D. R., KURMAN, R. J., RONNETT, B. M., STOLER, M. H., FERENCZY, A., HARPER, D. M., TAMMS, G. M., YU, J., LUPINACCI, L., RAILKAR, R., TADDEO, F. J., JANSEN, K. U., ESSER, M. T., SINGS, H. L., SAAH, A. J. & BARR, E. 2005. Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young

women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. *Lancet Oncol,* 6**,** 271-8.

VILLA, L. L. & DENNY, L. 2006. CHAPTER 7 Methods for detection of HPV infection and its clinical utility. *International Journal of Gynecology & Obstetrics,* 94**,** S71-S80.

VINOKUROVA, S., WENTZENSEN, N., KRAUS, I., KLAES, R., DRIESCH, C., MELSHEIMER, P., KISSELJOV, F., DURST, M., SCHNEIDER, A. & VON KNEBEL DOEBERITZ, M. 2008. Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. *Cancer Res,* 68**,** 307-13.

VIRMANI, A. K., MULLER, C., RATHI, A., ZOECHBAUER-MUELLER, S., MATHIS, M. & GAZDAR, A. F. 2001. Aberrant Methylation during Cervical Carcinogenesis. *Clinical Cancer Research,* 7**,** 584-589.

VISHNOI, K., MAHATA, S., TYAGI, A., PANDEY, A., VERMA, G., JADLI, M., SINGH, T., SINGH, S. M. & BHARTI, A. C. 2016. Cross-talk between Human Papillomavirus Oncoproteins and Hedgehog Signaling Synergistically Promotes Stemness in Cervical Cancer Cells. *Sci Rep,* 6**,** 34377.

WALBOOMERS, J. M., JACOBS, M. V., MANOS, M. M., BOSCH, F. X., KUMMER, J. A., SHAH, K. V., SNIJDERS, P. J., PETO, J., MEIJER, C. J. & MUNOZ, N.

1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol,* 189**,** 12-9.

WANG, H. K., DUFFY, A. A., BROKER, T. R. & CHOW, L. T. 2009a. Robust production and passaging of infectious HPV in squamous epithelium of primary human keratinocytes. *Genes Dev,* 23**,** 181-94.

WANG, Q., GRIFFIN, H., SOUTHERN, S., JACKSON, D., MARTIN, A., MCINTOSH, P., DAVY, C., MASTERSON, P. J., WALKER, P. A., LASKEY, P., OMARY, M. B. & DOORBAR, J. 2003. Functional Analysis of the Human Papillomavirus Type 16 E1 E4 Protein Provides a Mechanism for In Vivo and In Vitro Keratin Filament Reorganization. *Journal of Virology,* 78**,** 821-833.

WANG, Z., GERSTEIN, M. & SNYDER, M. 2009b. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet,* 10**,** 57-63.

WAXMAN, A. G. 2008. Cervical cancer screening in the early post vaccine era. *Obstet Gynecol Clin North Am,* 35**,** 537-48; vii.

WEIDNER, C. I. & WAGNER, W. 2014. The epigenetic tracks of aging. *Biol Chem,* 395**,** 1307-14.

WELSH ASSEMBLY. 2007. *Vaccination against cervical cancer to be introduced in Wales* [Online]. Available: www.wales.nhs.uk/documents/W070828-Hlt.pdf [Accessed].

WENTZENSEN, N., ARBYN, M., BERKHOF, J., BOWER, M., CANFELL, K., EINSTEIN, M., FARLEY, C., MONSONEGO, J. & FRANCESCHI, S. 2017. Eurogin 2016 Roadmap: how HPV knowledge is changing screening practice. *Int J Cancer,* 140**,** 2192-2200.

WENTZENSEN, N., SUN, C., GHOSH, A., KINNEY, W., MIRABELLO, L., WACHOLDER, S., SHABER, R., LAMERE, B., CLARKE, M., LORINCZ, A. T., CASTLE, P. E., SCHIFFMAN, M. & BURK, R. D. 2012. Methylation of HPV18, HPV31, and HPV45 genomes and cervical intraepithelial neoplasia grade 3. *J Natl Cancer Inst,* 104**,** 1738-49.

WENTZENSEN, N., VINOKUROVA, S. & VON KNEBEL DOEBERITZ, M. 2004. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. *Cancer Res,* 64**,** 3878-84.

WENTZENSEN, N. & VON KNEBEL, D. M. 2007. Biomarkers in cervical cancer screening. *Dis Markers,* 23.

WESTRICH, J. A., WARREN, C. J. & PYEON, D. 2017. Evasion of host immune defenses by human papillomavirus. *Virus Res,* 231**,** 21-33.

WILSON, R., RYAN, G. B., KNIGHT, G. L., LAIMINS, L. A. & ROBERTS, S. 2007. The full-length E1E4 protein of human papillomavirus type 18 modulates differentiation-dependent viral DNA amplification and late gene expression. *Virology,* 362**,** 453-60.

WILSON, V. G., WEST, M., WOYTEK, K. & RANGASAMY, D. 2002. Papillomavirus E1 proteins: form, function, and features. *Virus Genes,* 24**,** 275-90.

WIMARDHANI, Y. S., SASANTI, H., WARDHANY, II, SARSITO, A. S., PRADONO, S. A., SUBITA, G. P., SOEGYANTO, A. I., RAHMAYANTI, F., CHAMUSRI, N. & IAMAROON, A. 2015. Saliva-Based Screening of High-Risk Human Papillomavirus Strains: Detection in Female Indonesian and Thai Dental Students. *Asian Pac J Cancer Prev,* 16**,** 5525-9.

WINER, R., LEE, S., HUGHES, J., ADAM, D., KIVIAT, N. & KOUTSKY, L. 2003. Genital human papillomavirus infection: Incidence and risk factors in a cohort of female university students. *American Journal of Epidemiology,* 157**,**218-226.

WOJDACZ, T. K., DOBROVIC, A. & HANSEN, L. L. 2008. Methylation-sensitive high-resolution melting. *Nat Protoc,* 3**,** 1903-8.

WONG, M. L. & MEDRANO, J. F. 2005. Real-time PCR for mRNA quantitation. *Biotechniques,* 39**,** 75-85.

WOODMAN, C. B., COLLINS, S., ROLLASON, T. P., WINTER, H., BAILEY, A., YATES, M. & YOUNG, L. S. 2003. Human papillomavirus type 18 and rapidly progressing cervical intraepithelial neoplasia. *Lancet,* 361**,** 40-3.

YAJID, A. I., ZAKARIAH, M. A., MAT ZIN, A. A. & OTHMAN, N. H. 2017. Potential Role of E4 Protein in Human Papillomavirus Screening: a Review. *Asian Pac J Cancer Prev,* 18**,** 315-319.

YAN, M., WANG, L., ZUO, H., ZHANG, Z., CHEN, W., MAO, L. & ZHANG, P. 2011. HH/GLI signalling as a new therapeutic target for patients with oral squamous cell carcinoma. *Oral Oncol,* 47**,** 504-9.

YANG, H. J. 2013. Aberrant DNA methylation in cervical carcinogenesis. *Chin J Cancer,* 32**,** 42-8.

YE, J., COULOURIS, G., ZARETSKAYA, I., CUTCUTACHE, I., ROZEN, S. & MADDEN, T. L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics,* 13**,** 134.

ZHAI, L. & TUMBAN, E. 2016. Gardasil-9: A global survey of projected efficacy. *Antiviral Res,* 130**,** 101-9.

ZHANG, Z., FYE, S., BORECKI, I. B. & RADER, J. S. 2014. Polymorphisms in immune mediators associate with risk of cervical cancer. *Gynecol Oncol,* 135**,** 69-73.

ZHENG, Z.-M. & WANG, X. 2011. Regulation of cellular miRNA expression by human papillomaviruses. *Biochimica et biophysica acta,* 1809**,** 668-677.

ZHU, J., NORMAN, I., ELFGREN, K., GABERI, V., HAGMAR, B., HJERPE, A. & ANDERSSON, S. 2007. A comparison of liquid-based cytology and Pap smear as a screening method for cervical cancer. *Oncol Rep,* 18**,** 157-60.

Appendix 1

Prognostic/Predictive Biomarkers (BM) Roadmap

Figure apx1. Prognostic/Preductive Biomarker Roadmap as by Cancer Reaserch UK. This flow diagram details the stages of biomarker development. Potential Biomarkers which fulfil the criteria follow the progress (green arrows) or Biomarkers which fail follow the red arrows to be redefined or rejected.
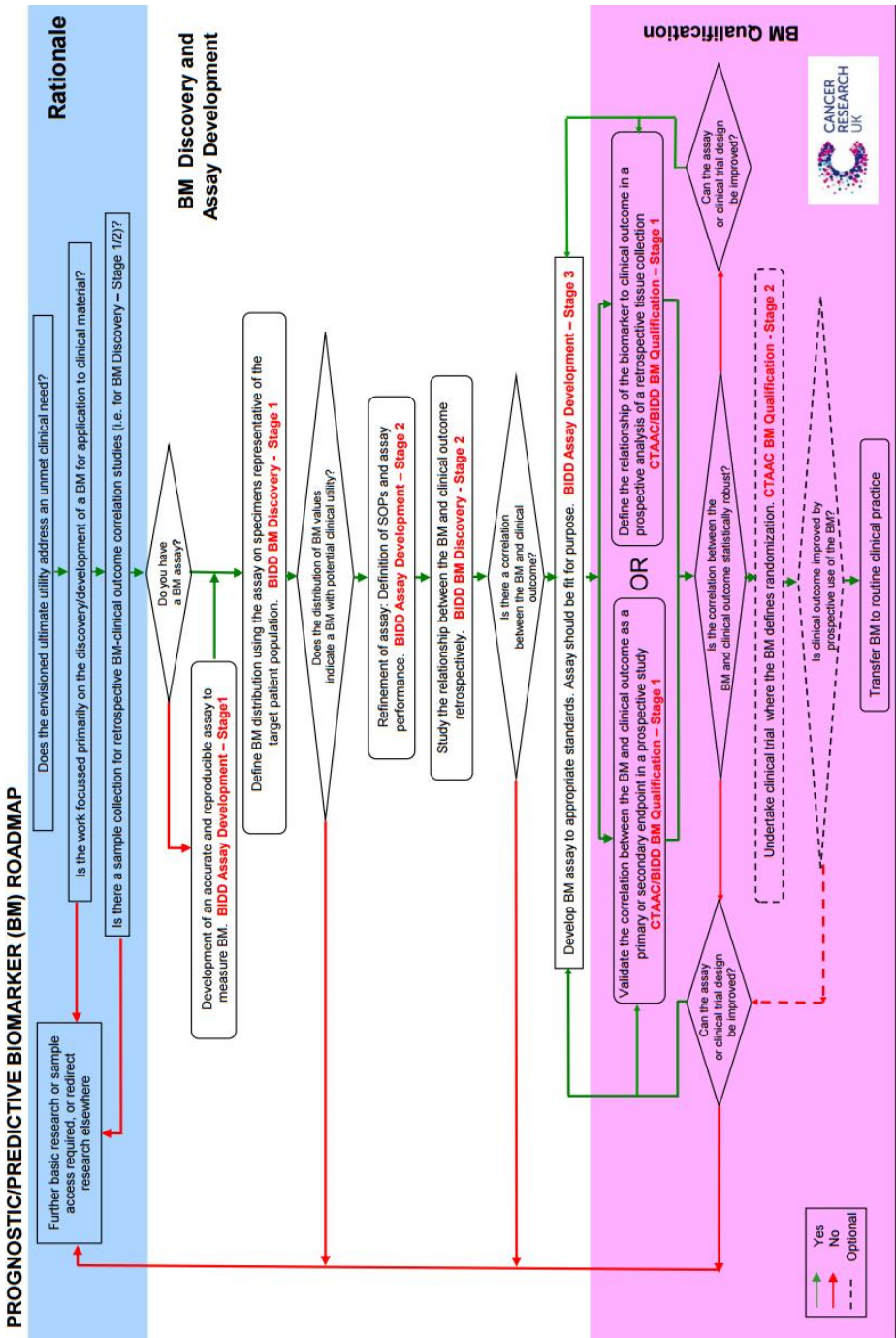
Appendix 2

Study Protocols

Samples from two studies were used in this study:

Study 1: Young women attending their first cervical smear

Study 2: HI-BCD Study Cohort

The Study protocols for each of these studies are included in this appendix.

10.1 Study 1: Young women attending their first cervical smear

Base HPV 2009: Characterisation of HPV infection in cervical smears from young women attending first cervical screening prior to implementation of HPV vaccination

Protocol Summary

Acronym: Base HPV 2009 (Baseline HPV Prevalence in women aged 20-22 yrs 2009)

Principle Clinical Investigator: Professor Alison Fiander

Other Investigators: Dr Sam Hibbitts, Dr Richard Peevor, Jo Jones, Dr Ned Powell, Dr Amanda Tristram, Dr Hilary Fielder, Dr Anne Hauke, Dr Nick Dallimore, Helen Beer, Jane McRea, Bryan Rose, Dr Malcolm Adams.

Study Site: HPV Research Laboratory, School of Medicine, Cardiff University and Cervical Screening Wales (Velindre NHS Trust).

Aim: The aim of the study is to assess and characterise the current HPV burden prior to the implementation of prophylactic HPV vaccination in Wales. This cohort will serve as a 'control' group to compare with vaccinated cohorts both as part of the catch up campaign and the main target group for vaccination.

Objectives Primary:

- To determine baseline Human Papillomavirus (HPV) genotype specific prevalence in women aged 20-22 years attending first cervical screening in Wales, prior to the introduction of an HPV vaccination programme.

Objectives Secondary

- To correlate HPV type-specific infection with cytology and histology where available.
- To correlate HPV type-specific infection with age.
- To correlate HPV type-specific infection with social deprivation score according to the Welsh Index of Multiple Deprivation (WIMD)

Objectives Tertiary:

- Development of 'bio-bank' to investigate potential prognostic biomarkers.

Methodology: This is a pseudo-anonymous prospective cohort study design

Number of Subjects: 250 samples per week up to 10,000 per annum

Inclusion Criteria:

- All women attending for their first routine cervical smear (20-22 yrs) within Cervical Screening Wales (CSW) in 2009.
- Samples from women who have an inadequate first routine cervical smear (20-22yrs) and are re-called for a repeat smear.

Exclusion Criteria:

283

- Women attending for any other routine cervical smear Inadequate Cytology result

- No information available for age, cytology or social deprivation score

Baseline HPV Prevalence in women aged 20-22 yrs 2009 (Base HPV 2009)

Characterisation of HPV infection in cervical smears from young women attending first cervical screening prior to implementation of HPV vaccination

Contents

1.0   Lay Summary

2.0   Introduction

2.1   Human Papillomavirus

2.2   Cervical Screening

2.3   HPV vaccination

2.4   Objectives

2.5   Rationale

3.0 Investigational Plan

3.1 Overall Study Design

3.2 Study Population

3.3 Inclusion and Exclusion Criteria

3.4 Sample Collection

3.5 Interventions

3.6 Blinding

3.7 Study Plan

3.8 Endpoints

4.0 Sample Management

4.1 Cytological analysis

4.2 HPV Testing

5.0 Data Management

5.1 Statistical Methods

5.2 Sample Size and Power Calculations

6.0 References

7.0 Project Management

8.0 Ethical Considerations

8.1   Administrative Procedures

8.2   Monitoring Procedures

8.3   Recording of Data and Retention of Documents

Lay Summary

Human Papillomavirus (HPV) is the underlying cause of cervical cancer and is found in 99.7% (Munoz, 2000),(Jacobs et al., 1997) of invasive cervical carcinomas. The introduction of a national HPV vaccination programme in 12-13 year old girls has been announced and is due to commence in Autumn 2008. A catch up campaign vaccinating 16 to 18 year olds will begin in Autumn 2009 and 15-17 year olds in Autumn 2010. The catch up programme will result in vaccinated women attending screening from 2011 (Welsh Assembly, 2007).

Vaccination will inevitably change the type-specific prevalence of HPV in the community and reduce the incidence of pre-invasive and invasive disease of the cervix (Villa et al., 2005, Harper et al., 2004).

This project is designed to assess and characterise the current HPV burden prior to the implementation of prophylactic HPV vaccination in Wales. This cohort will serve as a 'control' group to compare with vaccinated cohorts both as part of the catch up

campaign and the main target group for vaccination. Only a narrow window of opportunity exists in which to do this study, since delay may see ad hoc HPV vaccination outside of a National programme leading to contamination of the pre-vaccination cohort.

Introduction

2.1   Human papillomavirus

The Human Papillomavirus (HPV) plays a central aetiological role in cervical carcinogenesis, such that it is recognised as a necessary, if not a sufficient cause of cervical cancer (Bosch et al., 2002). HPV infection is very common with a 75% lifetime risk (Koutsky, 1997). It is particularly common in young women following onset of sexual activity (Hibbitts et al., 2006),(Peto et al., 2004a). The majority of infections are, however, transient and asymptomatic with less than 5% of those infected developing anogenital neoplasia.

2.2   Cervical screening

Cervical screening identifies women at increased risk of developing cervical cancer and is estimated to save around 5,000 lives per year in the UK (Peto et al., 2004c). Screening is currently based on cervical cytology identifying dyskaryotic cells indicating the need for further assessment and possibly treatment of cervical

intraepithelial neoplasia (CIN) to prevent potential progression to invasive disease. However, cytological screening is not infallible with up to half of the 1,000 women who die annually in the UK of cervical cancer having participated in Cervical Screening Programmes.

It is anticipated that HPV testing will be introduced into cervical screening, especially following implementation of prophylactic HPV vaccination. HPV testing has higher sensitivity for high-grade disease compared to cytology but since HPV infection is so common, this is at the expense of lowered specificity.

2.3   HPV vaccination

HPV vaccination represents a major public health breakthrough, which could potentially prevent 70-75% of cervical cancers in women vaccinated before contact with the virus, i.e. before sexual debut. Infection with HPV usually occurs decades before the development of cervical cancer and therefore the impact of HPV vaccines will take time to become apparent. Surrogate endpoints such as persistent type-specific HPV infection or high-grade intraepithelial neoplasia have been employed in phase III trials with near 100% efficacy for the vaccines in per protocol populations (Villa et al., 2005, Harper et al., 2004).

Monitoring and surveillance of HPV vaccine implementation and impact upon disease will be essential for the evaluation of this public health intervention. Changes

in HPV prevalence following vaccination may offer an early opportunity to assess impact of prophylactic vaccination especially where cervical screening programmes commence from the age of 20 years. A degree of cross protection against similar HPV types has been found, with the bivalent vaccine showing 60% efficacy for preventing 6-month persistent infections with HPV 45, 36% for HPV 31 and 32% for HPV 52. The 12-month efficacy against persistent infections for any non-vaccine oncogenic types was 27% (Paavonen et al., 2007). The possibility of type replacement has been raised following lowered prevalence of HPV 16 and 18 (also HPV 6 and 11 in the case of the quadrivalent) in the vaccinated population.

UK implementation of prophylactic vaccination will commence in September 2008, although acceptability, uptake and hence effectiveness in a community setting is at present unknown. Experience in Australia suggests uptake maybe in the region of 80% (S. Garland, personal communication).

A previous HPV prevalence study in women (20-65 years) attending cervical screening in South East Wales suggests an unusual pattern of HPV types, which may have implications for the effectiveness of prophylactic vaccination in Wales (Hibbitts et al., 2006). The current proposal is to characterise HPV type specific prevalence on an 'All Wales' basis amongst young women attending for their first cervical screen aged 20-22 years. This cohort represents a control or comparison group for subsequent vaccinated cohorts when they undergo first cervical screening upon reaching 20 years of age.

The two vaccines commercially available, Cervarix and Gardasil, are active against High Risk (HR) HPV types 16 and 18. Gardasil has additional protection against the Low Risk (LR) HPV types, 6 and 11. In June 2008, the Government announced that Cervarix™ had been chosen for the first 3 years of implementation within the UK. We should see a type specific reduction in HPV infection in vaccinated women and the magnitude of this reduction will give an indication of the efficacy of the vaccination programme and address both type replacement and cross protection. It is possible that transient cervical contamination by HPV may still occur but would not result in dyskaryosis. Linking HPV and subsequent cytology +/- histology will indicate clinically significant infections resulting in cervical dysplasia and neoplasia.

Ultimately, reduced prevalence of HPV infection and cervical intraepithelial neoplasia in the vaccinated population will affect the health economics of the screening programme and it is likely that change will be required. HPV testing is now commercially available, with ongoing clinical trials evaluating whether cervical screening can be improved by HPV testing and how best to incorporate HPV testing into the screening programme.

HPV testing has been shown to be more sensitive than cervical cytology for high-grade disease but less specific due to high prevalence of HPV infection within populations, especially in women less than 30 years of age. Additional 'prognostic'

biomarkers are being investigated to improve specificity for disease or the potential to progress to disease for example P16^{INK4}, MCM, TOP2A, viral integration, methylation, mRNA, etc. Storage of samples from this cohort will allow subsequent investigation of selected biomarkers to detect underlying high-grade disease or the potential to progress to high-grade disease when linked to future cytology and histology collected as part of the Cervical Screening Wales programme.

2.4   Objectives

The aim of the study is to assess and characterise the current HPV burden prior to the implementation of prophylactic HPV vaccination in Wales. This cohort will serve as a 'control' group to compare with vaccinated cohorts both as part of the catch up campaign and the main target group for vaccination. Only a narrow window of opportunity exists in which to do this study, since delay may see ad hoc HPV vaccination outside of a National programme leading to contamination of the pre-vaccination cohort.

Primary

- To determine baseline HPV genotype specific prevalence in women aged 20-22 years attending first cervical screening in Wales, prior to the introduction of an HPV vaccination programme.

Secondary

- To correlate HPV type-specific infection with cytology and histology where available.

- To correlate HPV type-specific infection with age.

- To correlate HPV type-specific infection with social deprivation score (SDS) according to the Welsh Index of Multiple Deprivation (WIMD)

Tertiary

- Development of 'bio-bank' to investigate potential prognostic biomarkers.

2.5   Rationale

Monitoring and surveillance of HPV infection following implementation of HPV vaccination, together with the impact upon cervical neoplasia will be essential for the evaluation of this major public health intervention. Changes in HPV prevalence following vaccination may offer an early opportunity to assess impact of prophylactic vaccination, especially where cervical screening programmes commence from the age of 20 years.

The organisation base in Wales is well positioned to carry out this early evaluation of HPV vaccination where cervical screening starts at 20 yrs; 5 years earlier than in England. In addition, within Wales data on HPV immunisation uptake, the incidence

of cervical cancer and pre-invasive disease are all collated by divisions of the same NHS Trust, thus facilitating a collaborative approach.

Subject to additional funding and regulatory approvals, the present cohort will be used in future work in comparison with vaccinated cohorts. Initially this will involve catch up cohorts (2011) and subsequently the target age group: girls aged 12-13 years as they reach the age of 20 years and attend first cervical screening (2015). This will allow us to address the issue of type replacement, comparison of HPV assays and assessment of biomarkers.

This work will form part of a comprehensive framework programme evaluating the HPV vaccination programme coordinated in collaboration with a working steering group that includes Cardiff University, Cervical Screening Wales (CSW), National Public Health Services, and the Health Protection Division of the Welsh Assembly Government.

3.0   Investigational Plan

3.1   Overall Study Design

This is a pseudo-anonymous prospective cohort study design. The primary outcome will establish a baseline HPV genotype specific prevalence in women aged 20-22 years attending first cervical screening in Wales, prior to the introduction of

prophylactic HPV vaccination for subsequent comparison with vaccinated cohorts of women. The analysis will be performed on residual liquid based cytology samples and the HPV typing result will not alter patient management.

3.2   Study Population

Approximately 27,000 20-22 year-old women are eligible and invited for first cervical screening in Wales annually. Screening uptake is approximately 39% in this age group and it is therefore estimated that ~15,000 residual Liquid Based Cytology (LBC SurePath) samples will be collected from this 'All Wales' unvaccinated cohort with information on current cytology result, age and social deprivation score (Welsh Index of Multiple Deprivation). Cytological and histological follow-up data will be collected on over the next two rounds (minimum 6 years) of screening.

3.3   Inclusion and Exclusion Criteria

All samples collected from women attending for their first routine cervical screen (20-22 yrs) at each of the Cervical Screening Laboratories in Wales will be flagged and included in the study and samples from women who have an inadequate first routine cervical smear (20-22yrs) and are then recalled for a repeat smear. Women attending for any other routine cervical smear will be excluded and those women with an inadequate cytology result or no information available for age, cytology or social deprivation score.

Data analysis will be performed on all samples that comply with the following inclusion criteria: $\Box$-globin PCR positive; complete information available on age, cytology result and SDS; within the target age group 20-22 years; first routine cervical smear.

3.4   Sample Collection

Sample collection will be initiated in March 2009 for a period of 18 months. LBC samples will be processed and analysed by the local cervical screening laboratories in Wales according to the guidelines of the British Society of Clinical Cytology (BSCC), standard operating procedures [12] and CSW guidelines. The residual samples will be anonymised by the originating laboratory prior to being forwarded to the HPV Research laboratory, School of Medicine, Cardiff University at room temperature. Two hundred and fifty samples per week will be processed (HR /LR HPV genotype) with completion of the full 15,000 within 18 months i.e. March 2009 - Sept 2010.

3.5   Interventions

This study is designed to evaluate baseline HPV prevalence in women attending for their first cervical smear in a real life setting and no additional samples will be taken from the participants. One additional test (HPV genotyping) will be carried out on a sample that has already been taken.

3.6   Blinding

All patient identifiable information will be removed from the residual liquid based cytology samples within the originating Cervical Screening Laboratory leaving only the lab ID on the specimen pot. Samples will be batched on a weekly basis and then forwarded to HPV Research Laboratory, School of Medicine, Cardiff University for HPV typing.

Each cytology lab will fax/ email the HPV lab when samples are dispatched giving the lab ID numbers and the originating cytology lab ID. The HPV lab will confirm receipt of the specified samples and assign a unique Study ID code. The Project manager will compile an excel datasheet that links the date of sample receipt by HPV lab, originating cytology lab, lab number and the HPV lab allocated study code from 00001 onwards. This information will be transferred back to the Cervical Screening Administration computer database to enable sample linkage with age, cytology result and social deprivation score (SDS).

CSW will have not have access to any HPV typing results and therefore this will not alter patient management and Cardiff University will not have access to any patient identifiable information (pseudo-anonymised).

3.7   Study Plan

October 2008-December 2008

Project sponsorship from Cardiff University will be established with application for the appropriate ethical and regulatory approvals.

Jan 2009-Feb 2009

Research Assistants appointed within the HPV Research laboratory, School of Medicine, Cardiff University.

March 2009

Sample collection initiated from all local cervical screening laboratories in Wales.

Work Process for each specimen:

Cervical cytology preformed according to the guidelines of the British Society of Clinical Cytology (BSCC) standard operating procedures and CSW Protocols [12] in the normal manner.

Flagging of targeted specimens defined as 'LBC smears from women attending for their first cervical screen 20-22yr' using date of birth and removal of all patient identifiable information within the originating Cervical Screening Laboratories in Wales.

Each cytology lab will fax/ email the HPV lab, Cardiff University, when samples are dispatched giving the lab ID numbers and the originating cytology lab ID.

Weekly collection of the targeted residual LBC specimen pots on an 'All-Wales' basis with subsequent transportation to the HPV Research Laboratory, Cardiff University.

The HPV lab will confirm receipt of the specified samples and assign a unique Study ID code. The Project manager will compile an excel datasheet that links the date of sample receipt by HPV lab, originating cytology lab, lab number and the HPV lab allocated study code from 00001 onwards.

This data will be transferred back to the Cervical Screening Administration computer database to enable subsequent sample linkage with age, cytology result and social deprivation score (SDS) (pseudo-anonymisation).

HPV testing following the laboratory standard operating procedures. Login receipt (hard copy and electronic copy of LBC samples and preparation of cell pellets from LBC material).

Extraction of DNA from cell pellets and quantification of extraction efficiency with PCR for the house-keeping gene b-globin.

HPV genotyping of samples using PCR-enzyme immunoassy with the GP5+/GP6+ primers [8].

Once the results of the HPV typing are available, data analysis will be performed for the primary outcome and secondary outcomes.

The samples collected at the end of the 18 months will form a sample bio-bank for future studies into potential biomarkers of disease progression.

3.8   Endpoints

Primary

Baseline HPV type specific prevalence in women attending for their first routine cervical smear prior to the introduction of the HPV vaccination programme.

Secondary

Correlation of HPV type-specific prevalence with cytology, age and SDS in women attending for their first cervical smear prior to the introduction of the HPV vaccination programme (2009).

Tertiary

Establish a sample bio-bank that will be available for future studies into potential prognostic biomarkers of disease progression.

4.0   Sample Management

Each LBC sample will be processed within the originating Cervical Screening laboratory according to the British Society of Clinical Cytology (BSCC) standard operating procedures [12] and CSW guidelines. Each sample will be given a unique study number and entered into a database with cytology result, age and social

deprivation score. Samples will be HPV genotyped according to standard operating procedures within the HPV Research Laboratory, Cardiff University. In brief specimens will be centrifuged and aliquoted before storage at -80ºC. HR and LR HPV typing using the PCR-EIA methodology reported by Jacobs *et al* (Jacobs et al., 1997) will be performed on all samples with full HR genotyping on HR positive specimens. To control for extraction efficiency each DNA sample will be screened for b-globin by PCR. HR genotyping will occur with 10% re-analysis of samples for quality control/assurance purposes.

4.1   Cytological analysis

Cervical samples are collected using a plastic broom device and placed into a vial of preservative fluid (SurePath CytoRich). The head of the device is detached and left in the vial which is then capped and transported to the cytology laboratory. The vials are then vortexed to re-suspend the cell pellet and an aliquot is placed into a centrifuge vial using the PrepMate device. The aliquot is treated through a density gradient centrifugation process to remove unwanted material and a concentrated pellet of cervical cells is produced. The pellet is then re-suspended and the PrepStain slide processor transfers an aliquot to a settling chamber mounted on a microscope slide. Cells sediment onto the slide to form a thin layer and the excess fluid and cells are discarded. The slides are routinely stained as part of the automated process. The cytology slides will be analysed in the normal manner, according to Cervical Screening Wales policy. They will undergo primary screening, checking and referral for consultant reporting as required. This will be completed

within four weeks. Residual samples are kept normally until a report has been issued with storage at room temperature. Flagged samples for use in this study will be separated from all specimens in each of the participating laboratories.

4.2   HPV Testing

HPV testing will be performed in the HPV Research Laboratory, Dept Obstetrics & Gynaecology, School of Medicine, Cardiff University following laboratory procedures.

In brief, the residual cell pellets will be washed and re-suspended in 1ml 10mM Tris pH 7.4 with DNA extraction using recombinant Proteinase K and subsequent PCR for the human b-globin gene to determine extraction efficiency as previously described [8].

The GP5/GP6+ HPV PCR-ELISA method of Jacobs *et al* [2] (Jacobs MV 1997)will be performed on all specimens in a 96-well format with minor modifications. Samples will be processed in batches of 96 to increase throughput and a two tier method applied:

An initial PCR-ELISA with a cocktail of HR or LR type specific probes

A second PCR-ELISA of all HR positive samples with genotyping using 14 individual HR HPV probes: HPV 16; 18; 31; 33; 35; 39; 45; 51; 52; 56; 58; 59; 66 and 68.

The initial 'HR/LR' and sub-typing PCR reaction will be performed in a final volume of 25µl and 100µl respectively. PCR cycling conditions were 94°C 4 mins, 40 cycles of 94°C 30 sec, 40°C 90 sec, 72°C 60 sec followed by 72°C 4 mins. Positive (CaSki) and negative (water) DNA extraction, PCR and ELISA controls will be included for every 94 samples.

4.3   Data Analysis

Results will be analysed in an Excel worksheet. Raw data from each ELISA reading at each time point is pasted into a specific position in the datasheet and results are given as positive (1) or negative (0). The negative extraction control included in every 96-well plate serves as the background reading for which all the other results are compared. A positive result is equivalent to 3 times background.

4.3   Quality Controls and Assurance

Positive (CaSki) and negative (water) DNA extraction, PCR and ELISA controls are included for every 94 samples for the GP5-GP6[+] HPV PCR-EIA. Re-analysis of 10% of all samples processed will be evaluated for quality assessment purposes.

5.0   Data Management

A hard and electronic (password protected) copy of raw HPV typing data will be maintained and kept in a central location (School of Medicine, Cardiff University

Office 6FT 157 main building, Heath Park). A central (password protected) database of results linking unique study number, age, cytology, SDS and HPV typing result will be maintained with a weekly back-up system of data on to an external hard drive.

5.1   Statistical Methods

Statistical analysis will be performed within Cardiff University on samples that comply with the following inclusion criteria: β-globin PCR positive; complete information available on age, cytology result and SDS; within the target age group 20-22 years; and first cervical smear. Social deprivation will be estimated by linking postcodes to the Welsh Index of Multiple Deprivation (WIMD) which describes levels of deprivation across Wales with higher scores indicating greater deprivation (WIMD, 2005). The overall deprivation scores in Wales range from 78.9 to 1.4. Geographical units are termed Lower Layer Super Output Areas (LSOA) and they are graded according to SDS as follows: 10% most deprived (41.91-78.9 SDS); 10% to 20% most deprived (32.71-41.9 SDS); 20% to 30% most deprived (26.31-32.70 SDS); 30% to 50% most deprived (17.91-26.30 SDS); 50% least deprived LSOA in Wales (1.4-17.9 SDS). The Chi-square test will be used to calculate a P value and 95% Confidence Intervals (CI) will be calculated where appropriate. All HR HPV positive cases will be differentiated into those with a single HR type and those with multiple HR types.

Sample Size and Power Calculations

All sample size calculation are based on the following assumptions:

Approximately 27,000 20-22 year-old women are eligible and invited for first cervical screening in Wales annually.

Screening uptake is approximately 39% in this age group

It is therefore estimated that ~15,000 residual LBC samples will be collected from this 'All Wales' cohort 1st March 2009 –1st Sept 2010 with information on current cytology result, age and social deprivation score (Welsh Index of Multiple Deprivation).

6.0 References

Advice for Cytopathology Laboratories on the Implementation of Liquid Based Cytology for Cervical Screening LBC (2004) Implementation Guide No2 Version 1.

Bosch FX, Lorincz A, Munoz N, Meijer CJ, Shah KV. The causal relation between human papillomavirus and cervical cancer. J Clin Pathol 2002 Apr;55(4):244-65.

Harper D, Franco E, Wheeler C, Ferris D, Jenkins D, Schuind A, et al. Efficacy of a bivalent L1 virus-like particle in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. Lancet 2004;364:1757-65.

Hibbitts S, Rieck GC, Hart K, Powell NG, Beukenholdt R, Dallimore N, et al. Human papillomavirus infection: an anonymous prevalence study in South Wales, UK. Br J Cancer 2006 Jul 17;95(2):226-32.

Jacobs MV, Snijders PJ, van den Brule AJ, Helmerhorst TJ, Meijer CJ, Walboomers JM. A general primer GP5+/GP6(+)-mediated PCR-enzyme immunoassay method for rapid detection of 14 high-risk and 6 low-risk human papillomavirus genotypes in cervical scrapings. J Clin Microbiol 1997;35(3):791-5.

Koutsky L. Epidemiology of genital human papillomavirus infection. The American journal of medicine 1997 May 5;102(5A):3-8.

Munoz N. Human papillomavirus and cancer: the epidemiological evidence. J Clin Virol 2000 Oct;19(1-2):1-5.

Paavonen J, Jenkins D, Bosch FX, Naud P, Salmeron J, Wheeler CM, et al. Efficacy of a prophylactic adjuvanted bivalent L1 virus-like-particle vaccine against infection with human papillomavirus types 16 and 18 in young women: an interim analysis of a phase III double-blind, randomised controlled trial. Lancet 2007 Jun 30;369(9580):2161-70.

Peto J, Gilham C, Deacon J, Taylor C, Evans C, Binns W, et al. Cervical HPV infection and neoplasia in a large population-based prospective study: the Manchester cohort. Br J Cancer 2004 Aug 31;91(5):942-53.

Peto J, Gilham C, Fletcher O, Matthews FE. The cervical cancer epidemic that screening has prevented in the UK. Lancet 2004 Jul 17-23;364(9430):249-56.

Villa LL, Costa RL, Petta CA, Andrade RP, Ault KA, Giuliano AR, et al. Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. Lancet Oncol 2005 May;6(5):271-8.

Welsh Assembly. Vaccination against cervical cancer to be introduced in Wales. 2007 [cited; Available from: www.wales.nhs.uk/documents/W070828-Hlt.pdf

7.0   Project Management

Principal Investigator: Prof Alison Fiander

With recognised expertise in HPV infection, the novel management of HPV associated neoplasia and clinical practice in the field of Gynaecological Oncology and Colposcopy Prof Alison Fiander will provide the central management for this project and assist in the interpretation of results as applicable to clinical practice.

Cytology Laboratory Coordinator: Jane McRea

Mrs Jane McRea is the Laboratory Manager for the Cytology Department, CSW at Llandough. She will develop the generic standard operating procedure for flagging, isolating and anonymising all required specimens and develop local amendments in collaboration with laboratory managers at each participating site in Wales or on the England/Wales border. Mrs McRea will be the central manager for the cytology laboratories involved in this study and ensure all specimens are identified and then transported to the HPV laboratory, School of Medicine, Cardiff University.

Scientific Project Manager: Dr Sam Hibbitts

Dr Hibbitts is the Scientific Manager for the HPV Laboratory, School of Medicine Cardiff University. She has demonstrated her abilities in handling large scale cross-sectional HPV genotyping studies (n=10,000) (Hibbitts *et al*, 2006) and will be responsible for coordinating the scientific study protocol and data management. This will include: maintenance; training and adherence to SOPs for sample login; sample storage; DNA extraction; PCR; ELISA; and quality control procedures.

Laboratory Technical Staff to be appointed: 1 x FTE and 1 x 21hr

250 samples are required to be processed in the HPV laboratory per week for a 12-month period. Two dedicated laboratory technical assistants will be appointed. One

part-time (21hrs) research assistant will be responsible for sample login, initial sample processing and storage. The full time technician will perform the DNA extraction, PCR and HPV typing ELISA.

Management and Coordinator for record linkage: Helen Beer

Helen Beer is the Information Analyst/ Manager for Screening Services within Cervical Screening Wales. She will oversee the management of the pseudo-anonymisation process required to confirm eligibility of all flagged specimens for inclusion into this study and linkage of samples with age, cytology and social deprivation score.

8.0   Ethical Considerations

Residual LBC samples will be anonymised but linked to age (years and completed months), cytology and social deprivation score. Anonymity will allow complete collection of samples in the cohort since individual consent will not be required. Patient identifiable information will be removed by the originating laboratory prior to their being forwarded to the HPV laboratory in the School of Medicine, Cardiff University. Each sample will be given a unique study number and entered into a database with cytology result, age and social deprivation score. The unique study number will enable future linkage with screening records over the next 2 rounds of screening (ie minimum 6 years). The HPV laboratory will not have access to patient identifiable information and CSW will not have access to individual HPV results,

thereby maintaining anonymity. Cervical Screening Wales will identify repeat samples from the same woman, thus avoiding confounding of HPV prevalence.

Approval will be sought for longitudinal follow up of cytology and histology over the next two rounds of cervical screening (minimum of 6 years) in order to correlate HPV type-specific prevalence and potential prognostic biomarkers for the development of high-grade cytology and/or histology.

8.1   Administrative Procedures

All potential women eligible for participation in this study will be identified by the originating laboratory i.e. 'first smear test and aged 20-22 yrs' and residual samples will be flagged for collection following routine cytological analysis. The Cervical Screening Administration Department (CSAD) computer database will confirm the eligibility of each sample (coordinated by the Information Analyst/ Manager). Patient identifiable information will be removed (anonymisation) by the originating laboratory and a unique study number will be assigned by the HPV Research Laboratory, School of Medicine, Cardiff University.

8.2   Monitoring Procedures

Mrs McRea will be the central coordinator for the cytology laboratories involved in this study and ensure specimens identification and transportation to the HPV

laboratory, School of Medicine, Cardiff University. Dr Hibbitts will oversee the project management of the HPV genotyping and ensure compliance with SOP and quality assurance of HPV typing results.

8.3   Recording of Data and Retention of Documents

A hard and electronic copy of HPV typing raw data will be maintained and kept in a central location (School of Medicine, Cardiff University Office 6FT 157 main building, Heath Park).   A central database of results linking unique study number, age, cytology, SDS and HPV typing result will be maintained with a weekly back-up system of data on to an external hard drive. No patient identifiable data will be held by the HPV Research Laboratory, School of Medicine, Cardiff University.

10.2 Study 2: HI-BCD Study Cohort

Protocol Summary

Acronym: HI BCD (HPV Integration as a Biomarker for Cervical Disease)

Principle Investigator: Sam Hibbitts, Senior Lecturer in Gynaecological Oncology, Cardiff University

Co-Investigators: Amanda Tristram, Senior Lecturer in Gynaecological Oncology, Cardiff University; Sue Ashman, Research Nurse Colposcopist, Cardiff University; Joanne Jones, Senior Technician in Gynaecological Oncology, Cardiff University; Helen Beer, Senior Information Manager and Research Specialist, Cervical Screening Wales

Study Advisors: Anne Hauke, Regional Programme Coordinator, Cervical Screening Wales; Bryan Rose, All Wales Programme Manager, Cervical Screening Wales; Dave Nuttall, Head of Laboratory Services, Cervical Screening Wales

Participating Institutions; Cardiff University; Cardiff and Vale University Health Board; Public Health Wales, Screening Division.

Aim: To pilot methods investigating Human Papillomavirus molecular biology and host cell biology to determine: if HPV viral integration can predict high-grade cervical abnormality in women with cytological abnormality the effectiveness of transformation zone (TZ) sampling by objectively demonstrating the presence and quantity of endocervical cells from the upper margin of the TZ. If there is a correlation between the presence and absence of host cell adhesion molecules and HPV infection and viral integration in cervical cells

Number of Subjects: 60

Study Design: Pilot, cohort, observational study.

Inclusion Criteria: Women referred to colposcopy following an abnormal cytology report

Exclusion Criteria: Women unable to give informed written consent.

Study Interventions: HPV testing, biomarker assays for viral integration on an additional smear taken during colposcopy.

Primary endpoint: Histologically proven high-grade disease at first colposcopy visit where biopsy taken.

Secondary endpoint: Histologically proven high-grade disease during subsequent visits (up to 4 years).

Contents

1.0   Lay Summary

This is a study to see if carrying out new tests on cervical screening samples could safely reduce the number of women referred to hospital following an abnormal smear result. Taking part in the study will involve having one additional smear taken but will not affect patient treatment. Identifiable personal details will be removed from the samples before they are tested. The result of the tests will only be used by the study team to evaluate the new tests for viral integration and host biology markers. The result will not be included in the women's hospital or screening records. The women will be invited to take part if they have been referred tothe colposcopy clinic following an abnormal smear. If they consent to being included in the study, a smear will be taken and tested for human Papillomavirus (HPV) before colposcopy.

HPV is a very common virus that is responsible for cervical cancer. HPV types 16 and 18 are considered to be high risk and, together with type 45, are the three commonest HPV types in cervical cancer, causing about 80% of cervical cancers. HPV 16, HPV 18 and HPV 45-positive samples will be tested further to examine how far the virus has integrated into the cervix. The results of these tests will be compared with available smear and biopsy results from consenting women at subsequent visits (up to 4 years). These results will be provided by Cervical Screening Wales part of the Screening Division, Public Health Wales. Results will be interpreted with the support of a data analyst.

2.0   Introduction

2.1   Cervical Screening

UK national cervical screening programmes have been estimated to prevent 80% of deaths from cervical cancer(Peto et al., 2004c).   The premise of the screening programmes is that women at risk of having cervical intraepithelial neoplasia (CIN) can be identified by cytological analysis of a cervical smear.   Women are then referred to colposcopy for diagnostic biopsy.   Those identified at colposcopy as having high-grade CIN (CIN2+) are then offered treatment in order to prevent future development of cervical cancer.

In Wales, women are referred for colposcopy as a result of the following cytological abnormalities:

- Persistent low-grade: either two consecutive samples with mild dyskaryosis or three with borderline nuclear change.

- One test reported as borderline nuclear change in endocervical cells

- One test reported as moderate dyskaryosis

- One test reported as severe dyskaryosis

- One test reported as possible invasion

- One test reported as glandular neoplasia

- One test reported as borderline change in endometrial cells

- One test reported as query high grade

- One test reported as other glandular cells

- Three tests reported as abnormal at any grade in a 10 year period

Whilst there is a high incidence of CIN2+ in biopsies taken following a moderate or severe cytology result (78% and 92% respectively), only approximately 24% of women with persistent low-grade cytology will have biopsy-assessed high-grade disease and require treatment(Wales, 2007). Therefore,approximately 50% of all women who are referred will have high-grade disease and require treatment. This means that many women are being referred that do not require hospital treatment. It is known that referral for colposcopy itself generates anxiety in women(Jones et al., 1996). Reducing the number of women referred would, in a high proportion of cases, spare women the unnecessary concern that they might have cervical cancer.

Moreover, this would permit valuable resources to be focused specifically only on those women that need them.

In this study we aim to investigate viral integration as a potential prognostic marker in women referred to colposcopy. Many women with low-grade cytological abnormalities are referred for colposcopy but do not have disease. Nevertheless, samples from women with all grades of cytological abnormality will be taken to evaluate integration status in relation to grade of abnormality. It may be possible to identify whether assays able to detect viral integration in low-grade smears are able to predict high-grade disease.

2.2   Human Papillomavirus

High risk Human Papilloma Virus (HPV) plays a central aetiological role in ano-genital neoplasia(Walboomers et al., 1999). Such is the strength of the association that there are increasing calls to incorporate oncogenic HPV testing into Cervical Screening or even to replace cervical cytology with a test for the virus(Cuzick et al., 2006, Cuzick et al., 2003). Consequently, many large randomized controlled trials are in progress to determine the most efficient and effective algorithm for detection of high-grade cervical intraepithelial neoplasia (CIN 2+) in primary screening (Davies et al., 2006, Mayrand et al., 2006). There is also evidence that testing for oncogenic HPV is a sensitive and cost effective measure in follow up of women treated for CIN

(Arbyn et al., 2005) and in triage of women with equivocal cytology (borderline and mild dyskaryosis) (Kulasingam et al., 2006, Moss et al., 2006, Solomon et al., 2001).

However HPV infection is common, with an 80% lifetime risk of infection. The majority of infections are transient and asymptomatic with less than 5% of individuals infected developing CIN. Because of the discrepancy between those infected and those developing disease, it is necessary to determine prognostic markers in order to identify women at risk of developing disease, thereby allowing valuable resources to be focussed on those most at risk. Regulation of HPV oncogene expression has emerged as a critical factor and this is intimately linked to the differentiation state of the host cell, and in undifferentiated cells, oncogene expression is repressed by modulation of chromatin structure by factors including CDP and YY1 (Stunkel and Bernard, 1999). Factors reported to effect expression include integration of the HPV genome into host cell DNA (Tan et al., 1994).

2.3   Viral Integration

Episomal HPV genomes exist during the normal viral life cycle and are maintained in basal cells of the squamous epithelium at approximately 50-100 copies per cell (Stanley et al., 1989b). Within HPV episomes the expression of viral oncogenes is highly regulated. E6 and E7 induce the unscheduled re-entry into the S-phase of the cell cycle and activate the host replication machinery in order to allow amplification of the viral genomes before virion synthesis(Cheng et al., 1995). Although viral

oncogenes E6 and E7 are actively produced within episomes they appear not to be functionally carcinogenic. This is because these events take place in cells that are ultimately lost from the cervical squamous epithelium as a result of the constant renewal process that occurs in this area. For cervical neoplastic progression to occur the viral oncogenes need to become expressed and maintained throughout the cervical epithelium.

Most cervical carcinomas have deregulated viral oncogene expression and their tumour cells contain truncated viral genomes integrated into the host genome. *In vitro*, viral integration increases cell proliferation (Wentzensen et al., 2004, Jeon et al., 1995) even though integration is not a normal part of the HPV life cycle and causes the deletion of viral genes that are essential for production of infectious virions. Numerous studies have characterised viral integrants and although many variants have been found, some consistent features have been defined. A predominant finding is the loss of the viral E2 gene. E2 recognises and binds to the origin of replication of the viral promoter and controls the decreased expression of E6 and E7. The loss of E2 is the first stage in transformation with subsequent increased expression of the viral oncogenes responsible for disruption in function of key cellular proteins (p53 and Rb).

A positive HR HPV result alone has a poor positive predictive value (PPV) due to the high number of transient infections. This project aims to establish whether viral integration can increase the PPV for detection of high-grade cervical disease in

women with cytological abnormalities. This is a pilot, prospective, cohort study and testing reflects the setting in which the biomarkers might be used if introduced into clinical practice.

2.4   Aims and Objectives

To pilot methods investigating Human Papillomavirus molecular biology and host

Cell biology to investigate if HPV viral integration can predict high grade cervical abnormality in women with cytological abnormalities.

Primary Outcomes:

Comparison of three viral integration methods for identification and prediction of high-grade disease.

To correlate results from the integration assays with histology obtained from this cohort at subsequent visits (up to 4 years) to validate long-term positive predictive value (PPV).

2.5   Rationale

This pilot, cohort, observational study has been designed to examine the the virus molecular biology. Benefits for patient management of demonstrating viral integration

following a HPV-positive result in a clinical cohort of patients attending colposcopy will be investigated.

3.0   Investigational Plan

3.1   Overall Study Design

This is a cohort study to investigate viral integration as a biomarker in a group of women referred for colposcopy with follow up of their screening results over the next round of screening (up to 4 years). The primary outcome will be whether viral integration following a HPV positive result can predict histological high-grade disease on biopsy. The analysis will not be undertaken until after the woman has attended colposcopy and consented and will not alter her management.

3.2   Discussion of Design

The design of the study has been chosen to reflect as closely as possible the setting in which the test might be used if introduced into clinical practice. As the efficacy has not been assessed in this setting before, it was considered appropriate for the study to be a pilot and observational, rather than interventional.

3.3   Study Population

The study population consists of those women attending colposcopy at a single site within Cardiff and Vale University Health Board. The sample will be a convenience sample, taken from women who give written informed consent.

3.4   Inclusion and Exclusion Criteria

Inclusion Criteria

- Women referred to colposcopy due to abnormal cytology
- Women undergoing treatment for high-grade CIN

Exclusion Criteria

- Women who are unable to give informed written consent.

3.5   Recruitment

Women will be invited to take part in the study when they attend for colposcopy. This study will include 15 patients with mild dyskaryosis, 15 with moderate dyskaryosis, 15 with severe dyskaryosis and 15 with cervical cancer. Approximately 20 women per month are expected to be eligible for the study within the single site. Recruitment is expected to be completed within nine months.

3.6   Interventions

This study is designed to evaluate viral integration in a real life setting: for this, one additional sample will be taken from the participants. Up to three additional methods for viral integration will be carried out on the sample taken.

3.7   Study Plan

Women will be informed of their smear test result and colposcopy referral according to Cervical Screening Wales (CSW) standard operating procedures (CSAD Quality Manual April 2011).

The colposcopy administration at Llandough Hospital will send the study information to women referred from CSAD with abnormal cytology along with their colposcopy appointment. This ensures women receive the study information in advance, giving them time to read and absorb the information. This study is a pilot and will only include 15 patients with mild dyskaryosis, 15 with moderate dyskaryosis, 15 with severe dyskaryosis and 15 with cervical cancer.

After confirming eligibility criteria, women will be given the opportunity to discuss the study and ask any questions when they attend the colposcopy clinic. Written informed consent will be obtained at this time.

Women will also be asked to consent to allow the results of cervical cytology and histology for one round of screening (up to 4 years from the date of the study sample) after entry to be made available for the study. These data will be retrieved by Cervical Screening Wales.

A cervical sample will be taken into sample transport medium optimal for DNA/RNA recovery, according to standard operating procedures, from all women who give written consent. The samples will be pseudo-anonymised in clinic: all patient identifiers will be removed and a unique study code allocated for transportation to the HPV Laboratory at Cardiff University. Once received in the lab the samples will be recorded electronically in an EXCEL worksheet linking study ID for linkage through the Informatics Team within Screening Division.

Colposcopy will be undertaken according to normal practice as set out in the current version of CSW's Quality Manual Version. When 60 eligible women have consented to the study, the HPV Research Team will notify CSAD to close recruitment. HPV testing and analysis of viral integration will be performed by the HPV Research Laboratories, Department of Obstetrics and Gynaecology at Cardiff University.

3.9 Endpoints

- Primary endpoint will be high-grade disease on histology at first colposcopy visit where biopsy taken.

- Secondary endpoints will be detection of viral integration.

Histological analysis will be carried out according to usual practice at each centre. Histological analysis is already subject to quality assurance as part of CSW standards and will not be repeated. This will ensure that participation in the trial will not change patient management and that it reflects true clinical practice. High-grade disease will be defined as the presence of CIN2+. The exact grade will be recorded and may be used in analysis of false negative samples.

4.0   Sample Management

4.1   Cytology

Samples will be taken from women with written informed consent within the Cardiff and Vale NHS Trust Colposcopy Clinics following standard operating procedures. The samples will be sealed and transported to the HPV Research Laboratories, Department of Obstetrics &Gynaecology, Cardiff University.

4.2   HPV Testing

Samples received by Cardiff University will be logged electronically and processed according to the standard operating procedures.

All specimens will be tested for HPV DNA using 2 methods:

The research based PCR-EIA method of Walboomers(Jacobs et al., 1997). Each sample will be divided into high-risk (HR) and low-risk (LR) HPV infection using a cocktail of type-specific probes and then HR HPV-positive samples will be sub-typed using type-specific probes.

The commercial Greiner Bio-one, PapilloCheck® HPV-Screen DNA-chip: detects 24 different HPV types (18 HR and 6 LR) in DNA-preparations from human cervical smears. The HPV types which can be detected and differentiated by PapilloCheck are HPV 6, 11, 16, 18, 31, 33, 35, 39, 40, 42, 43, 44/55, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73 and 81.

4.3   HPV Viral Integration

HPV 16, HPV 18 and HPV45 positive samples will be assessed using validated viral integration assays. There are a number of methods currently employed to detect HR-HPV integrants in the human genome. Amplification of Papillomavirus Oncogene Transcripts (APOT) detects transcripts of integrated HPV, Restriction Site PCR (RS-PCR), Southern blot and Detection of Integrated Papillomavirus Sequences (DIPS) detect integrated HPV DNA regardless of its transcriptional status. The methods used to detect HPV integration are important because a transcriptionally active integrant may contribute more to the malignant phenotype. However, integrants that are not transcribed may also contribute to malignant progression by regulating or disrupting the expression of genes that contribute to cervical malignant progression.

DNA and RNA based methods to detect HR HPV integration have been validated through a PhD studentship funded by the Welsh Office of Research and Development.

5.0   Data Management

5.1   Statistical Methods

This pilot study aims to investigate the potential of viral integration as a biomarker in predicting high-grade cervical abnormality in women referred to colposcopy.

5.2   Sample Size

The project will test 60 samples for HPV status and HPV 16, HPV 18 and HPV 45 positive cases will be investigated further for viral integration status. It is expected that approximately 60% of women in this group will be HPV positive and 60% of these cases will be HPV 16, HPV 18 or HPV 45 positive (n=22)(Jones et al., 2009).

Data Analysis and Dissemination

Using the unique study code CSW will identify the relevant screening data. The HPV Research Laboratories will link all results with a unique study code. The results of the histology, HPV testing and viral integration will be correlated for the primary outcome. Subsequent analysis will be performed over the follow-up period as data becomes available (up to 4 years). Findings will be disseminated through peer-

review publication. Favourable primary outcome will provide support for further funding proposals on increased sample size with prospective cohort studies and randomized control trials that aim to reduce cervical cancer mortality through decreased incidence with early detection of CIN2+.

6.0   Project Management

6.1   Investigators

Principal Investigator: Dr Sam Hibbitts

Dr Hibbitts is a Senior Lecturer in Gynaecological oncology and the Scientific Manager for the HPV Research Laboratory, School of Medicine, Cardiff University. She has demonstrated her abilities in handling large scale cross-sectional HPV genotyping studies (n=10,000) (Hibbitts *et al*, 2008) and will be responsible for coordinating the scientific study protocol and data management. Dr Hibbitts is the supervisor for a PhD studentship focusing on HPV viral integration.

Dr Amanda Tristram

Dr Tristram is a Senior Lecturer in Gynaecological Oncology and the clinical lead for the HPV Research Laboratory. Dr Tristram has experience in designing trials, obtaining regulatory approval and funding, recruiting patients, analysing results as well as publishing research and presenting it at International meetings.

Sue Ashman

Miss Ashman is a qualified nurse colposcopist and has previous experience in phase I and II clinical trials. Miss Ashman coordinates studies within the department of Obstetrics &Gynaecology and is experienced in study recruitment, organising clinics and taking informed consent. Miss Ashman has also been responsible for the

330

maintenance of study documentation and the preparation of protocols and patient specific documentation for clinical trials. Miss Ashman will be the primary link between the colposcopy clinic and the HPV laboratory.

Miss Helen Beer

Miss Beer is a Senior Information Manager and Research Specialist for Screening Division and has expertise in the linkage and statistical analysis of identifiable data for all screening programmes in Wales, concentrating on the cervical screening programme. Miss Beer has been involved in the evaluation of new technologies into the cervical screening programme, such as Liquid Based Cytology and Computer Assisted Screening.

6.2   Study Advisors

Dr Anne Hauke

Dr Hauke is the Regional Programme Coordinator for South East Wales, Cervical Screening Wales.

Mr Bryan Rose

Mr Rose is the All Wales Programme Manager for Cervical Screening Wales.

Mr Dave Nuttall

MrNuttall is the Head of Laboratory Services for Cervical Screening Wales.

7.0   Administrative Procedures

7.1   Changes to the Protocol

Any amendments to the protocol will be notified to relevant regulatory bodies.

7.2   Recording of Data and Retention of Documents

The research team will act to preserve patient confidentiality and will not disclose or reproduce any information by which patients could be identified.

7.3   Publication of Results

The results of the study will be submitted for publication in a peer reviewed journal and presented at national and international conferences.

8.0   Ethical Considerations and Consent

Women participating in this study will provide one extra sample however, the study will not influence their clinical management. They will be asked to permit study specific testing on the extra sample and to allow the study team access to results relating to their cervical smears and colposcopy for four years. The women will be sent information in advance and will be invited to consent during their visit to the

colposcopy clinic, having had an opportunity to discuss the study with an appropriately trained health professional.

The purpose of the extra testing is to evaluate a potential screening test. The results of the HPV viral integration tests will not change clinical management for the women and we will not inform women of the results of their study tests.

9.0   References

Arbyn, M., et al., Clinical utility of HPV-DNA detection: triage of minor cervical lesions, follow-up of women treated for high-grade CIN: an update of pooled evidence. Gynecol Oncol, 2005. 99(3 Suppl 1): p. S7-11.

Cheng, S., et al., Differentiation-dependent up-regulation of the human papillomavirus E7 gene reactivates cellular DNA replication in suprabasal differentiated keratinocytes. Genes Dev, 1995. 9(19): p. 2335-49.

Cuzick, J., et al., Management of women who test positive for high-risk types of human papillomavirus: the HART study. Lancet, 2003. 362(9399): p. 1871-6.

Cuzick, J., et al., Overview of the European and North American studies on HPV testing in primary cervical cancer screening. Int J Cancer, 2006. 119(5): p. 1095-101.

Davies, P., et al., A report on the current status of European research on the use of human papillomavirus testing for primary cervical cancer screening. Int J Cancer, 2006. 118(4): p. 791-6.

Jacobs, M.V., et al., A general primer GP5+/GP6(+)-mediated PCR-enzyme immunoassay method for rapid detection of 14 high-risk and 6 low-risk human papillomavirus genotypes in cervical scrapings. J Clin Microbiol, 1997. 35(3): p. 791-5.

Jeon, S., B.L. Allen-Hoffmann, and P.F. Lambert, Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. J Virol, 1995. 69(5): p. 2989-97.

Jones, J., et al., Comparison of the PapilloCheck DNA micro-array Human Papillomavirus detection assay with Hybrid Capture II and PCR-enzyme immunoassay using the GP5/6+ primer set. J Clin Virol, 2009. 45(2): p. 100-4.

Jones, M.H., A. Singer, and D. Jenkins, The mildly abnormal cervical smear: patient anxiety and choice of management. J R Soc Med, 1996. 89(5): p. 257-60.

Kulasingam, S.L., et al., Cost-effectiveness analysis based on the atypical squamous cells of undetermined significance/low-grade squamous intraepithelial lesion Triage Study (ALTS). J Natl Cancer Inst, 2006. 98(2): p. 92-100.

Mayrand, M.H., et al., Randomized controlled trial of human papillomavirus testing versus Pap cytology in the primary screening for cervical cancer precursors: design, methods and preliminary accrual results of the Canadian cervical cancer screening trial (CCCaST). Int J Cancer, 2006. 119(3): p. 615-23.

Moss, S., et al., Effect of testing for human papillomavirus as a triage during screening for cervical cancer: observational before and after study. Bmj, 2006. 332(7533): p. 83-5.

Peto, J., et al., The cervical cancer epidemic that screening has prevented in the UK. Lancet, 2004. 364(9430): p. 249-56.

Solomon, D., M. Schiffman, and R. Tarone, Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. J Natl Cancer Inst, 2001. 93(4): p. 293-9.

Stanley, M.A., et al., Properties of a non-tumorigenic human cervical keratinocyte cell line. Int J Cancer, 1989. 43(4): p. 672-6.

Stunkel, W. and H.U. Bernard, The chromatin structure of the long control region of human papillomavirus type 16 represses viral oncoprotein expression. J Virol, 1999. 73(3): p. 1918-30.

Tan, S.H., et al., The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. J Virol, 1994. 68(10): p. 6411-20.

Walboomers, J.M., et al., Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol, 1999. 189(1): p. 12-9.

Wales, C.S., KC53/61/65 Statistical Report 2008/2009. 2007.

Wentzensen, N., S. Vinokurova, and M. von Knebel Doeberitz, Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. Cancer Res, 2004. 64(11): p. 3878-84.

Appendix 3

PCR Primers

This appendix includes the details for all the PCR primers used in this thesis.

11.1 Beta-Globin PCR

Table Apx 3.1 *Beta-globin* PCR primers

Primer Name	F/R	Sequence	Product Size
BGPCO3	F	ACACAACTGTGTTCACTAGC	110
BGPC05	R	GAAACCCAAGAGTCTTCTCT	110

11.2 *E2* Tiling PCR

Table Apx 3.2 HPV *16* E2 tiling PCR Primers

Primer name	F/R	Sequence	Product size
PCR1	F	AGGACGTGGTCCAGATTAAG	419

	R	TCAAACTGCACTTCCACTGT	
PCR2	F	TAACTGCACCAACAGGATGT	341
	R	GCCAAGTGCTGCCTAATAAT	
PCR3	F	ATCTGTGTTTAGCAGCAACG	224
	R	TAAATGCAGTGAGGATTGGA	
PCR4	F	ACAGTGCTCCAATCCTCACT	244
	R	TCACGTTGCCATTCACTATC	
PCR5	F	GGCATTGGACAGGACATAAT	207
	R	CAAAAGCACACAAAGCAAAG	
E6 PCR	F	GAACAGCAATACAACAAACC	161
	R	GATCTGCAACAAGACATACA	

Table Apx 3.3 HPV 31 *E2* tiling PCR Primers

Primer Name	F/R	Sequence	Product Size
31E2_F_11	F	AGCAGACTGGTGGTTTTTACATTTCCA	352
31E2_R_01	R	TGGCCTTTGATACTGACAACGCTGG	

Primer Name	F/R	Sequence	Product Size
31E2_F_03	F	CCAGGTGGTGCCAGCGTTGT	
			124
31E2_R_03	R	TGTTTGCTGCATTGTCCAGTCCT	
31E2_F_04	F	AGGACTGGACAATGCAGCAAACA	
			184
31E2_R_04	R	TGCCCTTCCACAACAGTACATTGGC	
31E2_F_06	F	GGCCAATGTACTGTTGTGGAAGGGC	
			287
31E2_R_06	R	CGCCGCACACCTTCACTGGT	
31E2_F_08	F	AAAACCTGCGCCTTGGGCACC	
			135
31E2_R_08	R	GACACTGTCCACGGAGTCGCC	
31E2_F_09	F	TGTTGCGAGGCGACTCCGTG	
			323
31E2_R_09	R	TCCTGTTGACACTGATACTGTGTTAGGTA	
31E2_F_10	F	ACCTAACACAGTATCAGTGTCAACAGGA	
			160
31E2_R_10	R	ACCGACACAGACAGCACAAGTGG	

Table Apx 3.4 HPV 33 *E2* tiling PCR Primers

Primer Name	F/R	Sequence	Product Size
33E2_F_02	F	AGCACGTTTAAATGCAGTGCAGGA	178

33E2_R_02	R	TGCTAACAAAGAAGGCACCACCTG	
33E2_F_03	F	CACCAGGTGGTGCCTTCTTTGTT	155
33E2_R_03	R	GGTGGTTCACAAAGCCACACCTCT	
33E2_F_04	F	AGGTGTGGCTTTGTGAACCACC	276
33E2_R_04	R	CCTGACCACCCACATGTACTTCCCA	
33E2_F_05	F	TGGGAAGTACATGTGGGTGGTCAGG	204
33E2_R_05	R	GTCCAAGGCGGGGTCTGCAC	
33E2_F_06	F	ACCGACCACCACAAGCAGCG	281
33E2_R_06	R	ACTGGTCCAATGCCAGGTGGATGA	
33E2_F_09	F	GTGCAGACCCCGCCTTGGAC	358
33E2_R_07	R	AGCAGCACATATTGGCTTGTGATGT	
33E2_F_08	F	TCCCATTTGATGAAAATGGTAACCCAGTG	290
33E2_R_08	R	ACCTGGTGGCATAAATGTGAAAATCCCA	
33E2_F_E62	F	TGCACGACTATGTTTCAAGACACTGAGG	109
33E2_R_E62	R	TGCATTCCACGCACTGTAGTTCA	

Table Apx 3.5 HPV 35 *E2* tiling PCR Primers

Primer Name	F/R	Sequence	Product Size
HPV35_F_01	F	TCCTTTTTCTCAAGGACGTGGTGCAG	250
HPV35_R_01	R	GGCCTGCGTTGGAACCACTTG	
HPV35_F_02	F	ACCACCAAGTGGTTCCAACGC	121
HPV35_R_02	R	TGCAGTGTCCAGTCCTCTGTGCT	
HPV35_F_03	F	GCCAAAGCAATGCAAGCAATTGAAC	300
HPV35_R_03	R	TGTTTCTACACCCTGATGCACATAATAAATACC	
HPV35_F_04	F	ACTGTGGAAGCACAATTTGATGGTGA	295
HPV35_R_04	R	GCTGTGTAGCAATTTCAGCAGTGG	
HPV35_F_05	F	AGTGCATGTGGGTGGTCAGGT	289
HPV35_R_10	R	CGCACTCGCTTGGTGGGGTT	216
HPV35_F_11	F	CGAGGGGGTACCGAGCTCCC	225
HPV35_R_06	R	TGAAGCATCTTGATACAATGCTTTATATTTACCCA	
HPV35_F_07	F	AAAAGGTGATGCAAATACATTAAAGTGTTCAAGA	341
HPV35_R_07	R	GCTGAGTATAATGACACAGATAGCAATAGCG	

342

| HPV35_E6_F02 | F | ACATGTCAAAAACCGCTGTGTCCA | 100 |
| HPV35_E6_R02 | R | ACATACACCGACCTGTCCACCGT | |

Table Apx 3.6 HPV 51 *E2* tiling PCR Primers

Primer Name	F/R	Sequence	Product length
E2_Forward_01	F	TGGAGACCCTATGCCACCGTT	322
E2_Reverse_06	R	TGCTTGGGAGCCACACACCA	
E2_Forward_03	F	ACCTGCAGCGACGCGTTATCC	234
E2_Reverse_03	R	TGCACTTTGGTGCCCTCCAG	
E2_Forward_04	F	AGCACTAACACTGGAGGGCACC	210
E2_Reverse_04	R	TGTTTCCCGTTGATGTGCACTGT	
E2_Forward_05	F	TCCTCAACCTGGCATTGGACCAG	198
E2_Reverse_05	R	GCAGCACATATTGGCTTGTGACAGT	
E2_Forward_07	F	ATGGGGCACAACAGTGGGAGGT	154
E2_Reverse_07	R	CCCACGCAGGTGGTAAGGGGA	

E2_Forward_08	F	ATGTTATGAACTATGGTGTGTGGCTCCCAA	275
E2_Reverse_08	R	CATACATATAGACCTCCCACTGTTGTGCCC	
E2_Forward_09	F	TCCATTTGATAACAATGGGAATGCTGTGT	269
E2_Reverse_09	R	TCGTAAGTTTCTTTCCCGTGCTGC	
E6_Forward_03	F	ACGTTATAGCAGGTCTGTGTATGGTACT	176
E6_Reverse_04	R	ATTGCCCCGTCCAACGTCCC	

11.3 E1 tiling PCR

Table Apx 3.7 HPV 16 *E1* tiling PCR primers

Primer Name	F/R	Sequence	Product Size
HPV16 E1 1F	F	CTAGGAATTGTGTGCCCATCTG	1096
HPV16 E1 1R	R	CTTTCTATCCATTCTGGCGTGTCT	
HPV16 E1 2F	F	GATAGAGCCTCCAAAATTGCGT	995
HPV16 E1 2R	R	ACGTTGGCAAAGAGTCTCCATC	

11.4 **DIPS PCR**

Table Apx 3.8 HPV 16 DIPS PCR adapter primers

Name	Sequence
DIPS_AL 1	GGGCCATCAGTCAGCAGTCGTAGCCCGGATCCAGACTTA CACGTTG
DIPS_AS Taq	PO4-cgcaacgtgtaagtctg-NH2
DIPS_AS SauAl	PO4-gatccaacgtgtaagtctg-NH2
DIPS_AP 1	GGCCATCAGTCAGCAGTCGTAG

Table Apx 3.9 HPV 16 DIPS PCR Primers

Name	Sequence
DIPS_PCR1_16F1	ACAAAGCACACACGTAGACATTCG
DIPS_PCR1_16F2	AGTAATAAATCAACGTGTTGCGATTG

DIPS_PCR1_16F3	TTTG**G**TTACAACCATTAGCAGATGC
DIPS_PCR1_16F4	GTGCCAACACTGGCTGTATCAAAG
DIPS_PCR1_16F5	TACCAATTCTACTGTACCTAATGCCAG
DIPS_PCR1_16F6	ACTTATTGGGGTCAGGTAAATGTATTC
DIPS_PCR1_16F7	AGTAGATATGGCAGCACATAATGAC
DIPS_PCR1_16F8	GTTGGCAAGCAGTGCAGGTCAG
DIPS_PCR2_16F1	CGTACTTTGGAAGACCTGTTAATGG
DIPS_PCR2_16F2	GGACTTACACCCAGTATAGCTGACAG
DIPS_PCR2_16F3	AATAGGTATGTTAGATGATGCTACAG
DIPS_PCR2_16F4	ACAAGCAATTGAACTGCAACTAACG
DIPS_PCR2_16F5	GAGGTTAATGCTGGCCTATGTAAAG
DIPS_PCR2_16F6	CCCTGTATTGTAATCCTGATACTTTAGG
DIPS_PCR2_16F7	TGCGTGTAGTATCAACAACAGTAAC
DIPS_PCR2_16F8	TTAAACCATAGTTGCTGACATAGAAC

11.5 qRT-PCR

Table Apx 3.10 HPV 16 *HPRT* qRT-PCR Primers

HPRT	
Sense Primer	TGACACTGGCAAAACAATGCA
Antisense Primer	GGTCCTTTTCACCAGCAAGCT
Amplified DNA	4982 bp* (NT 133627546-133632438 of Chr 6)
Amplified RNA	94 bp (NT** 496-589 of M31642.1)
Reference	Allen et al, 2008

Table Apx 3.11 HPV 16 *TBP2* qRT-PCR Primers

TATA binding protein, 2nd pair of primers (TBP2)	
Sense Primer	TCAAACCCAGAATTGTTCTCCTTAT
Antisense Primer	CCTGAATCCCTTTAGAATAGGGTAGA
Amplified DNA	803 bp (NT 170880539-170881341 of chr 6)
Amplified RNA	122 bp (NT 1128-1224 of M55654.1)
Reference	Minner and Poumay 2008

Table Apx 3.12 HPV 16 *E6* qRT-PCR Primers

E6 Primer

Sense Primer	CTGCAATGTTTCAGGACCCA
Antisense Primer	TCATGTATAGTTGTTTGCAGCTCTGT
Amplified DNA/RNA	80 bp(NT 99-178 of NC001526.1***)

Table Apx 3.13 HPV 16 *E2* qRT-PCR Primers

E2 Primer

Sense Primer	AACGAAGTATCCTCTCCTGAAATTATTAG
Antisense Primer	CCAAGGCGACGGCTTTG
Amplified DNA/RNA	82 bp (nt 3362-3426 of NC001526.1)

11.6 **Methylation**

Table Apx 3.14 HPV 16 methylation PCR primers

Forward Primer	GTGAAATTATTAGGTAGTATTTGG

Reverse Primer	*BTN-CAACAACTTAATAATATAACAAAAA
Sequencing Primer	GTGAAATTATTAGGTAGTA
Sequenced CpGs	nt 3411 3414 3416 3432 3435 3447 3461 3472

Appendix 4

Presentation of Work in Thesis

Work in this thesis has previously been presented at the following events:

NCRI Conference Nov 2012 – Poster Presentation

UK HPV Conference Nov 2012 – Oral Presentation

WMA Winter Meeting Nov 2012 – Oral Presentation

Cardiff University Speaking of Science May 2013 - Oral Presentation

UK DNA Tumor Virus Meeting Jul 2013 - Oral Presentation

Cardiff University Medic Post Graduate Research Day Nov 2013 - Oral Presentation

MITReG Winter Symposium Dec 2013 - Oral Presentation