Human Papillomavirus Integration:

The Mechanism(s) Behind the High-Risk Associated with this Event and Cervical Disease Progression

A thesis submitted for the Degree of Doctor of Philosophy at

Cardiff University

By

Rachel Raybould

Supervisors: Dr. Sam Hibbitts

Professor Alison Fiander

Professor Gavin Wilkinson

Institute of Cancer Research and Medical Genetics
HPV Research Group

School of Medicine

Cardiff University

2013

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ..... (candidate) Date 7/3/2013

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of .....(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed ..... (candidate) Date 7/3/2013

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by explicit references.

Signed ..... (candidate) Date 7/3/2013

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate) Date 7/3/2013

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access previously approved by the Graduate Development Committee.**

Signed ..... (candidate) Date 7/3/2013

Acknowledgments

First of all I need to express my sincere gratitude to Sam Hibbitts. Not only did she give the opportunity to do a PhD, she has given me fantastic support throughout. Sam gave me the freedom and confidence to do my own research, was there for me through all the good times, and showed me how to remain positive and persevere when times were hard. Both Alison Fiander and Gavin Wilkinson provided me with "tag team" style supervision nonetheless they were there for me when I needed them. Alison spent the first two years of my PhD in Africa and although she was busy she had time to email me when I needed her. In my third year the supervision, support, encouragement and advice Alison gave me was second to none. Gavin provided me with excellent supervision in Alison's absence. Gavin's straight-talking, no-nonsense and encouraging approach to my supervision has inspired me and I feel I have matured as a scientist.

My time as a PhD student has not been easy but would have been pretty miserable without the support from my good friends and colleagues in the HPV Research Group. When work wasn't going to plan I had friends to talk to who gave me support and advice. I always knew I would have a good day even when my lab work was challenging. I could always rely on fellow PhD students Tiffany, Aine and Rach to cheer me up.

Last but not least I want to thank my family. To my family I dedicate all the hard graft, the very early mornings, the highs, the lows and the many tears and screams of frustration trying to get MS Word to format my work correctly. My Grandparents gave me a positive outlook on life. Gramps is an inspirational man and shows me how to make light out of bad moments. Nan always made me aware that family are **always** there: I may not always like them and they may not always like me, but I can rely on them.

My husband Dave and daughter Beth have been my motivators throughout. Without Dave and Beth a PhD would not have been easy: Dave has been my rock the whole time. Although Dave's eyes glaze over when I talk about science, spending time with him has been a great distraction from all the stress that comes as part of a PhD. Beth is a ray of sunshine and a little friend I can always rely on to make me smile. Her happy and positive personality has made most (but not all) weekends something to look forward to.

My Mum and Dad, brothers and sister have always been an inspiration to me. I would like to say thanks for giving me the motivation, confidence and encouragement to believe in myself and achieve.

Associated Publications

Raybould, R. et al. 2011. Human Papillomavirus Integration and its Role in Cervical Malignant Progression. *Open Clinical Cancer Journal* 5, pp. 1-7.

Contents

1 Introduction 2
1.1 Human Papillomavirus 3
1.1.1 Classification 3
1.1.2 Virion Structure 5
1.1.3 Genome 6
1.1.4 HPV Infection 11
1.2 Screening for Cervical Cancer 16
1.2.1 Cytology 16
1.2.2 Colposcopy 20
1.2.3 Improvements to screening 22
1.3 HPV Related Cancers 27
1.4 HPV Integration and Cervical Neoplasia 29
1.4.1 Integration at Fragile Sites 33
1.4.2 DNA damage 36
1.4.3 DNA Repair 36
1.4.4 HPV Integration and Selection 37
1.4.5 Methods to detect integration 42
2 Study Outline 45
2.1 Stage 1: Assay validation 46
2.2 Application of assays to clinical samples: 47
2.3 Mechanism of integration: 48
3 Materials and Methods 49
3.1 Samples 49
3.1.1 Cell lines 49
3.1.2 Cell Culture 51
3.1.3 Plasmids 55
3.1.4 Baseline Samples 58

3.1.5 HI-BCD Samples 61
3.1.6 Biopsy Samples 62
3.2 Nucleic Acid Preparation 62
3.2.1 DNA and RNA Extraction from Cell Culture 62
3.2.2 Extraction of HPV plasmid vector DNA 63
3.2.3 DNA extraction from Baseline samples 64
3.2.4 DNA and RNA extraction from HI-BCD samples 65
3.2.5 DNA and RNA extraction from Biopsy Samples 66
3.2.6 DNA and RNA Extraction Controls 66
3.2.7 Nucleic Acid Quantification 66
3.2.8 DNA Integrity 67
3.2.9 RNA Integrity 68
3.3 Molecular Methods 69
3.3.1 Polymerase Chain Reaction 70
3.3.2 HPV typing 77
3.3.3 Integration Assays 79
3.4 Statistical Methods 121
3.4.1 Power Calculations 121
3.4.2 Analysis of Variance 122
3.4.3 Chi-squared 122
3.4.4 Z test 123
3.4.5 Correction for multiple testing 124
3.5 Quality Control 124
4 Assay Validation 125
4.1 Assay Validation Results 125
4.1.1 E2 PCRs 125
4.1.2 RS-PCR 129

4.1.3 DIPS 137
4.1.4 APOT 145
4.1.5 Confirmation of integration by PCR 151
4.1.6 Quality control 152
4.2 Assay Validation: Discussion 152
4.2.1 Integration Detection 153
4.2.2 Detection of integration with presence of episomal HPV 154
4.2.3 Assay sensitivity and reproducibility 156
4.2.4 Integration and E1 and E2 Disruption 158
4.2.5 Unanswered questions: Assay Cross Reactivity 158
4.3 Assay Validation: Conclusion 159
5 Assay Validation Part 2: Evaluation in Primary Culture (PC) Cell Lines 160
5.1 Assay Validation Part 2: Results 160
5.1.1 Episomal Status 161
5.1.2 PC0 DIPS: Integration in DNA 162
5.1.3 PC0 APOT: Integrated Transcripts 166
5.1.4 Integration Within Genes 170
5.1.5 Confirmation of Integration Sites in PC0s 171
5.1.6 HPV E2, E6 and E7 Expression and Cell Growth 174
5.2 Assay Validation Part 2: Discussion 175
5.2.1 Assay Validation in PC08 and PC09 176
5.2.2 Integration and VIN and VAIN 177
5.2.3 PC08: Vulval Neoplasia 177
5.2.4 PC09: Vaginal Neoplasia 180
5.2.5 Future work 183
5.3 Assay Validation Part 2: Conclusion 183
6 Application of assays to Clinical Samples 185
6.1 Liquid Based Cytology Samples: Baseline Results 185

6.1.1 DNA Quality 185
6.1.2 E2 PCR 187
6.1.3 DIPS 190
6.1.4 Baseline: Summary 197
6.2 HPV Integration as a Biomarker of Cervical Disease: Results 197
6.2.1 Samples 198
6.2.2 HPV Integration 203
6.2.3 HI-BCD: Summary 209
6.3 Application of Assays to Cervical Cancer Biopsies: Results 211
6.3.1 HPV typing 211
6.3.2 DNA and RNA quality 211
6.3.3 Integration and E2 211
6.3.4 HPV Gene Expression 218
6.3.5 Biopsies: Summary 219
6.4 Integration in Clinical Samples: Discussion 219
6.4.1 E2 PCRs 220
6.4.2 Loss of episomal HPV is not associated with high grade neoplasia 221
6.4.3 Integration is not an early event in cervical neoplasia 223
6.4.4 Assay Comparison in Biopsy Samples 227
6.4.5 Integrated transcripts are not present in all cervical cancers and are not associated with episomal loss. 229
6.5 Integration in Clinical Samples: Conclusion 231
7 Mechanism of Integration 232
7.1 Mechanism of Integration: Results 232
7.1.1 Sites of HPV integration 236
7.1.2 Integration at Fragile sites 236
7.1.3 Integration at DNA repeat elements 237
7.1.4 DNA sequence at site of integration 238

7.1.5 HPV disruption 238
7.2 Mechanism of Integration and disease progression: Discussion 240
7.2.1 Integration was detected at fragile sites. 240
7.2.2 Hot-spots for integration are present 240
7.2.3 Integration is frequent at repeat elements 241
7.2.4 Integration commonly disrupts E1 and E2 ORFs 242
7.2.5 DNA sequence is similar between HPV and human DNA at the sites of integration 242
7.3 Mechanism of Integration: Conclusion 243
8 General Discussion 244
8.1 HPV Integration and Cervical Neoplasia 244
8.2 Integration and E2, E6 and E7 expression 247
8.3 Integration in Human Genes 248
8.4 Alternative mechanisms 248
8.4.1 Methylation 248
8.4.2 DNA instability by Replication of Integrated HPV 249
8.5 Clinical Implications 252
8.5.1 Integration in Vulval and Vaginal Neoplasia: Clinical Implications 252
8.5.2 Integration in Cervical Neoplasia: Clinical Implications 253
8.6 Unanswered Questions and Future Work 256
8.6.1 Integration and Vulval and Vaginal Neoplasia 256
8.6.2 Integration and Cervical Neoplasia 257
9 Conclusion 259
10 Appendices 261
10.1 Appendix 1: HI-BCD Study Protocol and Consent Forms 261
10.2 Appendix 2:Tiffany Onions' PC08 and PC09 Data 288
10.3 Appendix 3: Dean Bryant's Methylation Data 296
10.4 Appendix 4: Associated Publications 298

11 Bibliography 314

List of Figures

Figure 1 Deaths (thousand) due to cancer worldwide and in developed countries in the year 2008. Data taken from Ferlay et al. (2010). 3
Figure 2 Number of newly diagnosed cervical cancer cases in the UK in 2000 and 2010, grouped by age, taken from ONS (2000; 2010). 3
Figure 3 The Human Papillomavirus genome. Taken from Raybould et al. (2011). 7
Figure 4 HPV16 alternative splicing, adapted from Zheng and Baker (2006) and Milligan et al. (2007). 9
Figure 5 HPV16 oncoproteins and the cell cycle 13
Figure 6 Percentage of cytology test results by age in Wales 2011/2012. 18
Figure 7 Schematic diagram representing progression from mild dysplasia to invasive carcinoma adapted from Howley and Lowy (2007) and Schiffman et al. (2007). 21
Figure 8 HPV type distribution in cervical neoplasia in the UK; adapted from Castellsague et al. (2007). 23
Figure 9 Proportion of HPV related cancer worldwide; adapted from Parkin et al (2006). 28
Figure 10 Schematic representation of a productive HPV infection in a low grade lesion compared to a non-productive, transformed high grade lesion. 31
Figure 11 HPV integration in human genome; adapted from Wentzensen et al. (2004) 35
Figure 12 Non-homologous end joining adapted from Lange et al. (2011). 37
Figure 13 pBR322 vector adapted from Sutcliffe (1979). 56
Figure 14 pAT Vector adapted from Twigg and Sherratt (1980). 56
Figure 15 pGEM®-4 vector 58
Figure 16 Proportion of CIN grades in samples selected for integration analysis grouped by HPV type. 60
Figure 17 Nanodrop out-put screen. 67
Figure 18 Agilent Bioanalyser Output File. 69
Figure 19 The Polymerase Chain Reaction. 71
Figure 20 A schematic example of overlapping primers used for E2 PCR for HPV16, HPV18 and HPV45. 81
Figure 21 Steps involved in Restriction site PCR (RS-PCR). 84

Figure 22 Detection of Integrated Papillomavirus sequences (DIPS). 90
Figure 23 Amplification of Papillomavirus Oncogene Transcripts (APOT). 97
Figure 24 DNA sequencing chromatogram. 101
Figure 25 Expected BLAST output for an integration site. 107
Figure 26 BLAST alignment of data shown in Figure 25. 108
Figure 27 Output from UCSC BLAT alignment. 109
Figure 28 Output from NCBI "Gene" database when searched for "Fragile AND 8". 110
Figure 29 An example of a melting curve. Samples that generated secondary artefacts were excluded from further analysis.....120
Figure 30 HPV16 E2 PCR validation.. 126
Figure 31 RS-PCR detected integration in SiHa DNA. 130
Figure 32 All amplicons produced by RS-PCR were subjected to DNA sequence analysis. 132
Figure 33 RS-PCR can detect integration with presence of HPV episomes.....135
Figure 34 Electrophoresis of DIPS amplicons with a 2 hour restriction digestion time.....139
Figure 35 Validation of DIPS on cell line DNA. 141
Figure 36 Transcripts detected by APOT in cell lines. 146
Figure 37 Confirmation of novel integration sites by PCR with primers flanking integration sites. 152
Figure 38 Summary of PC08 and PC09 cloning. 161
Figure 39 Schematic alignment of PC08 (A) and PC09 (B) DIPS and RS-PCR data to HPV16 (NC_001526.1). 165
Figure 40 Non-contiguous sequence was confirmed in the P clone: 173
Figure 41 DNA Degradation of samples stored in SurePath™ cytology fluid. 186
Figure 42 Comparison of E2 as a marker of viral episome presence, with intact E2, or absence, with disrupted E2, in different cytology grades. 190
Figure 43 Integration detection by cytology groups.. 196
Figure 44 Integration detection by histology group: 196
Figure 45 Histology outcome (No CIN or no biopsy, CIN1, CIN2 CIN3 and CIN unknown) 200
Figure 46 HI-BCD DNA Integrity. 200
Figure 47 Log_{10} Calibrated Normalised Relative Quantities (CNRQs) of E2, E6 and E7 in HPV16 positive HIBCD samples. 209
Figure 48 Schematic alignment of biopsy A) RS-PCR and B) DIPS data to HPV16 (NC_001526.1). 215

Figure 49 Log_{10} Calibrated Normalised Relative Quantities (CNRQs) of E2, E6 and E7 in HPV16 positive cervical cancer biopsies. 218
Figure 50 Integration sites detected relative to Wentzensen et al. (2004). 235
Figure 51 Frequency of HPV disruption across E1 to E2 ORFs. 239
Figure 52 Model of replication of integrated HPV, adapted from Kadaja et al. (2009). 251
Figure 53 Multifocal clonal integration in vulval and vaginal neoplasia. 252
Figure 54 E2 PCR gel electrophoresis of W12p12, W12p32 and C33A. HPV16 plasmid DNA was used as a positive control and water as a negative control. 288
Figure 55 E2 PCR gel electrophoresis of Hela, C4I and C33A. 289
Figure 56 Optimisation of HPV45 E2 PCRs at differing $MgCl_2$ concentrations. 290
Figure 57 Transcripts detected in PC08 and PC09 by APOT. 291
Figure 58 HPV16 E2, E6 and E7 expression ratios, relative to CaSki and house-keeping genes TBP2 and HPRT for all PC08 clones (M, P and Y) and PC09 (A, D and H) clones at each passage. 292
Figure 59 Population doubling (PD) and doubling time (DT) for heterogeneous PC08 (A) and PC09 (B) cell lines. 293
Figure 60 Population doubling (A) and doubling time (B) in clones derived from PC08. ... 294
Figure 61 Population doubling (A) and doubling time (B) in clones derived from PC09. ... 295
Figure 62 HPV16 URR methylation and HPV E2, E6 and E7 expression in VIN biopsies. 296

List of Tables

Table 1 Association of HPV and clinical disease of genital tract and other lesions, adapted from Howley and Lowy (2007). 5
Table 2 Summary of HPV proteins and their functions; taken from Raybould et al. (2011). . 10
Table 3 Current screening intervals taken from NHSCSP (2010). 16
Table 4 Colposcopy management; taken from guidelines for NHS screening programme (NHSCSP 2010). 20
Table 5 Summary of HPV integration in cervical neoplasia detected by qPCR and Southern blot. Summary of available tabulated and published data. 32
Table 6 Prevalence of integrated and episomal transcripts in cervical neoplasia. Data taken from Klaes et al (1999) and Vinokurova (2008). 39

Table 7 A summary of the most common DNA and RNA based HR-HPV integration detection methods. 44
Table 8 Hypotheses for the PhD. 46
Table 9 Propagation and sub-culturing for cell lines. 54
Table 10 PCR conditions for all assays used in this PhD. 72
Table 11 Reagents used in all PCRs in this PhD. 74
Table 12 Primers for HPV16, HPV18 and HPV45 E2 PCR....82
Table 13 Taken from Ferber, Montoya et al (2003). RSO primers used for RS-PCR....85
Table 14 HPV16-specific primers used for RS-PCR....86
Table 15 HPV18-specific primers used for RS-PCR....87
Table 16 HPV45-specific primers used for RS-PCR....88
Table 17 Adapter Primers used in DIPS taken from Luft et al. (2001). 91
Table 18 HPV16 primers used in DIPS. From Luft et al. (2001). 92
Table 19 HPV18 primers used in DIPS. From Luft et al. (2001). 92
Table 20 Reagents used in Sau3AI digestion. 93
Table 21 Reagents used in Taq$^{\alpha}$I digestion.....93
Table 22 Reagents used in Sau3AI digestion. 94
Table 23 Reagents used in ligation of adapter primers to digested DNA.....94
Table 24 Reagents used in modified ligation of adapter primers to digested DNA. 95
Table 25 Expected amplicon sizes for DIPS for HPV16, HPV18 and HPV45 96
Table 26 Reagents used in reverse transcription.....99
Table 27 APOT primers for HPV16, HPV18 and HPV45. 99
Table 28 Reagents for Exonuclease I reaction....103
Table 29 Reagents for Alkaline phosphatase reaction....103
Table 30 Primers used for confirmation of integration or non-contiguous sequences. 112
Table 31 Reagents used in step 1 and 2 of reverse transcription for qPCR.....118
Table 32 Primer sequences for qPCR. 119
Table 33 Summary of E2 PCR validation results for all cell lines and plasmid encoding HPV18 and HPV45.....127
Table 34 E2 and E6 PCRs: lowest copy number of HPV DNA that produced PCR amplicons. 129
Table 35 Results of RS-PCR in validation on CaSki, SiHa, HeLa, C4I and HTB-34™ cell line DNA.. 131

Table 36 RS-PCR: the number of HPV copies required to detect integration in RS-PCR for HPV16, HPV18 and HPV45. 136
Table 37 Validation of DIPS in CaSki, SiHa, W12, HeLa, C4I and HTB-34™. 140
Table 38 Smallest number of integrated HPV copies that produced an amplicon with DIPS for HPV16, HPV18 and HPV45 with 16 or 2 hour digestion incubation. 144
Table 39 Transcripts detected by APOT. 147
Table 40 Relationship between DNA integration sites detected by DIPS and transcribed integration sites detected by APOT; 148
Table 41 PC0 RS-PCR, DIPS and E2 results. 164
Table 42 APOT results for PC08 clones M, P and Y and PC09 clones A, D and H at earliest and latest passage. 168
Table 43 Summary of integration detected in DNA by DIPS and transcripts by APOT for PC08 clones M and Y and PC09 clones D and H. 169
Table 44 Summary of integration sites detected by PCR on DNA from heterogeneous PC08, passages 5 and 11 of M, P and Y clones 172
Table 45 Summary of integration sites detected by PCR on DNA from heterogeneous PC09, passages 5 and 11 of A, D and H clones 172
Table 46 Comparison of E2 as a marker of viral episome presence, ... 189
Table 47 DIPS integration data for baseline samples. 194
Table 48 Summary of HI-BCD samples. 199
Table 49 Summary of HI-BCD E2 PCR and E6 PCR results. Age, infection type, HPV type, cytology and histology outcome are shown. 203
Table 50 DIPS and APOT results. Site of integration, accession number, site of HPV disruption and transcripts detected by APOT are shown. 207
Table 51 Calibrated Normalised Relative Quantities (CNRQs) of E2, E6 and E7 in HPV16 positive HI-BCD samples. 208
Table 52. E2 PCR and RS-PCR results for six cervical cancer biopsies. 214
Table 53 DIPS Biopsy Results 216
Table 54 APOT Biopsy results. 217
Table 55. Calibrated Normalised Relative Quantities (CNRQs) of E2, E6 and E7 in biopsy samples. 218
Table 56 Summary of integration sites detected. 233
Table 57 Summary of integration events detected by DIPS by Dean Bryant in VIN biopsy samples. 297

List of Abbreviations

APOT	Amplification Of Papillomavirus Sequences
CaCX	Cancer of the Cervix
CFS	Common Fragile Site
CIN	Cervical Intraepithelial Neoplasia
Ct	Threshold Cycle
DIPS	Detection Of Integrated Papillomavirus Sequences
DT	Doubling Time
EIA	Enzyme Immunoassay
ELISA	Enzyme Linked Immuosorbant Assay
EV	Epidermodysplasia Verruciformis
FISH	Fluorescent In-Situ Hybridisation
FRA	Fragile Site
HNSCC	Head and Neck Squamous Cell Carcinoma
HPV	Human Papillomavirus
HR	High Risk
ISH	In-Situ Hybridisation
LCR	Long Control Region
LINE	Long Interspersed Nuclear Element
LTR	Long Terminal Repeat
NHEJ	Non Homologous End Joining
NRQ	Normalised Relative Quantity
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PD	Population Doubling
PV	Papillomavirus
qPCR	Quantitative PCR
RACE	Rapid Amplification Of cDNA Ends
RIN	RNA Integrity Number
RS-PCR	Restriction Site PCR
RT-PCR	Reverse Transcriptase PCR
SINE	Short Interspersed Nuclear Element
SR	Serine Rich
URR	Upstream Regulatory Region
VAIN	Vaginal Neoplasia
VIN	Vulval Neoplasia

List of Abbreviations:Genes

Gene	Name
AP1S3	Adapter Related Protein 1
CDCP1	CUB Domain Protein
CLCNKB	Chloride Channel Voltage Sensitive Kb
GAPDH	Gyceraldehyde-3-phosphate dehydrogenase
hTERT	Human Telomerase
HPRT	Hypoxanthine phosphoribosyltransferase
LARGE1	Like Glycosyltransferase
LEPREL1	Leprecan Like Gene
MICAL2	Microtubule Associated Monoxygenase Calponin And LIM Domain
MYC	MYC Proto Oncogene Protein
pRB	Retinoblastoma Protein
RBBP8	Retinoblastoma Binding Protein
SLC7A11	Solute Carrier Family 7
TBP2	Tata Box Binding Protein
TGIF1	TGFB Induced Factor Homeobox

Summary

Cervical cancer is the second most common cancer among women worldwide. Infection with Human Papillomavirus (HPV) is essential but not the only contributing factor in cervical cancer development. HPV integration is reported to be present in over 80% of cervical cancers and disruption of HPV genome through integration leads to high levels of HPV oncogene expression. DNA damage and repair pathways are thought to induce HPV integration since HPV is detected at fragile sites in the human genome. There is controversy as to whether integration is an early or late event in cervical oncogenesis and there are no published studies to date that have investigated HPV integration using sensitive, DNA based, techniques at the nucleotide level in cervical precancers. This study aimed to test the hypothesis that integration is an early event in cervical neoplasia and episomal loss causes malignant transformation through transcription of integrated HPV. Also, this study served to pilot whether HPV integration can predict high-grade cervical disease in women with cytological abnormalities with an aim to improve current cervical screening methods. Assays to detect integration and E2 as a marker of episomal state were developed for HPV16, HPV18 and HPV45 and applied to cervical smears and biopsies from women with varying disease grades. The data presented in this thesis highlight that integration may not be essential for cervical cancer progression and different modes of disease progression may exist between young women and older women. Integration was detected at chromosome fragile sites but was more prevalent at SINE or LINE repeat elements; this implies a role for retroelements in the mechanism of integration. Finally, the data here suggest that integration induces a unique selective process in each individual and clonal selection may arise due to altered HPV oncogene expression and/or disruption to human gene expression.

1 Introduction

Cervical cancer is the second most common cancer in women worldwide with 500,000 women developing cancer and 274,000 women dying from cervical cancer each year (WHO 2008). In the UK, 2851 women are diagnosed with cervical cancer and 957 women die from cervical cancer per annum (ONS 2010). Figure 1 shows the number of cancer deaths world-wide in women in 2008 (Ferlay et al. 2010); cervical cancer accounts for the second highest incidence of cancer related deaths in women and the proportion of deaths due to cervical cancer is much higher in developing countries.

The age at which women are most at risk from developing cervical cancer in the UK is between the ages 25 to 54 years, whereas cervical cancer is rare in women under 25 years of age (Figure 2) (Sasieni et al. 2003). Due to the age at which cervical cancer affects women, it has a profound effect on a woman's family life, sexual health and career, thus cervical cancer is of great socioeconomic importance.

The primary cause of cervical cancer is Human Papillomavirus (HPV) infection; over 99% of cervical cancers contain HPV DNA (zur Hausen 1976; Durst et al. 1983). A very early, historical study of cervical cancer in married women, widowed women, virgins, nuns, and prostitutes postulated that there was a higher occurrence of cervical cancer in those women with higher sexual promiscuity (Rigoni 1987), but it was not until the early 1960's, during advances in bacteriology, that a connection between sexual infection and cervical cancer was made. In the 1970's, breakthrough studies linked HPV and cervical cancer (zur Hausen et al. 1974).

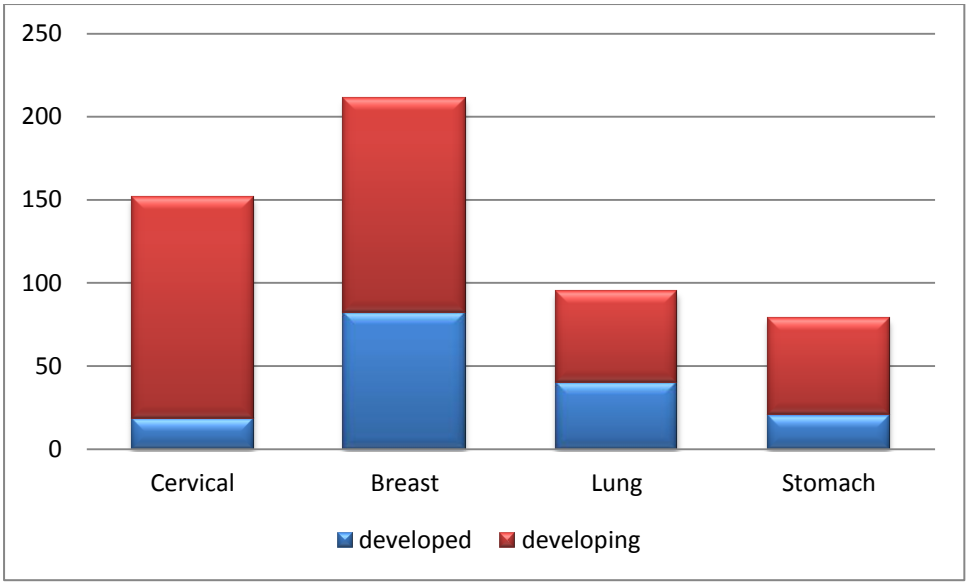


Figure 1 Deaths (thousand) due to cancer worldwide and in developed countries in the year 2008. Data taken from Ferlay et al. (2010).

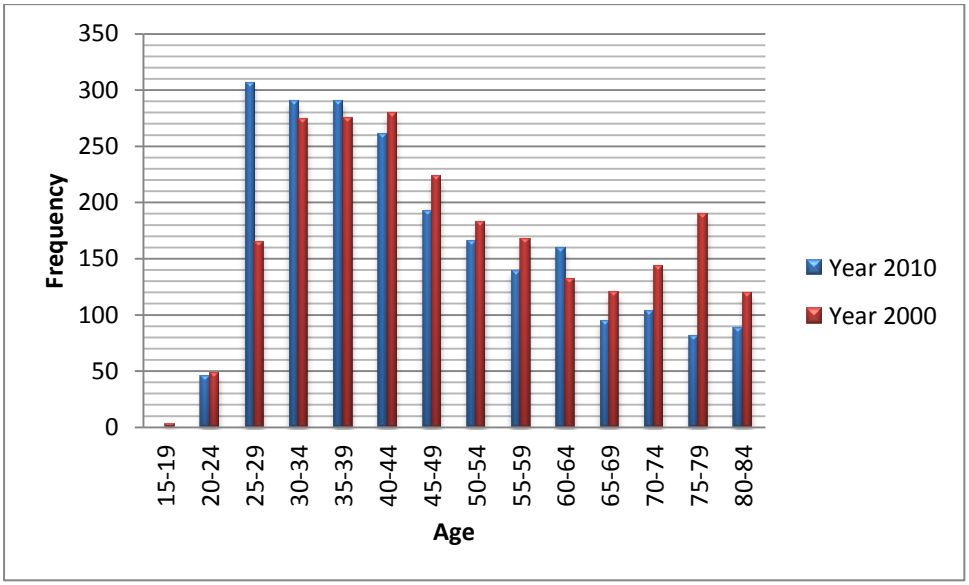


Figure 2 Number of newly diagnosed cervical cancer cases in the UK in 2000 and 2010, grouped by age, taken from ONS (2000; 2010).

1.1 Human Papillomavirus

1.1.1 Classification

Papillomaviruses (PVs) comprise a group of epitheliotropic DNA viruses that induce benign lesions of the skin (warts) and mucous membranes (condylomas). PVs are divided into 12

genera each one coded by a letter from the Greek alphabet. Each PV genus is classified according to the host species it infects and the sites and disease with which it is associated (Howley and Lowy 2007).

The PV phylogenetic classification is made according to the L1 Open Reading Frame (ORF). An ORF is a region of DNA, divided into exons that are spread throughout the genome that encode a specific polypeptide. The L1 ORF controls the expression of major structural viral proteins and is highly conserved between species of PV. Within a genus the L1 ORF DNA sequence is at least 60% similar in all members of that genus. A PV species is designated to members of a genus that share 60%-70% identity. A PV type is given to members of a species that share 71%-89% identity with other types in that species. Subtypes share 90%-98% identity and variants share more than 98% DNA similarity (Howley and Lowy 2007).

Clinical Association	HPV Type
Genital tract	
Subclinical infection	All genital HPV
Exophytic condyloma	HPV6, HPV11
Flat condyloma	HPV6, HPV11, HPV16, HPV18 HPV31, others
Bowenoid papulosis	HPV16
Giant condyloma	HPV6, HPV11
Cervical Cancer	
Strong Association	HPV16, HPV18, HPV31, HPV35, HPV45
Moderate Association	HPV33, HPV35, HPV39, HPV51, HPV52, HPV56, HPV58, HPV59, HPV68
Weak or no Association	HPV6, HPV11, HPV26, HPV42, HPV43, HPV44, HPV53, HPV54, HPV55, HPV62, HPV66
Vulval Cancer	HPV16
Penile Cancer	HPV16
Respiratory papilloma	HPV6, HPV11
Conjunctival papilloma	HPV6, HPV11
Oral and Tonsil cancer	HPV16

Table 1 Association of HPV and clinical disease of genital tract and other lesions, adapted from Howley and Lowy (2007).

HPV belongs to genera α, β, γ, μ and ν with the remainder of the genera being animal specific PVs (de Villiers et al. 2004). α HPV is the largest HPV genera and contains types that infect epithelial cells either of the skin or mucous membranes particularly of the anogenital tract and oropharynx. There are more than 30 α HPV types that infect cervical epithelia, a subset of which induce lesions that progress to malignancy (Table 1). The HPV types that are associated with malignant progression are known as high-risk HPV (HR-HPV) types: HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, and 82 (Munoz et al. 2003). α HPV types that are not typically associated with malignancy and are classed as low-risk (LR-HPV) depending on their prevalence in cancers. HPV types belonging to the β genera can cause latent infections, are predominantly associated with cutaneous infection in immunocompromised individuals, and are also associated with the development of non-melanoma skin cancer in Epidermodysplasia Verruciformis (EV) sufferers who have an abnormal susceptibility to HPV infection (Harwood et al. 2004). The remaining HPV genera that cause cutaneous papillomas and verrucas are not generally oncogenic.

1.1.2 Virion Structure

HPV is a small (52-55nm in diameter) non-enveloped, icosahedral DNA virus that replicates in the nucleus of squamous epithelial cells (Figure 3). One particle consists of one molecule

of circular double stranded DNA. The double stranded DNA is approximately 8kbp long and is contained within a protein coat known as a capsid. There are 72 capsomeres that complete the virion structure. The capsid consists of 2 structural proteins that, as discussed later, also play an important role in mediating virus infectivity. The first capsid protein is encoded by the L1 ORF that encodes 80% of the HPV viral capsid; the second capsid protein is encoded by the L2 ORF that encodes the minor capsid protein.

1.1.3 Genome

The ORFs of HPV are located on one strand of DNA and only one strand of double stranded DNA acts as a template for transcription, this is the coding strand. Eight ORFs are present on the coding strand and these ORFs are classified as early or late because of the location on the coding strand (Figure 3) and the point in the viral life cycle in which they express proteins. Key functions of HPV Early and Late proteins will be discussed in section 1.1.4, discussing HPV infection and HPV life cycle.

A region of the HPV genome, approximately 1Kbp in length contains no ORF. This region contains the origin of replication, transcription control elements and post-transcriptional control elements and is often referred to as the Long Control Region (LCR), Upstream Regulated Region (URR) and Non-Coding region.

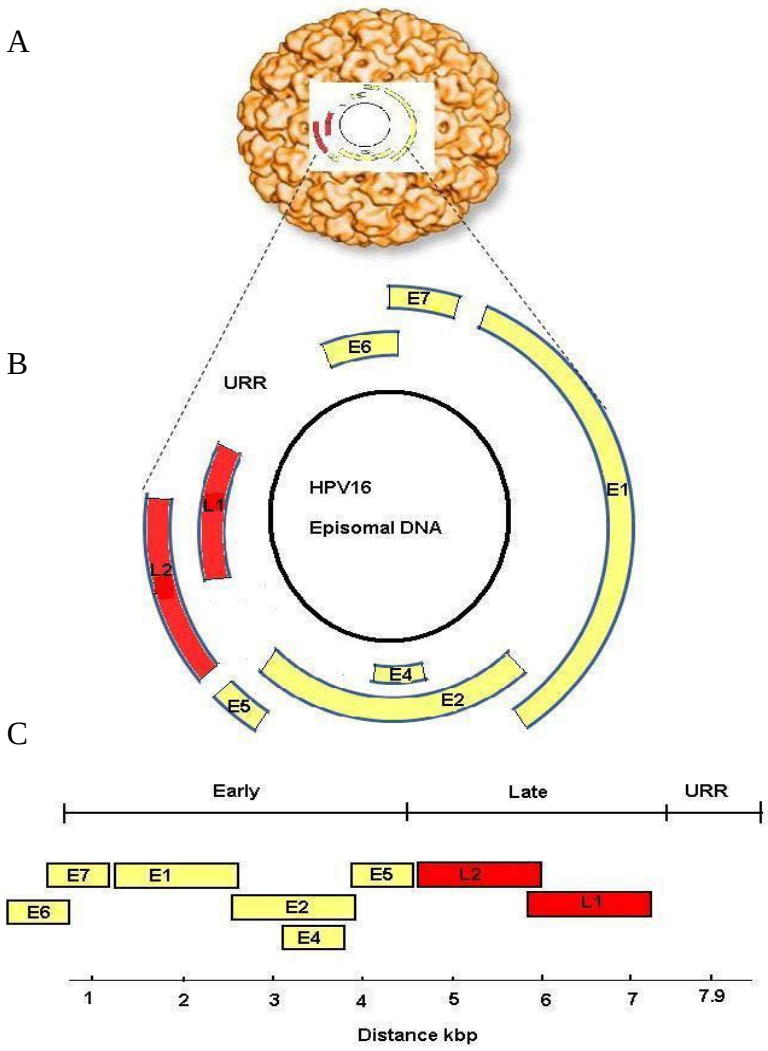


Figure 3 The Human Papillomavirus genome. **Taken from Raybould et al. (2011).** A) A representation of HPV16 episomal DNA contained within the HPV capsid, B) A schematic representation of HPV16 in episomal form, C) A schematic representation of linear HPV16.

1.1.3.1 HPV ORF expression

HPV has a small genome, the ORFs overlap and alternative splicing occurs to allow the differential expression of HPV proteins at different times in the HPV lifecycle (Figure 4). Table 2 summarises the function of the proteins produced by HPV16.

HPV ORF expression is tightly regulated; promoter regions are located relative to early and late protein production. HPV promoter regions are classified according to their position in base pairs within the HPV genome and the point at which they are activated in the HPV life cycle. The promoter positions vary in position between different HPV types. There are a

number of promoters that are active during the HPV life cycle but not all promoters are fully characterized for all HPV types; the two major early promoters are P_{97} and P_{105} for HPV16 and HPV18, respectively; the major late promoter for HPV16 is P_{670}. The P_{97} promoter lies upstream from the E6 ORF and is responsible for nearly all early ORF expression. The P_{670} promoter lies within the E7 ORF and is responsible for late ORF expression (Zheng and Baker 2006).

Polyadenylation is the addition of a string of adenine bases to the 3′ end of mRNA, known as a poly(A) tail. Polyadenylation is the maturation of mRNA for translation and plays an important role in the control of HPV gene expression by preventing mRNA from degradation in the cytoplasm and guiding mRNA out of the nucleus. Once transcription of a gene has ended, the 3′ end of the RNA is cleaved by a set of proteins and the proteins synthesise a poly(A) tail on the 3′ end of the RNA. Polyadenylation sites are coded regions within the genome where polyadenylation proteins act. Similar to regulation of HPV ORF expression from HPV promoters, mRNA polyadenylation is tightly regulated according to the HPV life cycle, the presence of polyadenylation sites, pA_E (Early) and pA_L(Late) are situated at the 3′ end of the Early and Late coding regions respectively.

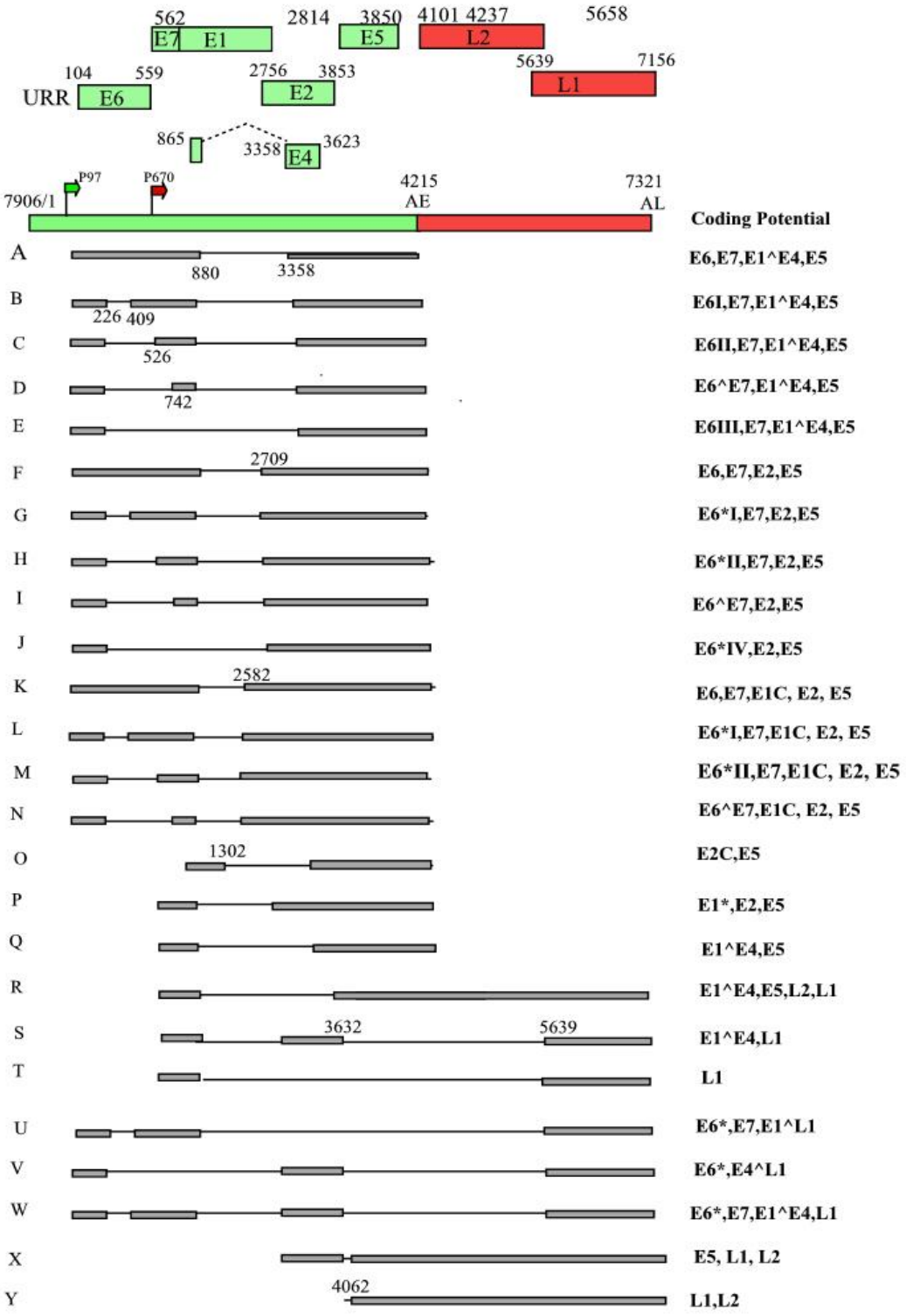


Figure 4 HPV16 alternative splicing, adapted from Zheng and Baker (2006) and Milligan et al. (2007). At the top, in green and red bars, linear HPV16 is represented with nucleotide positions according to NC_001526.1 shown above each ORF; HPV E4 spans 2 exons and requires splicing (dashed lines). Below linear HPV, each denoted by a letter, are the transcripts produced by HPV16; exons shown by grey bars linked by introns; coding potentials for each transcript are shown.

Protein	Function/s	Molecular Activity
E1	Maintains viral genome. Contributes to viral genome replication.	E1 interacts with cellular DNA polymerase and replication proteins to initiate viral DNA replication.
E2	Maintains viral genome. Anchors episomes to mitotic chromosomes in basal cells. Controls transcription.	E2 engages E1 to viral origin and binds E1 to cellular proteins essential for DNA replication. Low levels of E2 activate transcription and high levels of E2 inhibit transcription through interaction with viral early promoters. E2 represses hTERT promoter activity.
E4	Function unclear. E4 has potential roles in the late stage of viral lifecycle, apoptosis and cell cycle arrest.	E4 may bind to keratins and lead to breakdown of the cytokeratin network. E4 may bind to cdk/cyclin complexes leading to arrest of the cell cycle in G2 phase. E4 may interact with E2 and mediate relocation of E2 from nucleus to cytoplasm.
E5	Transformation (Bovine PV) Immune evasion	E5, with E6 and E7 enhances the transforming properties of HPV through interaction with epidermal growth factors and by interfering with cell cycling pathways. E5 may also inhibit apoptosis and cell-to-cell communication. E5 interacts directly with MHC/HLA class 1 and enhances immune evasion.
E6	Transformation	E6 increases cellular stability by binding to p300 and CREB binding protein to block acetylation of p53. HR-HPV E6 mediates ubiquitination and degradation of p53. HR-HPV E6 activates expression of hTERT.
E7	Transformation	E7 disrupts association of pRB with E2F transcription factors, activates cyclin/CDK complexes involved in viral DNA replication and pushes the cell from G1 phase to S phase. E7 stimulates cell proliferation through interaction with histone deacetylases, AP1 transcription complex, p21 and p27 cyclin-dependant kinase inhibitors.

Table 2 Summary of HPV proteins and their functions; taken from Raybould et al. (2011).

1.1.4 HPV Infection

As mentioned above, HPV has a specific tropism for squamous epithelial cells. The HPV life cycle is divided into early and late stages and HPV proteins are differentially expressed temporally and spatially as the life cycle progresses. To replicate and survive HPV must be able to:

1. Enter cell and translocate its genome to site of replication.
2. Replicate its genome and produce mRNA.
3. Generate viral proteins.
4. Assemble progeny.
5. Avoid host defences.
6. Disperse and persist in the environment.

1.1.4.1 Cell Entry

The basal cell is the only cell in squamous epithelial cells capable of undergoing replication and the virus must infect the basal cell to establish a persistent lesion. Primary infection occurs most probably by wound or microabrasion however, in the cervix the primary site of HPV infection is the transformation zone where basal epithelial cells are exposed. To enter the host epithelial cell successfully, HPV must be able to bind to cell surface receptors. Heparan sulphate proteoglycans on the cell surface are considered the initial binding receptor for HPV L1 and L2 (Joyce et al. 1999). Once in a cell the viral capsid disassembles and L2 viral genome then guides the HPV DNA through the cytoplasm into the nuclear domain (Day et al. 2004).

1.1.4.2 HPV DNA Synthesis

Following attachment and un-coating, the HPV genome is maintained in an episomal state in low copy numbers at approximately 10-200 copies per cell (Stanley et al. 1989). E1 and E2 are required to maintain low copy numbers (Frattini et al. 1996; Stubenrauch et al. 1998) to avoid host immune cell detection. As mentioned above, basal cells are capable of replication and E2 is responsible for facilitating correct segregation of HPV episomes during host basal cell division (Hamid et al. 2009).

Prior to HPV genome amplification, expression of HPV Early proteins is initiated. All HPV Early proteins are needed for HPV genome replication; E1 and E2 are responsible for viral DNA replication. E2 plays a vital role in HPV replication: E2 binds to the URR and guides E1 to the viral origin of replication (Sarafi and McBride 1995; Moscufo et al. 1999). E2 has the ability to stimulate or repress the expression of early ORF proteins through binding at a number of sites with a consensus recognition sequence of ACCg NNNN cGGT throughout the HPV genome (Hines et al. 1998). E2 has a dose dependent manner of transcription regulation: in low concentrations, E2 stimulates the promoter of HPV to initiate transcription, whereas in high concentrations it has the opposite affect and represses transcription at the promoter (Steger and Corbach 1997).

In an uninfected epithelial basal cell, once cell division has taken place, the cell exits the cell cycle, migrates to the suprabasal layers and host cell proteins are expressed that allow the cell to enter into a terminally differentiated state. Like many viruses, HPV relies on the host cell machinery to replicate its genome. HPV must trigger expression of the host cell DNA replication genes in differentiated epithelial cells. At this stage of the HPV life cycle, and epithelial cell life cycle, HPV E6 and E7 proteins are expressed. Both E6 and E7 target negative regulators of the cell cycle and have transforming properties (Figure 5).

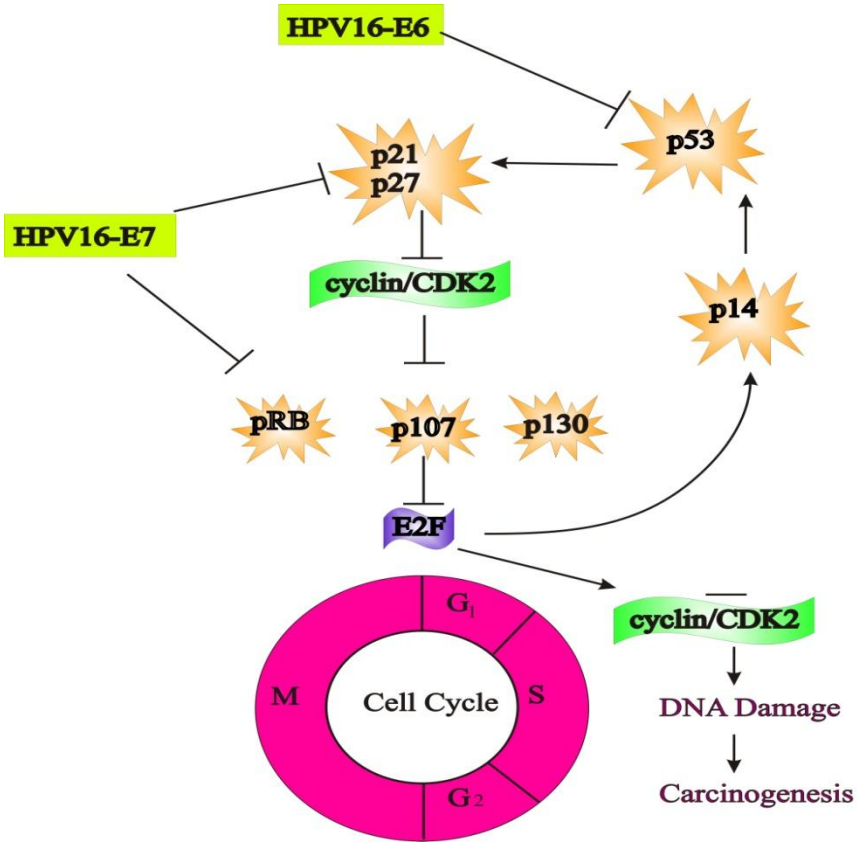


Figure 5 HPV16 oncoproteins and the cell cycle. HPV16 E7 oncoprotein binds and degrades the tumour suppressor retinoblastoma binding protein (pRb) and family members p107 and p130. This leads to E2F transcription factor expression resulting in activated cyclin/CDK complexes. HPV16 E6 induces degradation of tumour suppressor p53 and increases the proportion of cells with DNA defects. Replication of cells with undetected DNA damage at the G2/M cell cycle checkpoint, increases risk of carcinogenesis.

In HR-HPV and LR-HPV types, E7 is the protein primarily involved in inducing the expression of the host cell DNA replication genes. E7 interacts with members of the pocket protein family such as retinoblastoma protein (pRB) (Münger et al. 2001). pRB is a regulator of the cell cycle that interacts with external growth factor E2F; E7 binds to pRB, displaces E2F and initiates expression of host cell DNA replication genes. E7 also associates with other proteins involved in cell proliferation such as AP-1 transcription complex, histone deacetylases and cyclin dependant kinase inhibitors p21 and p27 (Zerfass-Thome et al. 1996; Funk et al. 1997; Jones et al. 1997; Münger et al. 2001). After cell replication, antigrowth signals maintain the cell in a quiescent state and prevent the cell from proliferating. pRB blocks proliferation by altering the function of E2F transcription factor and E7 inactivates pRB to drive quiescent cells back into their proliferative state, driving viral replication. E2F interacts with p14 and this leads to stabilisation of p53, a transcription factor that regulates cell cycle arrest, apoptosis, senescence, DNA repair and cell metabolism. p53 activity is

inhibited by ubiquitin ligase that ubiquitinates p53 to initiate p53 degradation. In HR-HPV, E6 regulates cell cycle activity through indirect interaction with p53: HR-HPV E6 binds to a cellular protein, known as the E6 Associated Protein (E6AP) that functions as an ubiquitin ligase. Ubiquitin-mediated degradation of p53 prevents premature apoptosis and allows the cell to proliferate (Scheffner et al. 1990; Scheffner et al. 1993; Scheffner 1998). Continued proliferation of differentiated cells allows the ends of chromosomes, telomeres to erode and chromosomal DNA becomes unprotected, chromosomes become fused and cell death occurs. To sustain cell viability, E6 also activates the expression of the human telomerase gene (hTERT) an enzyme that synthesises the telomeres (Klingelhutz et al. 1996); this in turn prevents apoptosis.

The exact function of HPV E5 is poorly understood. E5 is not present in every HPV type thus it is not vital for HPV survival and reproduction. It is hypothesised that E5, with E6 and E7 can enhance the transforming properties of HPV through interaction with epidermal growth factors and by interfering with cell cycling pathways (Venuti et al. 2011).

1.1.4.3 HPV Virion Synthesis and release

As mentioned above, L1 and L2 proteins constitute the HPV capsid. Expression of L1 and L2 is in a time dependent manner in order to avoid host immune cell detection and infectious HPV particles assemble in the upper layers of the epithelium (Schwartz 2000). The HPV genome is amplified next to promyelocytic leukaemia bodies (PMLs) that are nuclear multi-protein domains. The HPV DNA, recruited to PMLs by E2 (Day et al. 1998), is packaged into these sites upon virion synthesis. L2 gathers at PML bodies and recruits L1 to these domains. The viral DNA is packaged into the capsid in the upper epithelial layer and released. It is hypothesised that E4, as part of the E1^E4 transcript acts as a late protein, binds to keratins and leads to breakdown of the cytokeratin network and contributes to the release of the infectious particle from the epithelium by apoptosis (Doorbar et al. 1991; Wang et al. 2004; Nakahara et al. 2005).

1.1.4.4 Evasion of Host Defences

HPV has evolved pathways in which to evade destruction in order to replicate and some of these pathways are involved in malignant progression. In the majority of infections, HPV infection does not persist and infection is cleared. Persistent infection with HPV is a primary factor in the development of HPV related cancers. HPV has evolved to adapt its life cycle in order to avoid and interfere with the host immune system:

Firstly, HPV infects non-antigen presenting epithelial cells, and HPV is maintained in low copy numbers in the basal epithelial cells that are not at levels sufficient for detection by the host immune system. Also, HPV genome reproduction occurs in more distal layers, in cells that are not in close proximity to antigen presenting cells that are present in the basal layers.

HPV E5, E6 and E7 proteins have been reported to play a role in immune evasion. E5 directly down regulates the Major Histocompatibility Complex class I protein (Venuti et al. 2011). E6 binds to the Interferon Regulatory Transcription factor 3 and E7 binds to the Interferon Regulatory Transcription factor 1, p48 and membrane protein palmitoylated 3 (Mpp3) leading to disruption of the antiviral response (Yim and Park 2005).

Viral latency is the presence of viral DNA in the absence of differentiation dependent virion production (Stubenrauch and Laimins 1999) and is a means of host evasion allowing the virus to hide from the host immune response. Viral latency should not be confused with clinical latency where no symptoms are present because viral latency may produce symptoms (Moore and Chang 2010). After treatment for HPV infection there is a risk of infection recurrence due to latent infections (Lacey 2005). Studies of latent infections have reported that during latent infection, PV resides in a subset of basal cells and these cells may be epithelial stem cells that are capable of differentiation; PV E1, E2, E6 and E7 ORF expression can be present in latent infections too, with E1, E2, E6 and E7 transcription being essential for episome maintenance in a latent state (Maglennon et al. 2011).

1.1.4.5 Non-productive, transforming infection

HPV has been identified as the cause of certain cancers, detailed in sections 1.1.1 and 1.3 however, only a fraction of people that acquire HPV infection will go on to develop cancer. The major risk factor in the development of cancer is a persistent infection with a high risk type accompanied by disruption to the HPV genome and life cycle. In certain cancers HR-HPV DNA is integrated into the human genome and the HPV life cycle becomes disrupted. The relationship between integration and cervical cancer has been the focus of a great amount of research. To further the understanding of the role of HPV integration in the progression of cervical cancer, the focus of this PhD is an exploration of integration in cervical neoplasia.

1.2 Screening for Cervical Cancer

1.2.1 Cytology

The Papanicolaou test, also known as pap test or smear test, is the current method of screening and it is used to reduce the incidence and mortality of invasive cancer by identification and treatment of pre-invasive disease. Cervical screening in the UK has expanded and changed since its introduction in the 1960s. Early in the 1960s, women were screened when they attended family planning or when attending their GPs and the focus was on younger women. In 1985, a call-recall approach was initiated and all women in Wales between ages 20-64 were invited for screening every 3 years; in England all women aged 25-9 were invited every 3 years and women over fifty were invited every 5 years. Cervical screening saves hundreds of lives each year in the UK and studies modelled on projections of future cervical cancer rates suggest that if the screening program had not been introduced then there would be 11,000 invasive cancers and 5,500 deaths due to cervical cancer by the year 2030 (Peto et al. 2004).

Age Group (Years)	Frequency of Screening
20	First invitation in Wales
25	First invitation in England
25-49	Three yearly
50-64	Five yearly in England. Three yearly in Wales.
65+	Only screen those that have not been screened since age 50 or who have recent abnormal tests.

Table 3 Current screening intervals taken from NHSCSP (2010).

The age for first invitation to cervical screening is set at 25 years in England. This age limit is set due to reports that suggest cervical screening in women aged 20-24 has little or no impact on rates of invasive cervical cancer up to age 30 (Sasieni et al. 2009). Age of first screening has been under debate since the age was raised from 20-24 to 25 years in 2003 on the grounds that transient changes in the cervix could lead to unnecessary treatment with potentially negative consequences on child bearing (Sasieni et al. 2010) with these reports suggesting that screening in young women may do more harm than good (Sasieni et al. 2009). Screening is less effective in younger women (Sasieni et al. 2003) and models suggest that a change in the screening program for example, not screening women under the age of 25 years would have no effect on the life-time risk of cervical cancer (Canfell et al. 2004). Under the age of 25 years, cervical cancer is rare (Figure 2) but cytological abnormalities are common (Sasieni et al. 2003) (Figure 6), probably representing the acquisition of HPV infection and a susceptible transformation zone in the cervix of young women. However the incidence of cervical cancer in young women 25-29 years appears to be increasing compared to a decade ago and it is worrying that there may be insufficient time to initiate screening prior to development of invasive disease. Thus it could be argued that it would be more beneficial to improve the screening methods in young women, rather than cease screening those under 25 years of age.

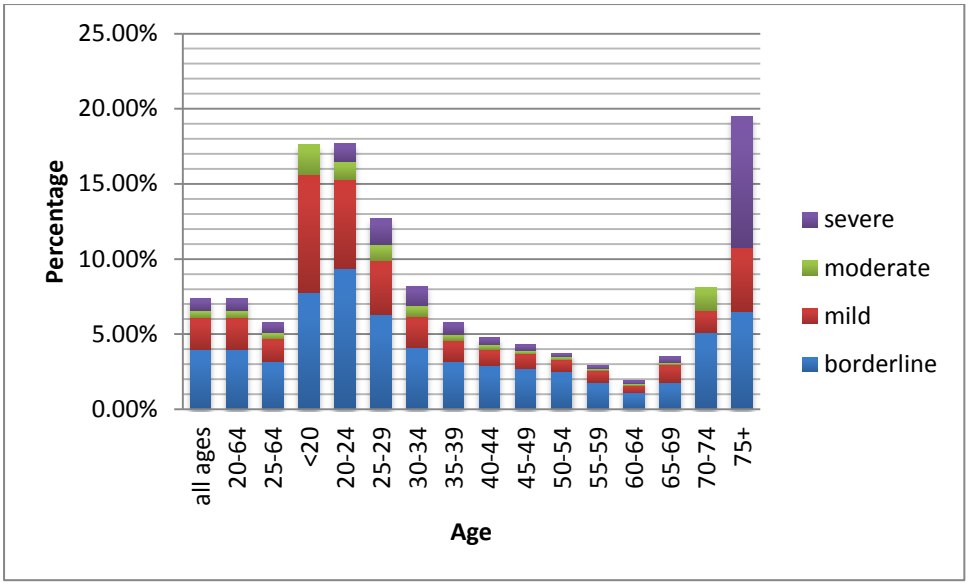


Figure 6 Percentage of cytology test results by age in Wales 2011/2012. Taken from Cervical Screening Wales Statistical report 2011/2012 (KC53/61/65 2012).

When NHSCSP raised the age for screening from 20 to 25 years in 2003 Cervical Screening Wales sought advice from the Academic Department of Obstetrics and Gynaecology, Wales College of Medicine, Cardiff University for the age at which to commence cervical screening. In response, a review was conducted to assess evidence for benefit and harm of undertaking cervical screening in women aged 20-24 years (Rieck et al. 2006). The report did not identify any literature that investigated the effectiveness in young women but indicated that screening in young women reduced the incidence of cervical cancer and the report suggested that the Welsh Cervical Screening should continue to screen women from 20 years of age; thus the age of cervical screening in Wales commences at 20. The age that a woman exits the screening program throughout the UK is set at 65 years of age although it may be safe to withdraw well screened women over the age of 50; this is being kept under constant review by NHSCSP (NHSCSP 2010).

Liquid based cytology (LBC) collection is the choice of cervical smear sampling. Liquid phase procedures frequently use a *Cervex* brush sampler (known as a broom) or an extended tip plastic spatula to collect cells from the cervix. Following sampling of the cervix, the broom or spatula is transferred to a vial containing a preservative fluid (a collection and transport medium), then sent to the laboratory for processing.

Cytological testing is the primary step in screening. The cells obtained from the cervix by LBC collection are observed under a microscope for cytological abnormalities. According to NSCCP guidelines (NHSCSP 2010), women who have dyskaryosis that infers invasive cancer or glandular neoplasia, are referred immediately to colposcopy as part of good clinical practice. Women with a moderate or severe dyskaryosis are seen within 4 weeks of referral, women with borderline nuclear change and mild dyskaryosis are seen in colposcopy within 8 weeks of referral and women that have provided inadequate samples in three consecutive smears are also referred to colposcopy because inadequate samples are associated with lesions that are not exfoliating. Under NHSCSP guidelines, women who have had 3 abnormal dyskaryosis results detected in a 10 year period should be referred to colposcopy as part of good clinical practice see Table 4.

Cytology is an established method to prevent cervical cancer and without it a woman would have a lifetime risk of cervical cancer in the UK of 1.7% compared to 0.46% with cytological based screening (Canfell et al. 2004). Cytology has limitations: cytology is low in sensitivity in detecting pre-cancers and has to be repeated frequently to be effective. Cytological testing is subjective, has low reproducibility and poor positive predicting value: it can miss disease when present and detect cellular abnormalities that may regress naturally without the need for treatment; this may result in overtreatment of women. Additionally, cytology is labour intensive and although the materials are relatively low in cost the reality of repeat testing means that cytology may not be the most cost effective method of screening. Finally, as explained above, cytological testing is least effective in women under 25 years of age (Sasieni et al. 2003; Canfell et al. 2004) and, with improved screening methods, it may be possible to better detect cervical abnormalities improving cervical cancer prevention, especially in younger women.

Cytology Outcome	Clinical Management (NHSCSP 2010)
Borderline dyskaryosis	Referred to colposcopy after 3 abnormal smear results; colposcopy: see and treat basis; return to normal recall screening after 3 negative smears 6 months apart.
Mild dyskaryosis	Referred to colposcopy after 2 mild smear results; colposcopy: see and treat basis; return to normal recall screening after 3 negative smears 6 months apart.
Moderate dyskaryosis	Referred to colposcopy after 1 abnormal smear result; colposcopy: see and treat basis; 62 day pathway to rule out cancer; then follow 18 week screening pathway.
Severe dyskaryosis	Referred to colposcopy after 1 abnormal smear result; colposcopy: see and treat basis; 62 day pathway to rule out cancer; then follow 18 week screening pathway.
Invasive	Referred to colposcopy after 1 abnormal smear result; GP referral to secondary care with a 2 week wait.

Table 4 Colposcopy management; taken from guidelines for NHS screening programme (NHSCSP 2010).

1.2.2 Colposcopy

Colposcopy is a method of illumination and magnification to view and detect premalignant and malignant lesions of the cervix, vagina and vulva. The cervix is stained with acetic acid and the cervix is viewed for lesions. Table 4 lists the guidelines for colposcopic management of patients: if a lesion is observed, a colposcopically directed biopsy is taken for histological examination. Confirmed High grade CIN is then treated. Currently, the most popular mode of treatment is LEEP (Loop Electrosurgical Excision Procedure) or LLETZ (Local Loop Excision of Transformation Zone). Treatment is followed by follow-up cytology and colposcopy six months later and then annual cytological follow-up for 10 years. Lesions that are removed are sent for histological evaluation. Figure 7 shows the progression of a cervical lesion from HPV infection to cervical cancer. The histological classification of biopsies taken in colposcopy can be by a 2 tier Bethesda classification: grouped into low grade or high grade squamous intraepithelial lesions (LSIL and LSIL, respectively) (Solomon et al. 2002), or a 3 tier system: grouped into cervical intraepithelial neoplasia 1, 2 and 3 (CIN1, CIN2 and CIN3, respectively) (Richart 1973).

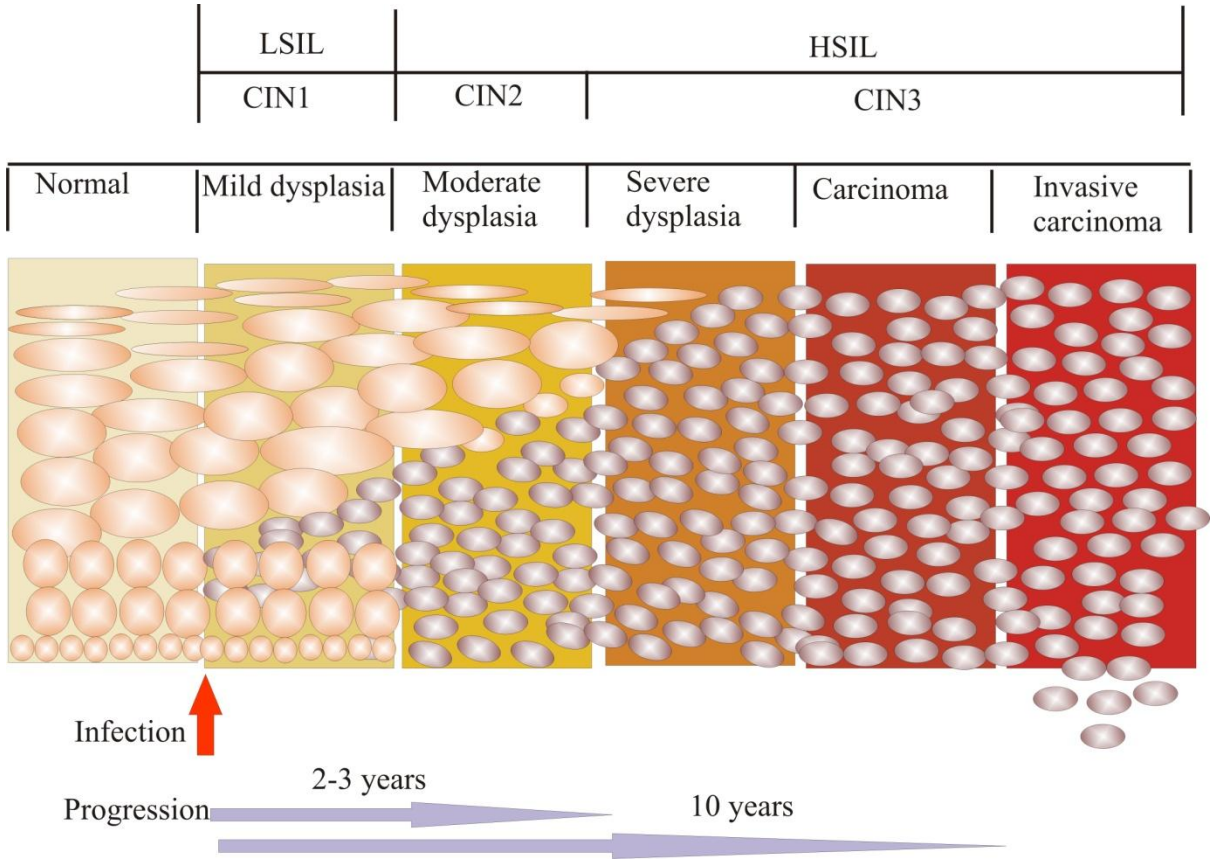


Figure 7 Schematic diagram representing progression from mild dysplasia to invasive carcinoma adapted from Howley and Lowy (2007) and Schiffman et al. (2007). Infection with HPV initially induces benign cervical lesions with mild dysplasia of cervical intraepithelial neoplasia grade 1 (CIN1). Progression to cervical cancer can take years and; cancers and precancers can be diagnosed by histological examination.

Cervical cancer usually takes many years to develop: the peak incidence of HPV infection being 20 years of age, where HPV changes or CIN1 is detected by histology; in the UK the peak incidence of CIN3 is 30 years of age and the peak incidence of cancer occurs at 40 (Kitchener et al. 2006), although this age may be decreasing. Based on mathematical models, in an unscreened population, the lifetime risk of cervical cancer is 1.7% compared to 0.77% with pre-2003 screening methods (Canfell et al. 2004). A woman is likely to recover from CIN1 without need for treatment whereas regression from CIN3 is uncommon. The methods used to treat lesions of the cervix need careful consideration: it is important to remove/excise CIN3 lesions whereas treatment of CIN1 may do more harm than good as treatment is linked with preterm labour (Kyrgiou et al. 2006; Arbyn et al. 2008). Also, attendance at colposcopy may have an effect on a woman's sexual health and cause stress. Based on current techniques utilised in colposcopy it is impossible to know which CIN1 lesions will progress to CIN3 or cancer and which will regress without the need for treatment. Research is needed to develop

suitable biomarkers that can predict high-grade disease to avoid over treatment and unnecessary stress to women, as well as focusing resources on those most at risk of progression of disease.

1.2.3 Improvements to screening

The following sections will describe ongoing changes that are being made to screening programmes and detail biomarkers that may prove useful in future to improve cervical screening sensitivity, specificity, negative predictive value and positive predictive value.

1.2.3.1 HPV typing

Persistent infection with high risk HPV types is essential for cervical cancer development (zur Hausen 1986) with high risk types HPV16 and HPV18 being detected most frequently in cervical cancer in the UK (Castellsague et al. 2007) (Figure 8). HPV16, HPV18 and HPV45 have the highest type specific risk of cancer in Wales (Powell et al. 2011). HPV typing has a high negative predictive value and can be used to rule out women that do not have an HPV infection and who are unlikely to progress to cervical cancer. Future screening of cervical cancer in the UK will utilise HR-HPV typing to triage women with mild and borderline dyskaryosis: women with cytological grades of moderate dyskaryosis or worse will be referred to colposcopy; women with mild or borderline dyskaryosis will be tested for HR-HPV; those that are positive will be referred to colposcopy and women who are negative for HR-HPV will go back to routine recall (Kelly et al. 2011).

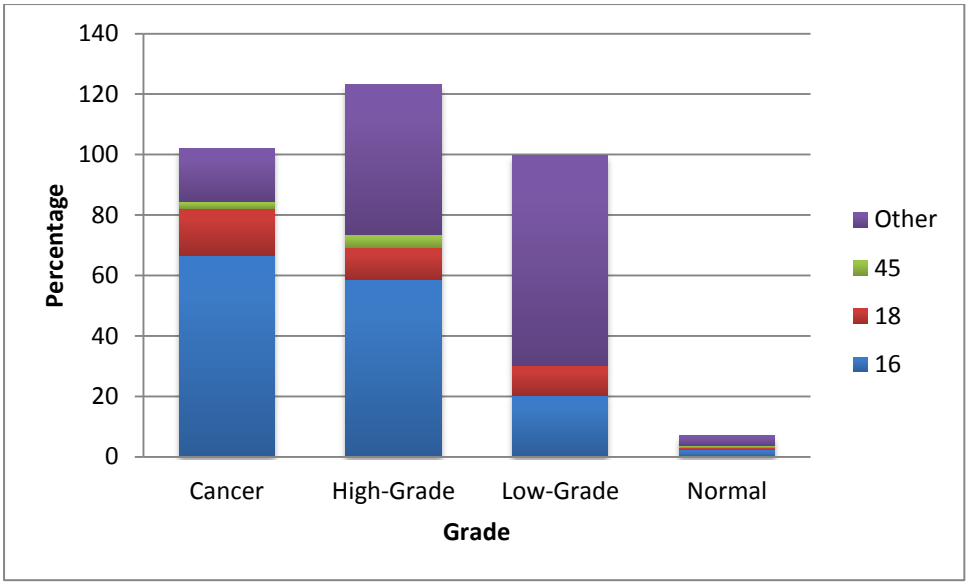


Figure 8 HPV type distribution in cervical neoplasia in the UK; adapted from Castellsague et al. (2007). High grade=HSIL; Low grade=LSIL; other=types other than HPV16, HPV18 or HPV45 (HPV6, HPV31, HPV33, HPV35, HPV39, HPV42, HPV51, HPV52, HPV54, HPV56, HPV58, HPV66 and HPV73).

HPV testing has an excellent negative predictive value: the majority of women that do not have an HPV infection have no risk of developing cervical cancer. Although HPV testing has good negative predictive value, it does remain possible that disruption to the HPV genome through integration can produce a false negative result; thus any HPV negative results need careful interpretation especially if a woman displays severe cellular dyskaryosis. Triage of women with mild and borderline cellular dyskaryosis by HR-HPV testing will reduce the number of women attending colposcopy and being treated unnecessarily. HR-HPV testing has low specificity and there will still be a number of women that have transient HR-HPV infections and do not have a lesion. HPV testing alone cannot determine which women will regress or which women will progress and a suitable biomarker is needed.

1.2.3.2 HPV Vaccination

Due to the strong relationship between cervical cancer and HR-HPV types, prophylactic vaccines were developed. From September 2008-2012, in the UK a bivalent vaccine, Cervarix™ (GSK) that protected against HPV16 and HPV18, was in use; From September 2012 in the UK and in the US a quadrivalent vaccine, Gardasil® (Merck) is used that protects against 2 additional types: HPV6 and HPV11, associated with genital warts. HPV vaccination is targeted primarily at girls aged 12-13 years of age, who have not yet reached sexual

maturity and have not yet been exposed to HPV infection in the genital tract. Clinical trials of the vaccines have reported vaccination reduces persistent and incident infection, high-grade lesions associated with HPV16 and HPV18, and low grade lesions associated with HPV6 and HPV11 (Schiller et al. 2012). Not only do the vaccines protect against cervical lesions, the vaccines potentially offer protection against cancers of the head and neck, vulva and vagina that are associated with HPV16 and HPV18 (see section 1.3). Vaccination of girls will prevent infection of men and boys through herd protection; this in turn will reduce HPV related cancers of the penis and head and neck cancers in males. There is much controversy about the vaccine targets being only girls: there is also a need to protect homosexual males that potentially will not benefit from herd protection against HPV related cancers. Australia has recently been the first country to introduce HPV vaccination for boys.

Vaccination against HR-HPV will not abolish the need for cervical screening: vaccination protects against HPV16 and HPV18 and may offer cross protection against other carcinogenic types, but there is still a risk that type replacement may occur (Saleem et al. 2009). Little is known about the effects of vaccination on cervical cancer and less common HPV types. However, screening will need to be adapted as cytology may become less effective in detecting cytological abnormalities; highlighting the future importance of biomarkers in cervical screening.

1.2.3.3 Biomarkers

The aim of screening is to reduce the incidence of cervical cancer and any biomarkers developed need to test positive when CIN2/3 is present and test negative when CIN2/3 is not present (Arbyn et al. 2009); a positive test in CIN1 is not clinically useful since the majority of CIN1 will regress without need for treatment. A biomarker must be reproducible, specific: produce a negative outcome when disease is absent, and sensitive: produce a positive result where disease is present. There is considerable interest in biomarkers for cervical cancer; the following sections will describe mRNA detection, viral load, p16 staining, methylation and HPV integration and their potential as biomarkers of cervical cancer.

1.2.3.3.1 Integration

HPV integration is widely assumed to be a primary cause of cervical cancer since integration is detected in the majority, but not all cervical cancers. HPV integration disrupts the HPV genome in regions vital for regulation of HPV oncogene expression leading to transformation. HPV integration has been widely studied in cervical cancer and its role as a biomarker to predict high-grade cervical disease has not been thoroughly assessed. The basis of this PhD is to assess the mechanism of HPV integration in cervical neoplasia and examine integration as a biomarker of cervical disease. Section 1.4 will describe HPV integration in detail.

1.2.3.3.2 HPV mRNA Detection

As mentioned above, HR-HPV DNA detection alone cannot predict high-grade disease because many women will have transient infections without disease. In high grade cervical lesions, expression of HR-HPV oncoproteins, E6 and E7, are increased. A number of studies have reported that HR-HPV E6 and E7 mRNA detection is strongly associated with high grade disease (Lie et al. 2005; Molden et al. 2005; Kraus et al. 2006; Molden et al. 2006; Castle et al. 2007) thus detection of E6 or E7 mRNA is a plausible biomarker to be used to predict high grade disease. Disrupted E6 and E7 expression is the endpoint for cervical disease progression; if integration precedes altered mRNA production, then it is possible integration detection is a more suitable biomarker.

1.2.3.3.3 Viral Load

Increased viral load has been reported to be associated with CIN2/3 (Moberg et al. 2004; Fontaine et al. 2005a; Fontaine et al. 2005b) however assays to detect viral load are not suitable in a clinical setting due to sample variability: a sample with combined low and high grade lesions may influence viral load, and multiple types may affect assay performance through cross reactivity. Additionally, if disease progression follows a pattern where viral integration occurs, followed by episome loss as part of disease progression, then viral load would be expected to be lower in high grade disease; viral load detection may therefore lack specificity and lead to false negative outcomes.

1.2.3.3.4 P16^{INK4a} Staining

p16^{INK4a} is a cyclin dependant kinase inhibitor that causes cell death in aged or damaged cells. As described in section 1.1.4, HPV E7 interacts with the Retinoblastoma Binding protein (pRB), this leads to an increased expression of p16^{INK4a}; therefore p16^{INK4a} is a plausible biomarker of cervical cancer. p16^{INK4a} is used as an immunohistochemical biomarker to stain cervical histopathological specimens (Dray et al. 2005) and liquid based cytology samples (Wentzensen et al. 2005). p16^{INK4a} does have limitations and may lack specificity because p16^{INK4a} can be detected in non-dysplastic cells that are dying naturally or in endocervical cells.

1.2.3.3.5 Methylation

DNA methylation is a biochemical process that serves to control gene expression as part of normal cellular development in higher organisms. Methylation is where a methyl group (CH_3) is added to the cytosine of DNA at CpG islands, regions of DNA where C and G nucleotides are situated next to each other in sequence. Studies of methylation in cervical cancer are technically feasible due to the wide availability of assays thus there has been much research into methylation of HPV and human genes:

Since HPV integration is not detected in all cervical cancers and some cancers possess HPV in its natural episomal form, an additional mode of HPV related carcinogenesis must exist. As detailed in section 1.1.3, HPV possesses a regulatory region, known as the Upstream Regulatory Region (URR). The HPV URR possesses CpG islands that have been reported to be hypomethylated or hypermethylated, depending on the stage of the HPV lifecycle (Kim et al. 2003; Kalantari et al. 2008). Changes to HPV methylation, resulting from cervical disease progression can be measured; this makes HPV methylation a plausible biomarker to predict high grade disease; further studies are required to determine the reproducibility, sensitivity and specificity of HPV methylation in a clinical setting.

Methylation in promoter regions of human genes has been implicated in cancer and there has been much interest in methylation of human genes in cervical cancer. A review of methylation of host genes in 52 studies of cervical cancer reported that methylation varied widely between studies and methylation in 7 genes (CDH1, FHIT, TERT, CDH13, MGMT, TIMP3, HIC1), were studied in most detail (Wentzensen et al. 2009). Methylation assays have not yet been widely validated and variation between studies may be due to the variation in assays used to quantify methylation. Wentzensen et al. (2009) also added that there was no evidence that methylation of host genes had good positive predictive value since methylation could not predict CIN2/3. Analysing methylation in panels consisting of a number of host genes may improve sensitivity and specificity.

1.2.3.3.6 Chromosomal Changes

Cancer progression leads to genetic abnormalities such as chromosome duplications insertions or deletions; these can be detected by fluorescent in situ hybridisation (FISH). The most common change, with detection in 77% of cervical cancers, is the increase of the human telomerase RNA component (TERC) this leads to an increase in the long arm of chromosome 3 (Heselmeyer et al. 1997). Although detection of chromosomal abnormalities has good sensitivity and specificity in cytology (Heselmeyer-Haddad et al. 2005), FISH is technically complicated and requires skilled interpretation.

1.3 HPV Related Cancers

HPV is found in all cervical cancers and is a worldwide health burden. The link between HPV infection and development of cancers at other sites in the human body is also of great importance (Figure 9).

Head and neck squamous cell carcinoma (HNSCC) is a global health burden with over 400 thousand new cases each year (Duvvuri and Myers 2009). Tumours occur in the oral cavity, larynx, oropharynx and other regions of the head and neck. Studies have reported an increasing frequency of head and neck cancers associated with HPV worldwide and the frequency of non-HPV related cancers of the head and neck are decreasing due to a reduction

in tobacco use. HPV16 is the predominant type in HNSCC, with HPV18, HPV6 and HPV11 being detected to a lesser extent. HPV related head and neck cancers have a different biology to non-HPV related head and neck cancers and HPV related HNSCC respond better to chemotherapy and radiotherapy and HPV is a strong prognostic factor of HNSCC (Fakhry et al. 2008). Compared to cervical cancer, HPV research related to HNSCC is a relatively new subject and little is known about the HPV pathophysiology in HNSCC.

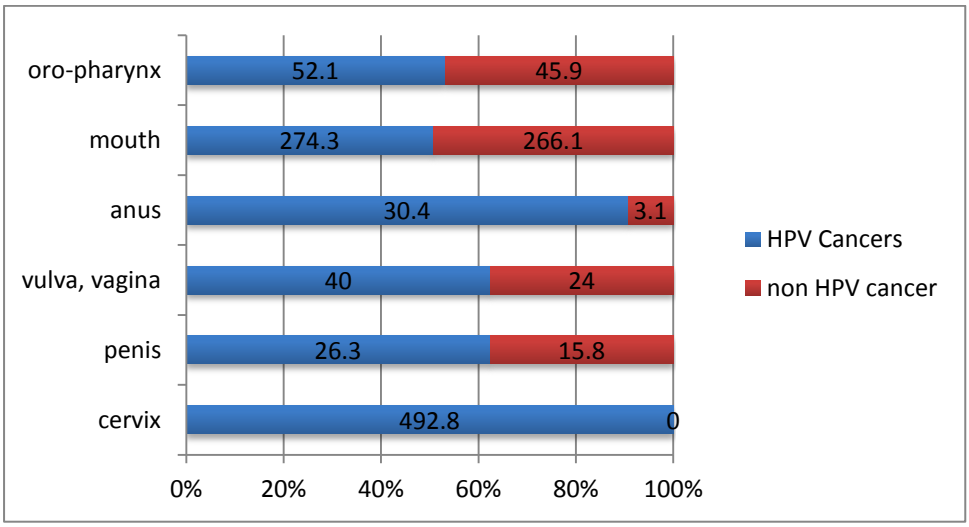


Figure 9 Proportion of HPV related cancer worldwide; adapted from Parkin et al (2006). Detection of HPV in oral or anogenital cancers; numbers of cases (thousand) are displayed on bars.

Transformation of the lower female genital tract occurs much less frequently compared to the cervix but a high proportion of women who develop vaginal or vulval intraepithelial neoplasia (VAIN and VIN, respectively) have previously received treatment for cervical neoplasia (Audet-Lapointe et al. 1990; Rome and England 2000; Dodge et al. 2001). Vulval cancer affects mainly pre-menopausal women with an average age of onset of vulval cancers and vaginal cancers being 10 years greater than cervical cancer (Diakomanolis et al. 2002). Treatment of both vaginal and vulval neoplasia involves the excision of the affected area; treatment is highly disfiguring and has a high recurrence rate. There is currently much interest in topical treatment for both vaginal and vulval cancer with an aim to reduce VIN and VAIN recurrence rates and reduce the need for surgery. The antiviral drug, Cidofovir, is widely used as a treatment in gynaecological conditions such as herpes and condylomata. A study of Cidofovir, in 12 women with VIN, reported complete regression in 4 women and a partial response in 3 women (Tristram and Fiander 2005).

The incidence of HPV-related penile and anal cancers relative to cervical cancer is low. Understanding the epidemiology of HPV penile and anal cancers and HPV transmission to partners is poorly understood, mainly due to under-developed sampling techniques. Although anal cancer is low in the general population, it is common in at-risk groups such as HIV patients and homosexual men.

An improved understanding of the pathophysiology of HPV in HNSCC, anal, penile, vulval and vaginal cancers would lead to better screening methods and therapeutics.

1.4 HPV Integration and Cervical Neoplasia

As mentioned above, HPV is maintained in episomal form in productive HPV infections, represented by low grade lesions in squamous epithelia (Figure 10) and HPV reproduction is synchronised carefully with epithelial differentiation; disruption to the HPV lifecycle leads to oncogenesis. A number of viruses are known to integrate into the host genome resulting in transformation of the host cell, for example, retroviruses express the protein integrase, which facilitates the integration of DNA into the host genome to support their lifecycle; this ultimately results in the disruption of certain host cellular oncogenes, leading to cancer. On the other hand, integration of DNA into the host genome in viruses such as HPV, Epstein - Barr virus (EBV) and Hepatitis B Virus (HBV) does not support the virus lifecycle, is deleterious to virus and results in the expression of viral oncogenes that contribute to the malignancy of the cell.

HPV integration has been detected in over 80% of cervical cancers, using sensitive techniques that detect integration at the nucleotide level (Klaes et al. 1999; Thorland et al. 2000; Luft et al. 2001; Thorland et al. 2003; Ziegert et al. 2003). To date, HPV integration into human DNA has been investigated in cervical neoplasia using quantitative PCR (qPCR) (Peitsaro et al. 2002; Cheung et al. 2006; Kulmala et al. 2006; Huang et al. 2008), or Southern blot (Evans et al. 2008) (Table 5). Overall, reports are consistent that integration is present in CIN1 combined with the presence of episomal HPV. Integration detected without

episomes is conflicting in studies: integration, without episomes, was reported to be present in CIN2 and CIN3 only (Peitsaro et al. 2002; Kulmala et al. 2006) implying integration is a late event, whereas Huang et al (Huang et al. 2008) reported that integration without episomes was detected in CIN1 through to invasive cervical cancer for both HPV16 and HPV18, implying integration was an early event. Another study, reported no integration without episomes, even in CIN3 and that lesions with integrated HPV were capable of regressing rather than progressing to cancer (Evans et al. 2008). Collins et al. (2009) used PCR across the E2 ORF to detect disrupted E2 as a marker of integration (Collins et al. 2009) (data not cited) and reported that E2 disruption was an early event in cervical carcinogenesis. Li et al. (Li et al. 2008) used a similar method of overlapping PCRs across E2 to detect integration in cervical biopsies; an increased frequency of integration was reported in CIN3 and cancer compared to CIN1; additionally this study reported that persistent infection in CIN was associated with integration rather than episomal HPV.

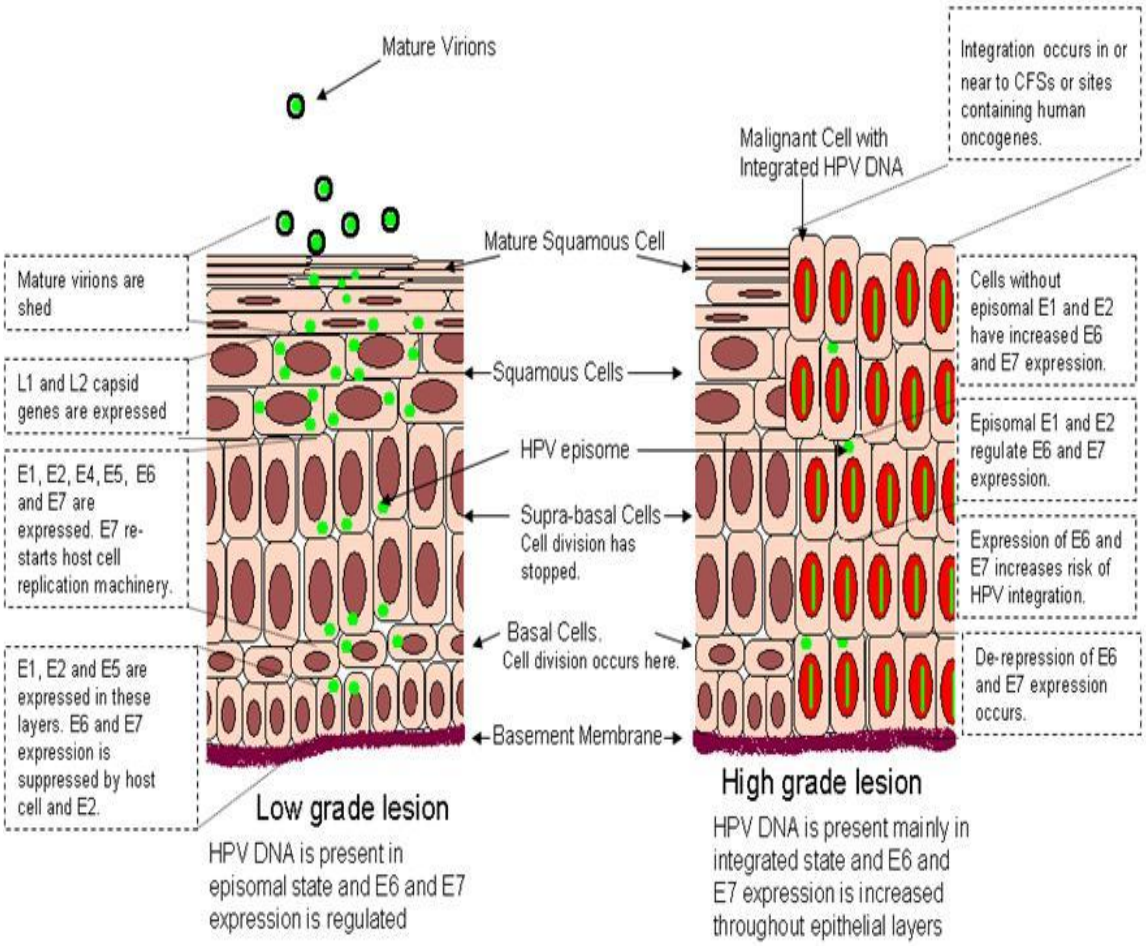


Figure 10 Schematic representation of a productive HPV infection in a low grade lesion compared to a non-productive, transformed high grade lesion. Taken from Raybould, Fiander et al. (2011).

A vast amount of knowledge of the mechanism of HPV integration has arisen from the study of integration in cervical cancer. Integration of HPV into the genome occurs in a number of steps, the first of which is damage to the host and viral DNA followed by DNA repair mechanisms that insert the HPV DNA into the genome. Selection of a cell, containing integrated HPV, forms an immortalised cell that multiplies clonally and progresses to invasive disease. The following sections of this chapter will review the mechanism of HPV integration and describe the molecular events that contribute to integration and lead to cancer.

Author	Assay[1]	HPV Type	N[2]	Histology[3]	Episomal n (%)	Integrated n (%)	Mixed n (%)
Peitsaro (2002)	qPCR (B)	16	31	NCIN	0(0)	0(0)	4 (100)
				CIN1	0(0)	0(0)	4 (100)
				CIN2	1(9)	2(18)	8(73)
				CIN3	0 (0)	11 (92)	1 (8)
Kulmala (2006)	qPCR (C)	16	164	Normal	1 (14)	0	6 (86)
				HPV no CIN	2 (29)	0	5 (71)
				CIN1	4 (29)	0	10 (71)
				CIN2	4 (29)	0	10 (71)
				CIN3	1 (16)	1 (16)	4 (68)
				CaCx	0	2 (33)	4 (67)
Cheung (2006)	qPCR (C)	16	104	Normal/CIN1	9 (31)	not given	19 (69)
				CIN2	5 (21)	not given	19 (79)
				CIN3	9 (33)	not given	18 (67)
				CaCx	7 (28)	not given	18 (72)
Evans (2008)	Southern (C)	16, 18, 43,51	95	Benign	27 (90)	0	3 (10)
				CIN1	7 (100)	0	0 (0)
				CIN2/3	28 (93)	0	2 (7)
Huang (2008)[4]	qPCR (B)	16	101	CIN1	(17)	(33)	(50)
				CIN2/3	(9)	(27)	(64)
				CCI	(17)	(26)	(57)
				CCII-IV	(6)	(47)	(47)
Huang (2008)[4]	qPCR (B)	18	101	CIN1	(50)	(17)	(33)
				CIN2/3	(44)	(44)	(12)
				CCI	(36)	(36)	(28)
				CCII-IV	(67)	(33)	(0)

Table 5 Summary of HPV integration in cervical neoplasia detected by qPCR and Southern blot. Summary of available tabulated and published data.

[1] Assay used and sample type: Southern=Southern blot; C= cervical smear; B= biopsy
[2] Number of samples assayed
[3] Histopathological grade. CIN= Cervical intraepithelial neoplasia, CC= cervical cancer followed by grade I to IV; CaCx= cervical cancer grade not cited; NCIN= HPV detected but no CIN
[4] Exact figures not cited.

1.4.1 Integration at Fragile Sites

Chromosome instability arises due to changes or faults in DNA replication and this instability occurs at chromosome fragile sites. Chromosome fragile sites are regions of genomic instability that are visible as gaps and breaks on metaphase chromosomes following partial inhibition of DNA synthesis. Fragile sites are categorised into two main classes based on their population prevalence and how they are induced in tissue culture:

Common fragile sites (CFSs) are the largest class of fragile sites (Durkin and Glover 2007) and are present on all chromosomes. Unlike rare fragile sites, CFSs are not the result of nucleotide expansion repeats and are induced by aphidicolin (Sutherland et al. 1998). CFSs are highly conserved between species (Durkin and Glover 2007) suggesting that they are key to species survival. CFSs are late replicating and may be the last sites to replicate and thus signal to the cell that replication is complete (Durkin and Glover 2007).

Rare fragile sites are seen in less than 5% of individuals and segregate in a Mendelian fashion (Sutherland et al. 1998). The major group of rare fragile is folate sensitive and the remainder of rare fragile sites are induced by bromodeoxyuridine. Folate sensitive fragile sites are associated with CGG-repeat expansion and this group includes FRAXA in the FMR1 gene, which is responsible for fragile X syndrome and FRAXE in the FMR2 gene, which is associated with non-specific mental retardation (Kremer et al. 1991). Non-folate sensitive fragile sites are characterised by expanded repeat regions of AT-mini-satellite repeats.

For decades it has been recognised that there is a link between fragile sites and cancer (Yunis and Soreng 1984). There are a number of chromosome fragile sites that reside within oncogenes: FRA11B resides within CBL2, a proto-oncogene and ubiquitin protein ligase (Jones et al. 1995). Gene AFF2 resides within fragile site FRAXE (Gecz et al. 1997); FRA7I contains the transforming immortalised mammary oncogene (TIM) linked with breast cancer (Chan et al. 1994). FRA7G has strong links to prostate and breast cancer (Hansen et al. 1997). There could be two possible ways that a fragile site may induce oncogenesis. Firstly, the fragile site may be able to inactivate a gene or be at a region of genome instability that causes deletions that result in alterations of a gene leading to altered transcripts. Secondly, the

fragile site may allow chromosomal breakage in response to DNA damaging agents (Sutherland et al. 1998). For HPV related oncogenesis, scenario 2 is most appropriate, where the fragile site allows breakage of the DNA and HPV is incorporated at the fragile site.

Early studies of HPV integration, using Southern blot and FISH, led to the general conclusion that HPV integration was random throughout the genome (Couturier et al. 1991) but more recently, studies have observed that DNA integration occurs at CFSs (Cannizzaro et al. 1988; Wilke et al. 1996; Thorland et al. 2000; Wentzensen et al. 2002; Thorland et al. 2003; Wentzensen et al. 2004; Yu et al. 2005; Dall et al. 2008; Kraus et al. 2008). Figure 11 shows a schematic representation of integration sites reported in a systematic review of HPV integration (Wentzensen et al. 2004). It remains unclear as to whether DNA integration is seen more commonly in CFSs because the DNA is more prone to breakage, because of factors such as increased expression of E6 and E7 or because the host DNA contains sequences that increase the likelihood of integration occurring. Integration occurs in clusters (Kraus et al. 2008) and clusters of integration may reflect chromosomal regions that are unstable.

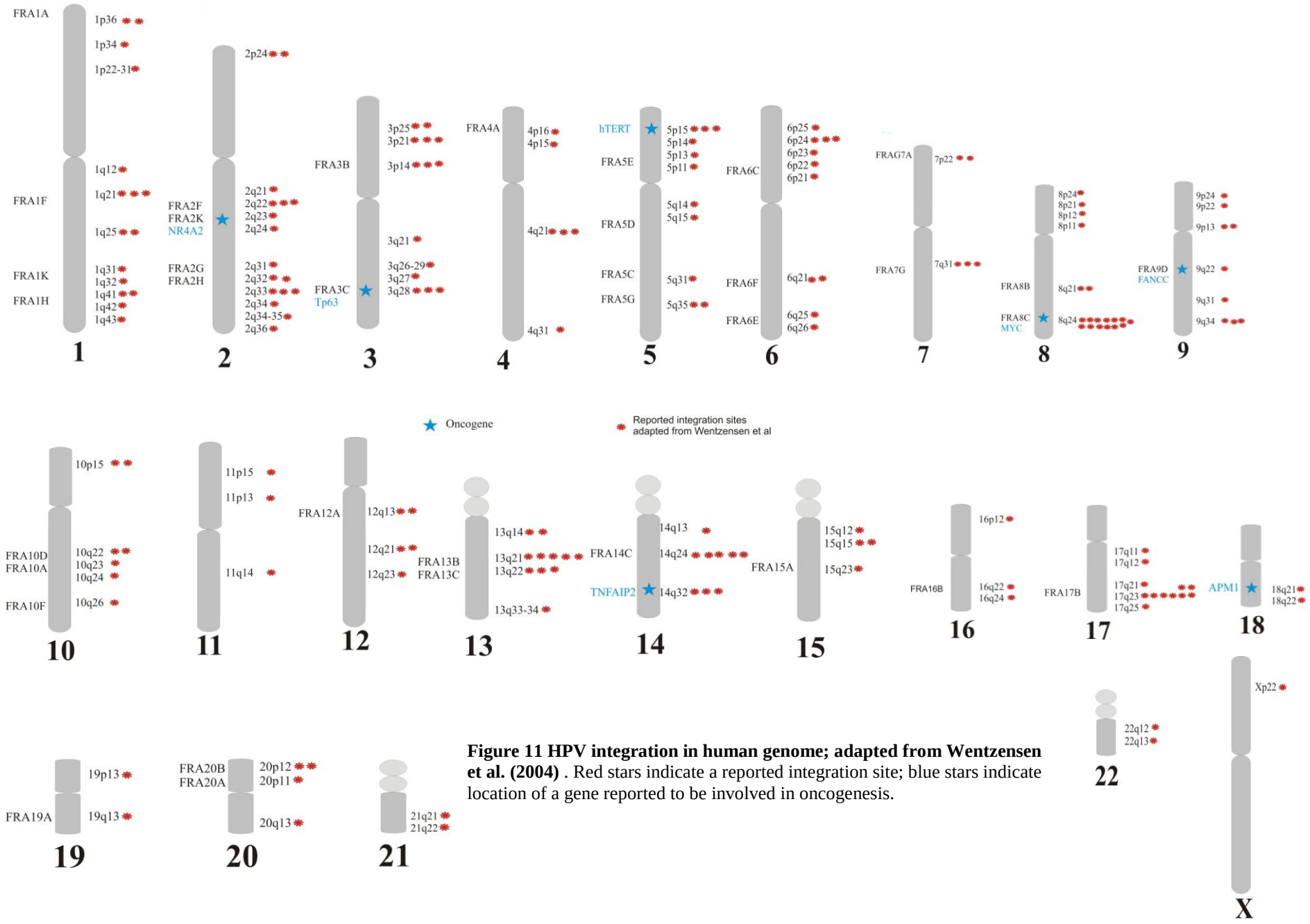


Figure 11 HPV integration in human genome; adapted from Wentzensen et al. (2004) . Red stars indicate a reported integration site; blue stars indicate location of a gene reported to be involved in oncogenesis.

1.4.2 DNA damage

DNA damage occurs when cells are exposed to various chemical agents, UV damage and ionising radiation. Cells have developed check points to detect DNA damage before the cell divides and cell division is delayed in response to any damage detected. E6 and E7 have the potential, independently of each other, to cause genomic instability through centrosome duplication errors and chromosome mis-alignment during mitosis (Duensing and Munger 2002; Duensing and Münger 2003). HR-HPV E6 and E7 expression in basal cells, for example, during latent infections, may increase the risk of HPV DNA integration into host genome and contribute to the malignant phenotype (Hudson et al. 1990; Kessis et al. 1996).

Although HPV integration risk is increased in the presence of DNA damaging agents, it is possible that integration occurs in an already unstable genome: a study of the W12 cell line, described in section 1.4.4, reported HPV16 integration is associated with the acquisition of genomic abnormalities: tetraploidy was observed at low passage in the W12; integration resulted in increased E7 expression and this resulted in aneuploidy and structural chromosome alterations (Pett et al. 2004). A similar observation was made *in-vivo:* biopsies from CIN and cancer were explored for aneuploidy and integration; expression of HPV oncogenes in the basal epithelial cells induced chromosomal aneuploidy and chromosomal instability this favoured integration and in turn increased the expression of E6 and E7 (Melsheimer et al. 2004).

1.4.3 DNA Repair

There are a number of DNA repair mechanisms that include non-homologous end-joining (NHEJ) (Figure 12) and homologous recombination. The mechanism of DNA repair depends primarily on the type of DNA break and the stage of the cell cycle at which DNA damage occurs. NHEJ is the most frequent method of DNA repair in the G1/G0 phase of the cell cycle, NHEJ occurs in response to DNA breakages that have complementary overhangs (Shrivastav et al. 2008). When ends cannot be precisely rejoined NHEJ aligns one or a few complementary bases to direct repair; this results in rearrangements such as small insertions or deletions at the site of repair (Shrivastav et al. 2008). Homologous recombination repair is more active in the G2/S phase of the cell cycle. Homologous recombination repair is considered a more accurate method of DNA repair and involves homologous sequences

elsewhere in the genome such as sister chromatids, repeated regions or homologous chromosomes. Homologous recombination and NHEJ are not independent pathways for DNA repair they both act together to maintain the integrity of the genome. There is also evidence to suggest that when one DNA repair method fails the other acts as a back-up (Valerie and Povirk 2003). NHEJ has been implicated in HPV integration (Durst et al. 1987; el Awady et al. 1987; Ziegert et al. 2003) because, at the site of integration, there is either overlapping sequences of homology between HPV and human sequence or an inserted region of DNA that bears no homology to human or HPV.

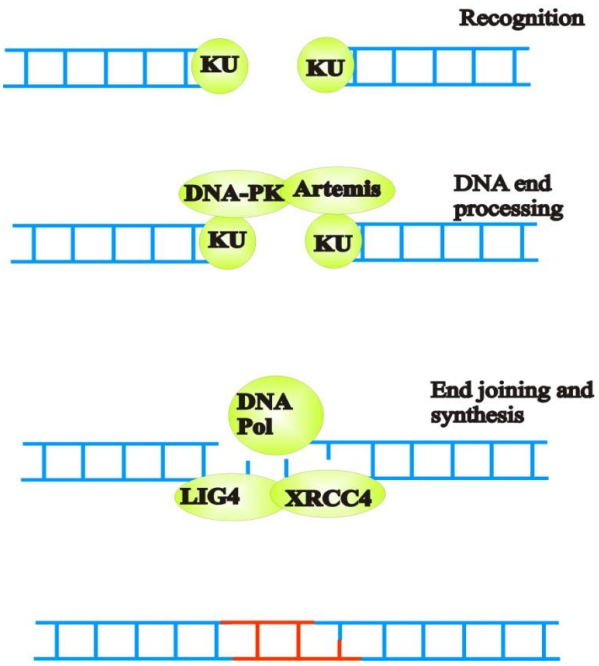


Figure 12 Non-homologous end joining adapted from Lange et al. (2011). Strand breaks caused by ionizing radiation or by enzymes that cleave DNA usually do not yield DNA ends that can be ligated directly. End-trimming and re-synthesis of bases is therefore required to join breaks.

1.4.4 HPV Integration and Selection

The W12 cell line was established from a low grade, warty lesion of the cervix (Stanley et al. 1989) and at low passage the W12 displays characteristics of a low grade disease and maintains episomes at approximately 100 copies per cell whereas at high passage the W12 is transformed and contains solely integrated HPV. A study of integration in the W12 cell line reported that much integration may occur at early stages of cancer progression, the integrants are selected, resulting in one integrant that is actively transcribed (Dall et al. 2008).

Regardless of number of integration sites, natural selection favours only one actively transcribed integrant that is localised to the perinuclear region of a cell (Van Tine et al. 2004). Selection of a cell can take place in the following ways:

1.4.4.1.1 Transcription of Integration Sites

A number of studies have reported increasing prevalence of integrated transcripts with increasing dysplasia (Table 6). Integrated transcripts are reported highest in cervical cancer, with low levels reported in cervical intraepithelial neoplasia (Klaes et al. 1999; Vinokurova et al. 2008). Episomal transcripts are reported at high levels in CIN1 to CIN3 but episomal transcripts are considerably reduced in cancer (Klaes et al. 1999). These studies imply that transcription of an integration event is a late event in the progression of cancer and that *selected* integrants are detected within transcriptionally active sites within the genome. Transcription of integration sites ultimately gives a cell a selective advantage. Also, episomal transcripts are lost in progression from CIN3 to cancer; implying episome loss plays a key role in cancer development.

Integrated transcripts are detected at significantly different levels between HPV types (Vinokurova et al. 2008): the highest level of integrated transcripts detected in HPV16, HPV18 and HPV45 and lowest levels of integrated transcripts in HPV31 and 33. This could mean integration sites are transcribed differently for different HPV types or integration into DNA occurs more frequently for different HPV types.

Finally, transcription of integrated HPV results in structurally different mRNA: integrated transcripts harbour human RNA at the 3′ end whereas natural, episomal transcripts possess an AU rich element. Integrated transcripts may have a longer half-life compared to mRNA of episomal transcripts and stability of mRNA would enhance the selection of a cell harbouring an integrant (Jeon and Lambert 1995).

Author	HPV Type	Histology	Episomal n (%)	Integrated n (%)
1[1]	16 and 18	Normal	19 (100)	0 (0)
		CIN1	10 (100)	0 (0)
		CIN2	21 (95.5)	1 (4.5)
		CIN3	54 (84.4)	10 (15.6)
		CaCx	5 (12.5)	35 (87.5)
2[2]	16, 18, 31, 33, 45	Normal	data not cited	0 (0)
		CIN1		0 (0)
		CIN2		5 (3)
		CIN3		36 (17)
		CaCx		95 (62)

Table 6 Prevalence of integrated and episomal transcripts in cervical neoplasia. Data taken from Klaes et al (1999) and Vinokurova (2008).

1.4.4.1.2 Increased E6 and E7 expression

Early reports (Matsukura et al. 1986; Pater et al. 1986) and many subsequent studies of HPV16 integration have demonstrated that upon HPV integration, variable parts of the HPV genome are lost. Fragments containing E1, E2, E4 and E5 ORFs are disrupted or missing whereas the entire E6 and E7 ORFs are integrated and retained. Over-expression of E6 and E7 genes, due to the disruption of E1 or E2 ORFs, upon integration, may provide a selective growth advantage (Romanczuk and Howley 1992). Jeon et al. (1995) reported, in W12 cell line, cells harbouring integrants possessed lower copy numbers of HPV than cells bearing extra-chromosomal HPV and HPV E7 expression was detected at higher levels in cells with integrated HPV (Jeon et al. 1995). Studies of HPV oncogene expression in relation to integration are limited in clinical samples but hypothetically integration gives a cell a selective growth advantage through disruption of HPV E1/E2 ORFs and leads to increased E6 and E7 expression. Increased production of HPV E6 and E7 oncogenic proteins allows transformation through interaction with hTERT, p53 and pRb. hTERT, p53 and pRb are proteins involved in cell differentiation and HPV replication and have well documented roles in human cell immortalisation (Munger et al. 2004).

[1] Klaes et al. (1999)
[2] Vinokurova et al (2008)

E6 and E7 expression has been demonstrated to be vital for ongoing cell proliferation of primary cervical carcinoma cell lines (Magaldi et al. 2012) and removal of E6 and E7 expression in culture induces senescence. However, integration into the host DNA alone is not sufficient to drive transformation, transcription of integrated HPV will be inhibited in cells possessing both integrated and episomal HPV thus loss of episomal E2 is a key role in the emergence of selectable integrated cells (Pett et al. 2006). A study of episomal loss and integration in the W12 cell line reported that loss of episomes is associated with the emergence of integrants and even the lowest levels of episome could exert an inhibitory effect on the expression of integrated oncogenes (Pett et al. 2006). Pett et al. (2006) reported that rapid loss of HPV episomes also correlated with expression of cellular IFN antiviral genes and loss of episomes through antiviral responses will result in the expression of E6 and E7; this has implications in treatment of HPV infections as described for VIN and VAIN in section 1.3.

Controversially, expression of HPV oncogenes *in vivo* may not follow the same pattern as that reported in the W12 cell line. A study of HPV E6 expression in cervical smear samples taken from women with varying disease grades reported that E6 expression was high in samples bearing only integrated HPV16, HPV18, HPV52 and HPV58; but levels of E6 expression were not significantly different between women who had cancer and women who did not, even if integration was detected (Ho et al. 2011). Integration may confer a higher level of oncogene expression but it may not be sufficient to maintain the transformed phenotype.

1.4.4.1.3 Genomic Instability

Selection of integrants may arise due to a number of factors. Firstly, integration within a cell may offer a selective advantage over other cells due to expression of HPV oncogenes. Secondly, integration may cause chromosomal alterations: as mentioned in section 1.1.4.2, HPV E1 and E2 are responsible for DNA replication and episomal E1 and E2, within a cell containing integrated HPV, has the potential to replicate the integrated HPV resulting in chromosomal alterations (Kadaja et al. 2009a); this offers an alternative mechanism for chromosomal alterations other than chromosomal damage due to increased E6 and E7 expression.

1.4.4.1.4 Integration within Oncogenes

The most frequent site of integration is at chromosome 8q24, at fragile site FRA8C, and corresponds to the MYC gene (Figure 11) that has strong links with cancer and functions in the cell cycle and apoptosis. Integration within a gene or near to a gene may disrupt the function of the gene or protein and give a selective advantage. For example, in a patient with fatal cervical cancer, integration was detected within and disrupted the Tumour Necrosis Factor gene (TNFAIP2). Integration in this region deleted 13kbp including the promoter and first two exons of TNFAIP2 (Einstein et al. 2002). In another study, integration was reported to be within the promoter of the human telomerase gene (hTERT), resulting in a 3-fold increase in the production of hTERT protein (Ferber et al. 2003a).

It is possible that large deletions near to integration sites may account for disruption to human genes. In a study of integration in cervical tumours, although integration was not within known human genes, integration induced large deletions and disrupted the expression of genes within the genomic region surrounding the integration site (Thorland et al. 2003).

On the other hand, integration may occur within or near to a gene and not induce any structural or functional abnormalities (Ferber et al. 2003a; Ferber et al. 2003b; Dall et al. 2008). Integration may have an affinity for transcriptionally active sites that allow the expression of HPV oncogenes after integration has occurred; this would give a selection advantage without disruption to gene structure or expression.

In a study of integration in the W12 cell line, Southern blot was used to predict which W12 clones had integration; clones with integration were subjected to micro-array technology to determine gene expression relative to integration (Alazawi et al. 2002). Eighty-five genes were observed to alter their expression following HPV integration into the host genome including anti-apoptotic genes and immunomodulatory genes, such as IFN-responsive genes. This study suggests that integration may alter expression of many genes, not just the

expression of genes at the site of integration. Interference with anti-apoptotic gene expression would increase chances of cell survival and lead to selection of the cell.

1.4.5 Methods to detect integration

There are many methods that can be employed to detect HPV integration (Table 7). The earliest investigations of HPV integration in cervical cancer used fluorescence in-situ hybridisation (FISH) where chromosomes were probed with a fluorescently labelled HPV probe to determine the chromosomal location of HPV integration. FISH is sensitive but requires fresh cellular material: it is useful in cell lines but is unfeasible for biopsies and cervical smear samples. Southern blot is a well-established method used in molecular biology: DNA is digested, electrophoresed and then probed with radiolabelled HPV. Southern blot method is very useful at distinguishing episomal HPV from integrated HPV and has been used widely in studies of the W12 cell line (Pett et al. 2004; Pett et al. 2006; Pett and Coleman 2007; Dall et al. 2008). Southern blot requires large quantities of DNA and is impractical for use in biopsy samples and cervical smear samples. In-situ hybridisation (ISH) is a method widely used in pathology especially in studies of head and neck cancer (Begum et al. 2005): sections of a biopsy are probed with HPV; when HPV is integrated the HPV probe forms punctuate regions and when HPV is in an episomal form, the probes appear diffuse. ISH requires no DNA extraction or nucleic acid preparation and it is quick to perform however status determination is purely subjective and does not indicate which chromosome contains the integrant.

PCR based methods such as E2 PCR and real-time qPCR aim to detect integrated HPV or episomal HPV. Real-time PCR compares the DNA quantification ratios of E2 to E6 (Kulmala et al. 2006) or E2 to E7 (Cheung et al. 2006). Samples with only episomal HPV will have equal quantities of E2 and E6 or E7 and have a ratio of 1; hypothetically, integrated HPV will only have E6 or E7 ORF and no E2 thus the ratio is 0. E2 PCR, PCR of overlapping segments across the E2 region, (Collins et al. 2009) is a quick and simple method that uses minimal DNA and is useful in samples where DNA availability is minimal such as cervical smear and biopsy samples; a failed PCR indicates disruption of E2. In low grade cervical disease, HPV may be integrated with a background of episomal HPV; episomal E2 would be amplified in

PCR and mask the detection of disrupted E2; therefore the presence of episomes is a crucial limitation to E2 PCRs and real-time PCR.

Methods such as Amplification of Papillomavirus Oncogene Transcripts (APOT) (Klaes et al. 1999), Restriction Site PCR (RS-PCR) (Sarkar et al. 1993; Thorland et al. 2000), and Detection of Integrated Papillomavirus Sequences (DIPS) (Luft et al. 2001), detect integration at the nucleotide level. RS-PCR and DIPS detect integration in the DNA and have been used widely in the detection of integration in DNA from cancer biopsies (Thorland et al. 2000; Luft et al. 2001; Ferber et al. 2003a; Ferber et al. 2003b; Thorland et al. 2003). APOT detects mRNA derived from integrated HPV; this method has been used on mRNA from biopsy and smear samples with varying disease grades (Table 6).

Method	Description	Strengths	Limitations
RS-PCR	Host/viral genomic regions are amplified by PCR using HPV specific primers and a primer designed to bind to restriction enzyme sites.	Can obtain DNA sequence of host/viral junctions.	Large concentrations of DNA required and labour intensive.
APOT	RT-PCR followed by PCR using HPV specific primer discriminates HPV mRNAs derived from integrated and episomal viral genomes.	Can obtain transcripts containing integration..	Labour intensive and expensive. Detects integration in RNA and cannot determine the nucleotide sequence or pin-point exact location of integration in DNA.
DIPS	Single-side-specific ligation-mediated PCR. Involves vectorette PCR and suppression PCR to detect integrated HPV DNA.	Can obtain DNA sequence of host/viral junctions.	Labour intensive.
Southern blot	Cellular DNA digestion and electrophoresis followed by hybridisation of labelled HPV DNA probes to determine the physical state (integrated or episomal) of HPV.	Can reliably distinguish episomal from integrated HPV DNA.	Uses large concentrations of DNA and labour intensive. The use of radio-labelled probes has health and safety implications.
E2 PCR	PCR of overlapping sections across E2 ORF can detect E2 disruption and indicate integration.	Uses small quantities of DNA and is quick to perform.	Background episomal DNA can mask the detection of disrupted E2.
Real-time qPCR	Physical state of HPV is estimated by calculating HPV E2:E6/E7 ratio by real-time PCR amplification of HPV E2 and E6/E7.	Uses small concentrations of DNA and is less labour intensive.	HPV E2:E6/E7 ratio may not reliably distinguish integrated DNA in a background of episomal DNA. Consumables expensive.
FISH	Chromosomes are fluorescently probed for HPV to determine chromosomal location of HPV integration.	Can sensitively detect HPV in chromosomes.	Requires fresh cellular material. Cannot determine the nucleotide sequence or pin-point exact location of integration.
ISH	Biopsy sections probed with HPV probe to determine integration status.	Detects integration status *in-situ* and less labour intensive than other methods.	Interpretation of integration status is subjective and less reliable.

Table 7 A summary of the most common DNA and RNA based HR-HPV integration detection methods. Adapted from Raybould et al. (2011). RS-PCR= Restriction Site PCR; APOT=Amplification of Oncogene Transcripts; DIPS=Detection of Integrated Papillomavirus; FISH= Fluorescent *in-situ* hybridisation, ISH=*in-situ* hybridisation.

2 Study Outline

HPV integration is detected in over 80% of cervical cancers and disruption of HPV genome through integration leads to high levels of HPV oncogene expression. To date, there are no published studies using sensitive, DNA based, techniques that detect integration at the nucleotide level in cervical precancers. Studies using RNA based assays, that detect transcribed integration events, have reported a higher frequency of integration in cancer neoplasia compared to CIN1 to CIN3; but studies using PCR and qPCR on DNA have reported integration as an early event in low grade neoplasia. *In-vitro* studies of cervical disease progression, using the W12 cervical cell line, have reported integration as a late event in the progression of cervical cancer and emphasised that episomal loss is key to malignant progression. This study aimed to test the hypothesis that integration is an early event in cervical neoplasia and episomal loss causes malignant transformation through transcription of integrated HPV. The mechanism of integration was explored to further understanding of why integration occurs: hot-spots for integration such as DNA fragile sites and similarities in HPV disruption were examined. This study also served as a pilot for larger future studies to determine whether integration is a plausible biomarker to be used as a positive predictor of cervical disease and explore the potential role of HPV integration in cervical screening.

The PhD was divided into three stages; the hypotheses are shown in Table 8:

Hypothesis	
1	Integration disrupts E1 or E2 ORFs.
2	Integration gives a selective growth advantage.
3	Integration results in loss of E2 expression and increased E6 and E7 expression and gives a selective growth advantage.
4	HPV integration is an early event present in CIN1.
5	Loss of episomes is a late event associated with CIN3, VIN3 or VAIN3.
6	Integrated transcripts are present in cancer and are linked with episomal loss.
7	Integration occurs within human genes.
8	Integration occurs at fragile sites.
9	DNA sequence similarities between human and HPV DNA at host viral junction indicate non homologous end joining in HPV integration.

Table 8 Hypotheses for the PhD.

2.1 Stage 1: Assay validation

In total, 4 different assays to detect integration were developed: DIPS, RS-PCR, E2 PCRs and APOT. DIPS, RS-PCR, E2 PCR and APOT have previously been developed for HPV16 and HPV18 (Klaes et al. 1999; Thorland et al. 2000; Luft et al. 2001; Ferber et al. 2003a; Collins et al. 2009) whereas development of HPV45 DIPS, RS-PCR and E2 PCR reported here is novel and not yet reported in literature. There are several methods that can be utilised to detect integration but a number of these were inappropriate for use in this PhD. Firstly, Southern blot utilises large quantities of DNA and is suitable for use on cell lines where vast amount of DNA are available, whereas LBC, smear and biopsy samples would have insufficient yields of DNA. qPCR has been used to detect integration according to ratios of E2 to E6 or E7; this method had already been trialled by the HPV research group at Cardiff University and although small quantities of DNA were required, it was difficult to detect integration where samples contain both episomal and integrated HPV. RS-PCR, DIPS and APOT have been broadly exploited in studies of HPV integration; these methods detect integration at the nucleotide level and were novel to the HPV research group, thus these assays were chosen for investigation in this PhD. E2 PCR was chosen to detect intact E2 as a marker of episomal HPV; this method required minimal DNA and was novel to the Cardiff University HPV research group.

Assay validation consisted of 2 stages:

The first stage of validation determined the ability of E2 PCRs to detect disrupted E2 and RS-PCR, DIPS and APOT to detect integration in cell lines with defined and published integration sites. Before assays were applied to clinical material it was essential to determine the strengths and limitations of each assay by comparing them in terms of integration detection and practicality. In low grade disease, HPV exists in episomal form and integration detection may be masked by the presence of episomal DNA. The majority of studies to date have not addressed the issue of integration in the presence of episomal HPV. Cervical cancer characteristically has only integrated HPV with no episomes and previous studies have performed DIPS and RS-PCR on samples that have only integrated DNA. There are no published reports of RS-PCR and DIPS in precancerous tumours and this study aims to determine whether integration can be detected within a background of episomal HPV. HPV copy numbers vary in clinical samples and the lowest number of HPV copies each assay could detect was determined. Finally, the assays used are technically complex therefore the reproducibility of each assay was determined by exploring the success of achieving the same outcome each time the assay was performed.

The second stage of assay validation compared the detection of integration by HPV16 DIPS, APOT and E2-PCRs on cell lines with unknown integration events in novel *in-vitro* cell lines of HPV16-associated vulval and vaginal neoplasia. Integration was observed in relation to quantities of HPV E2, E6 and E7 mRNA, to determine if integration and increased HPV16 E6 and E7 expression were essential for cell growth and doubling time.

Finally, it was important to determine that integration sites detected by DIPS or RS-PCR were true integration sites and not an artefact of DIPS or RS-PCR. Primers that flanked each integration site were designed and PCR was used to confirm the integration site.

2.2 Application of assays to clinical samples:

Stage 2 applied integration assays to cervical smears and biopsies from women with varying disease grades with an aim to determine how integration, transcription of integrants and

episomal loss contribute to cervical neoplasia. Integration in relation to episome loss was investigated by application of E2 PCR and DIPS to smear samples with varying cytological grades: normal, mild, borderline, moderate dyskaryosis, severe dyskaryosis and cervical cancer biopsies. The data from E2 PCR and DIPS on smear samples were correlated with histological outcome of biopsies taken at colposcopy. APOT was applied to cervical cancer biopsies and smear samples with varying cytological and histological grades to determine if integrated transcripts are present in CIN3 and cancer where episomes are lost. HPV16 qPCR was applied to cervical cancer biopsies and smear samples to quantify HPV E2, E6 and E7 expression in relation to integration and episomal loss. Additionally, primers that flanked each integration site were designed and PCR was used to confirm the integration site.

2.3 Mechanism of integration:

Stage 3 utilised bioinformatic analysis of sequence data to look at hot-spots for integration such as DNA fragile sites and similarities in HPV disruption. Integration sequence data was analysed to examine the site of integration for similarities between human and HPV DNA sequences.

3 Materials and Methods

This section describes materials and methods employed in the PhD including the statistical and quality control measures utilised. All work was performed by the author unless otherwise stated.

3.1 Samples

There were 4 different sample categories used in this PhD: cell lines, plasmid encoding HPV, smear samples and biopsy samples. This section will describe sample collection, sample storage and nucleic acid preparation for each of the sample categories.

3.1.1 Cell lines

Cell lines, with known integration sites were used as a source of DNA and RNA with which to develop the integration assays. Cell line DNA and RNA was used as positive controls in all assays employed. This section describes the cell lines used, culture, harvesting methods and nucleic acid preparation. Cell lines were selected according to their HPV type and commercial availability and were sourced from a number of different locations:

3.1.1.1 SiHa

SiHa is a cell line derived from a grade 2 uterine cancer lesion from a Japanese woman aged 55 years (Friedl et al. 1970). SiHa possess 2 copies of HPV16 integrated at 13q21 with disruption to HPV16 at 3132bp and 3384bp. SiHa cells were obtained from the American Type Culture Collection (ATCC-LGC), Teddington, Middlesex, UK.

3.1.1.2 W12

W12 is derived from a cervical lesion with warty atypia, displaying CIN1, from a 22 year old woman (Stanley et al. 1989). At early passage, W12 possesses HPV16 in episomal form in approximately 100 copies per cell and continued culture of W12, induces clearance of episomes with the emergence of cells possessing integrated HPV (Pett et al. 2004). W12 passage 12 (W12p12) and W12 passage 32 (W12p32) were cultured at the Department of

Pathology, University of Cambridge. DNA and RNA from W12p12 and W12p32 were sent to us by Mark Pett at the Department of Pathology, University of Cambridge.

3.1.1.3 CaSki

CaSki is derived from cervical carcinoma, metastatic to the small bowel of a 40 year old Caucasian woman (Pattillo et al. 1977). CaSki contains 600 copies of HPV integrated in concatenated form at 11 chromosomal sites including chromosomes 2, 3, 6, 7, 11, 12, 14, 20, and 21 (Yee et al. 1985; Mincheva et al. 1987). CaSki was obtained from the American Type Culture Collection (ATCC-LGC), Teddington, Middlesex, UK.

3.1.1.4 PC0

The PC0 cell lines were derived from grade 3 vulval (PC08) and vaginal (PC09) intraepithelial neoplasia lesions in a 46 and 31 year old woman, respectively. Women were recruited by consultant gynaecologists Professor Alison Fiander and Dr Amanda Tristram at the specialist VIN Clinic at Llandough Hospital in 2008. PC0 cell lines were established as a pilot study and ethical approval was obtained from South East Wales Research Ethics Committee (REC Reference number 08/WSE02/32). The pilot project was risk reviewed by the Joint Trust/University Peer/Risk Review Committee and approved by the Cardiff and Vale NHS Trust Research and Development Office. Heterogeneous PC08 and PC09 were established by Ned Powell, a Senior Lecturer, within the HPV research group at Cardiff University. Heterogeneous lines were cultured up to passage 21 for PC09, and 19 for PC08 and Southern blot performed on the DNA. Southern blot revealed entirely integrated HPV at passage 19 for PC08 and a mixed population of integrated concatenated HPV and episomal HPV for passage 21 PC09.

3.1.1.5 HeLa

HeLa is a cell line derived from adenocarcinoma of a 31 year old African American Woman (Gey et al. 1952; Jones et al. 1971). HeLa contains 3 sub genomic fragments of HPV18 on 8q24 (Mincheva et al. 1987). HeLa was obtained from The Health Protection Agency (HPC), Porton Down, Salisbury, UK.

3.1.1.6 C4I

C4I is established from carcinoma of the cervix from a 41 year old Caucasian woman (Auersperg and Hawryluk 1962) and contains HPV18 integrated on 8q24 with disruption to HPV18 at 2952bp and 5442bp (Cannizzaro et al. 1988; Luft et al. 2001). C4I was obtained from The Health Protection Agency (HPC), Porton Down, Salisbury, UK.

3.1.1.7 HTB-34™

HTB-34™, also known as MS751, is a cell line derived from cervical carcinoma of a 47 year old woman (Sykes et al. 1970). Initially, HTB-34™ was reported to have HPV18 DNA present (Pater and Pater 1985; Yee et al. 1985) but a subsequent study confirmed HTB-34™ solely positive for HPV45 (Geisbill et al. 1997). Integration sites are unconfirmed in this cell line but HPV45 disruption is reported at 1888bp and 6963bp (Geisbill et al. 1997). HTB-34™ was obtained from American Type Culture Collection (ATCC-LGC), Teddington, Middlesex, UK.

3.1.1.8 C33A

C33A is an HPV negative cervical carcinoma cell line from a 66 year old woman (Auersperg and Hawryluk 1962). C33A was obtained from American Type Culture Collection (ATCC-LGC), Teddington, Middlesex, UK.

3.1.2 Cell Culture

Heterogeneous PC08 and PC09 lines were cultured previously by Ned Powell, Senior Lecturer, in the HPV Research Group, Cardiff University. Single cell cloning of PC08 and PC09 was undertaken by Tiffany Onions, PhD student, HPV Research Group, Cardiff University.

All cell lines were grown as a monolayer in tissue flasks and handled in sterile conditions in a category II laminar flow cabinet. The growth media and supplements required to culture each

cell line are summarised in Table 9. All cell lines were incubated in an incubator at 37°C supplied with 5% CO_2. J2 3T3 mouse fibroblasts were used to supplement PC08 and PC09 as feeder cells in culture and although they are listed in Table 9, they were not used for any of the integration studies. Growth media was prepared in sterile conditions and for PC08 and PC09, all growth media was filter sterilised.

3.1.2.1 Sub-Culturing/Passage of cell lines

Cell lines were initially seeded into small flasks (25cm^2) and grown until the cells were 80% confluent. The cells were sub-cultured according to ratios in Table 9 and grown until at least 2 large flasks (125cm^2) were 80% confluent.

Cell lines were sub-cultured by removal of growth media followed by washing twice with sterile Phosphate Buffered Saline (PBS) at pH7.4. The cells were then detached from the flask by incubation with 0.25% trypsin, 0.03% EDTA solution (Sigma Aldrich®, New Rd, Gilllingham, Dorset, UK) at 37ºC for no longer than 10 minutes. The volume of trypsin depended on the flask size: 1ml of trypsin-EDTA was added to 25cm^2 flasks and 5mls of trypsin-EDTA added to 125cm^2 flasks. Trypsinised cells were then seeded into flasks according to ratios in Table 9 and supplemented with growth media.

3.1.2.2 Harvesting cell cultures for Nucleic Acid Preparation

Production of RNA is optimal when cells are 80% confluent and diminishes thereafter. Therefore, all cell lines, except J2 3T3s, were harvested for DNA and RNA when 80% confluent to ensure optimal RNA retrieval.

Cell lines were harvested by removal of growth media followed by washing twice with sterile Phosphate Buffered Saline (PBS) at pH7.4. Cells were then detached from the flask by incubation with 0.25% trypsin, 0.03% EDTA solution (Sigma Aldrich®, New Rd, Gilllingham, Dorset, UK) at 37ºC for no longer than 10 minutes. Five millilitres of growth media was then added to cells to inactivate the trypsin enzyme and this was followed by

centrifugation at 1000rpm for 5 minutes. Supernatant was removed, 5mls of sterile PBS were added and the cells were counted using a haemocytometer prior to nucleic acid preparation.

Cell line	Media[1]	Ratio[2]	Serum[3]	Antibiotics[4]	Supplements
PC08 and PC09	GEM	1:3 to 1:8	10% FBS	PS	2mM LGlutamine, 0.1%ug/mL hydrocortisone, 10^{-10} mol/l cholera toxin and $5x10^5/25cm^2$ lethally irradiated (60 Gray dose) J2 3T3 fibroblasts. Media was supplemented with 10ng/mLepidermal growth factor 48 hours after plating.
J2 3T3	DMEM	1:3 to 1:8	10% NCS	PS	none
SiHa	RPMI	1:3 to 1:8	10% NCS	PS	none
CaSki	RPMI	1:4 to 1:10	10% NCS	PS	2mM LGlutamine.
C4I	Waymouth's MB 752/1	1:4 to 1:10	10% NCS	PS	2mM LGlutamine.
HeLa	EMEM	1:3 to 1:10	10% NCS	PS	2mM LGlutamine.
HTB-34	EMEM	1:2 to 1:5	10% NCS	PS	none
C33A	EMEM	1:3 to 1:8	10% NCS	PS	none

Table 9 Propagation and sub-culturing for cell lines. All media and supplements obtained from Sigma Aldrich®, Dorset, UK.

[1] Culture media. GMEM= Glasgow's Minimum Essential Medium; DMEM=Dulbecco's Modified Eagle Medium; EMEM= Eagles Minimum Essential Medium. Ten millilitres of media was added to $25cm^2$ flasks and 30mL of media was added to $125cm^2$ flasks.
[2] Sub-culturing ratio.
[3] Serum. NCS= Newborn Calf Serum; FBS=Fetal Bovine Serum
[4] PS=Penicillin–streptomycin at 50,000U and 50mg per 500ml, respectively.

3.1.3 Plasmids

Plasmids are extra-chromosomal DNA molecules found in bacteria. The plasmids are replicated upon bacterial growth thus making plasmids ideal for cloning DNA. To clone DNA, DNA of interest is inserted into a plasmid vector possessing genetic markers that allow a selective growth advantage under certain conditions. The vector is then transformed into bacteria such as *E. Coli* and selective pressure, for example using antibiotics or blue/white selection, will allow *E. Coli* containing the plasmid with DNA of interest to grow. The DNA of interest can be cloned by further culture of a single colony.

Plasmid vector encoding HPV was used in the development of E2 PCR and as positive controls in E2 PCR. Plasmid vector encoding HPV were supplied from a number of sources:

3.1.3.1 HPV16 Plasmid

Complete intact HPV16 DNA, in vector was obtained from the HPV research group at Manchester University. Genomic DNA of HPV16 was originally obtained by the Manchester HPV research group from the World Health Organisation, inserted into vector pBR322 (Sutcliffe 1979) at BamHI restriction site (357bp) (Figure 13), transformed into HB101 *E. Coli* HB101 cultured in Luria Bertani (LB) Broth, supplemented with 100µg/ml ampicillin. Two millilitres of HPV16 plasmid in vector cultured in LB Broth was supplied; the culture was stored at -80ºC.

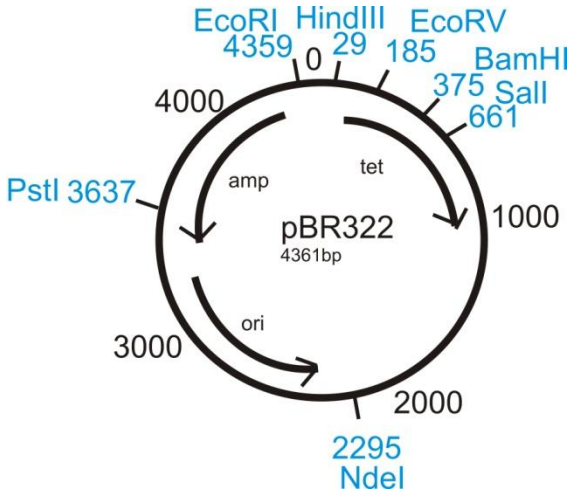


Figure 13 pBR322 vector adapted from Sutcliffe (1979).

3.1.3.2 HPV18 Plasmid

Complete intact HPV18 DNA, in vector was obtained from the HPV research group at Manchester University. Genomic DNA of HPV18 was originally obtained by the Manchester HPV research group from the World Health Organisation and inserted into vector pAT153 (Twigg and Sherratt 1980) at EcoRI site (3636bp) (Figure 14). The vector was transformed into HB101 *E. Coli* HB101, cultured LB Broth supplemented with 100µg/ml ampicillin. Two millilitres of HPV18 plasmid in vector cultured in LB Broth was supplied; the culture was stored at -80ºC.

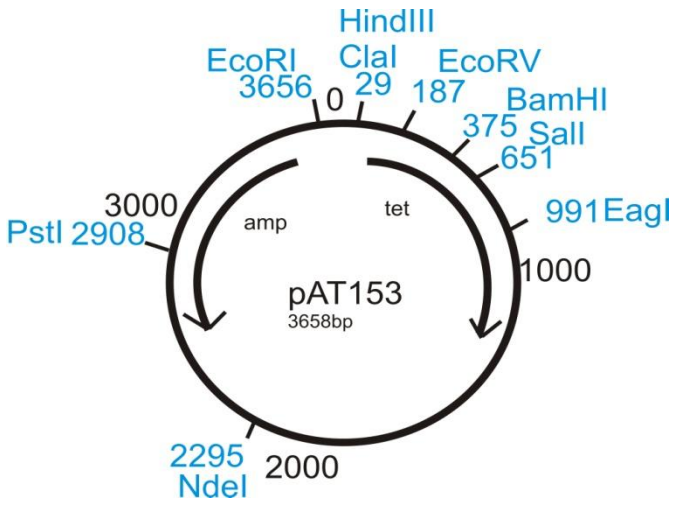


Figure 14 pAT Vector adapted from Twigg and Sherratt (1980).

3.1.3.3 HPV45 Plasmid Vector

Complete intact HPV45 DNA, inserted into pGEM®-4 vector (Promega Corporation, Southampton, UK) at EcoRI restriction site (10bp) (Figure 15), without host, was supplied by Professor de Villers at the German Cancer Research centre (DKFZ), Heidelberg.

On receipt HPV45 plasmid in pGEM®-4 vector was transformed into One Shot® Top10 chemically competent *E. Coli* (Invitrogen™, Life Technologies Ltd , Paisley, UK). Transformation was done following manufacturer's instructions by heat shock of *E. Coli* at 42ºC for 30 seconds. Transformed cells were then selected by blue/white selection. Blue/white selection is a method to identify *E. Coli* cells that possess the plasmid with vector containing inserted DNA. Certain vectors carry a sequence that codes for β-galactosidase that metabolise 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-Gal) and produce blue *E. Coli*. Expression of β-galactosidase is initiated from the Lac promoter. When DNA is inserted into a vector and disrupts the Lac promoter, the metabolism of X-Gal is diminished and *E-Coli* cells are white; thus white cells possess the cloned DNA of interest. Transformed *E. Coli* were plated onto LB agarose (Sigma Aldrich®, Dorset, UK) supplemented with 50µg/ml X-Gal (Sigma Aldrich®, Dorset, UK) and 100µg/ml Ampicillin (Sigma Aldrich®, Dorset, UK) and plates were incubated at 37ºC overnight. A single white colony was picked and grown in 5mLs of LB Broth (Sigma Aldrich®, Dorset, UK) supplemented with 100µg/ml ampicillin (Sigma Aldrich®, Dorset, UK) overnight at 37ºC in a shaking incubator at 5000g. One and a half millilitres of transformed *E. Coli* in LB broth were then mixed with 500µl of glycerol and stored at -80 ºC.

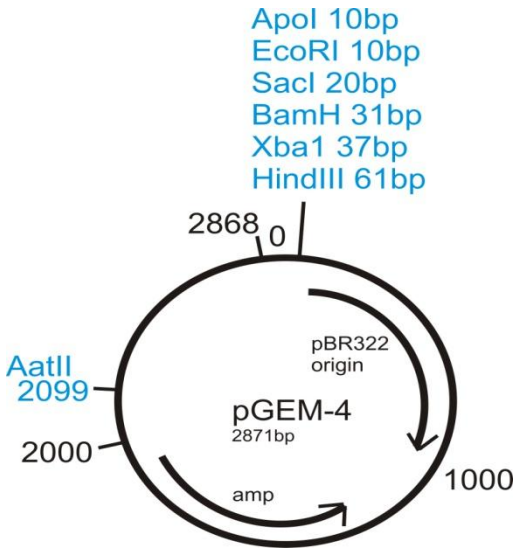


Figure 15 pGEM®-4 vector

3.1.3.4 Culture of HPV plasmid vectors

Plasmids encoding HPV16, HPV18 and HPV45 in *E. Coli* were cultured by incubating 20µl of the transformed *E. Coli*, in 5mL of LB broth, supplemented with 100µg/mL ampicillin, overnight at 37ºC in a shaking incubator at 150rpm. The *E. Coli* were centrifuged at 5000g for 5 minutes to pellet the *E. Coli* in preparation for plasmid DNA extraction, see section 3.2.2.

3.1.4 Baseline Samples

The Baseline samples were originally collected as part of a pseudo-anonymous study of HPV prevalence in 14,128 samples from women aged 20 to 22 years, who had not been vaccinated against HPV. Sample collection occurred between April 2009 and July 2010 in collaboration with Cervical Screening Wales. Women were eligible for the study if they resided in Wales and had attended their first smear between 20-22 years of age. Dyfed Powys Local Research Ethics Committee approved the study for all Wales recruitment.

3.1.4.1 Sample Collection

Samples were collected as liquid based cytology samples in BD SurePath™ preservative medium. Samples were processed by Cervical Screening Wales at participating cytology laboratories: Llandough, Singleton, Llandudno, Royal Gwent Hospital, Wrexham Maelor,

Ysbyty Glan Clwyd, Withybush, Royal Glamorgan, Princess of Wales, Prince Charles, West Wales General, and neighbouring English laboratories in Hereford, Chester and Shrewsbury. Samples were processed according to the British Society of Clinical Cytology guidelines and residual samples were transported to the HPV Research Laboratory, University Hospital of Wales, School of Medicine, Cardiff University, through the central transport system operating between hospitals in Wales.

3.1.4.2 Sample processing

Upon receipt at the HPV Research Laboratory in Cardiff, samples were processed by Research Scientists Jo Jones , Vasiliki Kiparoglou and Angharad Edwards. Residual sample was washed twice with 1ml of 10mM Tris (pH 7.4) by mixing, centrifugation and removal of supernatant. The cells were re-suspended in 2ml of 10mM Tris (pH 7.4), aliquotted and stored at -80ºC. DNA was extracted from samples by Jo Jones and Angharad Edwards, using proteinase K method, as described in section 3.2.3, and extraction efficiency was determined by Beta-globin PCR (section 3.2.8.1). Samples that were positive for Beta-globin PCR were subjected to HPV typing by GP5+/GP6+PCR-Enzyme Immunoassay method (EIA), described in 3.3.2.1.

3.1.4.3 Collection of Patient Information

All histological and cytological data used in this PhD were received and processed by Cervical Screening Wales and pseudo-anonomysed. All data that passed between Cervical Screening Wales and the HPV Research Laboratory in Cardiff was encrypted. The HPV Research Laboratory did not know the identity of the woman who gave the sample and Cervical Screening Wales did not know the HPV test result or the outcome of integration analysis. The Baseline study has ethical approval to collect follow up clinical screening information over two future screening rounds (up to 6 years).

3.1.4.4 Selection of Baseline samples for integration analysis

Only women that had abnormal cytological outcome and had attended colposcopy were selected for integration analysis because these women had histology data that was needed to investigate integration in cervical neoplasia. Women attended colposcopy more than once

during the period 2008-2012 thus the histology data linked with first attendance at colposcopy was used. Sample selection was prioritised by detection of a single HPV type and histological outcome. The study of Klaes et al. (1999) was used to estimate the number of Baseline samples required for this study: to achieve 80% power to detect a result of true significance, a total of 49 samples would have been needed, based on numbers of integrated transcripts versus numbers of episomal transcripts between high grade (CIN3 and cervical cancer) and low grade (no CIN, CIN1, CIN2) neoplasia (χ^2=26.03, df=1, φ=0.4, α=0.05). However, it was hypothesised that integration events in DNA may be more frequent in low grade CIN than those observed by Klaes et al. (1999), due to the reports of Huang et al.(2008) (Table 5). This would mean that more than 49 samples would be required to achieve 80% power. A total of 131 samples were selected (Figure 16) with approximately equal proportions of CIN grade diagnosed by histology. Women with no CIN were rare for HPV16 positivity and were not observed for HPV18 or HPV45 thus only 10 samples with no CIN were included.

An aliquot of Baseline cell suspension for each Baseline sample to be assayed for integration was retrieved from the freezer by Jo Jones or Angharad Edwards. The sample identification numbers were double checked by the author to eliminate the risk of assaying the incorrect sample.

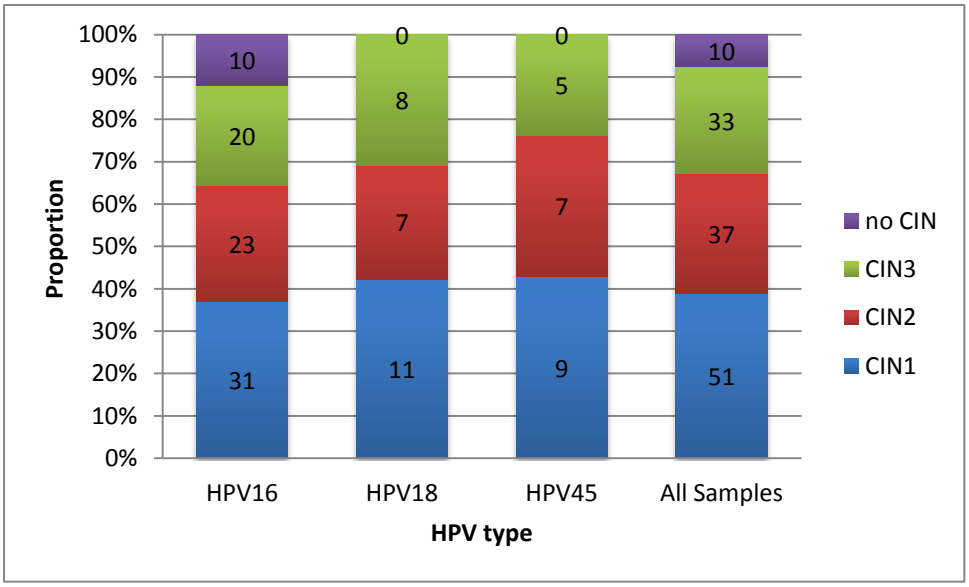


Figure 16 Proportion of CIN grades in samples selected for integration analysis grouped by HPV type.

3.1.5 HI-BCD Samples

The idea for collection of the HI-BCD samples was instigated by the author for the purpose of collecting both RNA and DNA from smear samples for integration assays. The "HPV Integration as a Biomarker of Cervical Disease" (HI-BCD) study was started in 2012 and is ongoing. The author contributed to study design and was a co-investigator in the project (see section 10.1, Appendix 1, for study protocol and consent form). Smear samples were collected from women attending colposcopy at Llandough Hospital, Cardiff, UK following detection of cellular dyskaryosis by smear. The intention was to collect a total of 45 samples from 15 women referred to colposcopy with mild, moderate and severe cellular dyskaryosis. Women referred to colposcopy due to abnormal cytology and undergoing treatment for high-grade CIN were included in this study. After confirmation of eligibility for the study, the study was discussed with the women and informed consent was given; women also gave consent to allow the results of cervical cytology and histology for one round of screening (up to 4 years) after entry to be made available for the study. Women that did not give informed consent were not eligible for the study. A total of 28 samples were collected during the period of this PhD. Ethical approval was given from Dyfed Powys Local Research Ethics Committee/ National Institute for Social Care and Health Research Permission Co-ordinating Unit and the HI-BCD study was approved by Cardiff and Vale R&D committee.

3.1.5.1 Sample Collection

Samples were collected by colposcopy nurse Sue Ashman and colposcopy gynaecologists Dr. Amanda Tristram, Dr. Sadie Jones and Dr. Jonathan Lippiatt at Llandough Hospital, Cardiff, UK. A single smear sample was collected by cytological brush in 2mL of RNAProtect® (QIAGEN Ltd, Manchester, UK) in a sterile 10ml Universal tube. The brush was agitated in RNAProtect® and the end of the brush was snapped off into the tube. Samples were transported by Sue Ashman, at room temperature, to the author at the HPV Research Laboratory, University Hospital of Wales, School of Medicine, Cardiff University. Pseudo-anonomysed sample identification numbers and cytological outcome were recorded electronically. All samples were stored at -80ºC until a batch of 10 was accumulated and then nucleic acids were purified from the samples (section 3.2.4).

3.1.5.2 Collection of Patient Information

All histological and cytological data included in this PhD were received by Cervical Screening Wales and pseudo-anonomysed. Cytological data and histological data were linked with the data from the assays performed in this PhD.

3.1.6 Biopsy Samples

Six biopsy samples were provided by Dr. Christopher Holmes, a senior lecturer in the Faculty of Medicine and Dentistry at Bristol University. Biopsy samples were collected from the colposcopy clinic at St Michael's Hospital, Southwell Street, Bristol between 1988 and 1993. The same colposcopist collected each biopsy but his/her name remains anonymous. Sample collection was approved by the Ethics Committee of the United Bristol Hospital Trust. Samples were collected as part of the screening program where cellular dyskaryosis, detected by cervical smear, indicated the possibility of cervical pathology.

All samples were snap frozen in liquid nitrogen within 1 hour of collection and have been stored in liquid nitrogen at Cardiff University according to the Human Tissue Act.

Two biopsies, 4T and 6W, were moderately differentiated adenocarcinoma from a 63 and 45 year old woman, respectively. Samples 1W, 2A, 3O and 5W were from squamous cell carcinoma of the cervix from a 23, 33, 46 and 30 year old woman, respectively. Samples 1W, 3O and 5W were poorly differentiated and 2A was moderately differentiated.

3.2 Nucleic Acid Preparation

Each sample category required different nucleic acid preparation techniques:

3.2.1 DNA and RNA Extraction from Cell Culture

Live cells were counted using a haemocytometer and DNA and RNA was extracted from between $6x10^{10}$ and $8 x10^{10}$ cells using AllPrep™ kit (QIAGEN Ltd, Manchester, UK) following manufacturer's instructions. AllPrep™ DNA/RNA Mini Kit allowed the

simultaneous purification of genomic DNA and total RNA from the same biological sample. Manufacturer's instructions were followed. In summary, the cultured cells were immediately lysed and homogenised in 750µl lysis buffer containing 10µl of 14.3M β-mercaptoethanol (Sigma Aldrich®, Dorset, UK) per 1ml of buffer, and 1µl of 2M dithiothreitol (DTT) (QIAGEN Ltd, Manchester, UK). β-mercaptoethanol was added to stabilise RNA and DTT was added to prevent excessive foaming of sample during homogenisation. Samples were homogenised by passing the sample through a sterile syringe and 20 gauge needle 5-10 times. The sample was then applied to an AllPrep™ DNA spin column and centrifuged at 12,000g for 1 minute to collect the "flow through" containing RNA. The "flow-through" that remained from passing the lysate through the DNA column was mixed with 750µl 70% ethanol and applied to an RNeasy® spin column, supplied with the AllPrep™ kit. The RNeasy® spin column was then washed by application of buffers to RNeasy® spin column and centrifuged at 12,000g for 1 minute to remove DNAses and RNAses. RNA was eluted from RNeasy® spin column by addition of 25µl sterile RNAse free water, incubation at room temperate for 1 minute, followed by centrifuged at 12,000g for 1 minute. Elution of RNA was repeated to maximise RNA yield. DNA was washed by application of buffers, that inactivate DNAses, to the AllPrep™ DNA spin column and centrifuged at 12,000g for 1 minute. DNA was washed again using an ethanol based buffer to remove salts, and centrifuged at 12,000g for 1 minute. DNA was eluted from the column by addition of 100µl elution buffer, incubation at room temperature for 1 minute followed by centrifugation at 12,000g for 1 minute. Elution of DNA was repeated to maximise DNA yield. DNA was stored at -20ºC and RNA was stored at -80ºC.

Extraction of RNA and DNA from PC08 and PC09 clones was done by Tiffany Onions and utilised the same method as described above.

3.2.2 Extraction of HPV plasmid vector DNA

Extraction of plasmid encoding HPV DNA from *E. Coli* was done using QIAprep® Miniprep kit (QIAGEN Ltd, Manchester, UK) designed to purify high-copy plasmid DNA from *E. Coli* in LB broth. The principle of the kit is based on alkaline lysis of bacterial cells followed by adsorption of DNA onto a silica membrane in a column. Manufacturer's instructions were followed. In summary: a bacterial lysate was produced by addition of a lysate buffer to the *E.*

Coli cell pellet. Two further buffers were added and mixed into to the lysate, by inversion of microfuge tube, to allow precipitation of genomic DNA and cell debris. The cell debris and E. Coli DNA were pelleted by centrifugation at 12,000g for 10 minutes and the supernatant, containing the plasmid DNA was applied to a QIAprep® column. Further application of buffers and centrifugation cleaned the plasmid DNA by removing endonucleases and co-purified contaminants such as salts. The plasmid DNA was then eluted from the column membrane by addition of 50µl elution buffer, incubation at room temperature for 1 minute followed by centrifugation at 12,000g for 1 minute. The DNA was stored at -20ºC.

3.2.2.1 Confirmation of Plasmid Vector DNA

To confirm the presence of HPV16, HPV18 and HPV45 in plasmid vectors, DNA was subjected to DNA sequence analysis, see section 3.3.3.6.4.

3.2.3 DNA extraction from Baseline samples

Initially, an experiment was performed to select the optimal method of extraction of DNA from the Baseline samples; QIAamp® (QIAGEN Ltd, Manchester, UK) and proteinase K methods were compared. AllPrep™, described above, was unsuitable for extraction of DNA from Baseline samples because viable RNA could not be extracted from Baseline samples.

QIAamp® is a DNA extraction method that can extract DNA from samples such as swabs and is based on DNA affinity binding to a purification column. Manufacturer's instructions were followed; in summary: 20µl QIAGEN Proteinase K was added to 100µl of the sample suspension and incubated at 56ºC for 30 minutes to digest proteins and lyse the sample. A buffer containing carrier RNA, to enhance binding of viral DNA was mixed with the lysate and incubated at 70ºC for 10 minutes; the lysate was then incubated with 300µl of 100% ethanol at room temperature for 5 minutes then bound to the membrane of a QIAamp® mini column by centrifugation at 6,000g for 3 minutes. The sample was washed by subsequent application of buffers and ethanol to the membrane and a number of centrifugation steps at 6,000g for 3 minutes. DNA was eluted by addition of 60µl of elution buffer to maximise the DNA concentration, incubation at room temperature for 1 minute, and centrifuged at 12,000g for 1 minute. DNA was stored at -20ºC.

Proteinase K allows crude extraction of DNA, it does not involve DNA washing or binding DNA to a membrane thus endonucleases and co-purified salts are present. The sample is lysed by addition of proteinase K followed by heat inactivation; centrifugation of the sample collects cellular debris as a pellet and the supernatant containing the DNA is removed from the pellet and stored. In summary, 10mg/ml proteinase K (Roche Diagnostics Ltd, West Sussex, UK) was added to 100µl of sample suspension and the sample was incubated for 2 hours at 56ºC in a shaking incubator at 300rpm; proteinase K was heat inactivated by incubation of the sample at 80ºC for 10 minutes. The sample was then chilled at 4 ºC for 10 minutes and centrifuged at 12,000g, at 4ºC for 10 minutes. The supernatant, approximately 60µl , containing crude DNA extract was removed by pipette into a sterile tube and stored at -20ºC.

Eight Baseline samples were extracted using QIAamp® and proteinase K and compared. DNA extracted using QIAamp® yielded less DNA compared to the Proteinase K method. High yield of DNA was important for the assays used in this PhD thus Proteinase K method was selected as the extraction method for the Baseline samples.

3.2.4 DNA and RNA extraction from HI-BCD samples

HI-BCD samples were collected in 2mL of RNAProtect® (QIAGEN Ltd, Manchester, UK) in a sterile 10ml Universal tube. Samples were transferred to a 2ml sterile, microfuge tube and centrifuged at 5,000g for 5 minutes to pellet cellular constituents. It was not possible to count the cells in RNAProtect® before extraction due to cell lysis in RNAProtect® but the numbers of cells were estimated to be $1x10^5$. AllPrep™ kit (QIAGEN Ltd, Manchester, UK), was used to extract DNA and RNA from HI-BCD samples following manufacturer's instructions described in section 3.2.1 with amendments: the HI-BCD cell pellet was immediately lysed and homogenised in 350µl RLT buffer containing 10µl of 14.3M β-mercaptoethanol per 1ml of buffer and 1µl of 2M DTT; RNA was eluted from RNeasy® spin column with 15µl sterile water, to increase concentration of RNA; DNA was eluted in 50 µl of elution buffer, to increase concentration of DNA. DNA was stored at -20ºC and RNA was stored at -80ºC.

3.2.5 DNA and RNA extraction from Biopsy Samples

Biopsies were supplied on dry ice to prevent RNA degradation. The biopsy weight was estimated according to the size of biopsy: 1.5mm cube weighs approximately 3.5 to 4.5mg according to AllPrep™ kit handbook. AllPrep™ kit (QIAGEN Ltd, Manchester, UK), was used to extract DNA and RNA from biopsies samples. AllPrep™ can extract DNA and RNA from a maximum of 5mg of biopsy thus samples that were estimated to be larger than 5mg were divided by a sterile scalpel and two extractions were done. AllPrep™ was done following manufacturer's instructions described in section 3.2.1 with amendments: biopsies were immediately homogenised in 350µl buffer containing 10µl of 14.3M β-mercaptoethanol per 1ml of buffer and 1µl of 2M DTT. Samples were homogenised using an electrical tissue ruptor until sample was completely homogeneous but for no longer than 1 minute to prevent DNA shredding. The homogenate was then centrifuged at 13,000rpm for 3 minutes and the supernatant applied to an AllPrep™ DNA spin column. RNA was eluted from RNeasy® spin column with 15µl sterile water, to increase concentration of RNA. DNA was eluted in 50µl of elution buffer, and elution performed twice to increase yield of DNA.

3.2.6 DNA and RNA Extraction Controls

A negative control of sterile water was used in each extraction. A positive control was extracted with each extraction done, with the exception of DNA and RNA extraction from cell lines. For Baseline extraction, a 100µl cell suspension of $4x10^3$cells/µl of CaSki, stored in PBS at -80ºC was used. For HI-BCD, a positive control of $4x10^5$ SiHa cells suspended in 2mL of RNAProtect®, stored at -80ºC were used. For biopsies DNA and RNA extraction, no biopsy positive control was available.

3.2.7 Nucleic Acid Quantification

RNA and DNA were quantified using a Thermo Scientific Nanodrop® 1000 Spectrophotomer (NanoDrop products, 3411 Silverside Rd, Bancroft Building, Wilmington, DE 19810, USA) following manufacturer's instructions. The Nanodrop® accurately measures DNA or RNA in a volume as little at 1µl without the need for cuvettes. The sample is applied to the end of a fibre-optic cable and xenon light is flashed onto the sample and a spectrophotomer measures the light after it has transmitted through the sample. The out-put gives sample concentration (ng/µl) and purity data: the 260/280 ratio determines the protein contamination level and the

260/230 ratio indicates the presence of co-purified contaminants such as salt. A pure DNA sample will have 260/280 ratio of ~1.8 and a 260/230 ratio of ~2. A pure RNA sample will have 260/280 ratio of ~2 and a 260/230 ratio of ~2. A pure DNA sample would expect to have an out-put image such as the one in Figure 17.

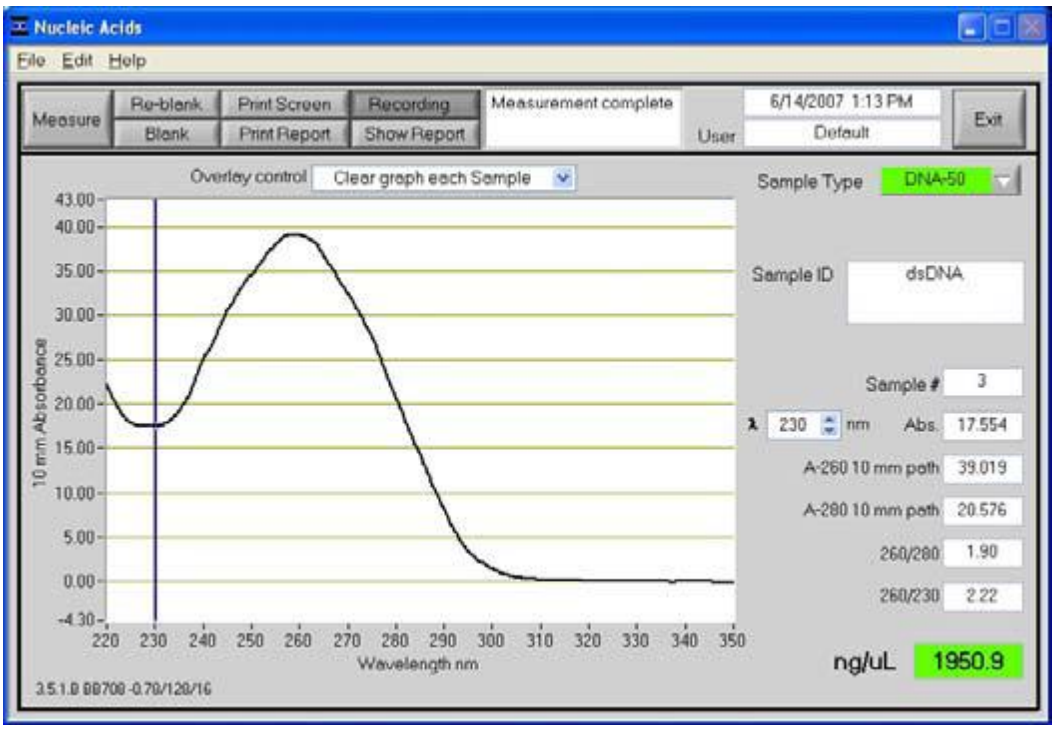


Figure 17 Nanodrop out-put screen.

3.2.8 DNA Integrity

There were two methods used to determine the integrity of DNA extracted from cell lines, smear samples and biopsies: Beta-globin PCR and electrophoresis of genomic DNA on an agarose gel.

3.2.8.1 Beta -globin PCR

Beta-globin is a conserved housekeeping gene encoding a peptide that forms the Beta chain of haemoglobin. Amplification of part of this gene, using PCR, can determine extraction efficiency and establish whether DNA extracted is suitable for PCR based methods. PCR is described in section 3.3.1. Five micro-litres of extracted DNA, diluted 1:10, were added to the PCR reagents summarised in Table 11. PCR conditions are shown in Table 10. Primers

were previously described by Huang et al. (1989) and obtained from Sigma Aldrich®, Dorset, UK. Amplicons were separated on a 2% (w/v) agarose gel. Samples that produced an amplicon of 110bp showed that DNA extraction was efficient and that the sample was suitable for PCR based methods.

3.2.8.2 Fine-pore Agarose Gel Electrophoresis

Gel electrophoresis was used to determine if DNA was degraded. Genomic DNA purified with the AllPrep™ procedure has an average length of 15–30kbp depending on homogenization conditions. If DNA was un-degraded then a single band of 15-30kbp would be observed when electrophoresing the DNA on an agarose gel. If DNA were degraded a smear would be observed on the gel due to fragments smaller than 15-30kbp being present. Large DNA fragments require electrophoresis on a low percentage agarose gel thus 5µl of extracted DNA was electrophoresed on 0.8% (w/v) AquaPor™ agarose gel (AGTC Bioproducts t/a National Diagnostics UK, Hessle, Yorkshire, UK).

3.2.9 RNA Integrity

Two microlitres of RNA was sent to Cardiff Biotechnology Services, at Cardiff University for analysis on an Agilent Bioanalyser (Agilent Technologies UK Ltd, Stockport, UK). An Agilent Bioanalyser is a microfluidics-based platform for sizing, quantification and quality control of DNA, RNA, proteins and cells. ribosomal ribonucleic acid rRNA is essential for protein synthesis in all organisms. Mammalian rRNA comprises of two subunits translated from 18S and 28S rRNA. RNA is applied to an Agilent Bioanalyser chip and RNA integrity is given as an RNA Integrity Number (RIN).The RIN is a software tool designed to estimate the integrity of total RNA, including 18S and 28S rRNA, in a sample. The software automatically assigns an integrity number to eukaryote total RNA sample. Using this tool, sample integrity is determined by the entire electrophoretic trace of the RNA sample not just the ratio of the ribosomal bands and includes the presence or absence of degradation products. An RIN of 10 indicates that RNA is un-degraded whereas an RIN of 1 indicates that RNA is degraded. The Agilent Bioanalyser also quantifies the total RNA. Figure 18 shows an example of Agilent Bioanalyser output: sample 1 has degraded RNA, with an RIN of 4.5; sample 3 has intact RNA and an RIN of 10.

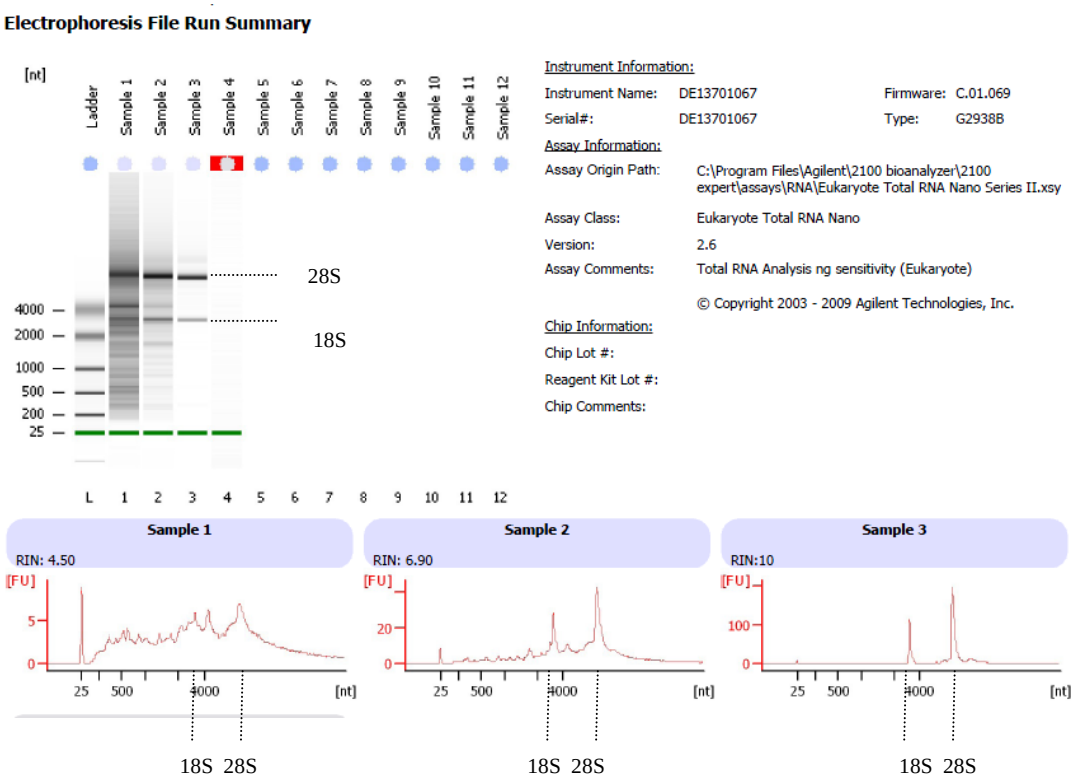


Figure 18 Agilent Bioanalyser Output File. Top Left hand size is electrophoretic separation of RNA and along the bottom are electropherograms of 18S and 28S rRNA. Sample 1 has degraded RNA, sample 3 has intact RNA and sample 2 has RNA that is partially degraded.

3.3 Molecular Methods

There were two assays used to type HPV and four main assays used to detect integration; each employed the polymerase chain reaction. The first method to detect integration is detection of disruption of HPV E2 ORF using PCR across the E2 region, the second method is Restriction Site PCR (RS-PCR), the third is Detection of Integrated Papillomavirus Sequences (DIPS) and the fourth is Amplification of Papillomavirus Oncogene Transcripts (APOT). All PCR reagents and PCR conditions for each assay are shown in Table 10 and Table 11. RS-PCR, DIPS and APOT use DNA sequence analysis; this will be described in section 3.3.3.6.

After the detection of an integration site it was essential to confirm that integration detection could be replicated; this was done by PCR using primers that flanked cellular and viral DNA sequences; see section 3.3.3.7.

Finally, qPCR was done to quantify HPV16 E2, E6 and E7 mRNA in HPV positive samples, where RNA was available, see section 3.3.3.8.

3.3.1 Polymerase Chain Reaction

The polymerase chain reaction (PCR) was developed in 1988 by Saiki et al. (Saiki et al. 1988). It is a fundamental technique used in most molecular applications. PCR allows the amplification of millions of copies of any region of a DNA sequence where at least some sequence information is known. PCR consists of 3 phases: denaturation, annealing and extension (Figure 19). Initially DNA is denatured and DNA strands are separated, by heating at 94-95ºC. Denaturation is followed by annealing of primers, otherwise known as oligonucleotides. Annealing temperature requires optimisation to ascertain optimal annealing temperature for successful PCR and annealing temperatures can be anywhere between 45-65ºC. A lower annealing temperature in PCR confers lower specificity for annealing the primers to target DNA and a higher annealing temperature in PCR confers greater specificity for annealing the primers to target DNA. Annealing temperature of primers, as a rule of thumb, are generally 2-3ºC lower than the melting temperature of the primers, calculated from the A+T and G+C content of the primers. Furthermore, the annealing temperature of the primers should not differ by more than 5ºC to ensure sufficient denaturation during each cycle of the PCR. Extension is the final step catalysed by Taq polymerase at 72ºC. Taq Polymerase is a thermostable enzyme that is active at 72ºC and it acts by extending the DNA fragment from the primer onwards by adding dioxynucleotide triphosphates (dNTPs: A, G, T and C) complementary to the DNA strand to form a double strand of DNA. Magnesium chloride is required as a cofactor for efficient Taq polymerase activity. The concentration of magnesium chloride can be altered in PCR during the optimisation process. Denaturation, annealing and extension are performed for a number of cycles and results in exponential amplification of the required DNA sequence. The PCR product, otherwise known as an amplicon, can be used for other applications such as sequencing, screening for mutations and genotyping. All PCR based methods were performed in a PCR hood, in sterile conditions to prevent PCR contamination. All PCRs were done on a thermocycler, with a heated lid to prevent evaporation during heating.

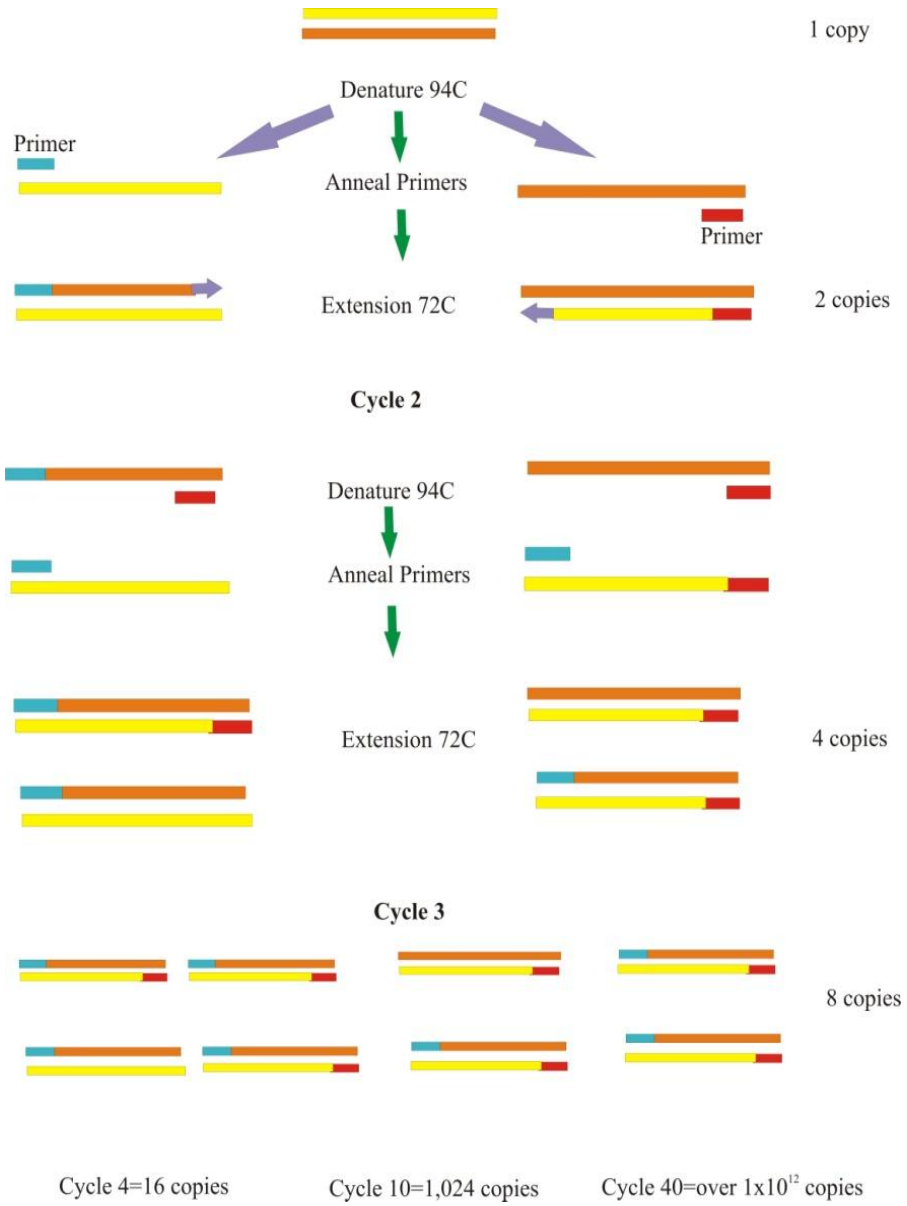


Figure 19 The Polymerase Chain Reaction. The Polymerase Chain Reaction allows the amplification of millions of copies of any region of a DNA sequence where at least some sequence information is known. DNA is denatured, primers are annealed and extension follows.

Assay	Temperature (ºC)	Time	Cycles
Beta-globin PCR	94	4min	
	94	30s	40
	55	30s	
	72	30s	
	72	4min	
GP5+/6+ PCR EIA	94	4min	
	94	30s	40
	40	90s	
	72	1min	
	72	4min	
PapilloCheck®	95	15min	
	95	30s	40
	55	25s	
	72	45s	
	95	30s	15
	72	55s	
E2 PCR	95	5min	
	95	30s	10
	65 -1 ºC per cycle	30s	
	72	1min	
	95	30s	30
	55	30s	
	72	1min	
	72	10min	
RS-PCR Primary	94	2min	
	94	30s	10
	45	30s	
	68	3min	
	95	30s	25
	45	30s	
	68	3min +10s per cycle	
	72	10min	
RS-PCR Nested	94	2min	
	94	30s	10
	55	30s	
	68	3min	
	95	30s	25
	55	30s	
	68	3min +10s per cycle	
	72	10min	

Table 10 PCR conditions for all assays used in this PhD.

Assay	Temperature (ºC)	Time	Cycles
DIPS Linear	95	15min	
	94	30s	40
	66	30s	
	72	3min	
	72	10min	
DIPS Nested	95	15min	
	94	30s	30
	66	30s	
	72	3min	
	72	10min	
APOT Primary	94	3min	
	94	30s	30
	66	3min	
	72	3min	
	72	10min	
APOT Nested	94	3min	
	94	30s	30
	67	30s	
	72	3min	
	72	10min	
Junction PCR	95	5min	
	95	30s	10
	(AT+10ºC) -1 per cycle[1]	30s	
	72	1min	
	95	30s	30
	AT[1]	30s	
	72	1min	
	72	10min	
qPCR	95	10min	
	95	10s	60
	AT[2]	10s	
	72	10s	

Table 10 continued. PCR conditions for all assays used in this PhD.

[1] AT= Annealing temperature; see Table 30.
[2] AT= Annealing temperature; see Table 32.

Assay	Reagent	Concentration	Volume (µl)
Beta-globin PCR	DNA	1:10 dilution	5
	Forward Primer	5µM	2.5
	Reverse Primer	5µM	2.5
	dNTP	2mM	2.5
	$MgCl_2$	50mM	0.875
	10X PCR buffer	1X	2.5
	Taq Polymerase	5U/µl	0.1
	Sterile Water	na	9.025
	Total Volume (µl)		25
GP5+/6+PCR EIA	DNA	1:10 dilution	20
	Forward Primer	5µM	10
	Reverse Primer 5'Biotinylated	5µM	10
	dNTP	2mM	10
	$MgCl_2$	50mM	7
	10X PCR buffer	10X	10
	Taq Polymerase	5U/ µl	0.4
	Sterile Water	na	32.6
	Total Volume (µl)		100
PapilloCheck®	DNA 1:10		5
	PapilloCheck® Master Mix		18.8
	Hot Star Taq® Polymerase	5U/µl	0.2
	Uracil-N-Glycosylase	1U/µl	1
	Total Volume (µl)		25
E2 PCR	DNA	20ng/ µl	5
	10X		2
	Forward Primer	10µM	2
	Reverse Primer	10µM	2
	$MgCl_2$	15mM	2
	dNTP	2mM	2
	Taq	5U/µl	0.2
	Sterile Water	na	4.8
	Total Volume (µl)		20

Table 11 Reagents used in all PCRs in this PhD. Taq polymerase PCR reagents and dNTPs obtained from Invitrogen™, Life Technologies Ltd, Paisley, UK. Hot Star Taq® Polymerase PCR reagents obtained from QIAGEN Ltd, Manchester, UK. Uracil-DNA Glycosylase obtained from Thermo Scientific* Fermentas (Northumberland, UK); PapilloCheck® reagents supplied in PapilloCheck®kit.

Assay	Reagent	Concentration	Volume (µl)
RS-PCR Primary	Primary HPV Primer	1µM	2
	RSO primer	10µM	2
	PCR Buffer	10X	2
	$MgCl_2$	15mM	2
	dNTP	2mM	2
	Taq Polymerase	5U/µl	0.1
	Sterile Water		4.9
	DNA	20ng/ µl	5
	Total Volume (µl)		20
RS-PCR Nested	Nested HPV Primer	1µM	2
	RSO primer	10µM	2
	PCR Buffer	10X	2
	$MgCl_2$	15mM	2
	dNTP	2mM	2
	Taq Polymerase	5U/µl	0.1
	Sterile Water		8.9
	Primary PCR amplicon		1
	Total Volume (µl)		20
DIPS Linear	Ligation product	30ng/µl	2
	PCR Buffer with 15mM MgCl2	10X	2.5
	dNTP	10mM	0.5
	HPV PCR 1 Primer	10µM	0.5
	Hot Star Taq® Polymerase	4U/µl	0.125
	Sterile Water		19.375
	Total Volume (µl)		25
DIPS Nested	Linear PCR amplicon	na	2
	PCR Buffer with 15mM MgCl2	10X	2.5
	dNTP	10mM	0.5
	HPV PCR 2 Primer	10µM	0.5
	Hot Star Taq® Polymerase	4U/µl	0.125
	Sterile Water		19.375
	Total Volume (µl)		25

Table 11 continued. Reagents used in all PCRs in this PhD. Taq polymerase PCR reagents and dNTPs obtained from Invitrogen™, Life Technologies Ltd, Paisley, UK. Hot Star Taq® Polymerase PCR reagents obtained from QIAGEN Ltd, Manchester, UK.

Assay	Reagent	Concentration	Volume (µl)
APOT Primary	cDNA		1
	10X buffer		2.5
	dNTP	10mM	0.5
	$MgCl_2$	50mM	0.75
	PCR1 Primer	10µM	0.5
	p3 reverse	10µM	0.5
	Taq Polymerase	5U/µl	0.2
	Sterile Water		19.05
	Total Volume (µl)		25
APOT Nested	Primary PCR amplicon		5
	10X buffer		5
	$MgCl_2$	25mM	5
	dNTP	25mM	1
	PCR2	10µM	5
	(dT)17-p3	10µM	5
	Taq Polymerase	5U/µl	0.2
	Sterile Water	na	23.8
	Total Volume (µl)		50
Junction PCR	DNA	10ng/ µl	5
	Forward Primer	10µM	2
	Reverse Primer	10µM	2
	PCR buffer	10X	2
	$MgCl_2$	15mM	See Table 30
	dNTP	2mM	2
	Taq Polymerase	5U/µl	0.2
	Sterile Water		To make volume 15
	Total Volume (µl)		20
qPCR	Template cDNA		2
	Forward Primer	5µM	2
	Reverse Primer	5µM	2
	FS Mix	10X	2
	$MgCl_2$ (25mM)	25mM	See Table 32
	Sterile Water		To make volume 18
	Total Volume (µl)		20

Table 11 continued. Reagents used in all PCRs in PhD. Taq polymerase PCR reagents and dNTPs obtained from Invitrogen™, Life Technologies Ltd, Paisley, UK. Hot Star Taq® Polymerase PCR reagents obtained from QIAGEN Ltd, Manchester, UK.

3.3.2 HPV typing

HPV typing was performed on the biopsy samples and HI-BCD samples; two methods were employed: GP5+/GP6+PCR-Enzyme Immunuassay (EIA) (Jacobs et al. 1997) and PapilloCheck® (Greiner Bio One Ltd, Gloucestershire, UK). Positive and negative controls from DNA extraction were included in each batch of samples typed.

3.3.2.1 GP5+/GP6+PCR- EIA

GP5+/GP6+PCR-EIA, as described by Jacobs et al. (1997), is a PCR based enzyme linked immunosorbent assay (ELISA) that utilises a solid phase enzyme immunoassay (EIA) to detect different types of HPV. In summary, GP5+/GP6+PCR-EIA involved PCR amplification of part of HPV L1 ORF using a forward primer and a biotinylated reverse primer. The amplicon was then bound to a solid support, hybridised to labelled HPV oligonucleotide probes and detected by immunohistochemistry. DNA samples were typed for high risk types HPV16, HPV18, HPV31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV56, HPV58, HPV59, HPV66, HPV68.

3.3.2.1.1 GP5+/GP6+ PCR

Twenty micro-litres of extracted DNA, diluted 1:10, were added to the GP5+/GP6+PCR-EIA PCR reagents summarised in Table 11 in a 100µl reaction; PCR conditions are shown in Table 10. Primers were described previously by Jacob et al. (1997).

3.3.2.1.2 GP5+/GP6+EIA ELISA

A separate ELISA reaction was performed for each HR HPV type using a digoxigenin (DIG)-labeled oligonucleotide probe specific for each HR HPV type (Jacobs et al. 1995) (Sigma Aldrich®, Dorset, UK). Two additional reactions were done using cocktail containing HPV16, HPV18, HPV31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV56, HPV58, HPV59, HPV66 and HPV68 digoxigenin (DIG)-labeled oligonucleotide probes at a final concentration of 10pMol.

Firstly, 5µl of amplicon were hybridised to a single well of a 96 well Streptawell plate (Roche Diagnostics Ltd, West Sussex, UK) by addition of 50µl of wash buffer (43.8g NaCl, 22.1g sodium citrate, and 25ml Tween® 20 in 1l of sterile distilled H_2O), diluted 1:5, and incubated at 37ºC for 1hr.

The plate was washed three times with 1:5 wash buffer. One hundred micro litres of 0.2M NaOH was added the well to denature the amplicon bound to the plate.

The plate was washed again three times with 1:5 wash buffer and 50µl (10pMol) of DIG labelled probe mixture was added. The plate was incubated at 37ºC for 1hr to hybridise the probes to the amplicon.

The plate was washed again three times with 1:5 wash buffer and 50µl of anti-DIG conjugate (Roche Diagnostics Ltd, West Sussex, UK) was added. The plate was incubated at 37ºC for 1hr and then washed 5 times with 1:5 wash buffer. One hundred micro-litres of substrate solution (Roche Diagnostics Ltd, West Sussex, UK) was added and the plate was incubated at 37ºC for 1hr.

Absorbance readings were made on a Bio-Rad 550 Microplate reader (Bio-Rad Laboratories Ltd, Hertfordshire, UK), with dual optical density (OD) settings 415nm and 630nm after 1, 2 and 24hrs of incubation at 37ºC. OD readings taken at 24hrs were taken into account. OD values that were at least three times that of the background value, as defined by the OD value of the negative sample, were considered positive for HPV.

3.3.2.2 PapilloCheck®

PapilloCheck® (Greiner Bio-one Ltd, Stonehouse, UK) identifies 18 types of HR HPV: HPV16, HPV18, HPV31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV53, HPV56, HPV58, HPV59, HPV66, HPV68, HPV70, HPV73 and HPV82 and low risk types: HPV6, HPV11, HPV40, HPV42, HPV43 and HPV44. PapilloCheck® is a DNA hybridisation

DNA array test. In summary, PCR, using reagents in kit provided, amplifies part of the E1 ORF, PCR amplicon is hybridised to a chip with an array of probes specific for the HPV types listed above; the chip is then scanned using a PapilloCheck® CheckScanner™(Greiner Bio-one Ltd, Stonehouse, UK). Data is analysed with specialist software supplied with the CheckScanner™. PapilloCheck® was performed following manufacturer's instructions:

3.3.2.2.1 PapilloCheck® PCR

DNA was diluted 1:10 and 5µl were added to the reagents listed in Table 11. PCR conditions PCR conditions are shown in Table 10.

3.3.2.2.2 PapilloCheck® Hybridisation, Washing and Chip Reading

PapilloCheck® Hybridisation buffer was added to PCR amplicon and then added to a compartment on the chip. The chip was incubated at room temperature in humid conditions then washed 3 times using the buffers provided. The chip was then scanned using a PapilloCheck® CheckScanner and data analysed with specialist software supplied with the CheckScanner™.

3.3.3 Integration Assays

3.3.3.1 Primer Design

Primers were designed by the author, unless otherwise stated. Primers were designed using a web-based program, Primer3 http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi (Rozen and Skaletsky 2000). Primer3 is a program that designs PCR primers according to melting temperatures and secondary structures that may be formed by the primers and inhibit the PCR reaction; amplicon sizes are also calculated. Accession Numbers NC_001526.1, X05015 and X74479 were used as HPV16, HPV18 and HPV45 reference sequences for primer design. The sequence of DNA containing the HPV region of interest was submitted to Primer3, parameters such as PCR product size in base pairs, primer annealing temperature and primer size were selected, regions of wanted DNA sequence were marked following web-site instructions; the remaining parameters were left as default. Suitable primers were then selected from a list suggested by the program.

The PCR could potentially be invalidated if non-specific sequences to HPV were amplified therefore, *all* primer sequences were subjected to alignment to human and "others" database sequences using NCBI megaBLAST (Basic Local Alignment Search Tool) at http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome to determine their specificity to HPV and human genomic sequences. Primers were obtained from Sigma-Aldrich®, Dorset, UK.

3.3.3.2 E2 and E6 PCR

As mentioned in the introduction, the E2 ORF of HPV is frequently disrupted upon integration. Collins et al. (2009) employed PCR to amplify overlapping fragments that spanned across the E2 region to detect E2 disruption caused by integration (Figure 20). To control for presence of HPV a set of primers that amplified the E6 ORF were used. Integration was detected if E6 primers produced an amplicon and one or more primer E2 ORF sets failed to produce an amplicon.

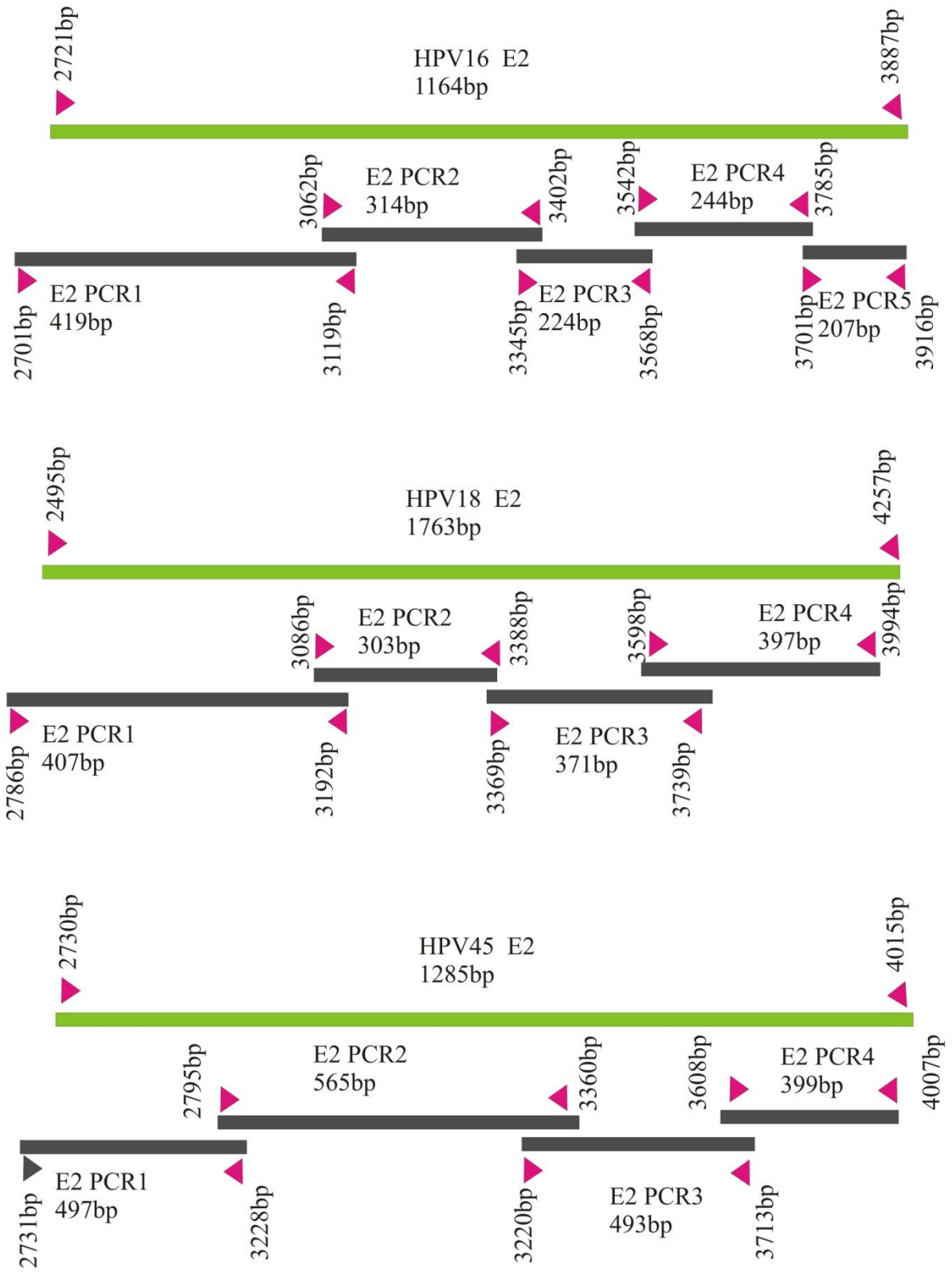


Figure 20 A schematic example of overlapping primers used for E2 PCR for HPV16, HPV18 and HPV45. Amplicon sizes are shown and primer annealing sites are shown as base pair number of viral sequence that is identical to the 5' base of the primer according to NC_001526.1, X05015 and X74479 for HPV16, HPV18 and HPV45, respectively. Green bar represents full length E2 amplicon; dark grey bars represent overlapping amplicons of E2 PCR.

3.3.3.2.1 Primers

Primers are listed in Table 12. Primers from Collins et al (2009) were used for HPV16 and HPV18. HPV45 E2 PCR primers were designed using method in section 3.3.3.1 with product size range selected from 100-300, 301-400, 401-500 and 501-600.

HPV16	**Forward Primer**	**Reverse Primer**	**Size (bp)**
E2	TTAAGTTTGCACGAGGACGA	CGCCAGTAATGTTGTGGATG	1167
PCR1	AGGACGTGGTCCAGATTAAG	TCAAACTGCACTTCCACTGT	419
PCR2	TAACTGCACCAACAGGATGT	GCCAAGTGCTGCCTAATAAT	341
PCR3	ATCTGTGTTTAGCAGCAACG	TAAATGCAGTGAGGATTGGA	224
PCR4	ACAGTGCTCCAATCCTCACT	TCACGTTGCCATTCACTATC	244
PCR5	GGCATTGGACAGGACATAAT	CAAAAGCACACAAAGCAAAG	207
E6 PCR	GAACAGCAATACAACAAACC	GATCTGCAACAAGACATACA	161
HPV18			
E2	TTAGATGATGCAACGACCAC	CGGTGGGATACCATACTTTT	1763
PCR1	TCCAGATTAGATTTGCACGA	CAATTGTCTTTGTTGCCATC	407
PCR2	ATACAAAACCGAGGATTGGA	ACTTCCCACGTACCTGTGTT	303
PCR3	AACACAGGTACGTGGGAAGT	TTTCGCAATCTGTACCGTAA	371
PCR4	GACCTGTCAACCCACTTCT	ACATGGCAGCACACATACAT	397
E6 PCR	TGTGTATGGAGACACATTGG	CTATAGTGCCCAGCTATGTTG	153
HPV45			
E2	TTGAAAGGACATGGTCCAGA	ACCAACAACCAAGCAAAAGC	1285
PCR1	TGAAAGGACATGGTCCAGATT	TGTCCCATATCCCTGTCTCAG	497
PCR2	CATTTCAAAAAGCAAAGCACA	TGCCCCCATATTGTACTTCC	565
PCR3	TATGGGACAAAACAGCAGCA	GTCTGCATATTTGCGTAGCC	493
PCR4	CAAAAGAAGGAAAGTGTGTAGTGG	CCAAGCAAAAGCACACACATA	399
E6 PCR	AAGCTTTGTGGAAAAGTGCAT	CTTGTGTTTCCCTACGTCTGC	498

Table 12 Primers for HPV16, HPV18 and HPV45 E2 PCR.

3.3.3.2.2 PCR Conditions

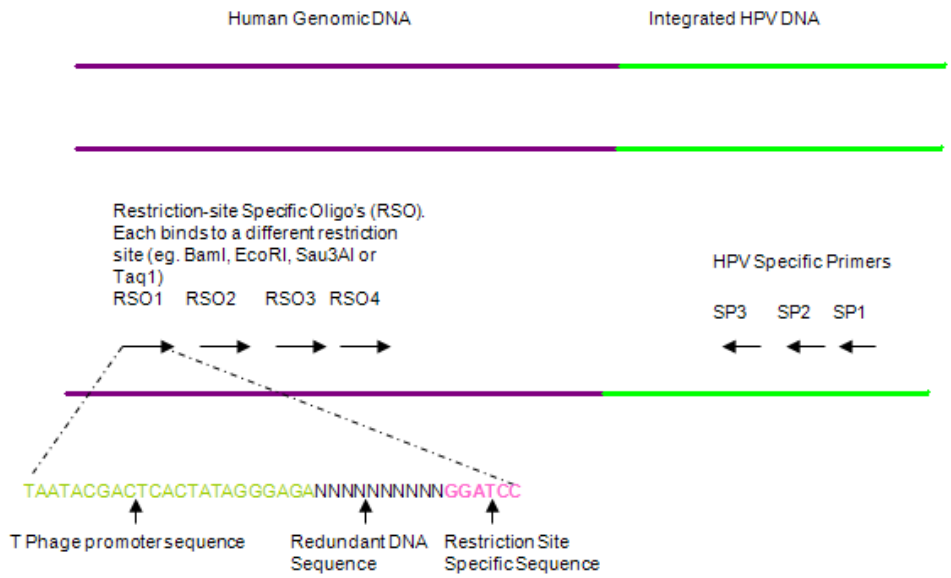
The reagents used for E2 and E6 PCRS are shown in Table 11. PCR conditions are shown in Table 10.

The amplicons for E2 and E6 PCRs were electrophoresed on a 2% (w/v) agarose gel stained with ethidium bromide and viewed under UV light. E2 was considered to be disrupted when E6 PCR produced an amplicon and no E2 PCR amplicons were detected, or when one or

more E2 PCR amplicons were not detected. E2 was considered to be intact when E2 PCR produced an amplicon for all PCR reactions. HPV was interpreted to be absent where E2 PCR and E6 amplicons were not detected.

3.3.3.3 Restriction Site PCR

RS-PCR (Sarkar et al. 1993) is a direct method that rapidly retrieves sequences of any DNA adjoining a known sequence; in this study the known sequence is HPV and the sequence of interest is the site of integration in the genome (Figure 21). RS-PCR includes a primary PCR reaction followed by a nested PCR reaction. The PCR reactions comprise a HPV specific primer and a primer designed to bind to DNA where the sequence of interest is unknown. Nested PCR products are separated by gel electrophoresis and selected fragments are subjected to DNA sequence analysis.



Step 1. First round of PCR using SP1 and all RSOs in separate PCR reactions.

Step 2. Second round of PCR using SP2 and all RSOs in separate PCR reactions.

Step 3. PCR products separated by gel electrophoresis in a 2% agarose gel.

Step 4. Samples with viable PCR products selected and subjected to DNA sequence analysis using SP3.

Figure 21 Steps involved in Restriction site PCR (RS-PCR). SP denotes HPV specific primers; RSO dentotes restriction site oligos. Purple bar represents human DNA and green represents HPV DNA.

3.3.3.3.1 Primers

RS-PCR uses primers (Restriction Site Oligonucleotides, RSOs) that recognize a given restriction enzyme recognition site, anneal to that site, and promote DNA synthesis in a PCR. Restriction recognition sequences are useful in RSOs because restriction sites are omnipresent across organisms, they have no repeat sequences in them and they are frequently repeated across an organism's genome (Sarkar et al. 1993). Amplification of PCR products greater than 1kbp in size can be difficult. To increase the chance of a restriction site being present within the range of PCR, RSOs were based on 4 to 6bp cutters because the chance of finding longer restriction sites within range PCR would be small. A restriction recognition site on its own would not be useful for PCR thus the RSO has an anchor sequence of 10bp of

redundant sequence attached to the 5′ end of the restriction recognition site. Attached to the 5′ end of the anchor sequence is a T7 phage promoter sequence to enable sequencing from the 5′ end of the sequence. Six RSOs were used (Ferber et al. 2003a) (Table 13). E6 and E7 ORFs are most likely to be retained after integration and so the HPV specific primers were designed to amplify regions of HPV genome, spanning out from the E6 and E7 genes at 750bp intervals. Primers sequences used for RS-PCR in HPV16 and HPV18 were described in published methods (Thorland et al. 2000; Ferber et al. 2003a) (Table 14 and Table 15). For HPV45, a novel primer set was designed using the method described in section 3.3.3.1 (Table 16).

Primer Name	**Primer Sequence**
RSO-*Bam*	TAATACGACTCACTATAGGGAGANNNNNNNNNN**GGATCC**
RSO-*Eco*	TAATACGACTCACTATAGGGAGANNNNNNNNNN**GAATTC**
RSO-*Nde*	TAATACGACTCACTATAGGGAGANNNNNNNNNN**CATATG**
RSO-*Sau*	TAATACGACTCACTATAGGGAGANNNNNNNNNN**GATC**
RSO-*Taq*	TAATACGACTCACTATAGGGAGANNNNNNNNNN**TCGA**
RSO-*Xba*	TAATACGACTCACTATAGGGAGANNNNNNNNNN**TCTAGA**

Table 13 Taken from Ferber, Montoya et al (2003). RSO primers used for RS-PCR.

Primer Name	Annotation	Primer Sequence
Primary RS-PCR	PCR1	
HPV16-768-24D	P1	ACAAAGCACACACGTAGACATTCG
HPV16-1545-26D	P2	AGTAATAAATCAACGTGTTGCGATTG
HPV16-2386-25D	P3	TTTG**G**TTACAACCATTAGCAGATGC
HPV16-2929-24D	P4	GTGCCAACACTGGCTGTATCAAAG
HPV16-5156-27U	P5	TACCAATTCTACTGTACCTAATGCCAG
HPV16-5883-27U	P6	ACTTATTGGGGTCAGGTAAATGTATTC
HPV16-6686-25U	P7	AGTAGATATGGCAGCACATAATGAC
HPV16-7591-22U	P8	GTTGGCAAGCAGTGCAGGTCAG
Nested RS-PCR	PCR2	
HPV16-790-25D	N1	CGTACTTTGGAAGACCTGTTAATGG
HPV16-1587-26D	N2	GGACTTACACCCAGTATAGCTGACAG
HPV16-2414-26D	N3	AATAGGTATGTTAGATGATGCTACAG
HPV16-2964-25D	N4	ACAAGCAATTGAACTGCAACTAACG
HPV16-5121-25U	N5	GAGGTTAATGCTGGCCTATGTAAAG
HPV16-5850-28U	N6	CCCTGTATTGTAATCCTGATACTTTAGG
HPV16-6651-25U	N7	TGCGTGTAGTATCAACAACAGTAAC
HPV16-7524-26U	N8	TTAAACCATAGTTGCTGACATAGAAC
Sequencing primers		
HPV16-839-25D	S1	TCTGTTCTCAGAAACCATAATCTAC
HPV16-1677-26D	S2	GGAATGGTTGTGTTACTATTAGTAAG
HPV16-2443-24D	S3	CCTGTTGGAACTACATAGATGACA
HPV16-2995-28D	S4	ACAATATATAACTCACAATATAGTAATG
HPV16-5069-27U	S5	GAGCTATATTAATACTATTATCATTAC
HPV16-5774-23U	S6	TCCAACTGCAAGTAGTCTGGATG
HPV16-6587-24U	S7	CTGTGCTCGTTGTAACCAATAAGG
HPV16-7467-24U	S8	CATGCAACCGAATTCGGTTGAAGC

Table 14 HPV16-specific primers used for RS-PCR, taken from Thorland, Myers et al. (2000). HPV16 denotes primers that are specific to Genbank accession number NC_001526.1 for HPV16. The number following HPV16 indicates the base pair number of viral sequence that is identical to the 5' base of the primer according to NC_001526.1. The number followed by D or U indicates the length of the primer. D indicates a sense primer, and U indicates an antisense primer. When compared to NCBI database sequences of HPV16 using BLAST, primer HPV16-2386-25D had one incorrect nucleotide. This was replaced with the correct nucleotide, highlighted in red. Primers were annotated to simplify microfuge tube labelling to avoid errors in RS-PCR.

Primer Name	Annotation	Primer Sequence
Primary		
HPV18-744D	P1	GACGAGCCGAACCACAACGTCAC
HPV18-1436D	P2	GGCAACAACAGCAGTGTAGACGG
HPV18-2266D	P3	ATTCCTGCGATACCAACAAATAGAG
HPV18-2996D	P4	CCAGGTGGTGCCAGCCTATAAC
HPV18-4145U	P5	CAATATAGCATGTATATGCAATAGTAAC
HPV18-5066U	P6	CTGAATCAGGAACATCACTACGAGG
HPV18-5749U	P7	CCAACAGTTAATAATCTAGAGCTGC
HPV18-6519U	P8	CAGAGGTAACAATAGAGCCACTTGG
HPV18-7129U	P9	CTGGCACGTACACGCACACGC
HPV18-7845U	P10	ATGTGCTGCCCAACCTATTTCGG
Nested		
HPV18-786D	N1	GTAAGTGTGAAGCCAGAATTGAGC
HPV18-1487D	N2	GTAAATCCACAATGTACCATAGCAC
HPV18-2412D	N3	TTATACAAGGAGCAGTAATATCATTTG
HPV18-3094D	N4	CCGAGGATTGGACACTGCAAGAC
HPV18-4029U	N5	CATACGCACACATACAGACAGATGG
HPV18-5035U	N6	TGTTAATGTAGTGTCCACAGGCTC
HPV18-5697U	N7	TCACATAATCATCGGTATTTAC
HPV18-6476U	N8	GCTGCCAGGTGAAGCAGGCATAC
HPV18-7093U	N9	GACGTAGTGGCAGATGGAGCAG
HPV18-7806U	N10	TGCACAGCTTAGTCATATTATAGTTC
Sequencing primer		
HPV18-826D	S1	AGCAGACGACCTTCGAGCATTCC
HPV18-1538D	S2	AATAAACAAGGAGCTATGTTAGCAG
HPV18-2470D	S3	GTTAACAGATACTAAGGTGGCCATG
HPV18-3129D	S4	CTATGGAATACAGAACCTACTCACTG
HPV18-3945U	S5	GTGATAACATATTGGTACTACAGCATA
HPV18-4956U	S6	TTAGCCACTGACACTTGTTGGTAG
HPV18-5655U	S7	GTGGAAGATATACGGTATTGTCAC
HPV18-6442U	S8	ATATATAAGGATTGAGGCACAGTGTC
HPV18-7047U	S9	GCTTGCGACGCAATCCAGCCT
HPV18-7742U	S10	TGTACAAGCCAAGTATGCAATTAGC

Table 15 HPV18-specific primers used for RS-PCR and sequencing, taken from Ferber et al. (2003a). HPV18 denotes primers that are specific to Genbank accession number X05015 for HPV18. The number following HPV18 indicates the base pair number of the viral sequence that is identical to the 5' base of the primer according to X05015. The number followed by D or U indicates the length of the primer. D indicates a sense primer, and U indicates an antisense primer. When compared to NCBI database sequences of HPV18 using BLAST, primer HPV18-1436D had 4 incorrect nucleotides. These were replaced with the correct nucleotides highlighted in red. Primers were annotated to simplify microfuge tube labelling to avoid errors in RS-PCR.

Primer Name	Annotation	Primer Sequence
Primary	PCR1	
HPV45_792_20D	P1	GTGACGGCAGAATTGAGCTT
HPV45_1537_20D	P2	TGGGCTGTCATTTACGGATT
HPV45_2334_23D	P3	CAGGAAAATCGTATTTTGGAATG
HPV45_3095_20D	P4	TACAGAACCGTCGCAGTGTT
HPV45_3847_21D	P5	AAATCTCGGTGGGATACATGA
HPV45_5552_25U	P6	CCATAAATAATATTGTGTGCCATGA
HPV45_6275_20U	P7	AATGGAACCTCGCACTTTGT
HPV45_7045_20U	P8	GACGTAACCCAGCCTGAACT
HPV45_7833_21U	P9	CAACCTTTTTCGGTTGCATAA
Nested	PCR2	
HPV45_815_20D	N1	GTAGAGAGCTCGGCAGAGGA
HPV45_1597_22D	N2	TTGGGTAATGGCTATATTTGGA
HPV45_2365_23D	N3	ACATTTCCTACAAGGTGCAATAA
HPV45_3146_20D	N4	CTTTGATGGCAACAAGGACA
HPV45_3939_21D	N5	TCTGTGTGCCTTTATGTGTGC
HPV45_5544_23U	N6	AATATTGTGTGCCATGAATACCT
HPV45_6266_20U	N7	TCGCACTTTGTATCCTGCAA
HPV45_7001_20U	N8	TGATCCAAATCGGAGGAAAA
HPV45_7764_23U	N9	CAGATTGTTGGATAAGAAAGTGG
Sequencing primer		
HPV45_861_20D	S1	TGAGCACCTTGTCCTTTGTG
HPV45_1622_20D	S2	AATCCAACGGTAGCAGAAGG
HPV45_2441_20D	S3	AAGGTAGCCATGTTGGATGA
HPV45_3209_21D	S4	TGAGACAGGGATATGGGACAA
HPV45_3969_20D	S5	CCGCTTGTGCAGTCTGTCTA
HPV45_5481_21U	S6	GCCACATAGGAGTATGGGATG
HPV45_6205_23U	S7	CACCATCCTCAATAATGGTGTTT
HPV45_6935_20U	S8	TCCTGCTTTTCTGGAGGTGT
HPV45_7724_24U	S9	ACAATTAGCACAGGTAAAAACAGA

Table 16 HPV45-specific primers used for RS-PCR and sequencing. HPV45 denotes primers that are specific to Genbank accession number X74479 for HPV45. The number following HPV45 indicates the base pair number of the viral nucleotide that is identical to the 5' base of the primer according to X74479. The number followed by D or U indicates the length of the primer. D indicates a sense primer, and U indicates an antisense primer. Primers were annotated to simplify microfuge tube labelling to avoid errors in RS-PCR

3.3.3.3.2 Primary PCR

PCR was done using one RSO and one HPV specific primer; this means for one sample 48 PCR reactions are performed for HPV16, 60 for HPV18 and 54 for HPV45. PCR conditions

have low annealing temperature to allow spontaneous annealing of the RSO. PCR conditions are shown in Table 10. See Table 11 for PCR reagents.

3.3.3.3.3 Nested PCR

The primary PCR amplicons were subjected to nested PCR using nested HPV specific primer and the same RSO that was used in the primary PCR reaction. PCR conditions are shown in Table 10. See Table 11 for PCR reagents. The annealing temperature in the nested PCR was 55ºC to increase the specificity of the PCR reaction.

3.3.3.3.4 Electrophoresis

Five micro-litres of nested PCR amplicon was electrophoresed on a 2% (w/v) agarose gel stained with ethidium bromide and viewed under UV light. Consistent with Thorland et al. (2000), all fragments were subjected to DNA sequence analysis using the HPV sequencing primer specific for the HPV region amplified by the nested PCR; for example, amplicons generated with HPV16 nested primer HPV16-790-25D would be sequenced with sequencing primer HPV16-839-25D; see section 3.3.3.6 for DNA sequence analysis.

3.3.3.4 Detection of Integrated Papillomavirus Sequences

Detection of Integrated Papillomaviruses Sequences (DIPS) (Luft et al. 2001) is based on a single side specific ligation mediated PCR and amplifies any sequence of DNA (cellular integration site) that is adjacent to a known sequence (HPV) (Figure 22). In summary, genomic DNA was digested with a restriction enzyme and a double stranded adapter primer, specific for the restriction enzyme used, was ligated to the digested DNA. This was followed by a linear PCR reaction that comprised a HPV specific primer. A second, nested, PCR reaction that comprised a HPV specific primer and a primer complementary to the missing part of the adapter primer was performed on the linear PCR product. The nested PCR amplicons were separated by gel electrophoresis and subjected to DNA sequence analysis.

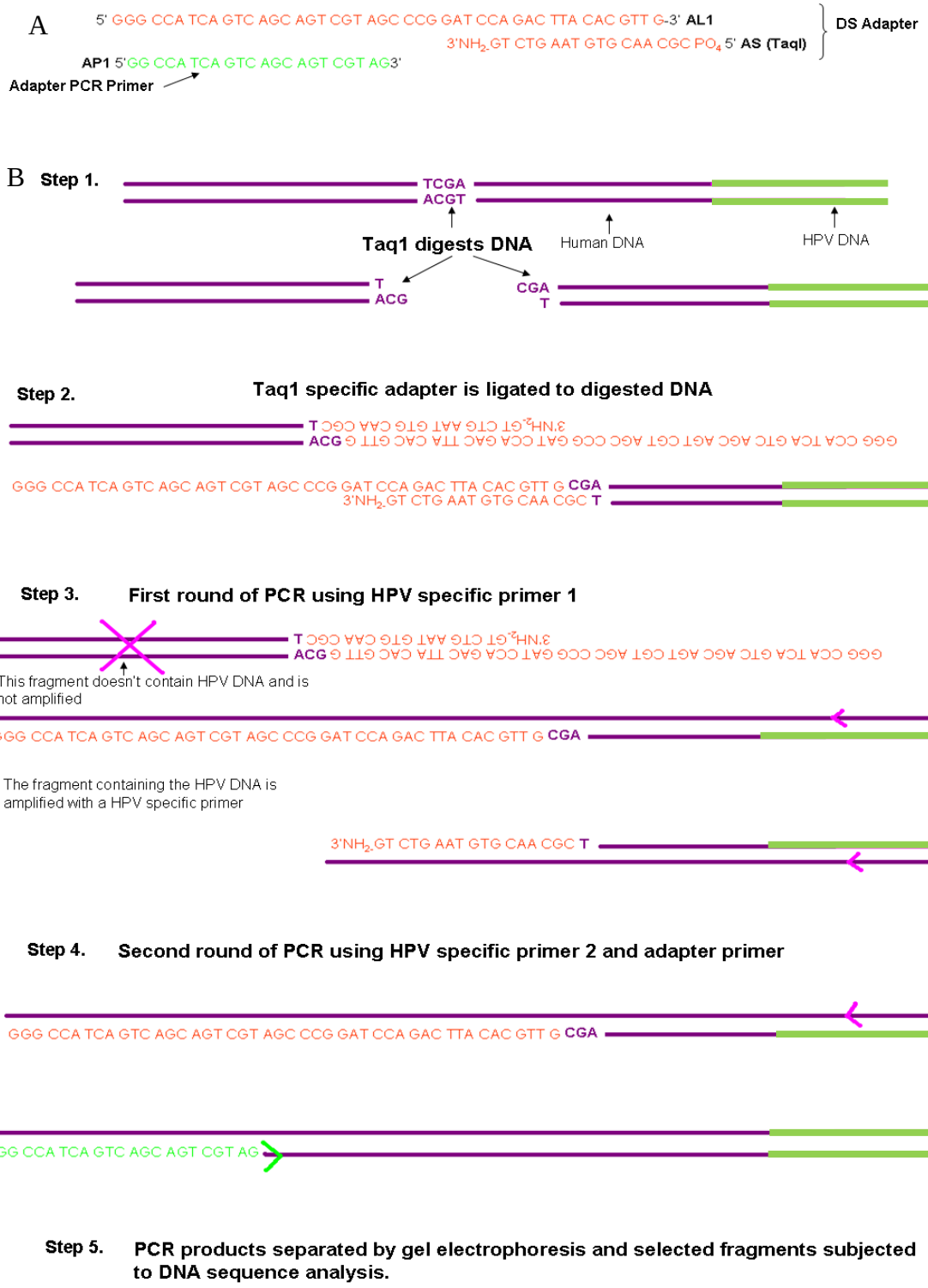


Figure 22 Detection of Integrated Papillomavirus sequences (DIPS). A) The double stranded adapter primer used in DIPS. B) The steps involved in DIPS.

3.3.3.4.1 DIPS Adapter Primers

The synthetic oligonucleotide adapter primers used in DIPS are known as "vectorettes". The vectorette consists of partial double stranded DNA and contains a central mis-matched region (Arnold and Hodgson 1991). Vectorettes are designed so they can be amplified only if they are attached to the DNA sequence of interest. DIPS involves "suppression PCR"; this evolved from the idea of vectorette PCR and uses some aspects of the method. Like vectorette PCR, it uses a partially double stranded synthetic oligonucleotide, the bottom of strand is complementary to the 3' end of the top strand and contains an amine group that prevents polymerase catalyzed extension of the lower adapter strand (Siebert et al. 1995) (Figure 22A).

Table 17 shows the adapter primers used in DIPS. A double stranded adapter primer specific for Taq$^{\alpha}$I and Sau3AI restriction enzyme used in the assay is constructed by mixing 25µl of 100µM AL primer and 25µl of 100µM AS primer in 50µl of 66mM trisHCL (pH 7.4), and gradually cooling the mix from 90ºC to 4ºC over a period of 16 hours on a thermocycler.

Name	**Sequence**
DIPS_AL1	GGGCCATCAGTCAGCAGTCGTAGCCCGGATCCAGACTTACACGTTG
DIPS_AS Taq	PO_4-CGCAACGTGTAAGTCTG-NH_2
DIPS_AS SauAI	PO_4-GATCCAACGTGTAAGTCTG-NH_2
DIPS_AP1	GGCCATCAGTCAGCAGTCGTAG

Table 17 Adapter Primers used in DIPS taken from Luft et al. (2001).

3.3.3.4.2 HPV Specific Primers

HPV16 and HPV18 specific primers used in DIPS are shown in Table 18 and Table 19, respectively. Primer sequences for HPV16 and HPV18 were obtained from Luft et al. (2001). The primary and nested HPV45 primers designed for RS-PCR (Table 16) were used in DIPS for HPV45. First attempts at DIPS using HPV16 primers specified by Luft et al. (2001) were unsuccessful. HPV16 primers specified by Luft et al. (2001) were substituted by HPV16 specific primers used for RS-PCR (Table 14).

Name	Annotation	Sequence
DIPS_PCR1_16F1	1F1	CGGGATGTAATGGATGGTTTTATG
DIPS_PCR1_16F2	1F2	GGGCGCCATGAGACTGAAACAC
DIPS_PCR1_16F3	1F3	CGCCAGAATGGATACAAAGACAAAC
DIPS_PCR1_16F4	1F4	GTTTGCACGAGGACGAGGAC
DIPS_PCR1_16F5	1F5	CACACCGGAAACAGAGCCAG
DIPS_PCR2_16F1	2F1	GGGGATGCTATATCAGATGACGAGAAC
DIPS_PCR2_16F2	2F2	GGTGGAAGTGGGGGTGGTTG
DIPS_PCR2_16F3	2F3	TGGTACAATGGGCCTACGATAATG
DIPS_PCR2_16F4	2F4	GACGAGGACAAGGAAAACGATGGAG
DIPS_PCR2_16F5	2F5	GGAAACCCCTGCCACACCAC

Table 18 HPV16 primers used in DIPS. From Luft et al. (2001).

Name	Annotation	Sequence
DIPS_PCR1_18F1	1F1	CCAGAAGGTACAGACGGGGAG
DIPS_PCR1_18F2	1F2	ATAGACAACGGGGGCACAGAG
DIPS_PCR1_18F3	1F3	CCACCAAAATTGCGAAGTAGTG
DIPS_PCR1_18F4	1F4	GAGGAAGAGGAAGATGCAGACAC
DIPS_PCR1_18F5	1F5	ACCTACAGGCAACAACAAAAGAC
DIPS_PCR1_18R1	1R1	CCAGTATCTACCATATCACCATCTTCCA
DIPS_PCR1_18R2	1R2	AGTGTCCACAGGCTCAAAGGC
DIPS_PCR1_18R3	1R3	GGAAATAGACACAGAGGTAGACGAAGG
DIPS_PCR1_18R4	1R4	GGGGACGTTATTACCACAATATACACA
DIPS_PCR2_18F1	2F1	CGGGTTGTAACGGCTGGTTTTATG
DIPS_PCR2_18F2	2F2	GGGGCACAGAGGGCAACAAC
DIPS_PCR2_18F3	2F3	TAATGGGAGACACACCTGAGTGGATAC
DIPS_PCR2_18F4	2F4	AAGATGCAGACACCGAAGGAAACC
DIPS_PCR2_18F5	2F5	CAGGCAACAACAAAAGACGGAAAC
DIPS_PCR2_18R1	2R1	TCTACCATATCACCATCTTCCAAAACTG
DIPS_PCR2_18R2	2R2	AGGATGGACGTGTAAGAAACTCAGGG
DIPS_PCR2_18R3	2R3	TCAAACCCAGACGTGCCAGTAAAC
DIPS_PCR2_18R4	2R4	ACAGACAGATGGCAAAAGCGGG

Table 19 HPV18 primers used in DIPS. From Luft et al. (2001).

3.3.3.4.3 Control Primers

As carried out by Luft et al. (2001), to confirm successful digestion and ligation of adapter, amplification of a genomic locus on chromosome 21 (1.4 kb, accession number ap001068) was done on the ligation product of each sample using primers DIPS_CON_1 (TTCTCTATGTGCGTTCTCTCCCTG) in the first linear PCR and DIPS_CON_2 (CAAACTCCAGGTCTCCAACCAG) together with AP1 (Table 17) in the nested PCR.

3.3.3.4.4 DNA Digestion

See Table 20 and Table 21 for reagents used. Genomic DNA (1.2µg) from each sample was digested with Taq$^{\alpha}$I enzyme and Sau3AI enzyme, in separate reactions, in a total volume of 20µl for 16 hours at 37°C. The volume of DNA added to each reaction varied due to differing DNA concentrations; a volume of sterile PCR grade water was added to make a final reaction volume of 20µl . Sau3AI and Taq$^{\alpha}$I enzymes were deactivated by incubation at 60°C and 80°C, respectively for twenty minutes.

Reagent	**Concentration**	**Volume (µl)**
DNA	Variable	To give 1.2µg
Sau3AI	4U/µl	2.5
Buffer	10X	2
BSA	100X	0.2
H_2O		To make total volume up to 20µl
Total volume (µl)		20

Table 20 Reagents used in Sau3AI digestion. Enzyme obtained from New England Biolabs (UK) Ltd, Hertfordshire, UK. Sau3AI recognition site is GATC.

Reagent	**Concentration**	**Volume (µl)**
DNA	Variable	To give 1.2µg
TaqI	20U/ µl	0.5
Buffer	10X	2
BSA	100X	0.2
H_2O		To make total volume up to 20µl
Total volume (µl)		20

Table 21 Reagents used in Taq$^{\alpha}$I digestion. Enzyme obtained from New England Biolabs (UK) Ltd, Hertfordshire, UK. Taq$^{\alpha}$I recognition site is TCGA.

3.3.3.4.5 Modifications to Digestion

Following advice from New England Biolabs, to reduce star activity, the digestion incubation time was decreased from 16 hours to 2 hours at 37°C for both Taq$^{\alpha}$I and Sau3AI digestions. Sau3AI is supplied in glycerol. A high glycerol ratio is linked with star activity thus the volume of Sau3AI digestion reaction was increased to 50µl to dilute the glycerol in the

digestion reaction (Table 22). Modification to Taq$^{\alpha}$I dissection was not needed due to low glycerol ratio in digestion reaction.

Reagent	Concentration	Volume (µl)
DNA	Variable	To give 1.2µg
Sau3AI	4U/µl	2.5
Buffer	10X	5
BSA	100X	0.5
H20		To make total volume up to 50µl
Total volume (µl)		50

Table 22 Reagents used in Sau3AI digestion. Enyzme obtained from New England Biolabs (UK) Ltd, Hertfordshire, UK. Sau3AI recognition site is GATC.

3.3.3.4.6 Ligation

Enzyme specific adapters for Taq$^{\alpha}$I and Sau3AI were ligated to the digested DNA by incubation at room temperature for 2 hours with T-4 DNA ligase (New England Biolabs (UK) Ltd, Hertfordshire, UK) (Table 23). The ligase reaction was heat inactivated by incubation at 65°C for 10 minutes. Sixteen micro-litres of sterile PCR grade water were added to the ligase product to achieve a final volume of 40µl .

Reagent	Concentration	Volume (µl)
Ligase Buffer	10X	2.4
T4 Ligase	400U/µl	1
Adapter	25µM	0.48
H2O		0.12
Digested DNA	60ng/µl	20
Total Volume		24

Table 23 Reagents used in ligation of adapter primers to digested DNA. T4 DNA ligase and buffer were obtained from New England Biolabs (UK) Ltd., Hertfordshire, UK.

3.3.3.4.7 Modification to ligation

The ligation reaction volume was increased to 60µl to compensate for increasing the volume of the restriction digest reaction for Sau3AI, described in section 3.3.3.4.5. Furthermore, the volume for the ligation reaction for Taq$^{\alpha}$I was increased to 60µl to reduce variability

between Taq$^{\alpha}$I and Sau3AI reactions (see Table 24). Incubation was as described in section 3.3.3.4.6.

Reagent	Concentration	Volume (µl) Sau3AI	Volume (µl) Taq$^{\alpha}$I
Ligase Buffer	10X	6	6
T4 Ligase	400U/µl	1	1
Adapter primer	25µM	1.2	1.2
H2O		1.8	31.8
Digested DNA	60ng/µl	50	20
Total Volume		60	60

Table 24 Reagents used in modified ligation of adapter primers to digested DNA. T4 DNA ligase and buffer were obtained from New England Biolabs® (UK) Ltd, Hertfordshire, UK.

3.3.3.4.8 Linear PCR

PCR conditions are shown in Table 10. In separate reactions for Sau3AI and Taq$^{\alpha}$I, 2µl of ligation product were added to the PCR reagents listed in Table 11. When using ligation product from the amended procedure in section 3.3.3.4.7, 3µl of ligation product and 18.37µl sterile water were added to the PCR reagents listed in Table 11.

3.3.3.4.9 Nested PCR

Two micro-litres of linear PCR amplicon were added to the PCR reagents listed in Table 11. PCR conditions are shown in Table 10.

3.3.3.4.10 Electrophoresis

Five micro-litres of each DIPS amplicon was electrophoresed on a 2% (w/v) agarose gel stained with ethidium bromide and viewed under UV light. All fragments that differed in size to predicted fragment sizes (Table 25) were subjected to DNA sequence analysis using the HPV specific primer that was used in nested PCR (see section 3.3.3.6).

HPV Type				Sau3AI Cut (GATC)		TaqI Cut (TCGA)	
NC_001526.1	**Primer**[1]	**Strand**[2]	**Nested**[3] **PCR**	**Sau3AI Cut**	**Expected size**	**Taq$^{\alpha}$I Cut**	**Expected size**
HPV16	P1	Sense	790	870	80		0
HPV16	P2	Sense	1587	3478	1891		0
HPV16	P3	Sense	2414	3478	1064		0
HPV16	P4	Sense	2964	3478	514		0
HPV16	P5	Anti-sense	5121	4537	584	505	4572
HPV16	P6	Anti-sense	5850	5233	617	505	5266
HPV16	P7	Anti-sense	6651	6150	501	505	6185
HPV16	P8	Anti-sense	7524	6950	574	505	7017
X05015							
HPV18	F1	Sense	948	4732	3833	3083	2160
HPV18	F2	Sense	1425	4732	3356	3083	1668
HPV18	F3	Sense	1929	4732	2852	3083	1224
HPV18	F4	Sense	2814	4732	1967	3083	279
HPV18	F5	Sense	3634	4732	1147	3083	0
HPV18	R1	Anti-sense	6217	5038	1223	3083	3140
HPV18	R2	Anti-sense	4982	4732	294	3083	1943
HPV18	R3	Anti-sense	4629	919	3754	3083	1619
HPV18	R4	Anti-sense	4016	919	3141	3083	989
X74479							
HPV45	P1	Sense	815	919	104	3453	2638
HPV45	P2	Sense	1597	2596	999	3453	1856
HPV45	P3	Sense	2365	2596	231	3453	1088
HPV45	P4	Sense	3146	4532	1386	3453	307
HPV45	P5	Sense	3939	4532	593	3453	486
HPV45	P6	Anti-sense	5544	4724	820	3453	2091
HPV45	P7	Anti-sense	6266	5844	422	3453	2813
HPV45	P8	Anti-sense	7001	6996	5	3435	3566
HPV45	P9	Anti-sense	7764	6996	768	7014	750

Table 25 Expected amplicon sizes for DIPS. Restriction sites (cut sites, nucleotide position shown in base-pairs) were predicted using New England Biolabs® "Webcutter" at http://tools.neb.com/NEBcutter2/ using Genbank accession numbers NC_001526.1, X05015 and X74479 for HPV16, HPV18 and HPV45, respectively.

[1] Annotated primer label. Note predicted fragment sizes for HPV16 are relative to primers from Table 14.
[2] DNA strand amplified by primer. Sense=coding strand. Anti-sense=template strand.
3 Base pair number of the viral nucleotide that is identical to the 5' base of the nested primer.

3.3.3.5 Amplification of Papillomavirus Oncogene Transcripts

Amplification of Papillomavirus Oncogene Transcripts (APOT) is an assay that allows the discrimination of HPV mRNAs derived from integrated and episomal viral genomes. APOT, first described by Klaes et al. (1999), is based on structural differences among 3′-ends of HPV oncogene transcripts and employs the rapid amplification of 3′ cDNA ends (3′-RACE) technique. Three prime-RACE reactions are used to isolate unknown 3′ sequences; in this PhD the unknown sequence was human sequence. APOT is summarised in Figure 23: cDNA is generated from mRNA using an adaptor linked oligo (dT)-primer. The cDNA is subjected to primary PCR with a HPV specific forward primer, specific for a region of E7 ORF, and a reverse primer consisting of the adapter primer sequence. The primary PCR is followed by a nested PCR with HPV E7 specific forward primer and a reverse primer consisting of the oligo$(dT)_{17}$ sequence. Nested PCR products are separated by gel electrophoresis and subjected to DNA sequence analysis.

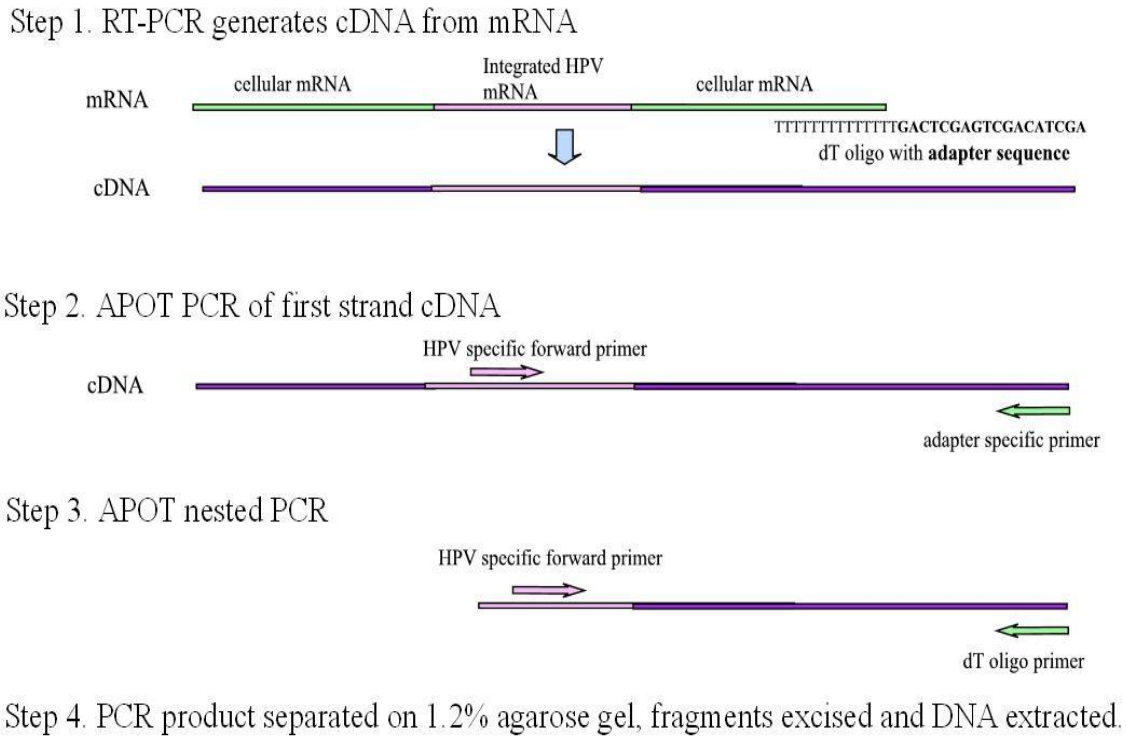


Figure 23 Amplification of Papillomavirus Oncogene Transcripts (APOT).

3.3.3.5.1 Reverse-transcription

Reverse transcription is the generation of complementary DNA (cDNA) from mRNA using reverse-transcriptase (RT) enzyme: RNA is denatured by heating at 70 ºC followed by rapid chilling on ice to maintain the RNA in linear, denatured form. dNTPs, primer and RT are added to the denatured RNA and incubated at a temperature optimal for the RT, to allow generation of cDNA by reverse transcriptase. Three prime-RACE utilises an adaptor linked oligo (dT)-primer, first described by Frohman et al. (1988), to generate cDNA. The adapter linked oligo $(dT)_{17}$ primer ((dT)17-p3) is a sequence of nucleotides consisting of restriction enzyme (XhoI (C/TCGAG), SalI (G/TCGAC) and ClaI (AT/CGAT)) recognition sites (adapter) followed by a sequence consisting of 17 thymine nucleotides (see Table 27) . The choice of RT enzyme for APOT is Superscript™ II (Invitrogen™, Life Technologies Ltd , Paisley, UK) that is identical to Moloney murine leukaemia virus reverse transcriptase (M-MLV RT), except that three acidic amino acids essential for RNase H activity of the gene have been mutated to neutral amino acids, thus eliminating the RNase H activity of the RT (Bustin 2000).

RNA is very sensitive to degradation and any work using RNA needs to be performed using strict aseptic techniques and equipment and consumables that are certified RNAse free. Reverse transcription was done in a PCR-hood sterilised using UV light and prior washing with Ambion® RNAseZAP®(Invitrogen™, Life Technologies Ltd , Paisley, UK) followed by washing with 70% ethanol. All pipettes and microfuge tube racks were wiped with Ambion® RNAseZAP® and 70% ethanol.

Residual DNA in an RNA sample could potentially invalidate APOT by amplification of DNA, not cDNA from a sample. A DNAse step was performed before reverse transcription to remove residual DNA: 1µg of RNA was incubated with 0.2 units of DNAse (New England Biolabs (UK) Ltd, Hertfordshire, UK) in a total volume of 10µl for 10min at 37ºC followed by heat inactivation at 75ºC for 10min.

The DNAse treated RNA was mixed with the reagents in Table 26, for step 1, and heated at 70ºC for 10min and put on ice. The RNA was then mixed with reagents in Table 26, for step

2, and heated at 42ºC for 10min. Two hundred units (1µl) of Superscript™ II (Invitrogen™, Life Technologies Ltd , Paisley, UK) were added to the RNA and incubated at 42ºC for 1hr followed by heat inactivation at 70ºC for 15min. The cDNA was diluted 1:50 by adding 980µl of sterile, RNAse free water.

Step	Reagent	Concentration	Volume (µl)
Step 1	RNA	100ng/µl	10
	(dT)17-p3 primer	500ng/ µl	1
	dNTP	10mM	1
	Total volume (µl)		12
Step 2	First strand buffer	5X	4
	DTT	0.1M	2
	Water (RNAse free)		1
	Total volume (µl)		19

Table 26 Reagents used in reverse transcription. Reagents supplied by Invitrogen™, Life Technologies Ltd, Paisley, UK.

3.3.3.5.2 Primers

APOT PCR comprised of two PCR reactions: a primary PCR and a nested PCR. See Table 27 for the primers used. HPV primer sequences were obtained from Vinokurova et al. (2008). p3 reverse and (dT)17-p3 sequences were obtained from Frohman et al. (1988).

Primer Name	Sequence
p3 reverse	GACTCGAGTCGACATCG
(dT)17-p3	GACTCGAGTCGACATCGATTTTTTTTTTTTTTTT
HPV16 PCR1	CGGACAGAGCCCATTACAAT
HPV16 PCR2	CTTTTTGTTGCAAGTGTGACTCTACG
HPV18 PCR1	TAGAAAGCTCAGCAGACGACC
HPV18 PCR2	ACGACCTTCGAGCATTCCAGCAG
HPV45 PCR1	CCCACGAGCCGAACCACAG
HPV45 PCR2	GAGAGCTCGGCAGAGGACCTTAG

Table 27 APOT primers for HPV16, HPV18 and HPV45. HPV primer sequences were obtained from Vinokurova et al. (2008). p3 reverse and (dT)17-p3 sequences were obtained from Frohman et al. (1988).

3.3.3.5.3 Primary PCR

PCR conditions are shown in Table 10. One micro-litre of cDNA was added to the reagents in Table 11.

3.3.3.5.4 Nested PCR

PCR conditions are shown in Table 10. Five micro-litres of primary PCR amplicon were added to the reagents in Table 11.

3.3.3.5.5 Electrophoresis

Nested PCR amplicons were electrophoresed on a 1.2% (w/v) agarose gel. All fragments were excised and subjected to DNA sequence analysis using the HPV specific primer used in the nested PCR reaction (section 3.3.3.6).

3.3.3.6 DNA Sequence Analysis

In this PhD, the Sanger dideoxy method of DNA sequencing was used (Sanger et al. 1977). In summary, a primer is annealed to a PCR amplicon in an asymmetric PCR reaction (a single primer used) containing normal dNTPs and a small proportion of dideoxynucleoside triphosphates (ddNTPs) labelled with a fluorescent tag. A different florescent label is used for each ddNTP nucleoside. If a ddNTP molecule is incorporated into the DNA strand during PCR, further extension along that strand is prevented. The PCR reaction produces fluorescently labelled amplicons that differ by one base pair. Amplicons are separated by electrophoresis usually using an ABI™ sequencer (Applied Biosystems™, Life Technologies Ltd, Paisley, UK). The ABI™ sequencer uses capillary technology and the amplicons are electrophoresed through capillaries containing acrylamide gel. As the amplicons are electrophoresed, the smaller fragments pass through the capillary faster and a laser excites the fluorophore on the amplicon, the fluorescence is detected and as subsequent amplicons are detected, a sequence comprised of a chain of fluorescence signals is obtained. The fluorescence data is converted into a file containing a chromatogram trace of the DNA sequence (Figure 24). Sanger sequencing was done off site by Source BioScience LifeSciences (Oxford, UK) and the following requirements were adhered to:

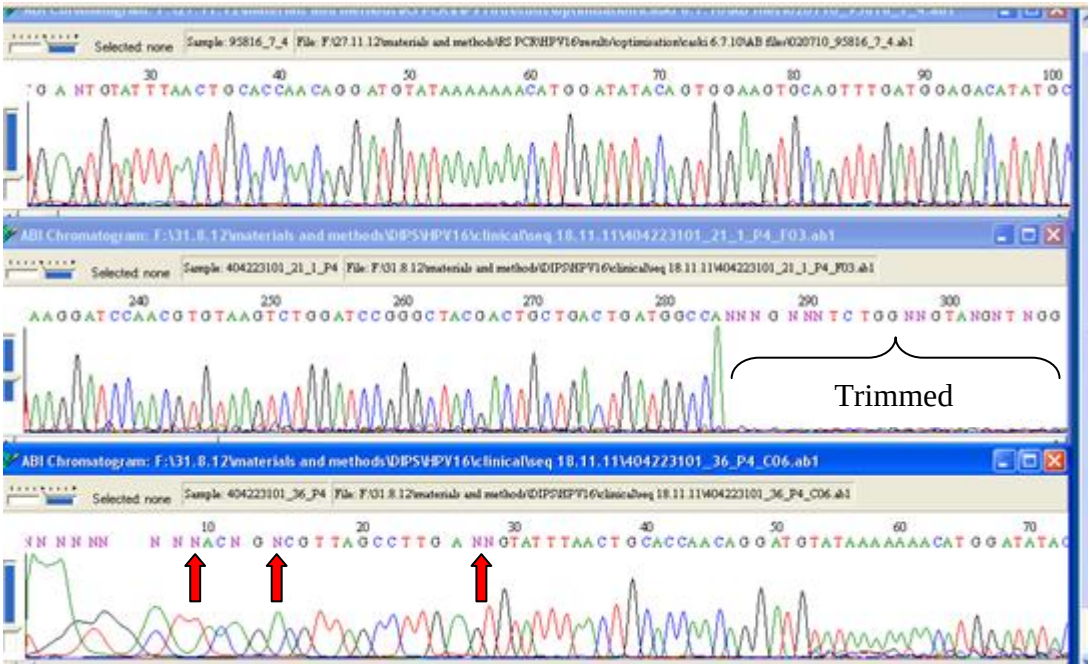


Figure 24 DNA sequencing chromatogram. Top chromatogram required minimal editing. Middle chromatogram required trimming after "GGCCA" to remove noise. The bottom chromatogram required editing to reduce the number of uncalled bases (N); the correct nucleotide was added.

3.3.3.6.1 Amplicon Requirements

One nanogram per micro-litre of DNA per 100bp of amplicon length was needed for sequencing. For amplicons with a single band observed by electrophoresis, the concentration of the amplicon was estimated by comparison to a size standard. A 100bp DNA ladder, with known concentrations, was used as a size standard and electrophoresed alongside the samples on the same gel. Amplicons with multiple bands observed by electrophoresis were subjected to gel purification. The concentrations of the purified amplicons were estimated by electrophoresis of 5µl of the purified amplicon alongside a size standard 100bp DNA ladder.

3.3.3.6.2 Primer Requirements

Amplicons were sequenced with a HPV specific primer as explained in each section for RS-PCR (3.3.3.3.4), DIPS (3.3.3.4.10) and APOT (3.3.3.5.5). Primers were sent at a concentration of 3.2pmol/µl.

3.3.3.6.3 PCR Purification

Before a PCR amplicon could be sent for Sanger sequencing, purification was essential to remove residual dNTPs and primers remaining in the amplicon that could interfere with the Sanger PCR reaction. Amplicons that were observed as a single band by electrophoresis, exonuclease I and alkaline phosphatase purification was used. Amplicons that were observed as multiple bands by electrophoresis, amplicons were excised from the gel and a gel purification kit was used.

Exonuclease I catalyses the removal of nucleotides from single-stranded DNA in the 3' to 5' direction and was applied to amplicons to remove residual primers. Amplicon size and concentration was estimated as described in 3.3.3.6.1 and sufficient DNA was added to the exonuclease reaction to allow 1ng/µl per 100bp in a total volume of 20µl (Table 28). The sample was incubated at 37ºC for 30min followed by heat inactivation at 80 ºC for 20min.

Alkaline phosphatase is an enzyme that catalyzes the removal of phosphate groups from nucleotides. An alkaline phosphatase reaction was used to prevent the PCR amplicons from ligating to each other and maintaining the amplicons in a linear form for sequencing. The entire 10µL of exonuclease treated amplicon was added to the reagents for alkaline phosphatase treatment in Table 29 and incubated at 37ºC for 30min followed by heat inactivation at 80 ºC for 2min. The amplicon was then sent for sequencing with the appropriate primer.

Reagent	Volume (µl)
Amplicon DNA	Variable
Alkaline Phosphatase Buffer	1
Sterile water	For a final volume of 10 µl
Exonuclease I 20U/µl	0.5
Total	10

Table 28 Reagents for Exonuclease I reaction (New England Biolabs (UK) Ltd, Hertfordshire, UK).

Reagent	Volume (µl)
Exonuclease I treated amplicon DNA	10
Alkaline Phosphatase 1U/µl	1
Alkaline Phosphatase Buffer	2
Sterile water	7
Total Volume	20µl

Table 29 Reagents for Alkaline phosphatase reaction (Roche Diagnostics Ltd, West Sussex, UK).

Samples that had more than one band observed by electrophoresis were subjected to gel purification. To maximise the concentration purified from the gel, the entire remaining amplicon was electrophoresed on a gel stained with ethidium bromide. The gel was viewed by UV light at lowest possible intensity to prevent DNA damage, and photographed. Each lane in the gel was separated by cutting with a sterile scalpel and the bands were excised, 1 lane at a time whilst being view under low intensity UV light. Each gel band excised was placed in a microfuge tube and DNA was extracted using an Illustra™ GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare Life Sciences, Buckinghamshire, UK). The Illustra™ GFX™ PCR DNA and Gel Band Purification Kit is designed to extract DNA from PCR products or agarose gels using adsorption of DNA to a silica membrane in a column, followed by washing and elution of the DNA from the membrane. Manufacturer's instructions were followed. In summary, the gel slice was dissolved in a buffer containing a pH indicator to ensure the solution was optimal for DNA adsorption to take place. The dissolved agarose was added to the column and DNA was bound to the silica membrane by centrifuging at 13,000rpm for 30s. The membrane was washed using a buffer containing ethanol followed by centrifuging at 13,000rpm for 30s. The DNA was eluted by addition of 25µl of sterile water and incubation at room temperature for 1 minute followed by

centrifuging at 13,000rpm for 1 minute. Five micro-litres of eluted DNA was electrophoresed on a 2% (w/v) agarose gel stained with ethidium bromide; the concentration of the purified DNA was estimated and if a single band was observed, 1ng/µl per 100 base pairs of amplicon were sent for sequencing with an appropriate primer.

3.3.3.6.4 Sequencing of HPV16, HPV18 and HPV45 DNA in Vector

To confirm the presence of HPV16, HPV18 and HPV45 DNA in the plasmid vectors, described in section 3.1.3, 100ng/µl of plasmid encoding HPV DNA were sent to Source Bioscience LifeSciences (Oxford, UK) for Sanger DNA sequence analysis. E2 forward primer was the primer used for sequencing and E2 was confirmed in all of the vectors.

3.3.3.6.5 Data Analysis

The sequence data from Source BioScience LifeSciences was downloaded through a link provided in an email. Files in "ABI" format were used for analysis. The ABI file was opened and edited using "BioEdit", a free sequence alignment editor (Figure 24). Edited files were saved with a new filename and the nucleotide sequences were saved in "FASTA" format in a txt file. FASTA format allows multiple DNA sequences to be analysed simultaneously and is ">" followed by sequence name and the sequence is entered on the following line for example:

>DNA SEQ_1

Agctagctagatagatagctgatcgta.....

>DNA SEQ_2

Agctagctagatagatagctgatcgta.....

The txt file containing FASTA data was uploaded to The National Centre for Biotechnology Information Basic Local Alignment Search Tool (BLAST) at http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome. BLAST allows a comparison and alignment to sequence data in the NCBI database. BLAST compares

nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. The data was compared to the human genome database, to the database of "others" and to DNA sequences from reference strains with Genbank accession numbers NC_001526.1, X05015 and X74479 for HPV16, HPV18 and HPV45, respectively; megablast program was selected. Figure 25 shows BLAST output: sequences that had both human and HPV matches were integrated. Figure 26 shows the alignment of the BLAST data; the alignment in Figure 26B was used to determine the site of HPV disruption (1756bp in Figure 26B). Alignments from both Figure 26A and Figure 26B were used to determine whether sequence was the same, inserted or flush between human and HPV DNA. Matches with over 95% identity between query and reference sequence were considered to be viable matches; matches with less than 95% identity between query and reference sequence were interpreted with caution and a sample was sent for sequencing again if necessary. Occasionally a sequence would have a partial match to HPV sequence, for example in Figure 25B and Figure 26B but no match would be made to human sequences. In these circumstances DNA sequences were submitted to basic local alignment tool (BLAT) at the University of California Santa Cruz (UCSC), described below.

Nucleotide sequences that had a match to human DNA using BLAST were also submitted to a basic local alignment tool (BLAT) at the University of California Santa Cruz (UCSC) online at http://genome.ucsc.edu/cgi-bin/hgBlat?command=start (Kent et al. 2002) . BLAT is an alignment tool that compares the query sequence to an index of the entire human genome. BLAT search results (Figure 27A) were used to determine the match that had the highest score, with highest identity and spanned the lowest number of nucleotides. BLAT browser (Figure 27B) was then used to determine the chromosomal band in which integration occurred and to gather information of genes or repeat elements present at the site of integration.

To determine if integration resided within a fragile site, NCBI "Gene" database at www.ncbi.nlm.nih.gov/gene/ was used. The chromosome number "AND" and "Fragile" were entered into the search engine and output was a list of fragile sites on the chromosome of interest. Figure 28 shows a search for fragile sites on chromosome 8. Note that common fragile sites are detected as breaks on metaphase chromosomes when treated with reagents

such as aphidicolin. A common fragile site spans region of a chromosome (Sutherland and Richards 1999), approximately 1Mbp (Thorland et al. 2003), and is not a precise location such as the one given by a search for HPV integration, described above. Integration was considered to be within a common fragile site if the chromosomal band from the UCSC browser was within 1Mbp of a common fragile site listed on the NCBI Gene database. Rare fragile sites are identified by presence of AT nucleotide repeats thus the location can be precisely defined. Integration was considered to be within a rare fragile site if the chromosomal band was identical between the output from the UCSC browser and a rare fragile site listed on the NCBI Gene database.

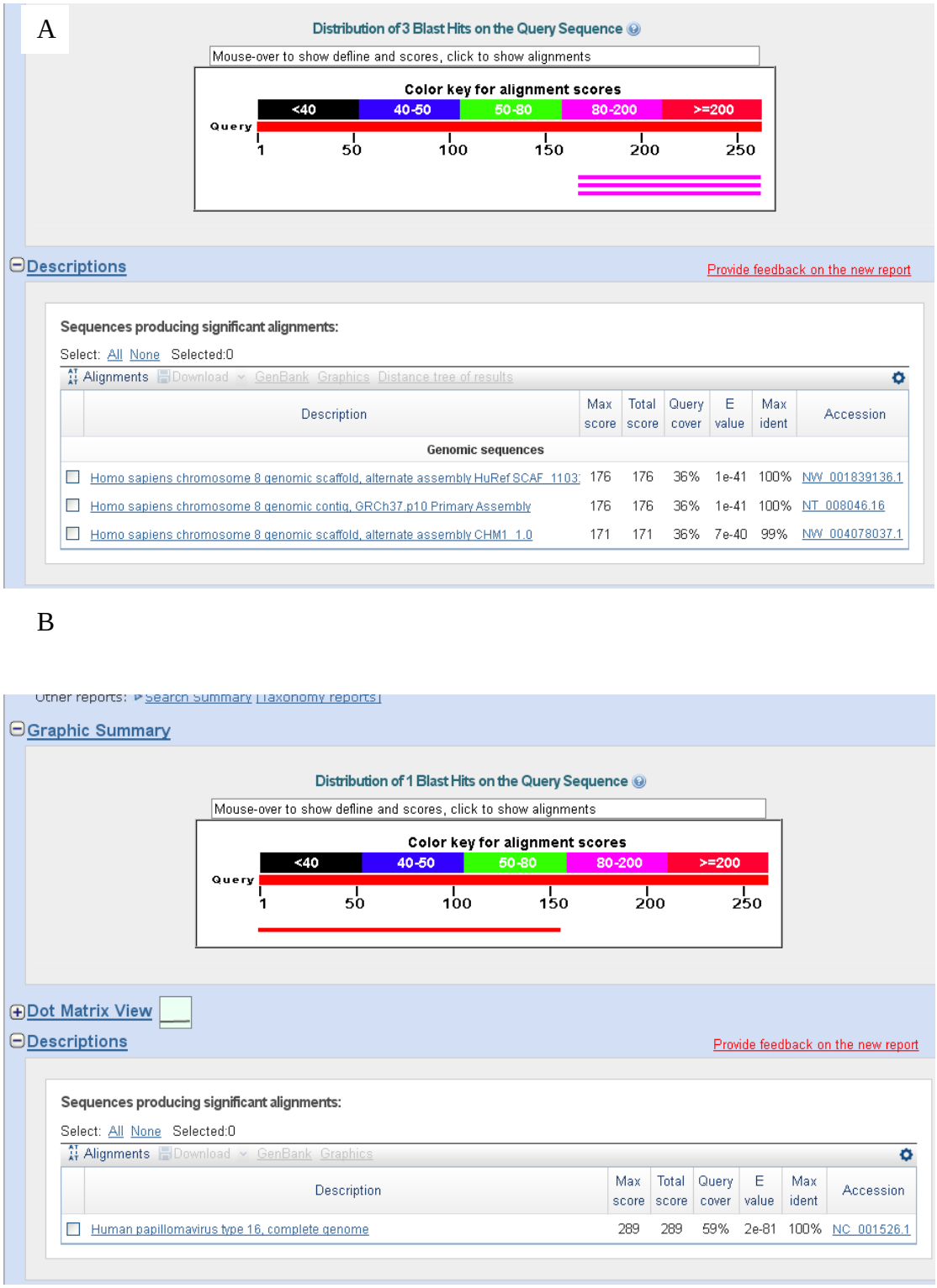
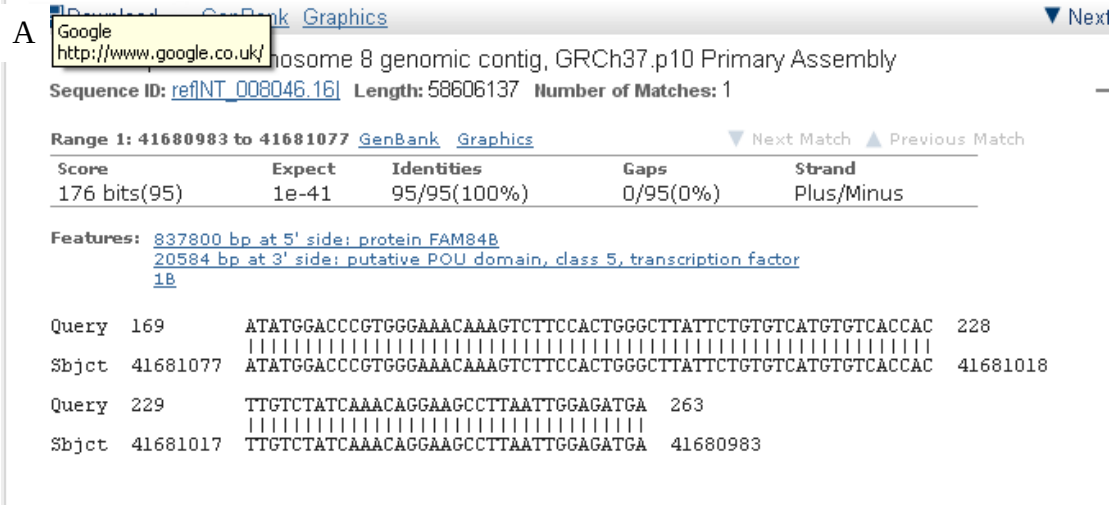


Figure 25 Expected BLAST output for an integration site. A) Comparison of integrated HPV to the human genome database; sequence after 150bp is a 100% match for human sequence on chromosome 8. B) Comparison of integrated HPV to HPV16 reference strain NC_001526.1; sequence up to approximately 150bp is 100% match to HPV16.

A



B

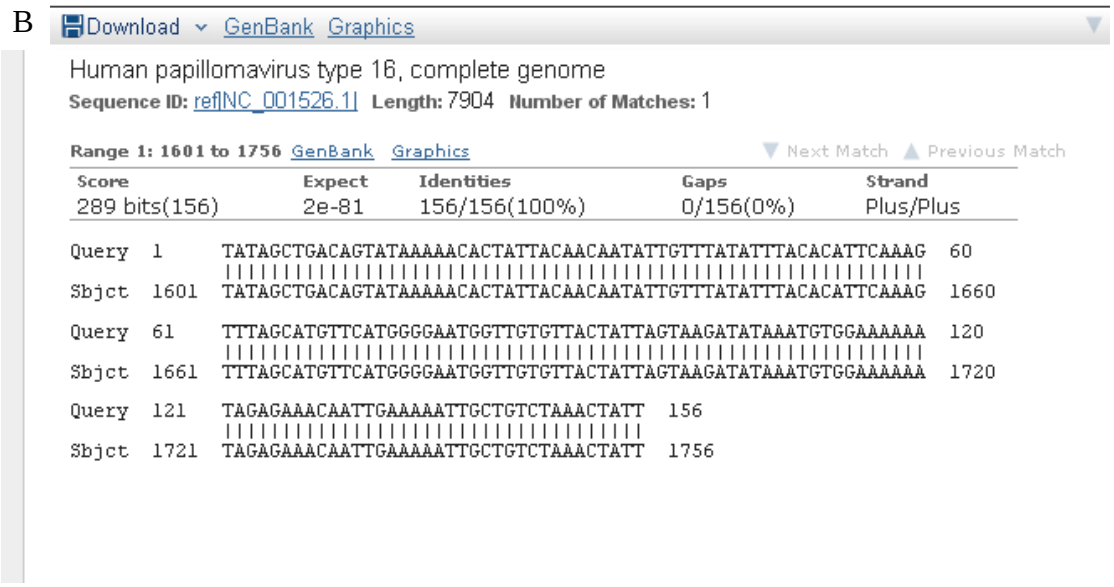


Figure 26 BLAST alignment of data shown in Figure 25. "Query" is the sequence uploaded to BLAST; "Subject" is the sequence the query is aligned to; plus/minus indicates antisense orientation; plus/plus indicates sense orientation. A) Alignment of query sequence to human NT_0088046_16: 169bp to 263bp matches to NT_0088046_16 from 41681077bp to 41680983bp. B) Alignment of query sequence to HPV16: 1bp to 156bp matches to NC_001526.1 from 1601bp to 1756bp.

A

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	95	169	263	263	100.0%	8	-	128407434	128407528	95
browser details	YourSeq	28	34	62	263	100.0%	4	-	54599379	54599801	423
browser details	YourSeq	24	101	132	263	72.0%	4	-	135015350	135015374	25
browser details	YourSeq	24	100	124	263	100.0%	2	+	34561568	34561601	34
browser details	YourSeq	22	116	139	263	95.9%	3	-	120704734	120704757	24
browser details	YourSeq	21	154	175	263	100.0%	6	-	47490544	47490566	23
browser details	YourSeq	20	153	172	263	100.0%	2	-	235507404	235507423	20
browser details	YourSeq	20	33	52	263	100.0%	2	-	191741164	191741183	20

B

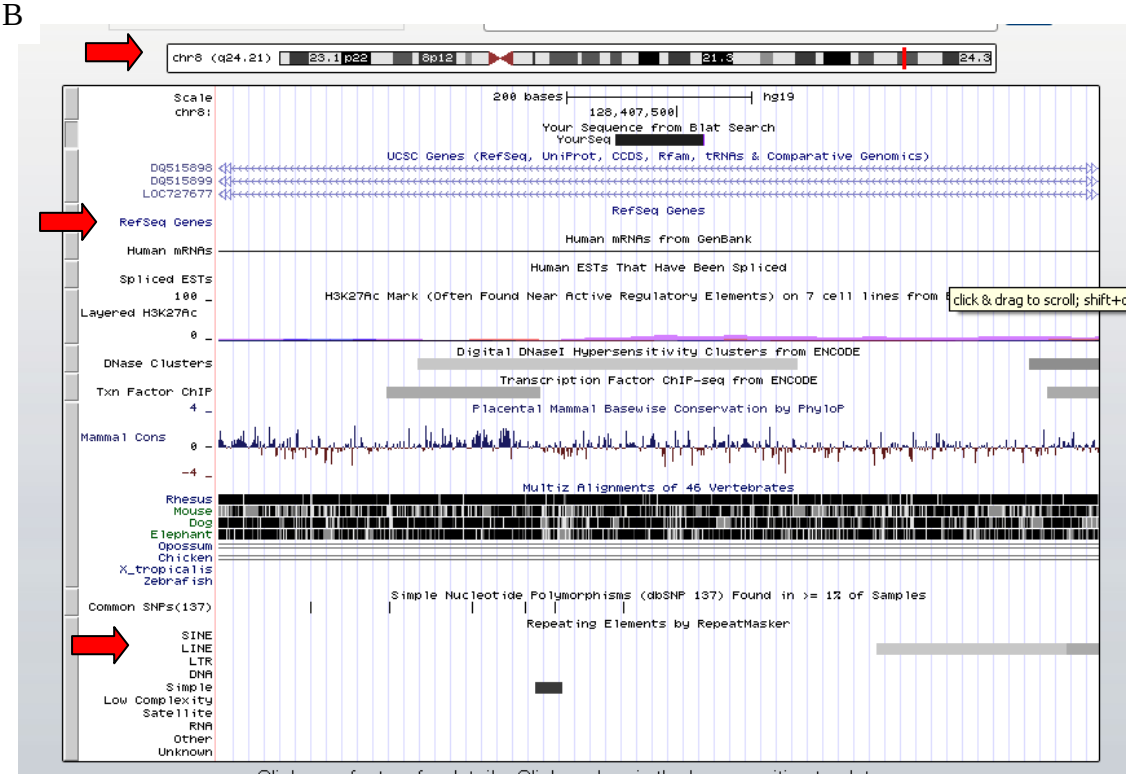


Figure 27 Output from UCSC BLAT alignment. A) BLAT search results. Red arrow indicates best match to chromosome 8, with highest score (95bp match with 100% identity that spans 95bp). B) BLAT browser output Top red arrow shows data of the chromosome band in which integration occurs; middle red arrow indicates data of integration occurs in relation to known genes; bottom red arrow shows data of integration in relation to repeat elements. Integration is within 8q24.21 and within 200bp of a DNA and LINE repeat element; there are no known genes within this region of integration.

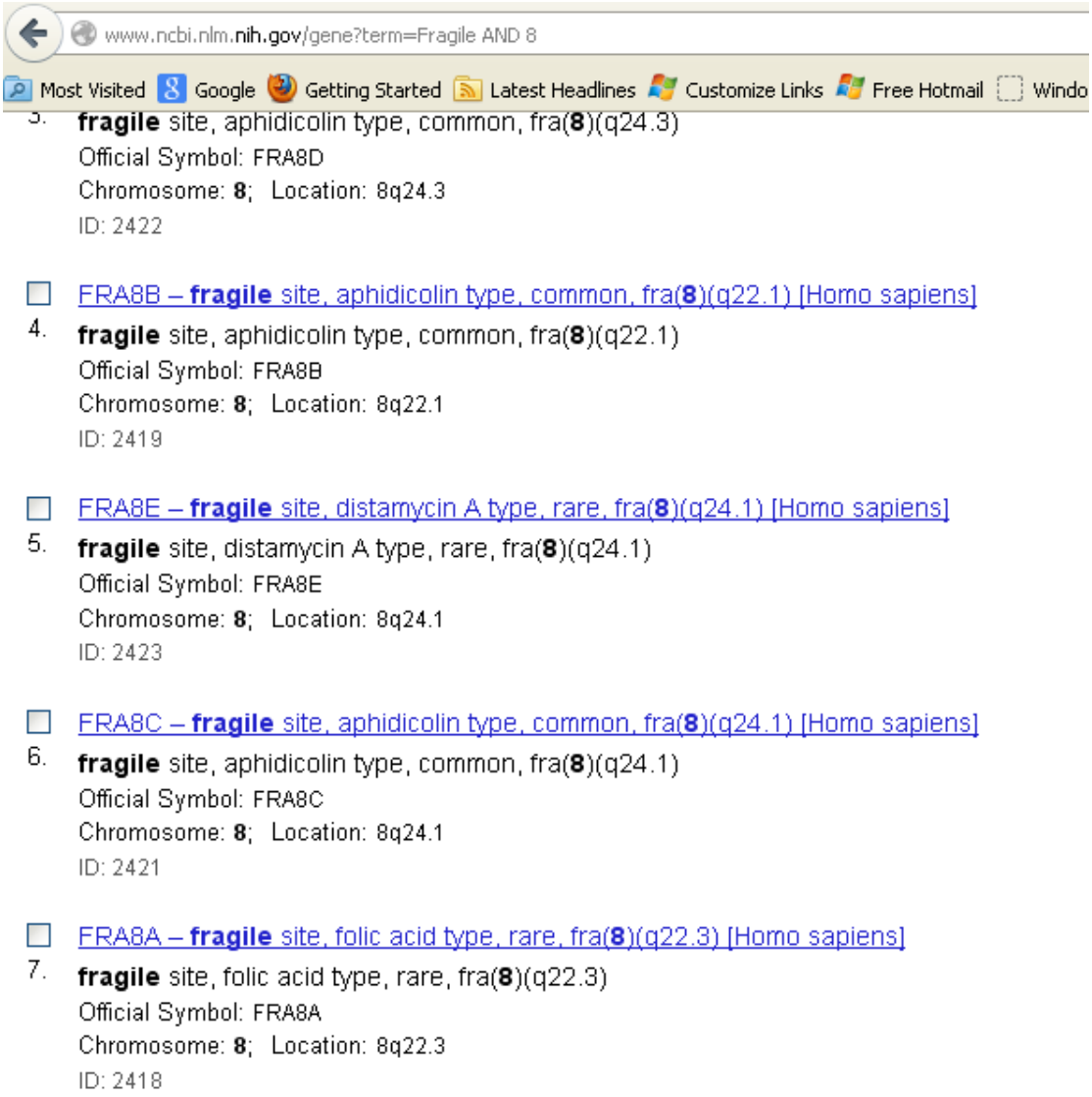


Figure 28 Output from NCBI "Gene" database when searched for "Fragile AND 8". Two fragile sites FRA8E and FRA8C reside in 8q24.1 and FRA8D on 8q24.3 but are not within 8q24.21 or within 1Mb of 8q24.21.

3.3.3.7 Confirmation of Integration sites

To verify the detection of integration sites, primers that flanked the host viral junction on HPV DNA and human DNA were designed. Primers were designed using the method described in section 3.3.3.1 with the following exceptions:

It was not possible to use chromatogram data to design primers because the sequence data were either too short, had too many uncalled bases or the host-viral junction was too close to the end of the chromatogram data to allow appropriate primers to be designed. An artificial sequence had to be made by obtaining 400bp either side of the host viral junction. Four hundred base pairs of HPV DNA preceding the host viral junction was obtained using the reference strain sequences (NC_001526.1 for HPV16 and X05015 for HPV18) and 400 base pairs of human DNA, beyond the host viral junction, was obtained using "View" "DNA" options on UCSC genome browser. It was important to consider the orientation of both the HPV and human sequences as this would affect the primer design. To verify that correct sequences for human and HPV had been obtained, the artificial HPV and human sequences were aligned against the edited chromatogram data (in section 3.3.3.6.5) for the integration event; BLAST was used. If the chromatogram data matched the obtained data then Primer3 was used to design primers. The host viral junction was marked with "[" and "]" so that primers were not designed over the host viral junction. A product size range of 200 to 500 base pairs was selected.

To verify non-contiguous sequences of HPV, the same principle as above applied but only HPV sequence was used to design the primers.

As described in section 3.3.3.1, the PCR could potentially be invalidated if non-specific sequences to HPV or human were amplified therefore, primer sequences were subjected to alignment to human and "others" database sequences using NCBI megaBLAST.

Primer ID	Sample[1]	Junction[2]	Sequence	Annealing[3]	$MgCL_2$[4]
J1455_caski_F	CaSki	J1455	CTGCACAGGAAGCAAAACAA	55	2
J1455_caski_R			TCATTGCCAAGGAAAACTCA		
Caski E1L1F	CaSki	Non- contiguous	TGAAATTTCTGCAAGGGTCTG	55	2
Caski E1L1R			GCAAACCACCTATAGGGGAAC		
Caski E2L1F	CaSki	Non- contiguous	ATGCGGGTGGTCAGGTAATA	55	2
Caski E2L1R			CCCATGTACCAATGTTGCAG		
Caski L1L1F	CaSki	Non- contiguous	TGCGTGCAACATATTCATCC	55	2
Caski L1L1R			GAACCATATGGCGACAGCTT		
SiHa J3132_F	SiHa	J3132	TGGATATACAGTGGAAGTGCAGTTTG	55	2
SiHa J3132_R			TGAGGCCACAAGACGTGGCA		
W12_J1756F	W12	J1756	AACGTGTTGCGATTGGTGTA	55	2
W12_J1756R			TTCCAACCTGAAACACACACA		
W12_J2749F	W12	J2749	AAGAAATGCATTGGATGGAAA	55	2
W12_J2749R			TCCTTGATCTGCCATGCTTA		
W12_J3726F	W12	J3726	CCCTGCCACACCACTAAGTT	55	2
W12_J3726R			ATAAGCAGGCTCGACCAAAA		
W12_J3197F	W12	J3197	ACAGTGGAAGTGCAGTTTGAT	55	2
W12_J3197R			TTGTAAGGCTCCTGCATGAAT		
Hela_J2497F	HeLa	J2497	TTCCTGCGATACCAACAAAT	55	2
Hela_J2497R			CCCGTCAGTTTCCTCATCTG		
Hela _J3030Chr8_F	HeLa	J3030	GCGTGCAGGACAAAATCATA	55	2
Hela _J3030Chr8_R			GCAGGGGGAGTACATTAGAGG		

Table 30 Primers used for confirmation of integration or non-contiguous sequences. Primer ID, sample type, junction name, primer sequences, PCR annealing temperatures and volume of 15mM $MgCl_2$ added to PCR regents is shown. Primer sets with "na" in annealing and $MgCl_2$ could not be successfully produce an amplicon following multiple attempts at optimisation.

[1] Sample number shown for clinical samples: biopsy, Baseline and HI-BCD. Clone ID shown for PC08 or PC09 cell lines.
[2] Junction ID is denoted as viral disruption number equal to last viral nucleotide before recombination to human sequence. Numbering of HPV sequence is according to Genebank accession numbers NC_001526.1 and X05015 for HPV16 and HPV18, respectively.
[3] Annealing temperature used in touchdown PCR in Table 10.
[4] Volume of 15mM $MgCl_2$ added to PCR reagents in Table 11.

Primer ID	Sample[1]	Junction[2]	Sequence	Annealing[3]	$MgCl_2$[4]
Hela _J3100Chr8_F	HeLa	J3100	GAAGGAAACCCTTTCGGAAC	55	2
Hela _J3100Chr8_R			CCCACTGGGAAGGATACAAA		
D1D D5D J5003 F	PC09 D	J5003	GCACCTATAGATTTTCCACTACG	55	2
D1D D5D J5003 R			ACAAGGTCGCTGCTTAGGG		
PCOD_J3303_f	PC09 D	J3303	TCAGTAACTGTGGTAGAGGGTCAA	68	2
PCOD_J3303_R			AGCTCTGTGGAGTCCGTGAT		
D5H J1848 F	PC09 H	J1848	ATGTTCATGGGGAATGGTTG	58	2
D5H J1848 R			CCTATGGGGCAGCATGATTA		
D1H D5H J2490F	PC09 H	J2490	TGAAATTTCTGCAAGGGTCTG	55	2
D1H D5H J2490R			CTCTCTGCCCACGGAAAATA		
PCO9 J6033F	PC09 Het	J6033	TGGTACATGGGGATCCTTTG	58	2
PCO9 J6033 R			GTTAGTCCTCCGAGGGAAGC		
PC09_J3632_F	PC09 Het	J3632	CCCTGCCACACCACTAAGTT	55	2
PC09_J3632_R			GTGGCTGTAGAGGTGGGAAA		
1M4M_J1194F	PC08 M	J1194	CTGCACAGGAAGCAAAACAA	58	2
1M4M_J1194R			ATGTTCCAGGGAGAACAGGA		
D1M non contig F	PC08 M	Non- contiguous	TGCCAGTACGCCTAGAGGTT	55	2
D1M non contig R			CGTGCCAAATCCCTGTTTT		
PC08_P_Noncontig_F	PC08 P	Non- contiguous	AAGGATTGTGCAACAATGTG	55	2
PC08_P_Noncontig_R			TGCACAAAATATGTTCGTATTCC		
D1Y J2116 F	PC08 Y	J2116	ATGCACAATTGGCAGACACT	58	2
D1Y J2116 R			ATCCCACCACGGTTGATTT		
D4Y J3167F	PC08 Y	J3167	TTTAACTGCACCAACAGGATG	58	4
D4Y J3167R			GTTGCCTCAATTCTGGGTGT		

Table 30 Continued. Primers used for confirmation of integration or non-contiguous sequences. Primer ID, sample type, junction name, primer sequences, PCR annealing temperatures and volume of 15mM $MgCl_2$ added to PCR regents is shown. Primer sets with "na" in annealing and $MgCl_2$ could not be successfully produce an amplicon following multiple attempts at optimisation.

[1] Sample number shown for clinical samples: biopsy, Baseline and HI-BCD. Clone ID shown for PC08 or PC09 cell lines.
[2] Junction ID is denoted as viral disruption number equal to last viral nucleotide before recombination to human sequence. Numbering of HPV sequence is according to Genebank accession numbers NC_001526.1 and X05015 for HPV16 and HPV18, respectively.
[3] Annealing temperature used in touchdown PC in Table 10.
[4] Volume of 15mM $MgCl_2$ added to PCR reagents in Table 11.

Primer ID	Sample[1]	Junction[2]	Sequence	Annealing[3]	$MgCl_2$[4]
1571_J1781_F	Baseline	J1781	ATGTTCATGGGGAATGGTTG	na	
1571_J1781_R	1571		GCTGCAGTGAGCTGAGATTG		
1571_J3205_f	Baseline	J3205	TGGAAGTGCAGTTTGATGGA	na	
1571_J3205_R	1571		TGTAGGTGGGAGGCGATATG		
1571_J3080F	Baseline	J3080	GCAATTGAACTGCAACTAACG	na	
1571_J3080R	1571		CTGGCAGGATGAAAATCCTAA		
913_J3098_F	Baseline	J3098	TTTAACTGCACCAACAGGATG	na	
913_J3098_R	913		TGTGGCTCGAACACAAACAT		
913_J6543_F	Baseline	J6543	GATATGGCAGCACATAATGACA	na	
913_J6543_R	913		CCATGGCTCTGGGTTTAGAT		
HIBCD8_J2431F	HIBCD8	J2461	TGGTGCAGCTAACACAGGTAA	na	
HIBCD8_J2431R			TTTTGCCAAAGGATTTCTGC		
HIBCD9_J1903F	HIBCD9	J1903	GCGTAGTACAGCAGCAGCAT	na	
HIBCD9_J1903R			GCTGGCTAACATGGCAAAAT		
1W E1-L1 F	Biopsy	Non- contiguous	AGGTACCAATGGGGAAGAGG	55	2
1W E1-L1 R	1W		GAACCATATGGCGACAGCTT		
1W L1-E6E7 F	Biopsy	Non- contiguous	GGACCGGTCGATGTATGTCT	55	2
1W L1-E6E7 R	1W		GCAACATATTCATCCGTGCTT		
1W L1-L1 F	Biopsy	Non- contiguous	TTTGCTACATCCTGTTTTTGTTTT	55	2
1W L1-L1 R	1W		AATGAAGGAGCTTGGTCAGTTA		
J2345bp 6N F	Biopsy	J2345	TGGTACAATGGGCCTACGAT	55	3
J2345bp 6N R	6N		ACCTGTGGATGTGCATGTGT		
J944 3O F	Biopsy	J944	ACTCTACGCTTCGGTTGTGC	55	2
J944 3O R	3O		CGGTCAGTTTCCTGCATTTT		

Table 30 Continued. Primers used for confirmation of integration or non-contiguous sequences. Primer ID, sample type, junction name, primer sequences, PCR annealing temperatures and volume of 15mM $MgCl_2$ added to PCR regents is shown. Primer sets with "na" in annealing and $MgCl_2$ could not be successfully produce an amplicon following multiple attempts at optimisation.

[1] Sample number shown for clinical samples: biopsy, Baseline and HI-BCD. Clone ID shown for PC08 or PC09 cell lines.
[2] Junction ID is denoted as viral disruption number equal to last viral nucleotide before recombination to human sequence. Numbering of HPV sequence is according to Genebank accession numbers NC_001526.1 and X05015 for HPV16 and HPV18, respectively.
[3] Annealing temperature used in touchdown PCR in Table 10.
[4] Volume of 15mM $MgCl_2$ added to PCR reagents in Table 11.

Primer ID	Sample[1]	Junction[2]	Sequence	Annealing[3]	$MgCl_2$[4]
5B J2673 F	Biopsy	J2673	TGAAATTTCTGCAAGGGTCTG	na	na
5B J2673 R	5B		CCCACATCATCCGCTAGATT		
6N J1368 F	Biopsy	J1368	GCACATGCGTTGTTTACTGC	na	na
6N J1368 R	6N		TACCCAGCAGTGGGATTGTT		
6N J2345 F	Biopsy	J2345	TGGTACAATGGGCCTACGAT	55	3
6N J2345 R	6N		ACCTGTGGATGTGCATGTGT		

Table 30 **Continued. Primers used for confirmation of integration or non-contiguous sequences.** Primer ID, sample type, junction name, primer sequences, PCR annealing temperatures and volume of 15mM $MgCl_2$ added to PCR regents is shown. Primer sets with "na" in annealing and $MgCl_2$ could not be successfully produce an amplicon following multiple attempts at optimisation.

[1] Sample number shown for clinical samples: biopsy, Baseline and HI-BCD. Clone ID shown for PC08 or PC09 cell lines.
[2] Junction ID is denoted as viral disruption number equal to last viral nucleotide before recombination to human sequence. Numbering of HPV sequence is according to Genebank accession numbers NC_001526.1, X05015 and X74479 for HPV16, HPV18 and HPV45, respectively.
[3] Annealing temperature used in touchdown PCR in Table 10.
[4] Volume of 15mM $MgCl_2$ added to PCR reagents in Table 11.

3.3.3.7.1 PCR Conditions

All primers used to verify host viral junctions or non-contiguous sequences are shown in Table 30.

All PCRs were optimised using a "Touchdown" protocol. Touchdown PCR is used to increase specificity of primer annealing and involves reducing the annealing temperature each time a PCR cycle is performed. Five micro-litres of sample DNA (10ng/µl) was added to the PCR reagents shown in Table 11. PCR conditions are shown in Table 10. All amplicons were electrophoresed on a 2% (w/v) agarose gel and subjected to DNA sequence analysis (sections 3.3.3.6.1, 3.3.3.6.2 and 3.3.3.6.5).

Where an amplicon could not be produced by PCR, further attempts were made to produce an amplicon by changing the $MgCl_2$ volume and/or increasing or decreasing the annealing temperature.

3.3.3.8 Real-time, quantitative Reverse-Transcription PCR

Real-time, quantitative Reverse-Transcription PCR (RT-PCR), also known as qPCR is a tool to quantify mRNA and used to detect transcripts and detect changes in gene expression. qPCR is based on the principle of PCR, described in section 3.3.1 but it involves the use of a fluorescent dye or probe the fluorescence of which can be quantified in real-time. In this PhD, SYBR Green I fluorescent dye was used. SYBR green is a fluorescent dye that binds to double stranded DNA and is ideal for use in qPCR as it eliminates the need for molecular probes (Morrison et al. 1998). In qPCR, with SYBR green, fluorescence increases during elongation and upon denaturation, fluorescence is diminished. Relative quantification was used and is based on levels of target gene expression versus expression levels of a housekeeping gene. Expression of housekeeping genes can vary between individuals. Two housekeeping genes, TBP2 and HPRT, were used to obtain target gene expression relative to the average of housekeeping gene expression. The target genes amplified were HPV16 E2, E6 and E7.

There are 3 steps in qPCR: firstly, cDNA is generated from mRNA using reverse transcription, as mentioned in section 3.3.3.5.1, cDNA is then subjected to qPCR and finally the data is analysed. CaSki was used as a positive control and to calculate PCR efficiencies for each gene analysed. An RNA negative (water) was used as a negative control.

3.3.3.8.1 Reverse Transcription

As mentioned in section 3.3.3.5.1, RNA is easily degraded and reverse transcription for qPCR was performed under the same aseptic, RNAse free conditions as described in section 3.3.3.5.1. Superscript®III (invirtrogen), is an M-MLV RT enzyme, and was used in reverse transcription for qPCR. Superscript®III is a proprietary mutant of SuperScript® II RT that is active at 50°C and has a half-life of 220 minutes, providing increased specificity. Random hexamer primers were used in reverse transcription to generate cDNA. A random hexamer primer is an oligonucleotide of 6 base pairs that is synthesised randomly and anneals to RNA at random points. The use of a random primer reduces the risk of secondary structures caused by RNA during the process of reverse transcription.

Contaminating DNA can potentially invalidate qPCR thus it was necessary to remove residual DNA from the RNA samples using DNAse treatment. One micro-gram of RNA was treated with DNAse using the method in section 3.3.3.5.1.

Although a DNAse step was performed before reverse transcription, it was still possible that contaminating DNA remained. As described in the next section, primers that annealed to separate exons were used to avoid amplification of contaminating DNA. For amplification of HPV E2, E6 and E7 mRNA, it was not possible to design primers on separate exons because E2, E6 and E7 do not have intron-exon boundaries like human genes. To overcome this, reverse transcriptase negative controls without Superscript®III, were generated for each sample to allow amplification and detection of DNA, rather than cDNA. The qPCR values generated by RT-positive samples were adjusted by deducting qPCR values generated by the RT-negative controls (section 3.3.3.8.3).

Half a microgram of DNAse treated RNA for RT positive reaction and 0.5µg of DNAse treated RNA for RT negative control was added to the reagents for step 1 in Table 31. The sample was incubated at 65ºC for 5min and then placed on ice. The entire sample was added to the reagents in step 2 in Table 31. The sample was incubated at 25ºC for 5min, followed by 50ºC for 50min and heat inactivated by incubation at 70ºC for 15min.

	Reagent	Concentration	Volume µl
Step 1	RNA	0.5µg	5
	Random primers	200ng/µl	1
	dNTP	10mM	1
	RNAse free water		6
	Total volume µl		13
Step 2	First strand buffer	5X	4
	DTT	0.1M	1
	RNAse™OUT		1
	Superscript®III	200U/µl	1
	Total volume		20

Table 31 Reagents used in step 1 and 2 of reverse transcription for qPCR. dNTPS were obtained from Invitrogen™, Life Technologies Ltd , Paisley, UK; random primers were obtained from Sigma Aldrich®, Dorset, UK.

3.3.3.8.2 qPCR

cDNA from experimental samples were diluted 1:10 and cDNA from CaSki was diluted 1:100. To obtain mean values for each sample experimental samples were subjected to qPCR in duplicate and CaSki was subjected to qPCR in triplicate. All qPCR assays were optimised by Dean Bryant, a PhD student in the HPV Research Group at Cardiff University. Housekeeping genes for qPCR were selected by Dean Bryant based on their stability and expression levels described in study of housekeeping gene expression in epidermal keratinocytes (Allen et al. 2008).

Primers (Table 32) that annealed to separate exons were used to avoid amplification of contaminating DNA. HPRT and TBP2 primer sequences were described by Allen at al.

(2008) and Minner and Poumay (2009), respectively. E6 and E7 primers were described by Wang-Johanning et al. (2002); E2 primers were described by Roberts et al. (2008).

LightCycler® Sybr Green I master mix (Roche Diagnostics Ltd, West Sussex, UK) was used for qPCR. Eighteen micro-litres of a mastermix of the reagents in Table 11 was added to a 20 µl LightCycler® capillary. Two micro-litres of diluted cDNA were added to the capillary. The capillaries were placed in a Roche LightCycler® Carousel Based System (Roche Diagnostics Ltd, West Sussex, UK). qPCR conditions are shown in Table 10.

Primer Name	Sequence	AT (ºC)	$MgCl_2$ (µl)
HPRT_F[1]	TGACACTGGCAAAACAATGCA	60	1.6
HPRT_R[1]	GGTCCTTTTCACCAGCAAGCT		
TBP2_F[2]	TCAAACCCAGAATTGTTCTCCTTAT	60	2.4
TBP2_R[1]	CCTGAATCCCTTTAGAATAGGGTAGA		
E6_F[3]	CTGCAATGTTTCAGGACCCA	60	1.6
E6_R[3]	TCATGTATAGTTGTTTGCAGCTCTGT		
E7_F[3]	AAGTGTGACTCTACGCTTCGGTT	62	2
E7_R[3]	GCCCATTAACAGGTCTTCCAAA		
E2_F[4]	AACGAAGTATCCTCTCCTGAAATTATTAG	58	1.6
E2_R[4]	CCAAGGCGACGGCTTTG		

Table 32 Primer sequences for qPCR. Annealing temperature (AT) and $MgCl_2$ volume for each qPCR reaction are shown.

3.3.3.8.3 qPCR analysis

Data from qPCR was viewed using LightCycler® software version 3.5. A limitation in using SYBR Green I is that any double stranded DNA will be detected, even primer dimerisation. To overcome this, a melting curve was plotted by increasing the temperature over the melting temperature of the amplicon produced, at the end of the qPCR cycles, to distinguish amplicons of qPCR from secondary artefacts such as primer dimers (Figure 29). Any data that was not consistent with amplicon melt curves were excluded from further analysis.

[1] Allen et al. (2008).
[2] Minner and Poumay (2008).
[3] Wang-Johanning et al. (2002).
[4]Roberts et al. (2008).

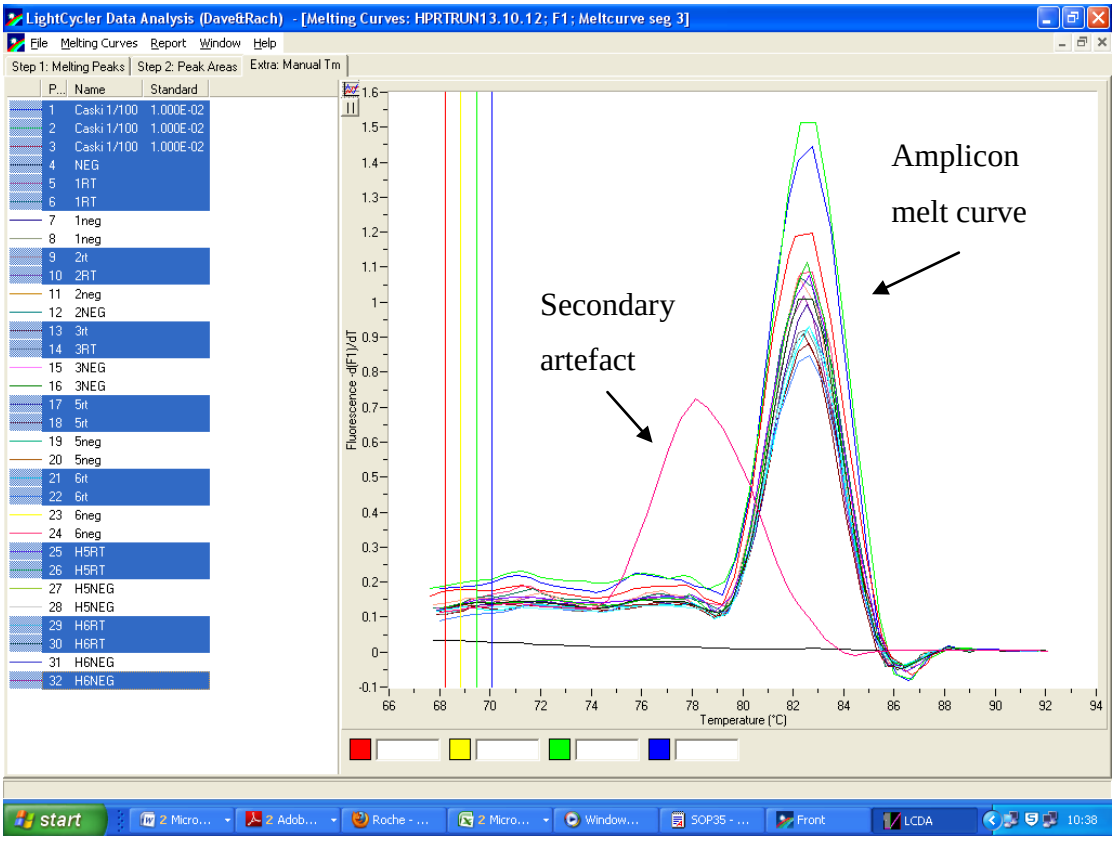


Figure 29 An example of a melting curve. Samples that generated secondary artefacts were excluded from further analysis.

Individual samples generate variable data primarily due to variation in starting material quantity or quality; this may induce errors in quantification. PCR efficiency calculations were included to compensate for sample variability. PCR efficiencies of each housekeeping gene and each target gene were made by Dean Bryant by performing qPCR, as described above on CaSki cDNA diluted 1/3, 1/9, 1/27, 1/81, 1/243, 1/729 and 1/2187. The purpose of a dilution series is to mimic as much as possible the experimental samples that are to be quantified. The data from the dilution series and experimental samples were uploaded on to qBase, software that utilises mathematical models to determine expression of target genes (Hellemans et al. 2007). Default parameters in qBase selected for analysis were: normalisation strategy was set at “reference targets”; and the average cycle quantity (Cq) was the arithmetic mean and target genes were scaled to the Cq average. RT positive and RT negative samples were indicated and cycle threshold (Ct) values from RT negative samples were deducted from RT positive sample by the qBase software. Ct value is defined as the number of cycles required for the fluorescent signal to exceed background levels and is inversely proportional to the amount of nucleic acid in the sample. Normalised relative quantities (NRQ) were calculated for each

sample by qBase. Sample Ct values were normalised to the PCR efficiency of CaSki for each target and housekeeping gene. NRQ calculations were based on qPCR Ct values and PCR efficiency according to Hellemans et al. (2007). qBase also incorporates quality control parameters that consider the stability of expression of the housekeeping genes; if there is variability in the stability of housekeeping gene expression across samples assayed, errors in NRQ calculation could be made. It was therefore essential to observe the stability of housekeeping genes and exclude samples that did not adhere to the housekeeping gene stability values. Gene stability values (M) and mean coefficients of variance (CV) were set at levels according to Hellemans et al. (2007): "mean CV and M values of lower than 25% and 0.5, respectively are typically observed for stably expressed housekeeping genes and in heterogeneous panels, mean CV and M values of lower than 50% and 1 are acceptable".

Furthermore, there was greater stability of housekeeping gene expression in biopsy samples compared to smear samples thus NRQ values were calculated separately for smear and biopsy samples.

3.4 Statistical Methods

For all statistical analyses two tailed tests were applied and the level of statistical significance (α) was 0.05.

3.4.1 Power Calculations

Power calculations were performed for the chi-squared test and for Z tests. Power calculations and power estimations for chi-squared were performed using "pwr" supplementary statistical package for "R". "R" defined variables such as sample size, effect size (denoted as "w"), degrees of freedom, significance level and power. The effect size was calculated using Phi (φ) (Equation 1) or Cramers Phi (φ_c) (Equation 2). For Z tests, power calculations were done using "qnorm" and "pnorm" functions in "R" statistical software.

$$\varphi = \sqrt{\frac{\chi^2}{N}}$$

Equation 1 Phi. Phi is the square root of the chi-squared value divided by the total number of observations (N).

$$\varphi_c = \sqrt{\frac{\chi^2}{N(k-1)}}$$

Equation 2 Cramer's Phi. Cramer's Phi is the square root of the chi-squared value divided by the total number of observations (N) multiplied by the smaller of either the number of rows or columns minus 1 (k-1).

3.4.2 Analysis of Variance

Paired sample T-test and Analysis of Variance (ANOVA) were used when performing assay sensitivity tests in assay development to compare the HPV copy numbers or RNA required for each assay between HPV types. ANOVA and T-tests were done using SPSS statistical package version 16.0.

3.4.3 Chi-squared

Chi-squared distribution was used to compare frequency of integration events between cytology groups and CIN groups. Chi-squared was performed using "R" statistical package. Yates's continuity correction was automatically applied by "R" where sample sizes were small.

Chi-squared "goodness of fit" tests were used to compare the frequency of integration detected at DNA repeat elements under the assumption that repeat elements constitute 70% of the human genome (de Koning et al. 2011) and transposable repeat elements comprise 45% of the human genome (Lander et al. 2001).

3.4.4 Z test

Normal approximation to the binomial distribution, using the Z test, was used to examine the relationship between chromosome fragile sites, HPV disruption sites and HPV integration (Equation 3).

For fragile sites the following assumptions were made: the average size of a fragile site is 1 Mbp (Thorland et al. 2003); half of the human genome (1500Mbp), the light stained G bands, are likely to possess fragile sites (Hecht 1988). There are 120 known fragile sites in the human genome (Durkin and Glover 2007; Lukusa and Fryns 2008) thus the probability of detecting integration in a single fragile site is 0.08.

For integration detection the following assumptions were made: There are a total of 315 entire bands on UCSC database (for example counting 22q11.1, 22q11.2 and 22q11.3 as one band on 22q11) and Wentzensen et al. (2004) detected integration in 118 of them. The probability of detecting integration in one of the same bands as Wentzensen et al. (2004) is 0.37.

For HPV disruption the following assumptions were made: 37% of the HPV genome is either E1 or E2 ORF therefore the probability of detection of an integration event with disruption to E1 or E2 is 0.37; 36% of the HPV genome constitutes either L1 or L2 ORF therefore the probability of detection of an integration event with disruption to L1 or L2 is 0.36.

Statistical significance was determined using the “pnorm” function in “R” statistical software.

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

Equation 3 Normal approximation to Binomial. X is the number of observations, np is the mean number of observations where n is the number of observations and p is the probability or proportion.

3.4.5 Correction for multiple testing

Correction for multiple testing was performed using Bonferonni correction (Equation 4) where n is number of tests and α is significance level.

$$\alpha_i = \frac{\alpha}{n}$$

Equation 4 Bonferroni correction.

3.5 Quality Control

A positive control was used when each assay was done. If a positive control failed to produce an amplicon/s of correct size, the assay was repeated. For E2 PCR, a positive control with un-disrupted E2 and E6 was used alongside clinical samples in each set of PCRs: for HPV16 CaSki DNA was used; for HPV18 and HPV45, plasmid DNA encoding HPV18 or HPV45 was used. For DIPS, APOT and RS-PCR, SiHa, HeLa and HTB-34™ were used for HPV16, HPV18 and HPV45 assays, respectively.

To overcome the risk of error in calling “disrupted” or “intact” E2 when analysing E2 PCR data, all gel electrophoresis images were analysed by me and then checked by Dean Bryant and again by Jo Jones.

4 Assay Validation

Validation aimed to determine if assays:

1. Could detect integration in cell lines equivalent to published data.
2. Could detect integration with presence of episomal HPV.
3. Produced the same result each time the assay was performed.
4. Had comparable sensitivity and could detect integration and/or HPV at low copy numbers.

4.1 Assay Validation Results

4.1.1 E2 PCRs

The E2 PCR assay is a series of PCR reactions that amplify overlapping sections of the E2 ORF of HPV (Figure 30A). Integration of HPV into the genome commonly disrupts E2 and PCR across E2 can determine whether E2 is intact or disrupted. E2 PCR was validated on DNA from W12p12, W12p32, SiHa, HeLa, C4I and HTB-34™ cell lines. Generated amplicons were consistent with the expected amplicon sizes specified in the materials and methods 3.3.3.2 (Table 12). Data for E2 PCR on cell line DNA were comparable to published data for each cell line (Table 33).

E2 was absent in HeLa, C4I and HTB-34™ cells as no E2 amplicons were generated. E6 PCR was used as a positive control for HPV; HPV was present in HeLa, C4I and HTB-34™ because E6 PCR produced amplicons (Figure 56 and Figure 55 in appendices). These data emphasise the importance of E6 PCR as a positive control where integration has removed E2. E2 and E6 were intact in plasmid DNA encoding HPV18 and HPV45 therefore plasmid, encoding HPV18 and HPV45 DNA, were used to optimise and validate HPV18 and HPV45 E2 PCRs.

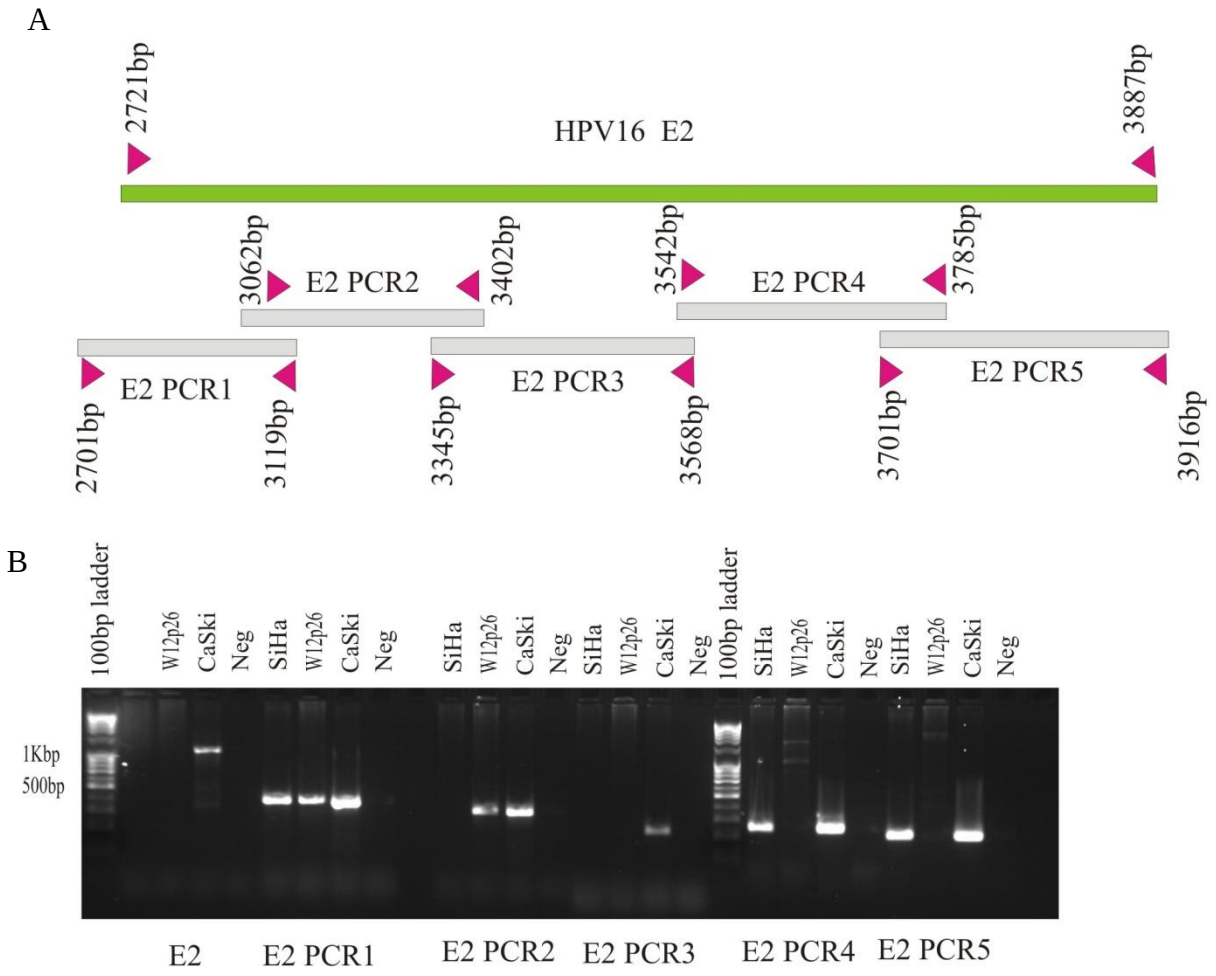


Figure 30 HPV16 E2 PCR validation. E2 is intact in CaSki cells and was used as a positive control. Water was used as a negative control. E2 was disrupted in SiHa and W12 cells**.** A) Not to scale, schematic representation of overlapping PCRs covering the E2 ORF of HPV16; red triangles show the location of primers with positions according to NC_001526.1 in base pairs. B) Electrophoresis of E2 amplicons: CaSki has intact E2 and produced amplicons for all PCRs; SiHa had disrupted E2 with no amplicons for E2, E2 PCR2 and E2 PCR3. E2 PCR results for W12p26 also shown; W12p26 did not have a defined input and was of unknown source and was not used further in this PhD.

DNA	HPV Type	E2 Status	E6 Status
W12p12	HPV16	Intact	Present
W12p32	HPV16	Intact	Present
SiHa	HPV16	Disrupted	Present
CaSki	HPV16	Intact	Present
HeLa	HPV18	Not detected	Present
C4I	HPV18	Not detected	Present
Plasmid	HPV18	Intact	Present
HTB-34™	HPV45	Not detected	Present
Plasmid	HPV45	Intact	Present
C33A	Negative	Not detected[1]	Not detected[1]

Table 33 Summary of E2 PCR validation results for all cell lines and plasmid encoding HPV18 and HPV45 (see Figure 54, Figure 55, Figure 56 in appendices).

4.1.1.1 E2 Disruption Detection in a Background of Episomes

The W12 cell line is derived from a low grade cervical lesion. Published data reports that at low passage, W12 cells contains HPV16 episomes at approximately 100 copies with no integrated HPV detected by Southern blot. Episomes are lost from W12 cell line with increasing passage, at passage 32 no episomes are detected by Southern blot and only integrated HPV16 is present. E2 PCRs on DNA from W12p12 cells did not detect disruption to E2; this is consistent with HPV being in episomal form. However, in section 4.1.3.3 (Table 37), DIPS detected integration in W12p12, with disruption to E2 gene: thus W12p12 harbours both integrated HPV16 and HPV16 in episomal form. These data illustrate that episomal E2 can mask the detection of disrupted E2 by E2 PCR.

4.1.1.2 E2 PCR and HPV copy numbers

To determine the lowest copy number of HPV that could generate amplicons for E2 PCR and E6 PCR, plasmid encoding HPV16, HPV18 and HPV45 DNA at tenfold serial dilutions were added to PCR. Table 34 shows the copy number of HPV that can generate an amplicon by PCR for each of the E2 primer sets and E6 primer set. There is slight variation between types in the copy number of HPV required to produce an amplicon for each set of primers but when mean copy number was compared for each primer set between types there was no significant difference in mass (ANOVA F=0.87, df=2, p=0.45). The true test of sensitivity is measured

[1] C33A and PCR negative included in E2 and E6 PCRs for HPV16, HPV18 and HPV45 and no E6 or E2 was detected for any type.

by the lowest copy number of HPV required at which all primer sets reliably produce an amplicon. The lowest copy number of HPV16 at which E2 and E6 PCRs produce an amplicon is approximately $8x10^7$ copies. The lowest copy number of HPV18 at which E2 and E6 PCRs produce an amplicon is approximately $9x10^7$ copies. The lowest copy number of HPV45 at which all E2 and E6 primer sets produce an amplicon is approximately $8x10^6$ copies. Thus HPV45 E2 and E6 PCRs have 10 times more sensitivity than those for HPV18 and 100 times more sensitivity than those for HPV16. These data show that E2 PCRs are poor in sensitivity. Sensitivity tests for all primer sets were repeated, the number of PCR cycles was increased and the same result achieved.

	Primers					
Type	**Set 1**	**Set 2**	**Set 3**	**Set 4**	**Set 5[1]**	**E6**
HPV16	$8x10^7$	810	$8x10^6$	810	$8x10^7$	810
HPV18	$9x10^4$	900	$9x10^7$	900	na	$9x10^6$
HPV45	$8x10^4$	$8x10^4$	$8x10^6$	$8x10^4$	na	$8x10^6$

Table 34 E2 and E6 PCRs: lowest copy number of HPV DNA that produced PCR amplicons. Na=not applicable.

4.1.1.3 E2 PCR Reproducibility

Reproducibility is defined as the number of times an assay produces the same result. E2 and E6 PCRs were performed on HPV16, HPV18 and HPV45 clinical samples (Chapter 6). CaSki and cloned HPV18 and HPV45 plasmid DNA were used as positive controls due to the presence of intact E2 and E6. HPV16 PCR was performed 13 times and intact E2 was detected in CaSki DNA every time; HPV18 and HPV45 E2 PCRs were done three times, and E2 was intact each time. E2 and E6 PCRs are 100% reproducible.

4.1.2 RS-PCR

RS-PCR was validated on HPV16, HPV18 and HPV45 cell line DNA with known input. RS-PCR generated a number of amplicons for each cell line. As an example, Figure 31A shows how the amplicons are generated by RS-PCR with HPV primers and BamHI. RS-PCR detected integration consistent with published data for SiHa, HeLa and HTB-34™ (Table 35). No integration was detected by RS-PCR in C4I and CaSki.

[1] Note only 5 primer sets were performed for HPV16 and 4 primer sets for HPV18 and HPV45 and so figures for primer set 5 are not given for HPV18 and HPV45.

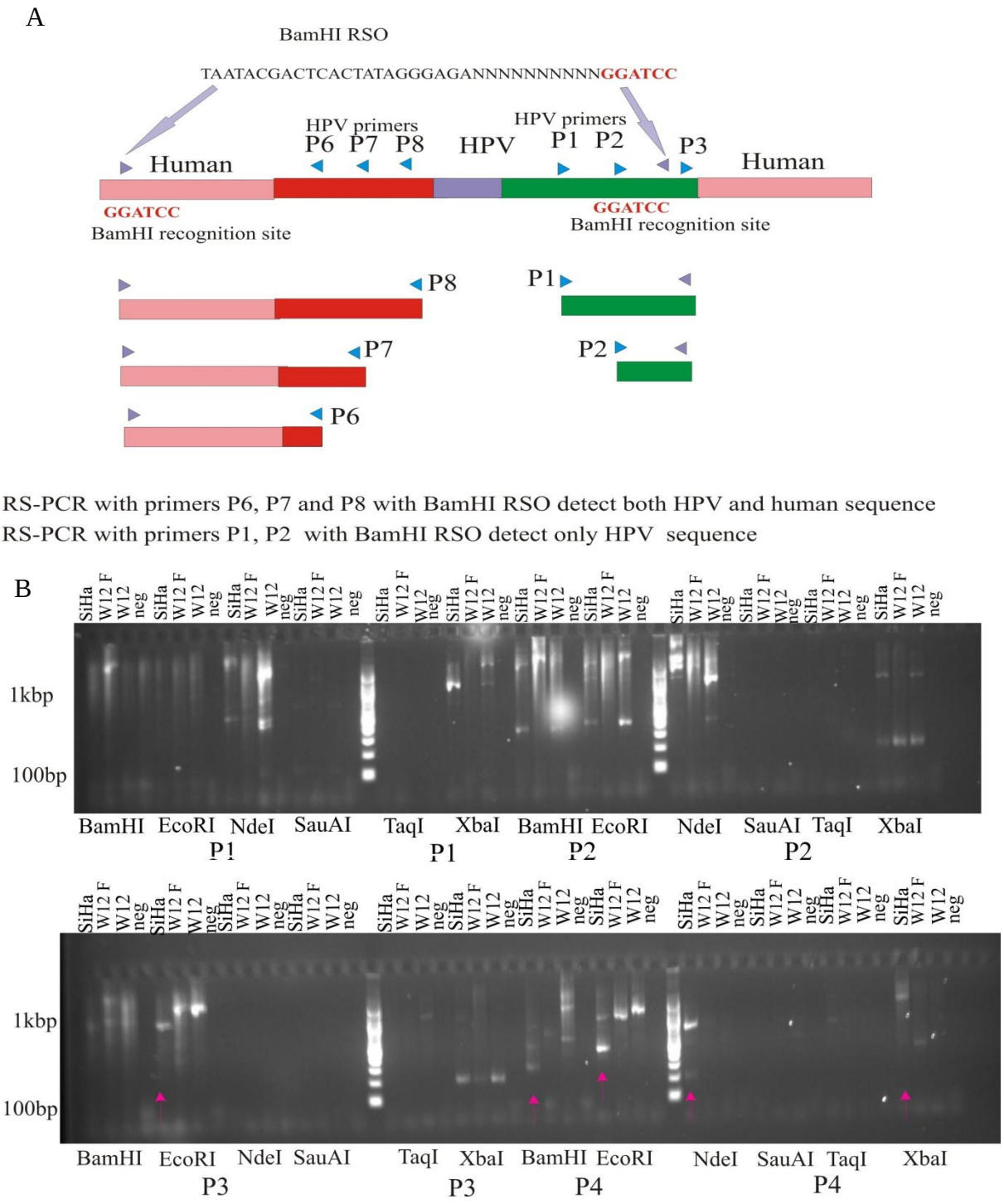


Figure 31 RS-PCR detected integration in SiHa DNA. SiHa was used as a positive control due to known HPV integration status; water was used as a negative control. A) A schematic diagram of hypothetical amplicons produced by RS-PCR: red, green and purple blocks represent HPV late ORFs, early ORFs and URR; red triangles represent HPV primer sets; blue triangles represent BamHI RSO. PCR with HPV primers and BamHI RSO can produce amplicons of different sizes that can detect human DNA adjoining HPV DNA. B) Gel electrophoresis of RS-PCR amplicons with HPV primers P1, P2, P3 and P4 and RSOs BamHI, EcoRI, NdeI, Sau3AI, TaqI and XbaI for SiHa. Uncharacterised W12 (W12F grown with 3T3 feeder cells) was also included in RS-PCR at this stage of the PhD but this clone had unknown input, was not used further assay development stage and results are not included. Pink arrows highlight amplicons with both human and HPV sequence.

Cell line	HPV	Primer[1]	RSO[2]	Viral Disruption[3]	Accession number[4]	Map[5] %	Match[6]	Dir[7]
CaSki	16	P2	BamHI	Non-contiguous 2573 (E1) to 6192 (L1)	HPV only			
		P4		Non-contiguous 3540 (E2) to 6315 (L1)	HPV only			
		P6		Non-contiguous 5610 (L2) to 6563 (L1)	HPV only			
SiHa	16	P3 P4	EcoRI XbaI, BamHI, NdeI, EcoRI	3132 (E2)	NT_024524.14	13q22.1	99%	AS
HeLa	18	P8	BamHI	5736 (L2)	NT_008046.16	8q24.1	99%	AS
C4I	18			Not detected	HPV only			
HTB-34™	45	P2	BamHI, NdeI, EcoRI	1878 (E1)	NT_010966.14	18q11	99%	S

Table 35 Results of RS-PCR in validation on CaSki, SiHa, HeLa, C4I and HTB-34™ cell line DNA. Integration sites detected are shown for each cell line; where a match to human sequence was detected, the accession number is given with percentage consensus and direction in which the HPV sequence is integrated.

[1] HPV specific primer set used in RS-PCR that detected integration.
[2] Restriction site oligonucleotide used in RS-PCR that detected integration.
[3] Viral disruption number denotes last viral nucleotide before recombination to human sequence. Numbering of HPV sequence is according to Genbank accession numbers NC_001526.1, X05015 and X74479 for HPV16, HPV18 and HPV45, respectively.
[4] EMBL Accession number for the NCBI database sequence with the most likeness to human sequence data.
[5] The genomic location of the site of integration, with respect to Giemsa-stained bands, was taken from the UCSC database (Kent et al. 2002).
[6] Percentage consensus of sequence data with NCBI database sequence.
[7] Direction of sequence. AS=antisense orientation, S= sense orientation.

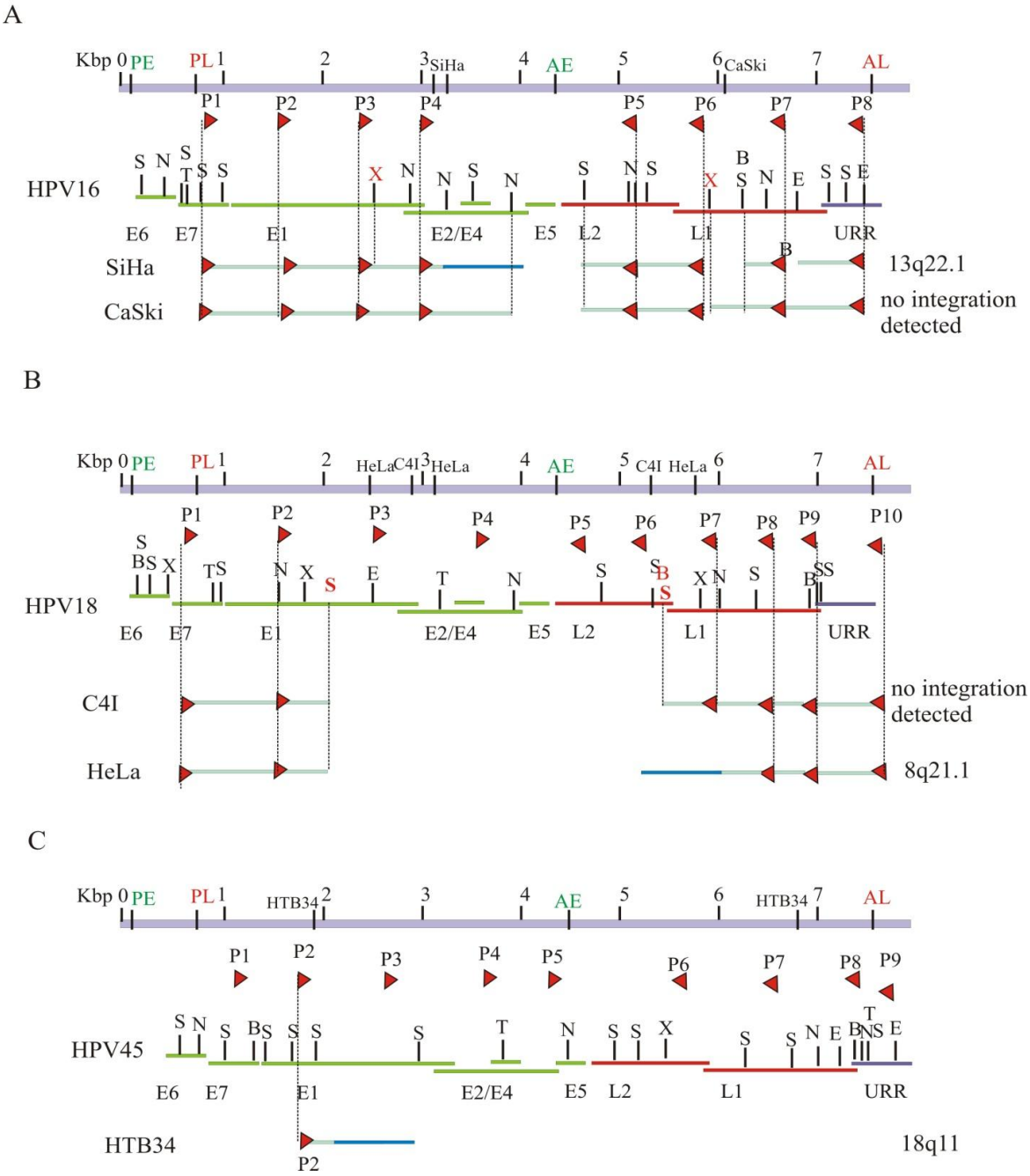


Figure 32 All amplicons produced by RS-PCR were subjected to DNA sequence analysis. Simplified schematic alignment of cell line RS-PCR sequence data to: A) HPV16: NC_001526.1, B) HPV18: X05015, C) HPV45: X74479. Purple bar shows an approximate scale of HPV in Kbp with the point of HPV disruption for each cell line, early promoter (PE), late promoter (PL), early polyadenylation (AE) and late polyadenylation (AL) sites shown. Red triangles show approximate location of RS-PCR primers in relation to HPV genome [green (early genes), red (late genes), dark purple (URR) bars, with BamHI (B), EcoRI (E), NdeI(N), SauAI (S), TaqI (T) and XbaI (X) restriction recognition sites indicated]; red letters are novel, not in the reference sequence. Turquoise bars show HPV DNA sequence data alignment. Integrated, human DNA shown by blue bars. The site of integration is given on right hand side, "no integration" denotes no integration detected. Dotted line denotes either start of sequence data in alignment with primer or end of sequence in alignment with restriction cut-site.

For SiHa, where amplicons were produced (Figure 31B), sequence data aligned to HPV16 (Figure 32). HPV specific primers P3 and P4 detected integration at 13q22.1, with disruption of HPV16 within the E2 ORF. SiHa has a single integration event, present on 13q22.1, and the above data are consistent with this.

CaSki has HPV16 integrated at several sites in tandem repeats, of hundreds of copies. RS-PCR did not detect any host viral junctions in CaSki. HPV DNA sequence that was disrupted and linked to HPV DNA, referred to as non-contiguous sequence, was detected in CaSki (Table 35). Although RS-PCR did not detect integration in CaSki, rearrangements were detected within HPV in CaSki indicating presence of HPV disruption. Failure to detect integration in CaSki with RS-PCR is likely due to CaSki having intact HPV16 integrated in tandem repeats at high copy number; RS-PCR lacks sensitivity to detect integration in this scenario.

HeLa is integrated with three sub-genomic HPV18 fragments on chromosome 8q24.1. RS-PCR in HeLa detected integration of HPV18 in antisense orientation at 8q24.1 with disruption to the L2 ORF (Table 35 and Figure 32). RS-PCR failed to detect the remaining HPV integration events. Sau3AI restriction sites are present in the region of integration so in theory, primers 2 and 3, with positions 1487 and 2412, respectively, should have detected integration with disruption at 2497bp; this shows RS-PCR has limitations to detect integration. The likely explanation for failure to detect disruption at 3100 and 5736 is the position of the HPV18 primers because primers 4, 5, 6 and 7 (nested primers) are specific for HPV sequence located at 3094, 4029, 5035 and 5697, respectively, beyond the sites of disruption.

In C4I, HPV18 is integrated at 8q21-22.3 with disruption at nucleotides 2952 and 5442. RS-PCR detected only HPV18 sequence in C4I and did not detect this integration event. The HPV primer 3 was specific for sequence upstream of 2952 and theoretically should have detected the disruption; HPV primer 7 produced amplicons with BamHI and Sau3AI RSOs and the site of disruption in C4I was beyond the BamHI and Sau3A restriction site (Figure 32). This highlights a limitation of RS-PCR's ability to detect integration: BamHI, EcoRI,

XbaI restriction sites are absent on 8q21-22.3, within the region of integration but Sau3AI and NdeI restriction sites are present; RS-PCR, with primer 3 and RSOs NdeI and Sau3AI, failed to detect integration as did RS-PCR with HPV primer 7 and RSO NdeI.

HTB-34™ does not have a defined integration site in human DNA in the literature but has HPV45 integrated into human DNA with disruption at nucleotides 1888 and 6963. RS-PCR detected integration in sense orientation at 18q11.2 in HTB-34™, within the Retinoblastoma Binding Protein 8 (RBBP8) (Figure 32 and Table 35). HPV45 specific primer set 2 with RSOs BamHI, NdeI and EcoRI detected HPV45 disruption at nucleotide 1878bp within the E1 ORF of HPV45. Primers 3, 4, 5, 6 and 7 did not produce any amplicons; this is due to the location of the primers relative to HPV45 disruption in HTB-34™. Primers 8 and 9 did not amplify any HPV45 sequence nor did they detect disruption at nucleotide 6963. Sau3AI, XbaI and TaqI restriction sites are present in the sequence at the site of integration so in theory these primers with Sau3AI, XbaI and TaqI RSOs should have detected integration; again this highlights the limitation of RS-PCR to detect integration.

4.1.2.1 RS-PCR: Detection of Integration With Episomal HPV

To determine the ability of RS-PCR to detect integration with presence of HPV in episomal form, 100ng SiHa DNA was spiked with varying copy numbers plasmid DNA encoding HPV16. RS-PCR was performed with a primer set that had previously detected integration in SiHa: HPV16 P4 and RSO NdeI. Figure 33 shows the chromatogram of sequence data for 100ng SiHa spiked with $6x10^7$, $1.2x10^7$ and $6x10^6$ copies of plasmid DNA encoding HPV16. SiHa DNA spiked with $6x10^7$ copies of HPV16 had a mixed chromatogram trace of human sequence and HPV16 sequence and beyond "TATGC" the sequence is unreadable, hence uncalled bases are present. SiHa with $1.2x10^7$ copies of background HPV16 has a chromatogram trace of HPV16 and human sequence beyond "TATGC"; the data peaks are low but sequence is readable and when aligned to human DNA using BLAST is a 96% consensus match to NT_024524.14 on chromosome 13q22.1. SiHa with $6x10^6$ copies of background HPV16 has a chromatogram trace of HPV16 and human sequence beyond "TATGC"; the sequence is readable and when aligned to human DNA using BLAST, is a 99% consensus match to NT_024524.14 on chromosome 13q22.1. This data show RS-PCR

can detect integration with presence of HPV in episomal form with approximately 750 copies of HPV episomes for every single copy of integrated HPV16.

Figure 33 RS-PCR can detect integration with presence of HPV episomes. Chromatogram data from RS-PCR of SiHa with varying copy numbers of HPV**.** To the left of the red, dotted line is HPV16 sequence and is the point at which, in SiHa, the HPV DNA links to the human DNA on chromosome 13q22.1. As the number of HPV 16 copies is reduced, the number of uncalled bases decreases and integration is detected.

4.1.2.2 RS-PCR and HPV Copy Numbers

DNA	HPV	RSO	HPV primer	Copy numbers[1]	Cells[2]
SiHa	HPV16	EcoRI	3	Failed	
		BamHI	4	10	5
		EcoRI	4	Failed	
		NdeI	4	1	0.5
		XbaI	4		
HeLa	HPV18	BamHI	8	20,000	400
HTB-34™	HPV45	BamHI	2	unknown	

Table 36 RS-PCR: the number of HPV copies required to detect integration in RS-PCR for HPV16, HPV18 and HPV45.

To determine the lowest number of integrated HPV copy numbers that are required for detection by RS-PCR, cell line DNA was diluted at tenfold serial dilutions ranging from 20ng/µl to 2x10-8ng/µl (Table 36). For SiHa DNA, HPV16 primers P3 and P4 with EcoRI and XbaI failed to produce any amplicons. For SiHa DNA, using HPV primer P4 with BamHI and NdeI RSOs produced an amplicon with 10 and 1 copies of integrated HPV, respectively.

For HeLa, HPV18 primer 8 with BamHI RSO detected integration with 20,000 copies of integrated HPV18.

In HTB-34™, HPV45 primer 8 and BamHI RSO detected integration with 10ng of HTB-34™ DNA. There is no documentation of HPV45 copy numbers in HTB-34™ thus copy number could not be estimated.

4.1.2.3 RS-PCR Reproducibility

Reproducibility is the rate at which the same result was produced each time. RS-PCR for HPV16 was performed 8 times using SiHa DNA as a positive control; on 6 occasions the

[1] HPV copy number calculated relative to estimated mass of DNA per cell and number of copies of HPV per cell for given cell line.
[2] Number of cells required to achieve given copy numbers: SiHa has 2 copies of HPV16 per cell; HeLa has 50 copies of HPV18 per cell.

same outcome was obtained; on two occasions, including the sensitivity test in section 4.1.2.2, RS-PCR failed to produce amplicons using EcoRI and XbaI RSOs. Thus for HPV16, RS-PCR produces the same result 75% of the time. RS-PCR detected integration every time in SiHa. RS-PCR was performed only once for HPV18 and HPV45 and therefore reproducibility not assessed for these types. In terms of integration detection, RS-PCR is 100% reproducible.

4.1.3 DIPS

DIPS amplifies and identifies an unknown sequence of DNA sequence that is adjacent to a known HPV sequence (Figure 34A). DIPS was validated on DNA from HPV16, HPV18 and HPV45 cell lines with known and published input: SiHa, CaSki, W12p12, W12p32, C4I, HeLa and HTB-34™.

4.1.3.1 HPV16 DIPS: Changes to Protocol

First attempts with DIPS using HPV16 primers and the protocol of Luft et al (Luft et al. 2001) on DNA from CaSki and SiHa, failed to yield satisfactory results: HPV16 was detected in CaSki DNA but no HPV16 was detected in SiHa DNA. When the sequence of the primers used in the DIPS protocol by Luft et al (2001) were compared to the sequence data generated by RS-PCR for SiHa, there was variation between the primers and the DNA sequence of SiHa (data not shown); this was indicative of variation in DNA sequence at the sites where the primers were located. The HPV16 primers used in RS detected integration in SiHa DNA thus the HPV16 primers for RS-PCR were used in the DIPS protocol.

DIPS, following the protocol of Luft et al (Luft et al. 2001) on HPV16 and HPV18 cell lines, did not product HPV amplicons consistent with the sizes predicted in Table 25(section 3.3.3.4). Sequencing of the amplicons indicated that star activity was occurring during digestion and ligation. The sequence at the end of the amplicon preceding the adapter primer was not GATC, as would be expected using a Sau3AI digestion, and was not TCG or AGC as would be expected using TaqI digestion. Following consultation with New England Biolabs (NEB), the digestion time was reduced to 2 hours for both Sau3AI and TaqI digests, and the

volume of the Sau3AI digestion was increased to 50µl to reduce the glycerol ratio in order to reduce star activity.

4.1.3.2 DIPS Control PCRs

To confirm the successful digestion and ligation of adapters, amplification of a genomic locus on chromosome 21 (1.4 kbp, accession number ap001068) was performed on the ligation product of each cell line DNA; all cell lines produced an amplicon of 1.4kbp (Figure 34B) and sequence analysis of the amplicon complied with a match to chromosome 21. Thus for all cell lines, DIPS assay was performing optimally.

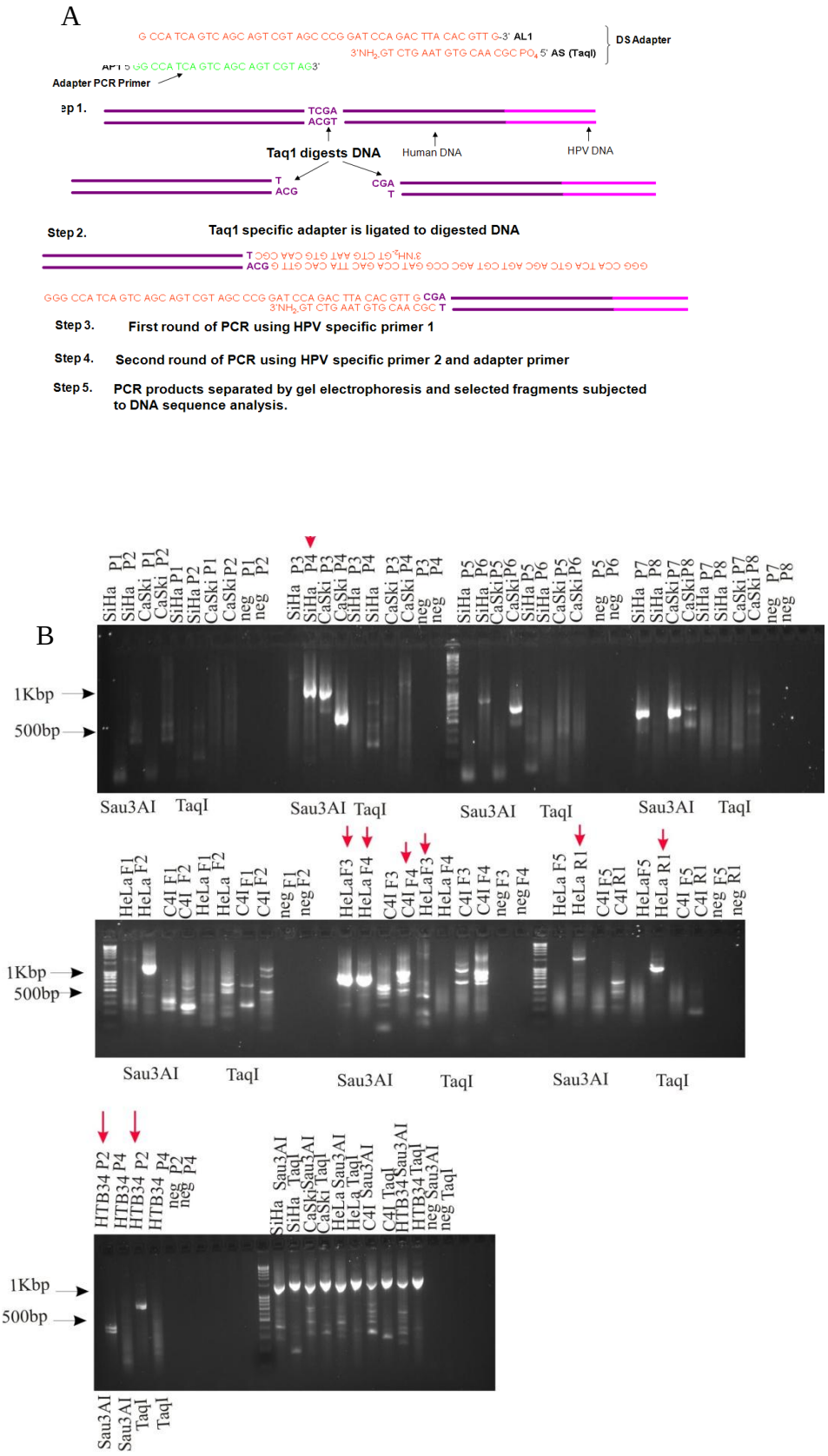


Figure 34 Electrophoresis of DIPS amplicons with a 2 hour restriction digestion time. A) A schematic figure of DIPS adapter primer and steps in the DIPS assay. B) Electrophoresis of SiHa and CaSki amplicons for primers P1 to P8, HeLa and C4I amplicons for primers P1 to P6 and HTB-34™ for primers P2 and P4; control amplicons are shown for SiHa, CaSki, C4I, HTB-34™. Red arrows show amplicons containing integrated sequences.

Cell line	HPV	Primer[1]	Digestion[2]	Viral Disruption[3]	Accession number[4]	Map[5]	% Match[6]	Orientation[7]
CaSki	16	P2	TaqI	1455 (E1)	NT_011681.16	Xq27.3	100%	S
		P4	TaqI	Non-contiguous 2857 (E2) to 6561 (L1)	HPV only			
		P4	TaqI	Non-contiguous 6901 (L2) to 470 (L1)	HPV only			
SiHa	16	P4	Sau3AI	3132 (E2)	NT_024524.14	13q22.1	99%	AS
W12p12	16	P2	TaqI	1756 (E1)	NT_008046.16	8q24.21	99%	AS
	16	P3	TaqI	2749 (E1)	NT_005403.17	2q35	99%	S
	16	P4	TaqI	3197 (E2)	NT_029419.12	12q14.3	99%	AS
W12p32	16	P2	TaqI	1756 (E1)	NT_008046.16	8q24.21	100%	AS
	16	P4	Taq1	3726 (E2)	NT_005403.17	2q31.1	100%	S
	16	P4	TaqI	3197 (E2)	NT_029419.12	12q14.3	99%	AS
HeLa	18	F3	Sau3AI[8] , TaqI	2497 (E1)	NT_008046.16	8q24.1	99%	AS
	18	F4	Sau3AI[8] , TaqI	3100 (E2)	NT_008046.16	8q24.1	97%	AS
	18	R1	Sau3AI[8] , TaqI	5736 (L2)	NT_008046.16	8q24.1	98%	AS
C4I	18	F4	Sau3AI[8]	2952 (E2)	NT_008046.16	8q21.3	99%	AS[8]
HTB-34™	45	P2	Sau3AI[8],TaqI	1878 (E1)	NT_010966.14	18q11	99%	S

Table 37 Validation of DIPS in CaSki, SiHa, W12, HeLa, C4I and HTB-34™. Integration sites detected are shown for each cell line; where a match to human sequence was detected, the accession number is given with percent consensus and direction in which the HPV sequence is integrated.

[1] HPV specific primer set used in DIPS that detected integration.
[2] Restriction enzyme used in DIPS that detected integration.
[3] Viral disruption number denotes last viral nucleotide before recombination to human sequence or viral sequence. Numbering of HPV sequence is according to Genbank accession numbers NC_001526.1, X05015 and X74479 for HPV16, HPV18 and HPV45, respectively.
[4] Genbank accession number for NCBI database sequence with the most likeness to human sequence data.
[5] The genomic location of the site of integration, with respect to Giemsa-stained bands, was taken from the UCSC database (Kent et al. 2002).
[6] Percentage consensus of sequence data with NCBI database sequence.
[7] Direction of sequence. AS=antisense orientation, S= sense orientation.
[8] Integration was detected when digestion incubation time was decreased to 2 hours.

4.1.3.3 DIPS Validation

DIPS detected integration consistent with published data in the DNA of 6/7 cell lines (Figure 35 and Table 37):

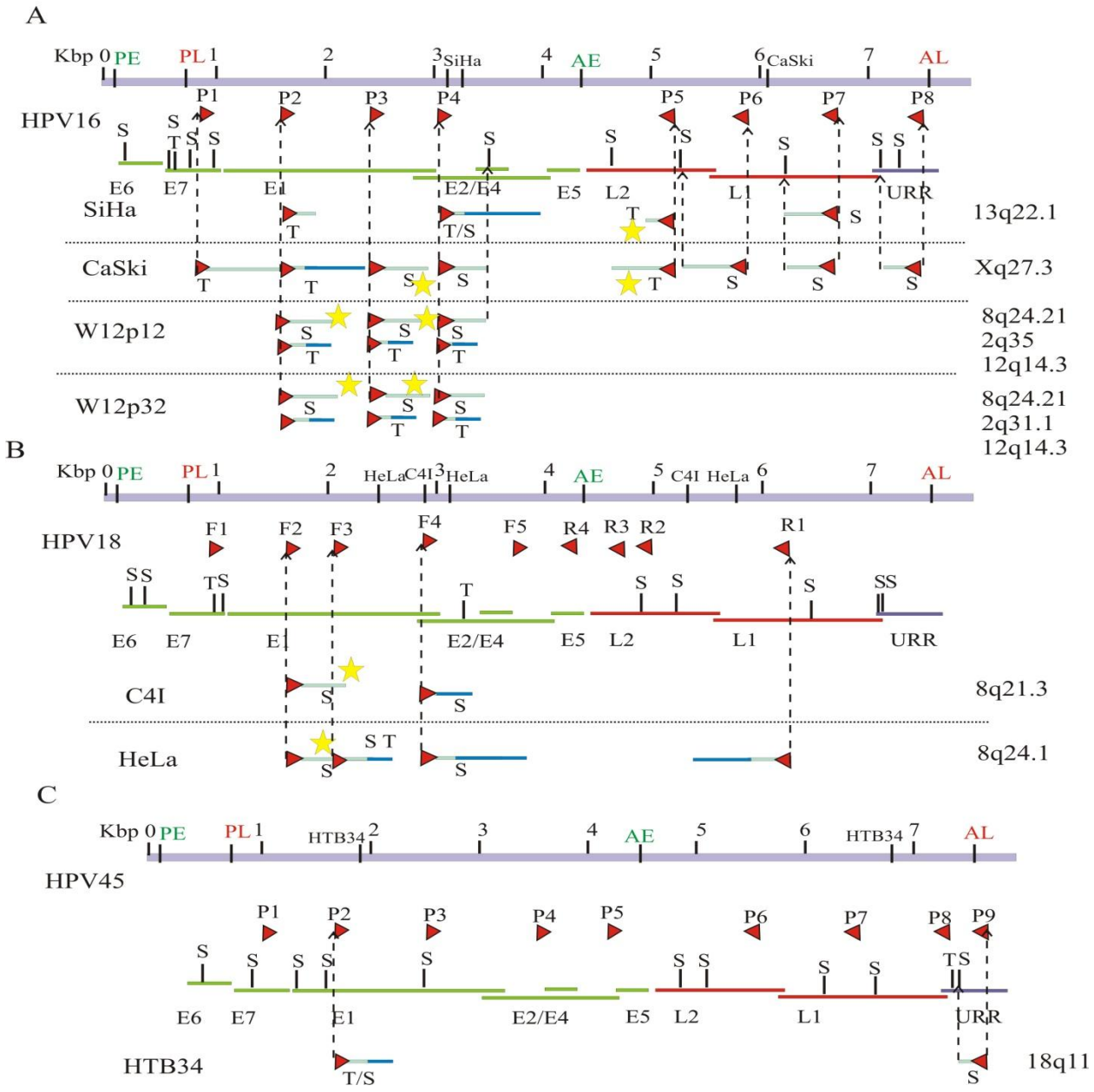


Figure 35 Validation of DIPS on cell line DNA. Schematic alignment of cell line DIPS data to: A) HPV16: NC_001526.1, B) HPV18: X05015, C) HPV45: X74479. Purple bar shows an approximate scale of HPV in Kbp with the point of HPV disruption for each cell line, early promoter (PE), late promoter (PL), early polyadenylation (AE) and late polyadenylation (AL) sites marked. Red triangles show approximate location of DIPS primers in relation to HPV genome in green (early genes), red (late genes), dark purple (URR) bars, with SauAI (S) and TaqI (T) restriction sites indicated. Turquoise bars show HPV DNA sequence data alignment with integrated, human DNA shown by blue bars. Stars show potential star activity in ligation or digestion reaction. The site of integration is given on right hand side, "no integration" denotes no integration detected.

In CaSki DNA, HPV16 is reported to be integrated at a number of sites in the HPV genome in tandem repeats. DIPS did not detect any of the published integration sites. However DIPS, with restriction time of 16 hours, using TaqI restriction enzyme and primers set 1 and 2 detected a novel site of integration in CaSki DNA in sense orientation on Xq27.3 with disruption to the E1 ORF at nucleotide 1455bp. Also, in CaSki, non-contiguous sequence was detected. DIPS HPV specific primer set 4 with TaqI restriction digest detected E2 ORF linked to L1 ORF by a 3bp sequence that was not viral or human sequence. DIPS HPV specific primer set 4 with TaqI restriction digest also detected L1 ORF linked to E6 ORF by two nucleotides that were neither viral or human sequence.

DIPS detected identical integration events in W12p12 and W12p32 DNA: integration of HPV16 at chromosome 8q24.21 and 12q14.3. The detection of integration in W12p12 is novel because reports indicate that early passage W12 is episomal with no integration detected by Southern blot. Additionally, as shown in section 4.1.1, E2 PCRs indicate that E2 is intact whereas DIPS showed that there is disruption to the E2 ORF in W12p12 indicating W12p12 DNA contains both integrated and episomal HPV.

DIPS in HeLa DNA detected integration of HPV18 in antisense orientation at 8q24.1 with disruption to HPV18 at 3 sites. HPV18 specific primer R1 detected HPV18 disruption within the L2 ORF. Disruption of HPV18 at nucleotide 5736 was also detected by RS-PCR; these data show that DIPS and RS-PCR data are reproducible.

In C4I DNA, HPV18 is integrated at 8q21-22.3 with disruption at nucleotides 2952 and 5442. Initially, DIPS detected only HPV18 sequence in C4I and did not detect integration. When the incubation time was reduced to 2 hours DIPS, with primer F4 and Sau3AI digestion, detected integration on 8q21.3 with disruption to E2 at 2952bp.

As mentioned in section 4.1.2, HTB-34™ does not have a published integration site but HPV45 is disrupted at nucleotides 1888 and 6963. DIPS detected integration at 18q11.2 in

sense orientation with HPV45 specific primer set 2; this site of integration was detected by RS-PCR.

4.1.3.4 DIPS: Detecting Integration With Presence of Episomal HPV

DIPS detected integration in W12p12 cell line DNA that has approximately 100 copies of HPV16 episomes per cell thus DIPS is sensitive enough to detect integration where episomes are present.

4.1.3.5 DIPS and HPV copy numbers

Sensitivity is given as the smallest number of integrated HPV copy numbers in which integration consistent with Table 38 was detected by DIPS. For DIPS with a 16 hour digestion time, sensitivity was determined for one primer set for SiHa, HeLa and HTB-34 ™. DIPS for HPV16 was more sensitive than HPV18 because it detected integration with 120 copies of integrated HPV16 in SiHa compared to 60,000 copies of integrated HPV18 for HeLa (Table 38). As mentioned in 4.1.2.2, there is no documentation of HPV45 copy numbers in HTB-34™ thus copy number could not be estimated.

For DIPS with a digestion time of 2 hours, all the primer sets that detected integration (Table 37) in SiHa, HeLa and HTB-34 were used in sensitivity analysis. The number of HPV copy numbers, to logarithm of base 10, needed to detect integration was compared between DIPS, with 2 hour digestion time and between DIPS with 16 hour digestion time; there was no significant difference in mean copy number ($t=0.69$, $df=8$, $p=0.51$; equal variances assumed); thus decreasing digestion time does not increase or decrease the sensitivity of DIPS. The number of HPV copy numbers, to logarithm of base 10, needed to detect integration was compared between SiHa and HeLa with 2 hour digestion time; there a significant difference in mean copy numbers ($t=-3.11$, $df=6$, $p=0.021$; equal variances assumed); this indicates that DIPS for HPV16 is more sensitive and can detect integration with lower copy numbers than HPV18.

DNA	HPV	Digestion	HPV primer	copy numbers[1]	Cells[2]
SiHa	HPV16	16hr Sau3AI	P4	1200	600
HeLa	HPV18	16hr TaqI	P4	60000	1200
HTB-34™	HPV45	16hr TaqI	P2	unknown	
SiHa	HPV16	2hr Sau3AI	F4	120	60
		2hr TaqI	F4	12	6
HeLa	HPV18	2hr Sau3AI	F3	6000	120
		2hr TaqI	F3	6000	120
		2hr Sau3AI	F4	60000	1200
		2hr TaqI	F4	6000	120
		2hr Sau3AI	R1	600	12
		2hr TaqI	R1	600	12
HTB-34™	HPV45	2hr Sau3AI	P2	unknown	
		2hr TaqI	P2	unknown	

Table 38 Smallest number of integrated HPV copies that produced an amplicon with DIPS for HPV16, HPV18 and HPV45 with 16 or 2 hour digestion incubation.

4.1.3.6 DIPS Reproducibility

Reproducibility is defined as the number of times DIPS produced either the same amplicon sizes and/or detected integration. Reducing the digestion time of DIPS to 2 hours from 16 hours increased the reproducibility of DIPS:

DIPS with 16 hour digestion incubation, was performed 10 times. DIPS detected integration every time in SiHa DNA. When observing each amplicon produced with each primer set, the same amplicon sizes were produced 51% of the time. For HPV18, DIPS was done twice on HeLa DNA and integration detected 100% of the time and the same amplicons produced 69% of the time. DIPS for HPV45, was done twice on HTB-34™ DNA; integration was detected and the same amplicons produced each time.

With 2 hour digestion incubation, DIPS was performed 8 times on SiHa DNA; DIPS detected integration each time and the same amplicon sizes were produced 87% of the time. For HPV18, DIPS was done twice on HeLa DNA with integration detected each time and the

[1] HPV copy number calculated relative to estimated mass of DNA per cell and number of copies of HPV per cell for given cell line. Copy number unknown for HTB-34™.
[2] Number of cells required to achieve given copy numbers: SiHa has 2 copies of HPV16 per cell; HeLa has 50 copies of HPV18 per cell.

same amplicons produced 91% of the time. DIPS for HPV45, was done twice on HTB-34™ DNA: integration was detected and same amplicons produced each time.

These data show DIPS is 100% reproducible in terms of integration detection.

4.1.4 APOT

Amplification of Papillomavirus Oncogene Transcripts (APOT) is an assay that allows the discrimination of HPV mRNAs derived from integrated and episomal viral genomes. APOT was validated on RNA from W12p12, W12p32, CaSki, SiHa, HeLa, C4I, HTB-34™ cell lines. Integration was detected in all cell lines consistent with published data and where integration had been detected in DNA by DIPS or RS-PCR, APOT detected integration at the same sites. Integration was detected in DNA either on the antisense strand or the sense strand; APOT detected transcripts mapping to the same strand of DNA. These data illustrate that DIPS, APOT and RS-PCR data are reliable and reproducible.

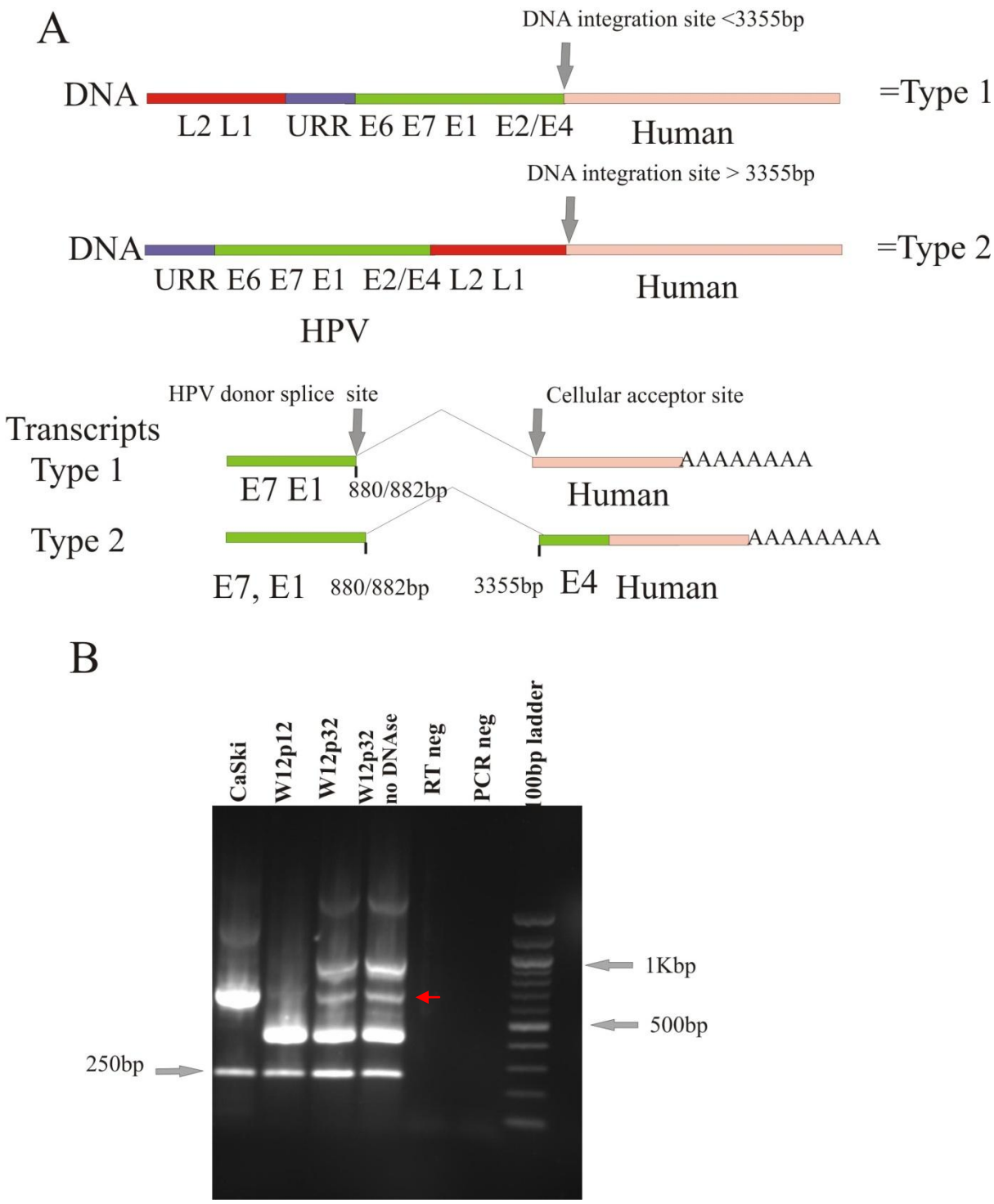


Figure 36 Transcripts detected by APOT in cell lines. A) Schematic diagram of transcripts detected in all cell lines and how they relate to integration in DNA. B) Amplicons produced by APOT for CaSki, W12p12 and W12p32. Different transcript patterns were seen between the cell lines. "250bp" indicates an amplicon produced by annealing of the APOT d(T) oligo to an adenosine rich region in the cDNA; this is not a transcript. The red arrow indicates a transcript that matches HPV16 sequence only.

Cell line	HPV	Transcript Size[1]	splice/fusion site[2]	Accession number[3]	Map[4]	% Match[5]	Orientation[6]
CaSki	16	700	3728 (E4/E2)	NT_007592.15	6p21.1	99%	AS
SiHa	16	300	880(E1)	NT_024524.14	13q22.1	100%	AS
SiHa	16	450	880(E1)	NT_024524.14	13q22.1	97%	AS
W12p12	16	400	880 (E1)	NT_008046.16	8q24.21	100%	AS
W12p32	16	400	880 (E1)	NT_008046.16	8q24.21	100%	AS
W12p32	16	600	E1^E4/E2 only	No human			
W12p32	16	900	3841(E4/E2)	NT_008046.16	8q24.21	99%	AS
HeLa	18	500	929 (E1)	NT_008046.16	8q24.1	97%	AS
HeLa	18	700	929 (E1)	NT_008046.16	8q24.1	99%	AS
C4I	18	400	929 (E1)	No match	8q21.3	100%[7]	AS
C4I	18	450	929 (E1)	NT_008046.16	8q21.3	97%	AS
C4I	18	600	929 (E1)	NT_008046.16	8q21.3	99%	AS
HTB-34™	45	400	931 (E1)	NT_010966.14	18q11.2	100%	S
HTB-34™	45	500	931 (E1)	NT_010966.14	18q11.2	99%	S

Table 39 Transcripts detected by APOT. Accession number of human sequence and orientation of HPV transcripts is shown for each transcript detected by APOT for each of the cell lines.

[1] Approximate transcript size according to DNA ladder in base pairs (bp).
[2] HPV splice site denotes last viral nucleotide before splicing to human sequence or viral sequence. Numbering of HPV sequence is according to Genebank accession numbers NC_001526.1, X05015 and X74479 for HPV16, HPV18 and HPV45, respectively.
[3] EMBL Accession number for the NCBI database sequence with the most likeness to human sequence data.
[4] The genomic location of the site of integration chromosomal reference of all viral-cellular fusion transcripts with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al. 2002).
[5] Percentage consensus of sequence data with NCBI database sequence.
[6] S=sense orientation, AS= Antisense orientation
[7] Percentage consensus of sequence data with UCSC database sequence.

			DIPS				APOT			
Sample	**HPV[1]**	**HPV**	**Map[2]**	**Locus[3]**	**Gene (I/E)[4]**	**Dir[5]**	**Map[2]**	**Locus[3]**	**Gene (E/I)[4]**	**Dir[5]**
CaSki	E1	16	Xq27.3	144775156	na	S	6p21.1	45659122	na	AS
SiHa	E2	16	13q22.1	74087563	na	AS	13q22.1	74031103	na	AS
W12p12	E1	16	8q24.21	128407528	na	AS	8q24.21	128406752	na	AS
	E1	16	2q35	216490303	na	S				
	E2	16	12q14.3	66050135	na	AS				
W12p32	E1	16	8q24.21	128407528	na	AS	8q24.21	128406752	na	AS
	E2	16	2q31.1	173036883	na	S	8q24.21	128451660	na	AS
	E2	16	12q14.3	66050135	na	AS				
HeLa	L2	18	8q24.1	128230632	CCAT (I)	S	8q24.1	128235915	na	AS
	E2	18	8q24.1	128233367	na	AS		128241328	na	AS
	E1	18	8q24.1	128241548	na	AS				
C4I	E2	18	8q21.3	87038854	EST	AS	8q21.3	87014293	na	AS
HTB-34™	E1	45	18q11.2	20604487	RBBP8 (I)	S	18q11.2	20606103	RBBP8 (E)	S

Table 40 Relationship between DNA integration sites detected by DIPS and transcribed integration sites detected by APOT; the exact point of integration in human sequence, in base pairs, is shown along with integration in relation to genes.

[1] ORF of HPV that was disrupted by integration.

[2] The genomic location of the site of integration chromosomal reference of all viral-cellular fusion transcripts with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al. 2002).

[3] Viral disruption number denotes last human nucleotide taken from UCSC database (Kent et al. 2002) before recombination to HPV sequence. Numbering of HPV sequence is according to Genebank accession number NC_001526.1.

[4] Gene at integration site. I= intronic integration; E= exonic integration; EST=expressed sequence tag.

[5] Direction of HPV sequence. AS=antisense orientation, S= sense orientation.

CaSki produced a single host viral transcript of type 2 (Figure 36) consisting of HPV16 E1ORF spliced at 880bp to E2/E4 acceptor site at 3355bp (Table 39). The E2/E4 ORF was fused to human sequence at 6p21.1; this is not the same site of integration detected by DIPS but is consistent with published data.

SiHa produced 2 integrant derived transcripts; both were type 1 and consisted of HPV16 E1 ORF spliced at 880bp to 13q22.1 in antisense orientation (Figure 36). This is the same site of integration as detected by RS-PCR and DIPS and is consistent with transcription of HPV16 E6 and E7 from integrated HPV16 on chromosome 13 in antisense orientation.

Identical transcript patterns were not detected in W12p12 and W12p32 RNA (Figure 36 and Table 39). W12p12 produced a single type 1 transcript consisting of HPV16 E1 spliced to human sequence that mapped to 8q24.21 in antisense orientation. W12p32 produced 3 transcripts: the 400bp transcript was identical to the transcript detected in W12p12, with a consensus of 99% identity. The transcript of 900bp consisted of E1 ORF to the E2/E4 ORF acceptor site at 3355bp then fused to human, cellular, sequence that mapped to 8q24.21 in antisense orientation. The transcript of 600bp was entirely HPV16 and consisted of E1 ORF spliced at 880bp to the E2/E4 ORF acceptor site at 3355bp and sequence terminated at 3837bp. Episomes are not present in W12p32 so the 600bp transcript is not an episome derived transcript; it is likely to be a non-integrated transcript derived from the same DNA integration site as the 900bp fragment. The site of 8q24.21 is the same site that was detected by DIPS in both W12p12 and W12p32. DIPS detected disruption to HPV16 at 1756bp, within the E1 ORF, but APOT has detected transcripts containing the E2 and E4 ORF; this is unusual because transcripts of E2 and E4 would be impossible if disruption at E1 were present. Table 40 shows that although both transcripts mapped to 8q24.21, the 600bp transcript in W12p32 mapped to a point downstream (128451660bp) of the integration event detected by DIPS (128407528bp). It is likely that another integration site is present in W12p32 on 8q24.21 but had not been detected by DIPS. The integration data acquired for W12p12 and W12p32 was with a 16 hour incubation time but there was insufficient DNA remaining to repeat DIPS with a 2 hour restriction digestion time to try and locate the undetected integration event on 8q24.1.

HeLa produced 2 type 1 transcripts (Table 39); both consisted of HPV18 E1 ORF of spliced to human sequence mapping to 8q24.1 in antisense orientation; this is the same site of integration as detected by RS-PCR and DIPS.

C4I produced 3 type 1 transcripts; all 3 transcripts consisted of HPV18 E1 ORF spliced to human sequence mapping to 8q21.3 in antisense orientation (Table 39). Note that C4I did not match NCBI database sequence data when BLAST was done but was a match for 8q21.3 when a BLAT alignment was performed. Both transcripts were transcribed from the non-template, coding, strand from the site of integration in the DNA downstream (87038854), within an expressed sequence tag (EST) (Table 40).

HTB-34™ produced 2 type 1 transcripts both mapping to an exon of RBBP8 gene. Both transcripts consisted of HPV45 E1 ORF spliced to human sequence at 18q11.2 in sense orientation; this is the same site of integration detected by RS-PCR and DIPS. Transcription in HTB-34™ is occurring on the template strand from the integration site upstream (20604487bp, within RBBP8 intron) with two different transcripts that map to the same point within an exon of RBBP8 (20606103bp).

4.1.4.1 APOT and HPV copy numbers

APOT is an RNA based assay and sensitivity tests cannot be based on copy numbers of HPV in DNA. To determine the lowest possible mass of cellular RNA that can be added to RT-PCR to allow host-viral transcript detection, SiHa RNA was diluted in tenfold serial dilution so that between 1µg and 0.01ng of RNA would be added to RT-PCR. APOT PCRs were then performed on cDNA generated from the RT-PCR with diluted RNA; the lowest possible amount of RNA that could produce a host-viral transcript was 0.1ng; this is approximate to 1/25 of a SiHa cell total RNA. These data illustrate that RT-PCR is sensitive enough to allow detection of APOT transcripts on very low starting quantities of RNA.

4.1.4.2 APOT Reproducibility

APOT was performed 12 times on SiHa and 10 times on CaSki; reproducibility was 100% for detection of integration and the same amplicon sizes were produced each time. These data show that APOT data is reproducible.

4.1.5 Confirmation of integration by PCR

Integration sites detected in cell lines were confirmed by PCR and sequence analysis using primers that flanked integration sites. A novel site of integration was detected in CaSki on Xq27.3; this was confirmed by PCR and by DNA sequence analysis (Figure 37). Integration was detected on 8q24.1, 2q35 and 12q14 in W12p12 DNA, previously reported to have no integration. Figure 37 shows confirmation of integration W12p12 DNA. DIPS detected integration in on 8q24.1, 2q31 and 12q14 in W12p32; PCR with flanking primers confirmed the presence of integration on 8q24.1 and 12q14. These data illustrate that in cell lines, where integration is clonal and present in every cell, PCR will confirm integration detection. The integration site on 2q31 could not be confirmed by PCR (data not shown) but DIPS sequence data had 100% consensus when analysed using BLAST (Table 37). Failure to confirm integration on 2q31 in W12p32 DNA can be interpreted in two ways. It may be possible integration on 2q31 is not a true integration site in W12p32 DNA or integration on 2q31 may not be clonal in W12p32.

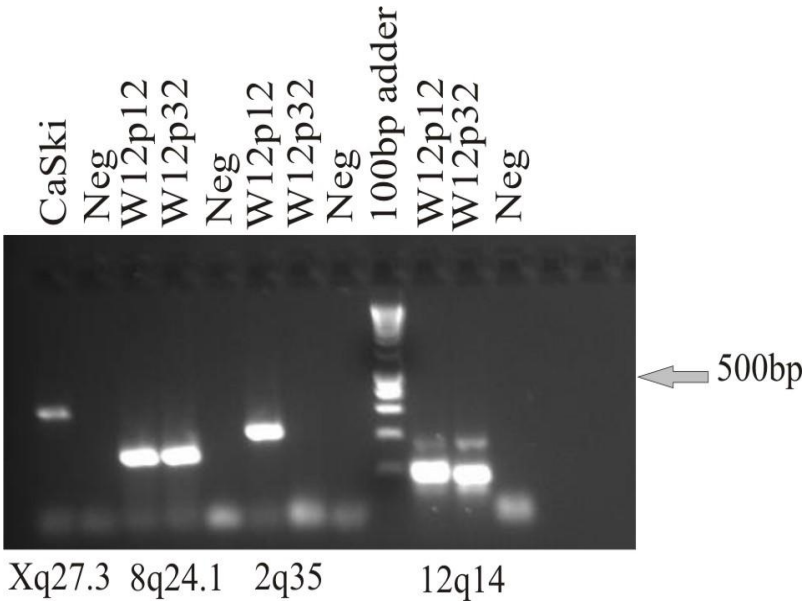


Figure 37 Confirmation of novel integration sites by PCR with primers flanking integration sites. Electrophoresis of amplicons produced with PCR using primers that flanked integration sites. Note only novel integration events are shown: Xq27.3 in CaSki and 8q24.1, 2q35, 12.14 in W12p12.

4.1.6 Quality control

Cell line DNA and RNA had low levels of protein and co-purified contaminants and had ratios consistent with those specified in section 3.2.7. All cell line DNA was intact and not degraded because a single band of 15–30kbp was observed and smears were not observed when DNA was electrophoresed on a 0.8% fine pore gel. All cell line DNA was positive for Beta-globin PCR thus was of good quality for PCR based assays. All cell line RNA had an RNA integrity value of 10 thus the RNA was intact and not degraded.

E2 PCRs, RS-PCR, DIPS and APOT did not produce amplicons or detect HPV16, HPV18 or HPV45 DNA sequences in the HPV negative cell line C33A thus there was no non-specific binding of the HPV primer sets to human DNA. PCR negative controls generated no amplicons and no contaminating DNA was present in any of the PCRs.

4.2 Assay Validation: Discussion

Before integration assays could be applied to clinical samples it was vital to understand assay strengths and weaknesses and determine assay capability to detect integration. Clinical

samples are highly variable since HPV levels are likely to differ considerably between samples. For example, a single copy of integrated HPV may be present in a single cell from cervical cancer whereas hundreds of copies of HPV are likely to exist in a productive infection. It was paramount to learn how sensitive the assays were and due to the complex nature of the assays, it was essential to examine the reliability of the assays in producing the same outcome each time.

4.2.1 Integration Detection

Cell lines were selected on their commercial availability and a number of cell lines were used for validation of HPV16, HPV18 and HPV45 assays.

For E2 PCR, detection of integration was good where episomes were not present and E2 PCRs produced data on CaSki, SiHa, W12p12, W12p32 and HeLa DNA consistent with Collins et al. (2009). E2 PCRs were not done on C4I by Collins et al. (2009) nonetheless E2 PCR data was consistent with Cannizzaro et al. (1988) and Luft et al. (2001). E2 PCR, developed for HPV45, used novel primer sets and produced data comparable to Geisbill et al (1997). DIPS detected integration, with disruption to E2, in W12p12 but E2 PCR did not detect it; this was due to the presence of episomal HPV. These data show that E2 PCR has limited capabilities at detecting integration where episomes are present. Furthermore, CaSki has HPV integration on chromosomes 2, 3, 6, 7, 11, 12, 14, 20, and 21, without disruption to E2 (Mincheva et al. 1987; Van Tine et al. 2004). It remains likely that integration can exist without disruption to E2; E2 PCR would not be useful at detecting integration in this scenario.

RS-PCR detected integration in SiHa, HeLa and HTB-34 but not in CaSki or C4I indicating RS-PCR does have limitations at detecting integration. DIPS detected integration in all the cell lines studied and, with the exception of CaSki, integration sites detected were consistent with published data summarised in section 3.1.1. DIPS did not detect integration on chromosomes 2, 3, 6, 7, 11, 12, 14, 20, and 21 in CaSki but DIPS detected a novel site of integration on Xq27. This integration event was confirmed by PCR using flanking primers thus the integration event is clonal and not an artefact of the DIPS assay. A reason for failing

to detect integration in C4I by RS-CR can be attributed to location of the primers relative to HPV disruption. HPV disruption is random but is more likely to be situated within the E1 or E2 ORF. Primers that are situated after the site of disruption will yield no amplicons and so integration detection may be missed. Also, the nature of HPV at the site of integration may prevent the detection of integration. In CaSki, HPV is integrated in concatenated form, with copy numbers of over 200 at each integration site. Integration in CaSki was originally detected using FISH (Mincheva et al. 1987) and the fact that HPV is concatenated in CaSki would allow easy detection of integration by FISH. DIPS and RS-PCR failed to detect reported integration sites in CaSki because high copy numbers of concatenated integrated HPV can mask the detection of integration. Failure to detect integration does not mean that integration is not present; this needs consideration when interpreting data from any integration assay.

APOT detected integrated transcripts that mapped to sites consistent with published reports in all cell line RNA. A limitation of APOT is that integration is only detected where integrated HPV is being transcribed. For example, in CaSki RNA, APOT detected integration on chromosome 6 despite integration being reported on chromosomes 2, 3, 6, 7, 11, 12, 14, 20, and 21 (Mincheva et al. 1987; Van Tine et al. 2004). It is possible that integration can be missed by APOT, if an integration site is not being transcribed. Hypothetically in a low grade lesion, integration may be present but transcription of the integration site may be suppressed by HPV in episomal form; in this case APOT would not detect it but a DNA based assay such as RS-PCR or DIPS would. CaSki does not possess HPV in episomal form. The mechanism of transcription suppression in CaSki is methylation, as discussed in section 8.4.1.

4.2.2 Detection of integration with presence of episomal HPV

Characteristically in a productive infection, a lesion possesses episomal HPV at approximately 100 copies per cell. It was hypothesised that integration detection may be hampered in low-grade lesions such as CIN1 due to presence of episomal HPV.

This hypothesis was tested in RS-PCR, by performing a novel experiment and integration can be detected by RS-PCR in the presence of episomal HPV. However, the detection of integrated HPV against episomal HPV16 needs cautious interpretation because, in a

productive infection, where HPV DNA is being replicated in upper epithelial layers, copy numbers maybe more than 750 and integration may not be detected; thus it remains possible that RS-PCR may not detect integration where a productive infection is present.

To explore whether DIPS could detect integration when HPV in episomal form is present, DIPS was performed on W12p12 DNA. W12p12 is a cell line that is representative of a low-grade infection and contains approximately 100 copies of HPV16 per cell (Stanley et al. 1989). Previous studies of W12p12 have used Southern blot to determine integration status and have reported no integration (Pett et al. 2006; Dall et al. 2008; Gray et al. 2010). DIPS detected three integration sites in W12p12, with disruption of E1 or E2 ORFs, all of which were confirmed by PCR using flanking primers. E2 PCR showed E2 was intact indicating presence of episomes. These data illustrate that W12p12 contains both integrated and episomal HPV and DIPS can detect integration with at least 100 copies of episomes per cell. DNA from W12p12 was obtained from cells grown in monolayer. Cells grown in monolayer do not have the same productive infection characteristics of a HPV infection in squamous epithelia and have the phenotype of basal epithelial cells. Episome copy numbers may be considerably less in W12p12 compared to a productive infection thus DIPS may not be sensitive enough to detect integration in a productive infection. These are novel data and this is the first time a study has explored whether DIPS and RS-PCR can detect integration with a background of episomes. To ascertain whether DIPS could detect integration in a productive infection, DIPS was applied to samples taken from low-grade lesions, discussed in Chapter 6.

APOT is an assay that differentiates episomal transcripts from integrated transcripts thus, in theory, is able to detect integration within a background of episomes. When APOT was performed on W12p12 and W12p32, integrated transcripts were detected in both but unexpectedly no episomal transcripts were detected in W12p12. Interpretation of DIPS and APOT data imply that integration, in W12p12, is an early event since W12p12 represents CIN1. These data contradict previously published APOT data (Klaes et al. 1999; Vinokurova et al. 2008) that reported integrated transcripts being present only in CIN3, cervical cancer and not present in CIN1. An explanation for the contradiction in data may be explained by the fact that W12p12 was grown in monolayer. Cells grown in monolayer lack epithelial layers that HPV requires for reproduction, thus episomal HPV, present in W12p12, did not

produce episomal transcripts or E2 transcripts that would normally repress the expression of integrated transcripts. This would explain why integrated transcripts but no episomal transcripts were detected by APOT in W12p12 RNA. Cells grown in raft culture maintain epithelial layers and support the HPV lifecycle. Gray et al. (2010) performed APOT on W12 cells grown in raft cultures from passage 20 to passage 60: both episomal and integrated transcripts were detected at passage 20 and only integrated transcripts were detected beyond passage 20. Gray et al. (2010) reported that raft cultures support expression of episomal transcripts but Gray et al. (2010) did not perform APOT on W12 before passage 20, so it remains unknown at what passage integrated transcripts are detected in W12. To determine whether integration is an early event and being transcribed in CIN1, attempts were made in this PhD to apply APOT and DIPS to samples obtained from varying grades of cervical neoplasia, discussed in section Chapter 6.

4.2.3 Assay sensitivity and reproducibility

Cell line DNA and RNA were used as positive controls every time assays were performed and this allowed assay reproducibility to be calculated. All assays detected integration every time and were therefore 100% reproducible. E2 PCR and APOT produced the same amplicon sizes each time the assay was performed with integration being detected every time. DIPS and RS-PCR produced more variation in amplicon sizes between runs but integration was detected every time. DIPS and RS-PCR are much more labour intensive than E2 PCR and APOT, where PCR preparation is simpler. DIPS and RS-PCR are complex to prepare and involve a total of 18 PCR reactions and 96 PCR reactions per sample, respectively; this may explain variation between runs.

First attempts at DIPS, using the method of Luft et al. (2001), were unsuccessful. Reducing restriction digestion time and increasing the reaction volume for Sau3AI was required to reduce star activity and detect integration in HeLa and C4I. There are a number of studies that have used DIPS making amendments to the protocol: Zeigert et al. (2003) increased the starting mass of genomic DNA from 0.6ng to 1.2ng; Schmitz et al. (2012a) shortened the adapter primer and postulated that adapter storage at -80ºC was paramount. This emphasises that DIPS has limited inter-lab reproducibility and that potential differences in integration reported in the literature could be due to assay development issues.

The same integration sites were detected by different assays and indicates that the assays are reproducible. This is further supported by the fact that integration can be confirmed by PCR using primers that flank integration sites. Integration was not confirmed in all the cell lines since integration on 2q31 was not confirmed in W12p32 DNA. Integration on 2q31 is unlikely to be a result of poor sequence data quality since consensus to human sequence was 100%. Potentially, DIPS can produce false integration sites where digestion and ligation can create HPV DNA adjoining human DNA. It is unlikely that integration on 2q31 is an artefact of the DIPS assay since overlapping sequence at the site of integration was not a match for a TaqI restriction site. It does remain plausible that integration on 2q31 is not clonal in W12p32 cells but was detected by chance; this would explain failure to replicate it. This emphasises that all integration sites may not have the same pathological importance: non-clonal integration sites are not likely to be contributing to disease whereas clonal integration sites are likely to be contributing to disease since each cell possesses the same integration event.

To determine the lowest amount of HPV RNA and DNA required to successfully detect integration, DIPS, APOT and RS-PCR, were applied to serial dilutions of DNA and RNA. E2 PCRs required high copy numbers of HPV to generate amplicons. Nonetheless, data presented here is comparable to that of Collins et al (2009) who reported 10^5 HPV copies were required for PCR success. HPV16 E2 PCRs required more HPV copies to generate an amplicon compared to HPV18 and HPV45; this means that HPV16 data is more likely to be inaccurate where small amounts of HPV16 are present because a PCR is more likely to fail due to lack of DNA, rather than disruption of the E2 ORF. E2 PCR data needs careful consideration because HPV presence in a clinical sample is likely to vary, if a PCR fails due to lack of DNA it can result in a false negative result. Detection of E2 disruption by E2 PCR requires follow-up with a technique such as RS-PCR or DIPS to confirm disruption by obtaining nucleotide sequence.

RS-PCR and DIPS were comparable in HPV copy numbers required to detect integration and both were more sensitive than E2 PCR and data confirms that DIPS and RS-PCR are sensitive and hypothetically likely to detect integration even where as few as ten cells each

containing the same integration events are present in a sample. The difference in sensitivity between E2 PCR and RS-PCR and DIPS is due to the nature of the assays since both RS-PCR and DIPS utilise nested PCR reactions, which is likely to result in higher sensitivity.

For APOT, as little as 1/25 total cell RNA was required for successful amplification of an integrated transcript. Transcription of an integration site is likely to vary between cell lines and between clinical samples; nonetheless APOT is a sensitive method.

4.2.4 Integration and E1 and E2 Disruption

The data presented in this section here support hypothesis 1. Integration disrupted the E1 or E2 ORF in DNA of all the cell lines investigated, with the exception of CaSki. HPV disruption is important to consider because it may affect the transcripts produced. Disruption of the HPV genome before the early polyadenylation site (A_E) results in generation of type 1 integrated transcripts as detected by APOT in SiHa, HeLa, C4I, W12p12 and W12p32 RNA. In CaSki, integration does not disrupt the HPV genome before A_E thus type 2 integrated transcripts were detected by APOT. Type 1 transcripts will not permit the production of E2 protein whereas all HPV early proteins have potential to be translated from a type 2 transcript. Theoretically, a cell with a type 1 transcript will not have any E2 protein resulting in high levels of E6 and E7 expression; this will give the cell a selective growth advantage. For CaSki, E2 is expressed (Figure 47 and Figure 49, in Chapter 6) because HPV disruption is after A_E nonetheless CaSki are derived from cervical cancer and E6 and E7 are expressed. These data highlight different transcript types exist but transformation remains. Therefore, there may be differing mechanisms for transformation between cell lines other than disruption to HPV E1 or E2 due to integration.

4.2.5 Unanswered questions: Assay Cross Reactivity

In a clinical sample from either a smear or biopsy, a mixture of HPV types are likely to be present. This may result in cross-reactivity of PCR primers between different HPV types and reduce assay sensitivity. The cell lines used in this PhD for assay development did not contain multiple types and so DIPS, RS-PCR, APOT and E2 PCR were not tested for cross reactivity. When E2 PCRs were performed on clinical samples, in chapter 6, cross reactivity

was observed. DIPS was performed on clinical samples where multiple types present and cross reactivity was not observed (section 6.4.3.1). In future, it would be an advantage to test assay performance where multiple HPV types are present by spiking cell lines with cloned HPV plasmids of multiple types.

4.3 Conclusion

Overall, RS-PCR, DIPS and APOT detected integration sites in cell lines that were consistent with the literature and E2 detected disruption to the E2 ORF due to integration. RS-PCR, DIPS and APOT detect integration within a background of episomal HPV and so were suitable for use on clinical samples that could potentially have a combination of episomal and integrated HPV. Due to cost and labour implications, DIPS was selected for use on clinical samples to detect integration in DNA because DIPS was less labour intensive, less expensive, used less DNA and involved less sequence data analysis compared to RS-PCR.

5 Assay Validation Part 2: Evaluation in Primary Culture (PC) Cell Lines

The Primary Culture (PC) cell lines were originally derived from VIN3 (PC08) and VAIN3 (PC09) biopsies. Currently the only HPV16 cell line model containing HPV in episomal and integrated forms is the W12 line derived from a cervical lesion. The aim of this research section was to characterise PC08 and PC09 single cell clones by determining site of integration and HPV episome status in relation to early and late passages of single cell clones of the PC0 cell lines and confirm whether these vulval and vaginal primary cell lines are an *in vitro* model of vulval and vaginal neoplasia. Exploring integration and HPV episome loss in VIN3 and VAIN3 would determine if episome loss is a late event associated with VIN3 and VAIN3 (hypothesis 5 in Table 8). E2 PCRs were performed by the author to determine the E2 status, as a marker of episomal presence or absence. DIPS and RS-PCR were undertaken by the author to detect integration. Integration analysis on PC08 and PC09 DNA aimed to test whether integration disrupts E1 and E2 ORFs and gives the cell a selective growth advantage (hypotheses 1 and 2 in Table 8). qPCR was done by Tiffany Onions to quantify E2, E6 and E7 mRNA at each passage to determine how integration influences HPV mRNA expression and cell growth. E2, E6 and E7 quantification would determine if integration disrupts E2, increases E6 and E7 expression, and gives a cell a selective growth advantage (hypothesis 3 in Table 8). Finally, APOT was performed by Tiffany Onions to explore integration events that are transcribed and give the cell a growth advantage. DIPS and APOT data were compared to determine reproducibility of DIPS and APOT at detecting integration

5.1 Assay Validation Part 2: Results

Figure 38 summarises single cell cloning work performed by Tiffany Onions. For PC08, 3 clones survived in culture and were characterised in this PhD. For PC09, 9 clones survived in culture. Three of the fastest growing PC09 clones were characterised in this PhD.

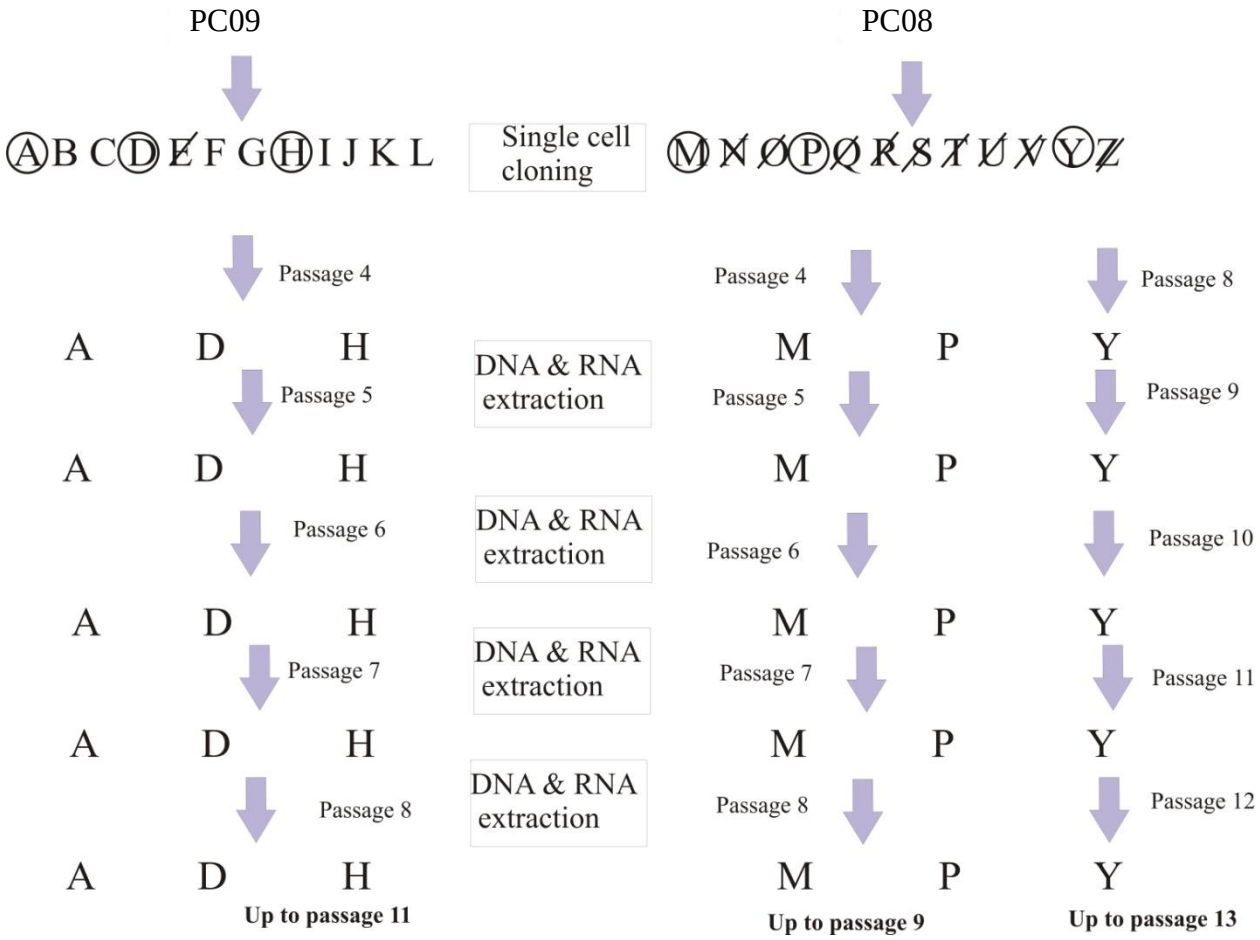


Figure 38 Summary of PC08 and PC09 cloning. Twelve single cell clones were derived from each heterogeneous primary cell (PC) lines PC08 (M to Z) and PC09 (A to L). All were passaged when 90% confluent; DNA and RNA was prepared at each passage up to passage 11, 9 or 13. PC09 clones grew at different rates and single clones that grew rapidly, A, D and H were maintained in culture; clone E was disposed due to contamination. PC08 clones M, P and Y survived in culture and were maintained in culture.

5.1.1 Episomal Status

The aim of this section of work was to determine whether intact E2 was present as a marker of episomal presence. If disrupted E2 was detected then it was assumed that integration was present with no episomes.

E2 PCR detected disrupted E2 in DNA from PC08 clones M and P at early and late passage whereas the Y clone DNA had intact E2 at early and late passage. Heterogeneous PC08 cell DNA line had intact E2; episomal E2 is absent in M and P clones but present in clone Y (Table 41).

Intact E2 was detected in DNA from both early and late passages for PC09 clones A, D and H. Heterogeneous PC09 DNA had intact E2; this indicates episomal E2 may be present in PC09 clones (Table 41).

These data do not support that episomal loss is associated with high grade neoplasia (hypothesis 5) since E2 PCR in both PC08 and PC09 DNA detected intact E2, indicative of the presence of episomal HPV.

5.1.2 PC0 DIPS: Integration in DNA

The data presented here show that each of the PC08 and PC09 clones were unique (Table 41). Integration disrupted E1 or the E2 ORF; these data support hypothesis 1.

DIPS detected differing integration events in DNA of PC08 clones M and Y. Passages 5 and 9 of clone M had disruption to HPV16 within the E1 ORF and integration mapped to 3q28. Clone Y had integration mapping to 3p21.31 with disruption of HPV16 within the E1 ORF, in passage 9 and with disruption within E2 ORF, in passage 13. As mentioned above, E2 PCR on the Y clones showed that intact E2 was present at both early and late passage whereas DIPS showed that there was integration, with disruption to E2 present in the same clone; this indicates HPV in integrated and episomal forms. DIPS also detected non-contiguous sequence in passage 5 of clone M with L2 ORF linked to the URR. No integration was detected by DIPS in P clone but a non-contiguous sequence disrupting E2 ORF was detected; this is novel and interesting because integration generally disrupts E2 ORF but in this case it is breakage of HPV and repair to form a non-contiguous sequence that has disrupted E2.

For PC09, DIPS detected integration in DNA from clones D and H, each with unique sites of integration. In DNA of Clone D, integration was detected at 18p11.3, in passages 5 and 11, with disruption within the L2 ORF. For DNA of clone H, HPV16 integration mapped to 11p15.3 in passage 5 and passage 11 with disruption of HPV16 within the E1 ORF; an additional site of integration on 22q12.3 was detected in passage 11 with disruption to the E1

ORF. In heterogeneous PC09 at late passage 21, DIPS detected integration at 2q36.1 with disruption to L1 ORF. DIPS did not detect integration in the A clone of PC09. DIPS or E2 PCR did not detect disruption to the E2 ORF in any of the PC09 clones. To determine if HPV was present in integrated and episomal form, by detecting disruption to E2 in PC09 clones, RS-PCR, using HPV primer 4 and BamHI, NdeI, EcoRI and XbaI RSOs, was performed. Integration on 18p11.31, with disruption within the E2 ORF was detected in the D clone DNA and integration mapping to 2q36, with disruption to E2 ORF was detected in heterogeneous PC09. These data illustrate that intact and disrupted E2 are present in DNA from heterogeneous PC09 and clone D, with both episomal and integrated HPV present

Clone	Biopsy	Passage	E2 PCR[1]	Primer[2]	Viral Disruption[3]	Accession[4]	Map[5]	Dir[6]
M	PC08	5	disrupted	P5 Taq1	4880 (L2)/7673 (URR)	HPV only	non-contig	
	PC08	5	disrupted	P1Taq1	1194 (E1)	NT_005612.16	3q28[7]	AS
M	PC08	9	disrupted	P1Taq1	1194 (E1)	NT_005612.16	3q28[7]	AS
P	PC08	5	disrupted	P2 Sau3AI	2201(E1)/3147 (E2)	HPV only	non-contig[7]	
P	PC08	9	disrupted	P2 Sau3AI	2201(E1)/3147 (E2)	HPV only	non-contig[7]	
Y	PC08	9	intact	P2 Taq1	2116 (E1)	NT_0022517.18	3p21.31	AS
Y	PC08	13	intact	P4 Taq1	3167 (E2)	NT_0022517.18	3p21.31	AS
Heterogen	PC08	19	intact		Not detected			
A	PC09	5	intact		Not detected	HPV only		
A	PC09	11	intact		Not detected	HPV only		
D	PC09	5	intact	P5 Taq1	5003 (L2)	NT_010859.14	18p11.31	AS
	PC09	5	intact	P4 BamHI	3303 (E2)	NT_010859.14	18p11.31[8]	AS
D	PC09	11	intact	P5 Taq1	5003 (L2)	NT_010859.14	18p11.31	AS
H	PC09	5	intact	P3 Taq1	2490 (E1)	NT_009237.18	11p15.3	S
H	PC09	11	intact	P2 Sau3AI	1848 (E1)	NT_011520.12	22q12.3	AS
				P3 Taq1	2490 (E1)	NT_009237.18	11p15.3	S
Heterogen	PC09	21	intact	P7 Taq1	6033 (L1)	NT_005403.17	2q36.1	S
	PC09	21	intact	P4 BamHI	3632 (E2)	NT_005403.17	2q36.1[8]	S

Table 41 PC0 RS-PCR, DIPS and E2 results. E2, RS-PCR and DIPS data is given for earliest and latest passage of PC08 and PC09 clones and latest passage of heterogeneous PC08 and PC09. E2 PCR outcome (intact or disrupted), site of viral disruption, accession number and site of human integration is shown. All data had good consensus with HPV and/or human database sequence with matches greater than 95%.

1 E2 PCR outcome: disrupted or intact.
2 Restriction enzyme and HPV primer used in DIPS or RSO and HPV primer used in RS-PCR that detected integration.
3 Viral disruption number denotes last viral nucleotide before recombination to human sequence. Numbering of HPV sequence is according to Genebank accession numbers NC_001526.1, X05015 and X74479 for HPV16, HPV18 and HPV45, respectively.
4 EMBL Accession number for the NCBI database sequence with the most similarity to human sequence data. "No human" is where no match to human sequence was observed.
5 The genomic location of the site of integration chromosomal reference of all viral-cellular fusion transcripts with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al . 2002).
6 Orientation of HPV DNA within human sequence. S= Sense orientation, AS=Antisense orientation.
7 Data were detected by reducing DIPS digestion time to 2 hours.
8 Integration detected by RS-PCR.

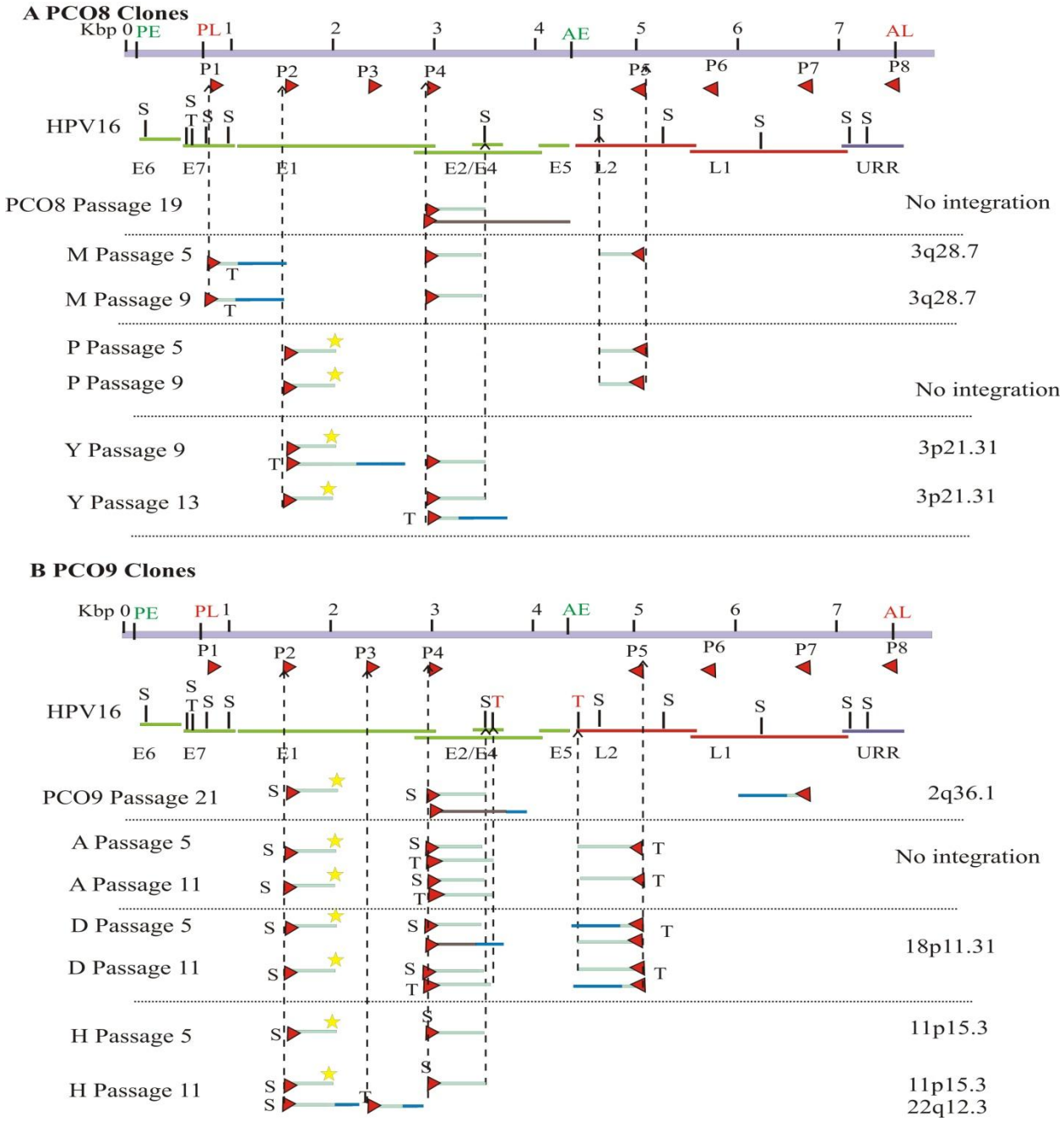


Figure 39 Schematic alignment of PC08 (A) and PC09 (B) DIPS and RS-PCR data to HPV16 (NC_001526.1). Purple bar shows an approximate scale of HPV in Kbp with early promoter (PE), late promoter (PL), early polyadenylation (AE) and late polyadenylation (AL) sites marked. Red triangles show approximate location of HPV primers in relation to HPV genome in green (early genes), red (late genes), dark purple (URR) bars, with Sau3AI (S) and TaqI (T) cut sites indicated. A red S or T denotes restriction site differing to those predicted in NC_001526.1. Turquoise and brown bars show DIPS and RS-PCR HPV DNA sequence data alignment, respectively with integrated, human DNA shown by dark blue bars. Stars show potential star activity in ligation or digestion reaction. The site of integration is given on right hand side, "no integration" denotes no integration detected.

These data do not support the hypothesis that episome loss is a late event in neoplasia (hypothesis 5, Table 8) since both VIN3 and VAIN3 possess episomal HPV. Furthermore, integration was not detected in all of the clones and data do not support the hypothesis that integration gives the cell a selective growth advantage (hypothesis 2, Table 8): there may be a mechanism other than integration that gives a cell a selective growth advantage.

5.1.3 PC0 APOT: Integrated Transcripts

APOT was performed on RNA from PC08 and PC09 clones to determine which integration events were actively transcribed (Figure 57 and Table 42). APOT was not done on heterogeneous PC08 and PC09 because RNA from heterogeneous PC08 and PC09 was not available. APOT data had excellent consensus with DIPS data illustrating that the data for both DIPS and APOT are reproducible (Table 43).

In summary, there were 4 different transcripts types detected: E7 spliced directly to human sequence (type 1), E7 spliced to E4 that was joined to human sequence (type 2), transcripts with E7 spliced to E4 (type 3), and episomal transcripts with E7 spliced to E4 and E5 (type 4) (Figure 57A). Type 3 transcripts do not possess human sequence but are likely to be due to integrated transcripts since there is no polyadenylation site after the E4 ORF. All electrophoretic transcript patterns were identical in early and late passage with the exception of PC09 clone H that gained an extra transcript at late passage and PC08 clone Y that lost a transcript at late passage. These data illustrate that integrated transcripts are identical in RNA between early and late passages and there is no change in the transcription at integration sites with increasing passage.

In RNA from PC08 clone P, type 3 episomal transcripts were detected with E7 spliced to E4 ORF in both early and late passages. RNA from PC08 clone M had a number of transcripts all of transcript type 1 and clone Y had 2 transcripts of type 1 at passage 9; these data are consistent with DIPS data that indicated disruption to HPV genome before A_E at 3355bp (Table 41). PC08 Clone M had a number of transcripts all mapping to 3q28 at different sites (Table 43); this is shown in Figure 57B and Table 43, with a number of amplicons of

differing sizes produced. Clone Y had 2 transcripts in passage 9 mapping to 3p21.31 at 2 different sites and a single transcript mapping to 3p21.31 at passage 13.

At passage 5, RNA from PC09 clone A had type 3 episomal transcripts of E7,E1 spliced to E4. At passage 11, clone A, gained an additional transcript of type 4. PC09 clone D RNA had a single type 2 transcript at early and late passage, mapping to 18p11.31. RNA from clone H had type 2 transcripts mapping to 5q11.2 at early passage and, at late passage, had an extra type 1 transcript mapping to 1p36.13.

Clone	Line	Passage	Viral Disruption[1]	Accession number[2]	Map[3]	% Match[4]	Orientation[5]
M	PC08	5	880 (E7,E1); type 1	NT_005612.16	3q28	100%	AS
M	PC08	9	880 (E7,E1); type 1	NT_005612.16	3q28	100%	AS
P	PC08	5	episomal E7,E1^E4; type 3	NC_001526.1	na	99%	na
P	PC08	9	episomal E7,E1^E4; type 3	NC_001526.1	na	99%	na
Y	PC08	9	882 (E7,E1); type 1	NT_022517.18	3p21.31	100%	AS
Y	PC08	13	882 (E7,E1); type 1	NT_022517.18	3p21.31	100%	AS
A	PC09	5	episomal E7,E1^E4; type 3	NC_001526.1	na	100%	na
A	PC09	11	episomal E7,E1^E4; type 3	NC_001526.1	na	100%	na
		11	episomal E7,E1^E4,E5; type 4	NC_001526.1	na	100%	n
D	PC09	5	3716 (E4); type 2	NT_010859.14	18p11.31	99%	S
D	PC09	11	3716 (E4); type 2	NT_010859.14	18p11.31	100%	S
H	PC09	5	882 (E7,E1); type 1	NT_006713.15	5q11.2	99%	S
H	PC09	11	882 (E7,E1); type 1	NT_006713.15	5q11.2	99%	S
			3494 (E4); type 2	NT_004610.19	1p36.13	99%	AS

Table 42 APOT results for PC08 clones M, P and Y and PC09 clones A, D and H at earliest and latest passage. Data shows HPV splice site adjoining the human acceptor site, the accession number of closest match to human database sequence, and orientation of the sequence.

[1] Viral disruption number denotes last viral nucleotide before recombination to human sequence. Numbering of HPV sequence is according to Genebank accession number NC_001526.1. Episomal detection indicates that only episomal transcripts were detected. Transcript type according to Figure 57A is given.
[2] EMBL Accession number for the NCBI database sequence with the most similarities to human sequence data.
[3] The genomic location of the site of integration chromosomal reference of all viral-cellular fusion transcripts with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al. 2002).
[4] Percentage consensus of sequence data with NCBI database sequence.
[5] Direction of HPV sequence. AS=antisense orientation, S= sense orientation.

			DIPS				APOT			
Clone	**Biopsy**	**Pass**	**Map[1]**	**Locus[2]**	**Gene (I/E)[3]**	**Dir[4]**	**Map[1]**	**Locus[2]**	**Gene (E/I)[3]**	**Dir[4]**
M	PC08	5&9	3q28.7	189689519	LEPREL1(I)	AS	3q28	189684664	LEPREL1 (I)	AS
							3q28	189688682	LEPREL1(E)	AS
							3q28	189681562	LEPREL1(E)	AS
							3q28	189679290	LEPREL1(E)	AS
Y	PC08	9&13	3p21.31	45152730	CDCP1 (I)	AS	3p21.31	45151731	CDCP1 (I)	AS
		13	3p21.31	45165841						
		9					3p21.31	45152297	CDCP1 (E)	AS
D	PC09	5	18p11.31	3462320	TGIF1 (I)	S	18p11.31	3440819	TGIF1 (I)	S
D	PC09	11	18p11.31	3462320	TGIF1 (I)	S	18p11.31	3440819	TGIF1 (I)	S
	PC09	5	18p11.31	3457595	TGIF1 (E)	AS				
H	PC09	5	11p15.3	12166499	MICAL2 (I)	S	5q11.2	52615288	no gene	S
H	PC09	11	11p15.3	12166499	MICAL2 (I)	S	5q11.2	52615288	no gene	
			22q12.3	33736314	LARGE (I)	AS	22q12.3	33733795	LARGE (E)	AS
							1p36.13	16381373	CLCNKB (I)	AS

Table 43 Summary of integration detected in DNA by DIPS and transcripts by APOT for PC08 clones M and Y and PC09 clones D and H. Integration site for both DIPS and APOT data, the locus in base pairs on the chromosome that integration is detected, the gene in which integration occurs and the direction of integrated HPV sequence is shown.

[1] The genomic location of the site of integration chromosomal reference of all viral-cellular fusion transcripts with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al. 2002).
[2] Last human nucleotide, taken from UCSC database (Kent et al. 2002), before recombination to HPV sequence.
[3] Gene at integration site. I= intronic integration. E= exonic integration.
[4] Direction of HPV sequence. AS=antisense orientation, S= sense orientation.

DIPS and APOT produced data that were consistent for all the clones with exception of the H clone (Table 43). DIPS did not detect integration in the P or A clone DNA; APOT data is consistent with this as only episomal transcripts were detected in RNA by APOT. Integration on chromosome 22q12.3 was detected in DNA and RNA by DIPS and APOT for clone H however there were discrepancies: DIPS detected integration on 11p15.3, APOT did not detect this. APOT detected integration on 5q11.2 and 1p36.13 in RNA, DIPS did not detect these in DNA. TaqI and Sau3AI restriction sites are present within 10Kbp of the integration sites detected by APOT on 5q11.2 and 1p36.13; thus the failure of DIPS to detect integration in clone H DNA cannot be explained by lack of TaqI and Sau3AI restriction sites.

5.1.4 Integration Within Genes

These data support the hypothesis that integration occurs within human genes (hypothesis 7) since all integration sites detected by DIPS in the PC0 clones were within genes (Table 43). Integration in clone M DNA at passages 5 and 9 was intronic of leprecan-like gene (LEPREL1) that functions in collagen formation. Integration was in antisense orientation and produced four integrated transcripts (Figure 57B), detected by APOT, that mapped to 4 different sites within the LEPREL1 gene, 3 of which were exonic; human mRNA was in antisense orientation thus transcription was taking place on the coding strand, rather than the template strand.

In clone Y DNA, at passages 9 and 13, integration was intronic of CUB domain containing protein (CDCP1), involved in cell adhesion and cell matrix association. A single integrated transcript was detected by APOT, intronic of CDCP1, in antisense orientation; transcription taking place on the coding strand.

In DNA from clone D, passage 5, a single integration event was detected on 18p11.31, intronic of TGFB-induced factor homeobox (TGIF1), involved in cellular retinoid protein binding. The site of integration detected within DNA differs to the transcribed site detected by APOT. In D clone, the chromosomal location of integration was the same between APOT and DIPS but when the exact location, in base pairs, was compared between DNA and

transcripts the DNA location was downstream from the transcription integration site. Since it is not possible for transcription to be initiated from the integration site downstream; another site of integration must be present although DIPS has failed to detect it. TaqI and Sau3AI restriction sites are present within 10Kbp of the integration sites detected by APOT on 18p11.31 thus the failure of DIPS to detect integration cannot be explained by lack of TaqI and Sau3AI restriction sites. Multiple attempts at DIPS and APOT did not resolve the discrepancies in D clone.

In DNA from H clone passages 5 and 11, DIPS detected integration intronic of microtubule associated monoxygenase, calponin and LIM domain containing gene (MICAL2). Integration in this gene was not detected by APOT thus integrated transcripts are not being produced from this site. DIPS in passage 11 detected integration in an intronic region of like-glycosyltransferase gene (LARGE), involved in glycosylation. An integrated transcript mapping to an exon of LARGE was detected by APOT. APOT detected integration within chloride channel, voltage-sensitive Kb (CLCNKB); DIPS did not detect this integration event.

For heterogeneous PC09, integration was intronic of adaptor-related protein complex 1, sigma 3 subunit gene (AP1S3); this gene encodes a protein involved in the golgi network and endosomes. As mentioned above, no RNA was available for heterogeneous PC09 and APOT was not done.

5.1.5 Confirmation of Integration Sites in PC0s

Integrated transcripts were confirmed by PCR on cDNA using primers that flanked the host viral transcript junction by Tiffany Onions. All integration and non-contiguous events detected by DIPS were confirmed by PCR and sequencing using primers that flanked the host viral junction. To determine whether integration events were present in DNA from PC08 and PC09 clones at early and late, PCR reactions for each integration event and non-contiguous event were performed. For integration events detected in PC08 clone DNA PCR was performed on M, P and Y clones and heterogeneous PC08 (Table 44). PCR reactions for each

integration event detected in PC09 were performed on DNA from early and late passage A, D and H clones and heterogeneous PC09 (Table 45). These data further illustrate that the clones derived from heterogeneous PC08 and PC09 are each unique with different integration sites seen in each:

Integration site: **Disruption:** **Gene**	**3p21.31** **2116bp** **CDCP1**	**3p21.31** **3167bp** **CDCP1**	**L2/URR** **non-contig**	**3q28.7** **1194bp** **LEPREL1**	**E1/E2** **non-contig**
M, passage 5			Y	Y	
M, passage 9			Y	Y	
P, passage 5			Y		Y
P, passage 9			Y		Y
Y, passage 5	Y		Y	Y	
Y, passage 13	Y	Y	Y		
PC08, passage 21	Y		Y	Y	

Table 44 Summary of integration sites detected by PCR on DNA from heterogeneous PC08, passages 5 and 11 of M, P and Y clones using primers flanking integration sites detected by DIPS. Y denotes an amplicon produced.

Integration site: **HPV disruption:** **Gene**	**22q12.3** **1848bp** **LARGE1**	**11p15.3** **2490bp** **MICAL2**	**2q36.31** **6033bp** **AP1S3**	**2q36.31** **3632bp** **AP1S3**	**18p11.31** **5003bp** **TGIF1**	**18p11.31** **3303bp** **TGIF1**
A, passage 5						
A, passage 11						
D, passage 5	Y	Y			Y	Y
D, passage 11		Y			Y	Y
H, passage 5	Y				Y	
H, passage 11	Y				Y	
PC09, passage 19			Y	Y	Y	

Table 45 Summary of integration sites detected by PCR on DNA from heterogeneous PC09, passages 5 and 11 of A, D and H clones using primers flanking integration sites detected by DIPS. Y denotes an amplicon produced.

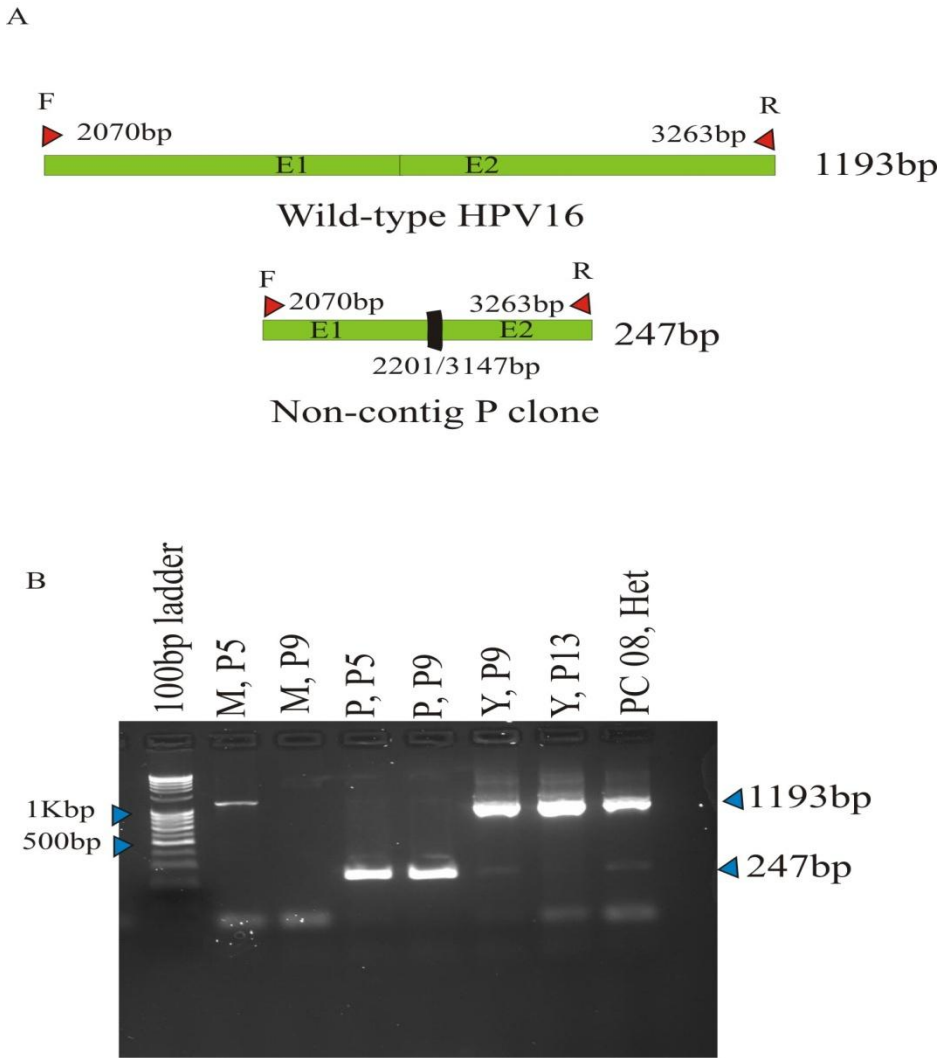


Figure 40 Non-contiguous sequence was confirmed in the P clone: A) A schematic representation of wild-type, un-disrupted HPV and non-contiguous sequence detected in P clone by DIPS; red triangles mark forward and reverse primers and HPV nucleotide positions are according to NC_001526.1. B) Electrophoretic separation of amplicons produced using primers designed to amplify across non-contiguous sequence for early and late passages of M, P, Y and heterogeneous PC08**.**

Non-contiguous sequence, with disruption to E2, was observed in P clones this was confirmed by PCR and by DNA sequence analysis at all passages (passage 5 and 9 shown in Figure 40). M clone at passage 5, Y clones and heterogeneous PC08 had an amplicon size of 1193bp and did not posses non-contiguous HPV with disruption to E2.

For PC08 clone DNA, PCR detected integration on 3p21.31, with disruption at 2216bp, in both passages of Y clones whereas 3p21.31, with disruption at 3167bp, was only present in passage 13 of Y clone. Integration on 3q28.7 was detected in both passages of M clones and

in passage 5 of Y clone but not present in the P clone or in passage 13 of Y clone. L2/URR non-contiguous sequence was present in all clones.

PCR, on DNA from PC09 clones, revealed that clone H, at both passage 5 and 11, had integration sites present on 22q12.3 and 18p11.31.Clone D at passages 5 and 11 had integration on 18p11.31and 11p15.3 and at passage 5 had integration at 22q12.3. Clone A, at both passages, had no integration events detectable by PCR; these data further support absence of integration in A clone.

DNA from late passage heterogeneous PC08 and PC09 bear one or more integration sites that are similar to their cloned counterparts suggesting that these integration events are needed to drive selection and transformation. Integration at 3p21.31, in CDCP1, with disruption at 2216bp was detected in PC08 heterogeneous DNA; this was also detected in clone Y. Integration on 3q28.7, in LEPREL1, was present in both passages of M clones and in passage 5 of Y clone and in heterogeneous PC08 DNA. L2/URR non-contiguous sequence was detected in heterogeneous PC08 and was present in all clones. An integration event in heterogeneous PC09 with disruption to E2 (3032bp) and L1 (6033bp) on 2q36.31 in AP1S3; this is a single integration event with disruption to both ends of HPV being detected. Heterogeneous PC09 had integration on 18p11.31, within TGIF1, present in D and H clones; this shows that integration in TGIF1 and AP1S3 are present in both heterogeneous PC09 and its cloned counterparts indicating that these events are essential for selection and cell growth. Early passage heterogeneous PC08 and PC09 DNA and RNA were not available and any integration events could not be confirmed at early passage in heterogeneous PC08 or PC09.

5.1.6 HPV E2, E6 and E7 Expression and Cell Growth

HPV16 E6, E7 and E2 expression was measured by qPCR relative to CaSki and housekeeping genes HPRT and TBP2 by Tiffany Onions. With the exception of clone D, all PC0s had steady levels of HPV E2, E6 and E7 expression (Figure 58, in Appendix 2). Clone D had small increases in E7 expression with increasing passage, E2 and E6 expression increased 4-fold and 5-fold, respectively, between passage 5 and passage 10. At passage 11,

expression of E6, E7 and E2 decreased to a ratio similar to clones A and H. qPCR was repeated for all D clones by Tiffany Onions and the same outcome observed.

Population doubling (PD) and doubling time (DT) were calculated for heterogeneous cell lines, by Ned Powell, and for each PC08 and PC09 clone by Tiffany Onions. PD refers to the number of times a cell population doubles over a particular time period, and DT refers to the time taken for a cell population to double over a particular time period. Overall there was no increase in PD or DT with increasing passage (Figure 59, Figure 60 and Figure 61 in Appendix 2). There was an increase in DT for clones M and P between passages 1 and 3. For PC09 clones there was no increase in PD or DT with increasing passage.

These data do not support hypothesis 3: Although the pattern of integration and/or episomal presence were different in each clone, HPV E2, E6 and E7 expression was steady. Integration did not result in changes in E2, E6 or E7 expression, with the exception of clone D. These data illustrate that E6 and E7 levels are not rising with increasing passage, but are maintained.

5.2 Assay Validation Part 2: Discussion

VIN and VAIN are chronic conditions that are very difficult to treat and may be recurrent. Previous studies of therapeutic agents for VIN and VAIN have been performed in immortalised cell lines (Andrei et al. 1998). Immortal cell lines do not show the characteristics of a precancerous lesion and the data from transformed cell lines may not be representative of a precancerous lesion. Therefore a novel study developed primary cultures, PC08 and PC09, to determine the viability of precancerous models to study VIN and VAIN.

Primarily, this PhD was an investigation of HPV integration in cervical neoplasia. PC08 and PC09, from vulval and vaginal biopsies, respectively, were included to allow a comparison of DIPS and APOT, to characterise the cell lines in terms of integration and episomal HPV and to gain insight into the role integration plays in anogenital intraepithelial neoplasia.

5.2.1 Assay Validation in PC08 and PC09

When performed on PC08 and PC09, DIPS and APOT had excellent consensus in integration detection and integration sites detected in PC08 and PC09 clones were confirmed by PCR. These data show that DIPS and APOT produce valid integration data that is reproducible.

As part of her PhD project, Tiffany Onions performed deep sequencing on RNA from PC08 clones M and P. Integrated transcripts containing HPV16 and LEPREL1 were detected in clone M RNA. In clone P RNA, no integrated transcripts were detected; again this replicates the findings from this study using another technique. Data from deep sequencing was not yet fully assembled but will be available as a future publication. These data emphasise the reliability of APOT and shows that the data presented in this thesis are supported by other techniques.

Although DIPS and APOT data show reproducibility, there were a few discrepancies in the data. For clone H, there were differences detected and multiple attempts at DIPS and APOT failed to resolve these, highlighting limitations of DIPS and APOT assays in detecting certain integration events. As discussed in section 4.2.1, the location of HPV specific primers in relation to restriction sites and/or the host viral junction at the site of integration can affect the detection of integration by DIPS; it is possible that primer location could explain the failure to detect integration on 5q11.2 and 1p36.13 by DIPS. Furthermore, DIPS detected integration on 11p15.3 but APOT did not detect integrated transcripts from this locus; it is possible that integration is not being transcribed on chromosome 11p15.3 and was not detected by APOT. Integration on 11p15.3 was within an intron of gene MICAL2, between exons 2 and 3. GeneCard database shows that there are 16 splicing variants for MICAL2 and 14/16 RNA species do not possess exons 2 or 3; this means that it is highly likely that integrated HPV is not transcribed due to alternative splicing of the MICAL1 gene. For clone D, both DIPS and APOT detected integration on 18p11.31. When the exact location of integration on chromosome 18 was observed, the integration events detected by DIPS did not correlate with the integrated transcript detected by APOT: the integration site detected by APOT was over 21Kbp upstream from the integration sites detected by DIPS and the DNA sequence orientation of the APOT data was sense. This implies that an integration event may be located

upstream from the transcribed integrant, detected by APOT and was not detected by DIPS in DNA. It also remains possible that genomic rearrangements may have occurred at the locus on chromosome 18p11.31 in the D clone resulting in inconsistency between DIPS and APOT data. This would be a limitation of comparing sequence data to the human genome data on UCSC, rather than a limitation of the assays.

5.2.2 Integration and VIN and VAIN

The observation that only 2 lines survived beyond passage 10 emphasises that producing a cell line from biopsies is very hard to achieve, with less than a 10% success rate. Both PC08 and PC09 were from biopsies with grade 3 neoplasia; this means that the biopsies had already transformed from a productive infection into a precancerous lesion and potentially already had genomic alterations that are commonly seen in cancer which would give them a growth advantage in culture. DNA was extracted from part of the biopsies at the start of the pilot study for the purposes of HPV typing but integration status or human genomic instability was not investigated. Knowledge of integration status and human genomic instability of each biopsy before it was subjected to culture would have given insight into events that occurred *in-vivo.* In turn this would have allowed better understanding of the role of genomic instability and integration status in successful culture of PC08 and PC09.

To gain an understanding of how single cells within each biopsy contributed to grade 3 neoplasia in PC08 and PC09, single cell cloning was performed:

5.2.3 PC08: Vulval Neoplasia

5.2.3.1 Integration and Cell Selection

Single cell cloning of PC08 showed that heterogeneity existed between cells within the biopsy because each clone was unique. The data, for PC08, presented here supported hypothesis 1 because E1 or E2 ORFs were disrupted due to integration. Theoretically, in a high-grade lesion, episomes are lost, integration disrupts E2 or E1, low levels of E2 expression are detected and E6 and E7 expression would be high. In M and P clones episomal HPV was not detected by E2 PCR and integration was present in clone M. Integration was detected in Y clone, alongside HPV in episomal form. Despite being different with regards to

integration status, clones from PC08 displayed little difference in HPV E2 expression, all having extremely low and steady levels of E6 and E7 expression; this does not support the hypothesis that integration results in increased E6 and E7 expression. Also, HPV gene expression analysis was undertaken by Ned Powell on passage 18 of heterogeneous PC08 in the pilot study: patterns of E2, E6 and E7 expression were similar to the cloned counterparts. This indicates that E6 and E7 are unlikely to be required to maintain the transformed phenotype in heterogeneous and cloned PC08. There were also similarities in doubling time and population doubling between PC08 clones when grown in culture. Initially cell doubling time was high and population doubling was low; suggesting that although a different HPV status was detected in each clone, the growth of the cells did not differ between clones. These data indicate neoplasia could not be attributed to differences in integration status or changes in HPV gene expression.

The hypothesis that integration gives a cell a selective growth advantage is not supported by the data in PC08 clones because integration was detected in M and Y clones and not in clone P. Also, there was no apparent growth advantage in the clones possessing the integrants. These data illustrate integration is likely to be influencing cell survival in culture in clones M and Y but not clone P.

Integration was detected in human genes in PC08 clones. In DNA from clones M and Y, integration was within genes. Integration within genes may give the cells selective advantage and offer an alternative pathway facilitating cell growth in culture that is not dependant on HPV E6 and E7 oncogene expression:

Integration in clone M was within an intron of LEPREL1. LEPREL1 otherwise known as P3H2, is a collagen forming protein in basement membranes (Tiainen et al. 2008) that is expressed in the majority of tissues. LEPREL1 expression has been reported to be down-regulated in breast cancer cell lines; the mechanism of down regulation was due to hyper-methylation of CpG islands around exon 1 (Shah et al. 2009). Loss of collagen is an early event in the development of some epithelial cancers (Ikeda et al. 2006). It is possible that

disruption of LEPREL1, due to integration, contributes to oncogenesis and clonal selection in PC08 because the integration event was detected at late passage 21 in heterogeneous PC08.

Integration in LEPREL1 was detected by PCR in passage 9 of clone Y but not detected in later passages of clone Y. Integration in LEPREL1 was not transcribed in the clone Y, and not detected by APOT. Integration was detected by DIPS and APOT in passage 9 and 13 of Y clone within an intron of CDCP1; this indicates integration within CDCP1 is clonal and contributing to cell survival in culture. Integration within CDCP1 may have a stronger selection advantage than LEPREL1 in the clone Y because cells containing integrants in CDCP1 outgrew the cells containing integration in LEPREL1. These data illustrate selection of an integrant in clone Y and may explain why integration in LEPREL1 was not detected in passage 13. CDCP1 is a cell surface glycoprotein and resides within a chromosomal region on 3q that is frequently deleted in a number of cancers. CDCP1 is significantly up-regulated in colon cancer (Scherl-Mostageer et al. 2001) and increased expression correlates with poor prognosis, increased metastasis and relapse rate in lung cancer patients (Ikeda et al. 2009). It is consistently reported in literature that CDCP1 has strong links with cancer and it is plausible that alteration of CDCP1, due to integration, contributes to cell survival in PC08 because integration in CDCP1 was detected by DIPS and APOT in clone Y and in passage 21 heterogeneous PC08.

In DNA from PC08 clone P, no integration was detected by DIPS or APOT. E2 disruption was detected by E2 PCR; this is interesting because non-contiguous HPV sequence was detected and E2 was disrupted by a mechanism other than integration. Non-contiguous sequence was also detected in CaSki (section 4.1.3.3) and reported for CaSki in the literature (Meissner 1999). Little is known or reported about non-contiguous sequence; this phenomenon will be discussed in section 8.4.2. E6 and E7 expression levels were not different to those of M and Y clones and HPV E6 and E7 expression may not contribute to cell survival in clone P. Non-contiguous sequence that disrupted E2 was confirmed by PCR and detected in clone P only and not in passage 21 of heterogeneous PC08; this indicates that non-contiguous disruption of E2 does not contribute to cell survival in PC08. APOT detected episomal transcripts of E7, E1^E4 for clone P; the PC08 cell lines were grown in monolayer and, as mentioned in section 4.2, cells grown in monolayer do not retain the HPV gene

expression characteristics of cells grown in epithelial layers; it remains possible that clone P may exhibit very different APOT transcripts when grown as a raft culture and this requires further exploration in future studies of PC08 clones.

5.2.4 PC09: Vaginal Neoplasia

In PC09, 12 clones were established, all survived in culture and 3 of the fastest growing clones were selected for this study. PC09 clones displayed differences in integration events between clones with integration being detected in D and H and no integration detected in clone A; this suggests heterogeneity within the PC09 biopsy. Where integration was detected, E1 or E2 were disrupted; these data support hypothesis 1.

For A and H clones, HPV E6, E7 and E2 expression was steady between earliest and late passages; this implies that HPV E6 and E7 do not contribute to cell survival in these clones; this does not support the hypothesis that integration results in increased E6 and E7 expression. For clone D, E2 and E6 expression increased between passage 5 and 10 and E7 increased the least. At passage 11 there was a large decrease in E6, E7 and E2 expression; these data suggest that HPV E6, E7 and E2 may contribute to transformation in the clone D but beyond passage 10, selection favours cells with decreased E6, E7 and E2 expression levels. Beyond passage 10 high levels of E6, E7 and E2 expression may no longer be required for cell survival or may even be deleterious to the cell. It is also possible that E6 and E7 expression contributes to genomic instability up to passage 10 and, after passage 10, the DNA is damaged enough to enhance cell survival. These data are novel and it would be interesting in future studies to investigate genomic changes to human DNA that are occurring in clone D up to passage 11. Furthermore these data are supported by an observation reported in W12 grown in raft culture (Gray et al. 2010): E6, E7 and E2 expression levels increased up to a passage where invasiveness was observed and beyond the point of invasiveness, E6, E7 and E2 expression levels decreased. Epigenetic mechanisms such as methylation of the HPV URR may inactivate HPV E6, E7 and E2 expression (Van Tine et al. 2004) and enhance cell survival.

Although differences were detected in integration status and HPV gene expression, cell population doubling and doubling time did not differ between clones or increase with increasing passage; it is likely that since the clones originated from grade 3 neoplasia the cells were already dividing at their peak rate; if the clones were from a lower grade lesion then one may expect to observe an increase in population doubling and doubling time with increasing passage.

Where integration was detected, integration was within human genes; these data support hypothesis 7. In PC09 clones D and H, integration was detected in genes MICAL2, LARGE, TGIF1, CLCNKB and in heterogeneous PC09, integration was detected in AP1S3:

MICAL2 is an enzyme that acts as a cytoskeletal regulator and is involved in axon guidance. The development of the nervous system is an invasive process and it is thought that MICAL2 contributes to invasiveness of cancer and is reported to be over expressed in prostate cancer (Ashida et al. 2006). Integration was detected within MICAL2 in clone D but not in late passage heterogeneous PC09, therefore integration within MICAL2 was present in the original biopsy, but selective pressure in culture allowed the cells containing MICAL2 integrant to be outgrown by cells with a stronger selective advantage. Clonal selection has also been reported in W12 where cells possessing episomal HPV outgrew the cells containing integrated HPV (Gray et al. 2010).

TGIF1 encodes a transcription factor that interacts with ubiquitin ligase to degrade SMAD proteins (Seo et al. 2004). SMAD proteins are tumour suppressors that inhibit cell proliferation and are reported to be down-regulated in colorectal cancer (Xie et al. 2003). HPV16 E6 acts as an ubiquitin ligase and degrades p53. It is possible that integration within TGIF1 disrupts either the gene or the function of the TGIF1 protein. Furthermore, as mentioned above, D clone displayed increased expression of E6 with increasing passage; hypothetically, disruption of TGIF1, due to integration, combined with degradation of SMAD through ubiquitination of TGIF1 protein by HPV16 E6 would provide a selective growth advantage and lead to selection of cells with increased E6 expression and TGIF1 integrants. Integration within TGIF1 was also detected in the heterogeneous PC09 and this event is

likely to be important for cell selection and survival. High levels of E6 expression were detected in passage 21 heterogeneous PC09 when qPCR was performed in the pilot study, thus E6 expression is likely to contribute to cell selection and survival.

CLCNKB encodes a voltage-gated chloride channel protein that is involved in cell volume regulation, membrane potential stabilization and trans-epithelial transportation. Mutations in this gene have been reported to be linked with Bartter's syndrome; a renal condition where excess salt is removed from the body; there are no documented links with cancer. As mentioned above, integration in CLCNKB was detected by APOT in H clone. Primers that flanked the DNA integration site could not be designed and it remains possible that integration within CLCNKB1 resides in A, D and heterogeneous PC09 but it could not be explored further as integration was not detected in the DNA by DIPS.

LARGE, as the name suggests, is the largest gene in the human genome and encodes a protein involved in glycosylation of alpha-dystroglycan. Mutations in LARGE are linked with muscular dystrophy and a large number of muscular dystrophy patients develop cancer. Integration in LARGE was not detected in heterogeneous PC09, this suggests that it did not have a strong clonal selective advantage when grown in a heterogeneous culture.

Integration in AP1S3 was detected in PC09 heterogeneous DNA but not detected in DNA from any of the clones. This is unusual but it is plausible that integration in AP1S3 resided within one or more of the 9 clones (B, C, E, F, G, I, J, K and L) that were not continued in culture. AP1S3 encodes a part of a protein involved in the golgi-network and endosomes and does not have documented links with cancer. APOT was not performed on heterogeneous PC09 so it remains unknown if integration on AP1S3 is transcribed and contributing selection at late passage.

5.2.5 Future work

PC08 and PC09 integration was detected within strong candidate genes for human cancer, however, hypotheses that imply integration is disrupting human gene expression or functions are speculative. For PC08, integration in CDCP1 and LEPREL1 was detected within introns and in antisense orientation and it is unlikely that integration produces a fusion protein containing HPV and CDCP1 or LEPREL1. In PC09, several integration events were detected within introns and exons of TGFI1 in both sense and antisense orientation; the mechanism of disruption to TGFI1 may be difficult to pinpoint. Further work is required to understand how, or if, HPV integration disrupts these genes. It is possible to determine whether integration disrupts gene expression or gene function and this has been performed in cervical cancer biopsies (Schmitz et al. 2012); it would be ideal in future research to determine whether integration disrupts TGIF1, CDCP1 and LEPREL1 by performing gene expression analysis, Western blot and ISH. It would also be interesting to determine whether reduction of E6 in PC09 D clone induces senescence or apoptosis through SMAD pathways.

E2 PCRs were chosen to confirm episomal presence with intact E2 indicative of episomal presence. As mentioned in section 4.2.4, E2 is not disrupted due to integration in all cases, for example in CaSki, integration is present in concatenated form but E2 remains intact. Southern blot on late passage heterogeneous PC08 and PC09 revealed entirely integrated HPV at passage 19 for PC08 and a mixed population of integrated concatenated HPV and episomal HPV for passage 21 PC09. It does remain possible where intact E2 is detected in PC09 clones, integration is in concatenated form. Southern blot is useful in confirming the presence of episomes and determining if integration is present in concatenated form. In future work, Southern blot will be applied to DNA from PC08 and PC09 clones to determine episomal status and if integration is in concatenated form.

5.3 Conclusion

In conclusion, data from DIPS and APOT are reproducible. Clones derived from PC08 and PC09 were unique in terms of integration and there was little difference between early and late passages implying that conditions required for selection were stable. There was evidence

in the PC08 or PC09 clones that integration disrupts E2 or E1, but no evidence to suggest that episomes are lost resulting in increased E6 and E7 expression. Furthermore, there was no evidence to suggest that integration is required for cell survival in culture since integration was not detected in all clones. Where integration was detected in a clone, there was evidence to imply disruption to human genes. Disruption to human genes may give the cell a selective growth advantage in culture and this requires verification in future experiments.

6 Application of assays to Clinical Samples

Hypothetically, integration commonly disrupts E1 and E2 ORFs and loss of episomal E2 leads to increased E6 and E7 expression and carcinogenesis. It is unclear as to whether HPV integration is an early or a late event in cervical carcinogenesis and integration has not yet been evaluated as a biomarker for predicting cervical disease. The aim of this section was to determine if integration is an early event, present in CIN1 (hypothesis 4), and to determine if episomal loss is a late event, present in CIN3 (hypothesis 5). Also qPCR was applied where RNA was available to test the hypothesis that integration results in elevated E6 and E7 expression and reduced E2 expression (hypothesis 3). RS-PCR, E2 PCR, DIPS, APOT and qPCR were applied to clinical samples comprising cervical smears with varying grades of disease and cervical cancer biopsies:

6.1 Liquid Based Cytology Samples: Baseline Results

The "Baseline" samples are a collection of cervical smear samples taken from women aged 20 to 22 years attending their first cervical smear test in Wales. Fifteen thousand samples were collected as part of a previous study to determine the HPV prevalence of HPV types in young women prior to the commencement of HPV vaccine administration. Single smear samples were placed into SurePath™ preservative fluid and cytology data obtained. Follow-up screening was available for a further 2 screening rounds and histology data recorded from all women who attended colposcopy. The aim of this section was to compare HPV integration and E2 as a marker of viral episome presence or absence between different cytological and histological grades to determine whether integration and episome clearance is a key event in the development of cervical cytological abnormality and neoplasia.

6.1.1 DNA Quality

6.1.1.1 DNA Degradation and Surepath™

Where DNA was available (n=55), 15µl of Baseline DNA was electrophoresed onto a fine pore 0.7% agarose gel. All of the samples failed to produce any band although there should have been at least 600ng visible as a single band, similar to that of the positive control in Figure 41. Subsequently, to test the ability of SurePath™ to preserve DNA, SurePath™ preservative fluid was spiked with a known amount of SiHa cells representative of what

would be expected in a clinical LBC specimen ($1x10^4$). If DNA were un-degraded then a single band would be observed as in lane 9 and lane "+" in Figure 41. After 48 hours in SurePath™ the SiHa DNA had degraded substantially with a visible smearing of the DNA electrophoresed on a fine pore gel. When SiHa was stored in SurePath™ for 1 month, SiHa DNA was degraded further (lane 1M in Figure 41). Storage of cervical smear cells and cell line cells in SurePath™ results in DNA degradation.

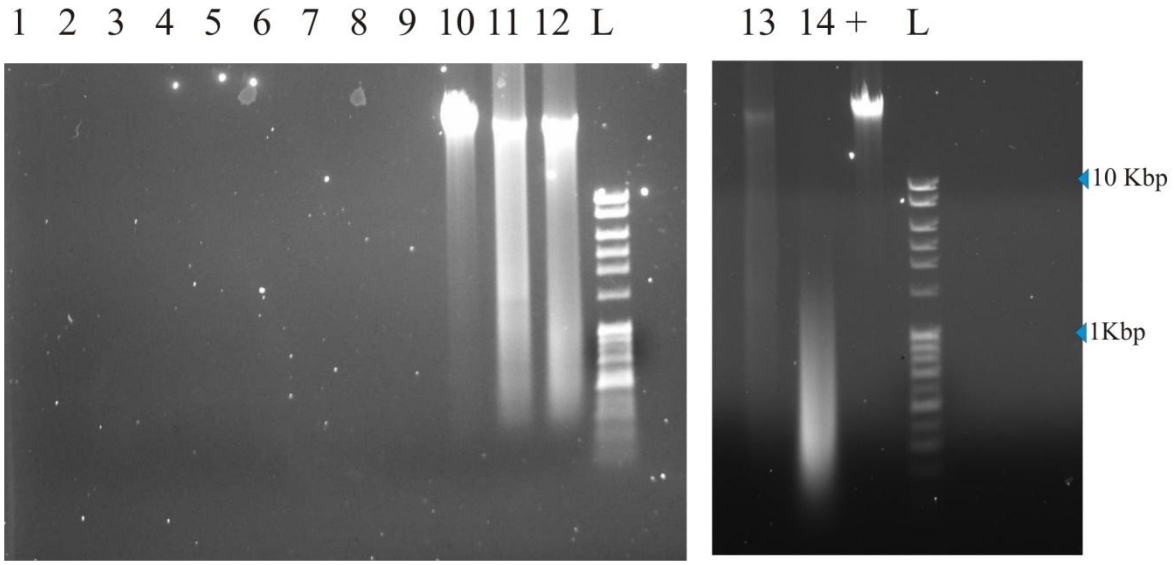


Figure 41 DNA Degradation of samples stored in SurePath™ cytology fluid. Lanes 1 to 9 contain 15µl of Baseline sample DNA; lane 10 contains 1.2µg SiHa DNA positive control; lanes 10 and 11 contain 1.2 µg SiHa DNA stored at room temperature in SurePath for 48 hours. Lane 13 contains DNA from SiHa stored in Surepath for 48 hours, lane 14 contains DNA from SiHa stored in SurePath for 1 month and lane "+" is DNA from SiHa that has not been stored in SurePath.

6.1.1.2 Beta-globin PCR

All samples selected for DIPS and E2 PCR were subjected to Beta-globin PCR to determine whether Baseline DNA was suitable for PCR methods. Eighty-five percent of the samples had a positive PCR result indicating good PCR quality DNA. The DNA from samples that failed the Beta-globin PCR were diluted 1 in 10 to dilute potential PCR inhibitory contaminants and Beta-globin PCR was repeated; the repeated samples did not provide a positive result. When E2 PCRs and DIPS were performed on Baseline samples, assays were more successful on HPV16 than HPV18 and HPV45. There was no significant difference in Beta-globin failure between HPV types (χ^2=3.63, df=2, p=0.16; data not shown) indicating DNA quality was equal between samples grouped by HPV type.

6.1.2 E2 PCR

To test the hypothesis that loss of episomes is associated with CIN3 (hypothesis 5), E2 PCR was applied to the Baseline samples. Analyses were performed on all samples and then repeated on data excluding borderline and mild cytology data.

As mentioned in 3.1.4.4, to achieve a maximum chance of a false positive outcome of 20%, 49 samples would need to be assayed. However, it was hypothesised that integration events in DNA may be more frequent in low grade CIN than those observed by Klaes et al. (1999), due to the reports of Huang et al.(2008); this would mean that more than 49 samples would be required to achieve 20% chance of a false positive result. Initially, thirty one HPV16, 25 HPV18 samples and 21 HPV45 samples were selected for E2 and E6 PCRs. A pilot set of 6 of each HPV16, HPV18 and HPV45 samples were subjected E2 and E6 PCRs. Of the HPV16 samples 100% were successful. For HPV18 and HPV45, only 50% of the samples yielded positive results: the positive control was positive and the failure of the E2 and E6 PCRs could not be attributed to PCR failure. Attempts to improve the results for HPV18 and HPV45 by diluting DNA before adding to the PCR were not successful.

The HPV16 samples were from patients with single infections and the HPV18 and HPV45 were from patients with multiple infections. Due to high failure rate in HPV18 and HPV45 baseline samples, further E2 PCRs and E6 PCR were not done on HPV18 and HPV45 Baseline samples. Instead, to prevent wasting time, reagents and to yield more meaningful data, a total of 83 HPV16 samples, comprising the entire HPV16 Baseline set with single infections that had attended colposcopy, were selected for E2 and E6 PCRs.

There was no evidence to support hypothesis 5 because detection of intact E2 did not decrease with increasing disease severity:

Glandular neoplasia and negative cytology group were removed from Chi-squared analyses to reduce the number of groups with a low number and to allow a more accurate test. There was no significant difference in intact E2 frequency between cytology groups (χ^2=0.28, df=3,

p=0.96) (Figure 42A). When cytology was grouped into high (severe and moderate) and low (mild and borderline) grade cytology, intact E2 was not significantly different between cytology groups (χ^2=0.046, df =1, p=0.83; data not shown).

E2 status was compared between histology data obtained from the first attendance at colposcopy following their first smear. There was a significant difference in the distribution of E2 status between histology groups (χ2=8.99, df=3, p=0.029) (Figure 42B), with a higher frequency of intact E2 than disrupted E2 in CIN1 and in CIN3. When Bonferroni correction for multiple testing was applied, the result was not statistically significant (p>0.05). The power of the study to detect a true association of intact E2 with CIN3, was 40% (0.4) with φ_C=0.23, N=74, 2 degrees of freedom at the 0.05 significance level. This highlights the likelihood of a false positive result at a rate of 60% in the number of samples tested and indicates a larger sample size is required to reduce the rate of detection of a false positive difference.

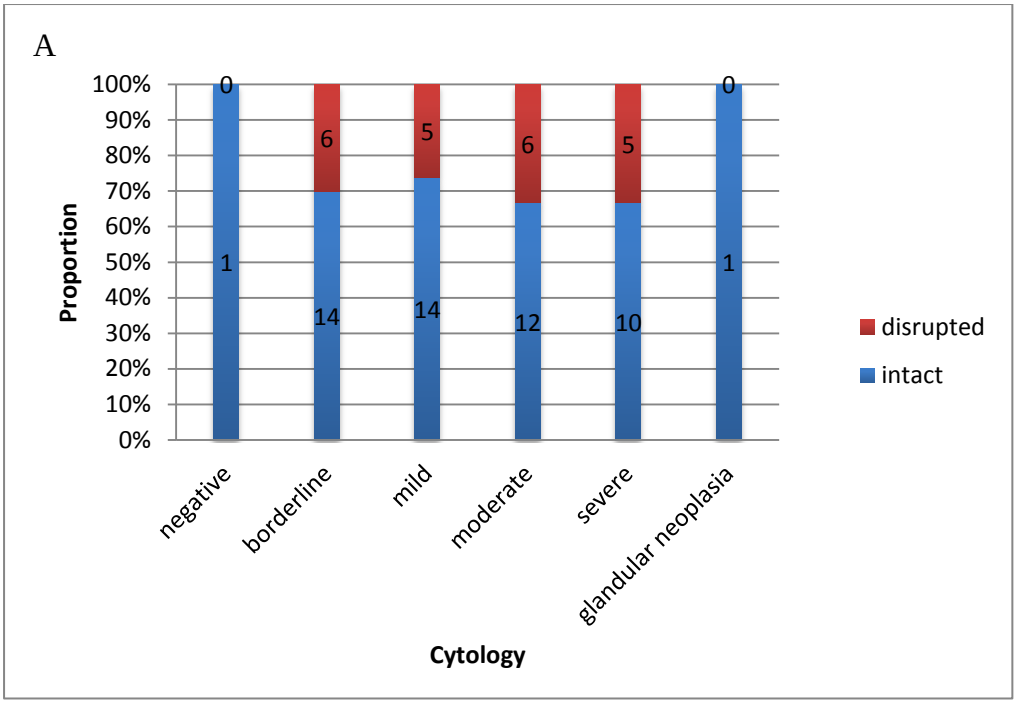
When histology was grouped into high grade (CIN2 and CIN3) and low grade (CIN1 and no CIN) histology, there was no significant difference in E2 status (χ^2=2.96, df=1, p=0.085, adjusted for Yates' continuity correction) although the frequency of intact E2 was higher in low grade histology with an odds ratio of 0.30 (95% CI=0.14-0.64) (Table 46); this indicated that women with intact E2 had 3/10 less chance of high grade neoplasia. The power of the study comparing E2 status between high-grade and low-grade disease with φ=0.20, N=74, 1 degree of freedom to achieve significance at the 0.05 level was 40%. This implies that the sample size was too small to accurately detect a significant difference. A total sample size of 196 would be needed to achieve power of 80% to correctly detect a significant difference, when α=0.05, between cases and controls when φ=0.20 at 1 degree of freedom.

E2 status was compared between CIN1 and CIN2/3 with mild and borderline cytology data removed; there was no significant difference in E2 status between high grade histology and low grade histology (χ^2=0.11, df=1, p=0.74).

	Low	High	χ^{21}	p-value	df	OR	95%CI
Intact E2	22(42%)	30 (58%)	2.96	0.085	1	0.30	0.14-0.64
Disrupted E2	4 (18%)	18 (82%)					

Table 46 Comparison of E2 as a marker of viral episome presence, with intact E2, or absence, with disrupted E2, between low-grade disease (Low: no CIN and CIN1) and high-grade disease (High: CIN2 and CIN3).

[1] Yates' continuity correction applied



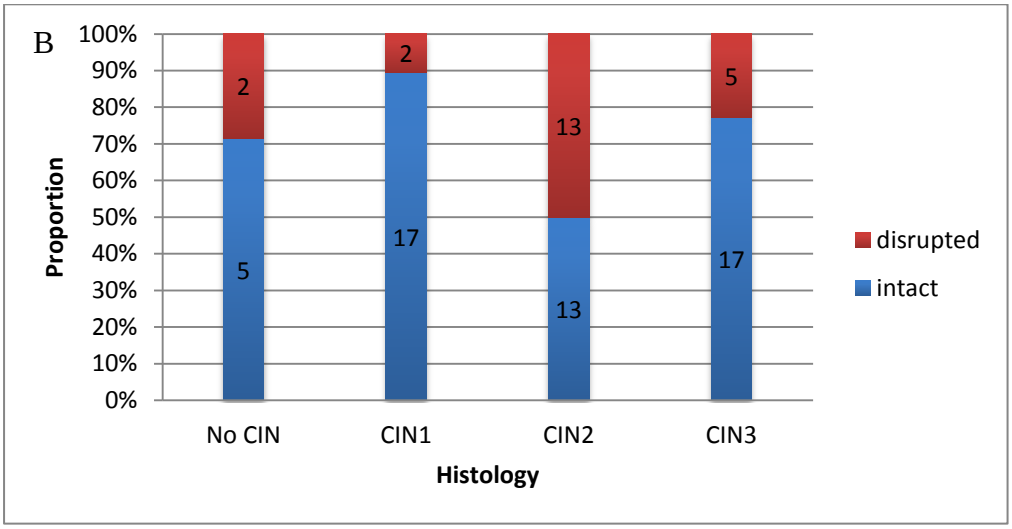


Figure 42 Comparison of E2 as a marker of viral episome presence, with intact E2, or absence, with disrupted E2, in different cytology grades. Proportion and frequency of E2 disruption versus intact E2 is shown for A) Negative, borderline, mild, moderate, severe dyskaryosis and glandular neoplasia (χ^2=0.28, df=3, p=0.96; when glandular neoplasia and negative dysplasia were removed from analysis). B) No CIN, CIN1, CIN2 and CIN3 (χ^2=8.99, df=3, p=0.029).

6.1.3 DIPS

6.1.3.1 DIPS Controls

SiHa, HeLa and HTB-34™ were used as HPV16, HPV18 and HPV45 positive controls, respectively; C33A was used as a HPV negative control. DIPS control PCR yielded amplicons of 1.2kbp for both Taq1 and Sau3AI digests in SiHa, HeLa, HTB34 and C33A.

Integration sites consistent with those in section 4.1.3 (Table 37) were detected for SiHa, HeLa and HTB-34™. No HPV16, HPV18 or HPV45 amplicons were generated for C33A.

DIPS control PCR was unsuccessful in the majority of baseline samples with 9% of the 103 samples tested producing an amplicon of 1.2kbp for either Sau3AI or Taq1. There was no significant difference in DIPS control success between HPV types (χ^2=0.616, df=2, p=0.74; data not shown) nor was there any significant difference in DIPS control success between multiple or single infections (χ^2= 0.742, df=1, p=0.39, Yates` continuity correction applied; data not shown). Thus DIPS performance cannot be attributed to difference in HPV type or difference in infection status.

Nanodrop data indicated that sufficient DNA had been put into the DIPS assay and the Beta-globin PCR indicated that the human DNA was good PCR quality. SiHa DNA that had been stored in SurePath™ for 1 month (Figure 41) did not produce an amplicon for the DIPS control PCR and it was likely that DNA degradation could explain DIPS control PCR failure in the Baseline samples. When the DIPS PCR using HPV16-specific primers was performed on SiHa DNA that had been stored in SurePath™ for 1 month, integration was detected. This illustrates that DIPS can successfully detect integration in degraded DNA.

DIPS was performed on the baseline samples before the problem of star activity was identified and a digestion time of 16 hours used. Subsequently, a selection of 8 HPV16, 8 HPV18 and 8 HPV45 samples were subjected to DIPS with a digestion time of 2 hours; the control PCR outcome was not improved therefore poor DIPS control PCR performance could not be due to digestion time or star activity.

6.1.3.2 DIPS on Baseline Samples

Initially, DIPS was performed on 31 HPV16, 25 HPV18 samples and 21 HPV45 samples. DIPS was also done on an additional 25 HPV16 samples that had either disrupted E2 or CIN3 samples with intact E2 to confirm the data generated by HPV16 E2 PCRs.

DIPS detected HPV16 in 51% of the HPV16 samples, HPV18 in 19% of the HPV18 samples and HPV45 in 28% of the HPV45 samples: there was a significant difference in HPV detection by DIPS between the different types (χ^2= 7.44, df=2, p=0.024); the power of which was 99% when φ_C=0.73, N=103, 2 degrees of freedom at the 0.05 significance level. All of the HPV16 samples were single infections whereas 13% of HPV45 samples and 63% of HPV18 samples were single infections. There was no difference in the detection of HPV by DIPS between single infection or multiple infections (χ^2= 0.26, df =1, p = 0.61) thus the difference in HPV detection between the HPV types cannot be attributed to single or multiple infections. There was no association between DIPS control success and the detection of HPV by DIPS (χ^2= 1, df=1, p=1; Yates` continuity correction applied (data not shown)): a successful DIPS control PCR, that detects human DNA, does not determine detection of HPV by DIPS.

Following DIPS application to the Baseline samples E6 PCR was applied to the HPV18 and HPV45 samples, where DNA were available. When the E6 data for HPV45 and HPV18 were combined with E6 data for HPV16 (n=83) there was a significant association between detection of E6 and detection of HPV by DIPS (χ^2= 4.93, df=1, p=0.026; Yates` continuity correction applied). For 90% of samples, that had HPV detected by DIPS, E6 PCR was positive. However, in 34% of samples that had no HPV detected by DIPS, E6 PCR was negative. This means that the positive predictive value of E6 PCR to detect HPV by DIPS is good but the negative predictive value of E6 PCR is poor. The power was 99% when ϕ =0.54, N=83, 1 degree of freedom at the 0.05 significance level; this indicates that the probability of a false positive result is 1% in the number of samples compared although the difference is no longer significant at the 0.05 level when adjusted for multiple testing.

Ten samples showed HPV integration, each with a unique integration site (Table 47). For all the samples where integration was detected, intact E2 was detected by E2 PCR. Nine integration events presented with disruption to the E2 ORF; it is likely that integrated and episomal HPV are present in these samples. Detection of integration was more frequent in HPV16 samples with 8 integration events occurring in HPV16 compared to 2 in HPV18

samples and none in HPV45 samples. Four samples had multiple integration events: 265, 913, 13290 and 1571 but this did not confer a higher severity of disease. Samples 265 and 913 had 2 integration events detected in the first smear sample and CIN1 was confirmed by histology when a biopsy was taken at colposcopy. Sample 13290 had 2 integration events detected in the smear sample and CIN2 was confirmed by histology. Sample 1571 had 3 integration events detected in the smear sample and CIN2 was confirmed in this patient by histological examination.

Samples 10295 and 12282 had integration that disrupted HPV at 1551bp, within E1. Sample 12282 had 4bp overlapping sequence, CAGC that was both a match to human and HPV sequence. Sample 10295 had 4bp overlapping sequence, GAGC that was both a match to human and HPV sequence. Sample 1571, and both the integration events in sample 13290, had HPV18 disrupted at 3080bp and had an overlap of 4bp GATG. HPV disruption occurs at a unique site every time when integration occurs and the event seen in the above samples is interesting. The similarity in the overlapping sequence at the sites of integration is the same, complementary or similar to the sequence preceding the adapter primer seen where star activity and incorrect ligation had occurred, GTCG, GCAG and ATCG; it is possible that the DNA has been digested and ligated back together at these positions and that these are not true integration events.

Sample 3220 had multiple infections when typed by GP5+/6+PCR-EIA, with both HPV16 and HPV45 present; there was no HPV45 detected by DIPS or by E6 PCR, but when *HPV16* DIPS was applied, integration was detected. Likewise, sample 10295 was HPV45 and HPV18 positive when typed by GP5+/6+PCR-EIA; when HPV45 DIPS and E6 PCR were applied no HPV45 was detected by DIPS; when *HPV18* DIPS was applied, integration was detected. These results illustrate the importance of multiple infection and E6 PCR: if a transient infection with a certain HPV type is present, then E6 or DIPS will not detect it because of low HPV copy numbers, on the other hand, where a certain HPV type causes a productive infection, the HPV type will be present at higher copy numbers and detected by both DIPS and E6 PCR.

Study ID	Type	E2[1]	CIN[2]	Primer[3]	Viral disruption[4]	Map[5]	Accession[6]	Match[7]	Orientation[8]
265	HPV16	intact	1	P4 Sau3AI	3376 (E2)	5q22.5	NT_034772.6	98%	S
265	HPV16	intact	1	P4 Sau3AI	3420 (E2)	15q15.1	BLAT only	100%	S
696	HPV16	intact	3	P4 Sau3AI	3206 (E2)	repeat	BLAT only	97%	Unknown
913	HPV16	intact	1	P4 Sau3AI	3098 (E2)	Xp11.1	NT_011630.14	99%	AS
913	HPV16	intact	1	P7 Sau3AI	6543 (L1)	9p13.3	NT_008413.18	97%	AS
1513	HPV16	intact	2	P7 Sau3AI	6478 (L1)	8q21.13	NT_008183.19	99%	S
1571	HPV16	intact	2	P2 Sau3AI	1781(E1)	10p12	BLAT only	100%	AS
1571	HPV16	intact	2	P4 Sau3AI	3205 (E2)	17q21.33	NT_10783.15	100%	AS
1571	HPV16	intact	2	P4 Sau3AI	3080 (E2)	8q11.23	NW_001839132.1	100%	AS
3577	HPV16	intact	0	P4 Sau3AI	3206 (E2)	3p14.1	NT_022459.15	100%	S
13290	HPV16	intact	2	P4 Sau3AI	3080 (E2)	16q11.2	NT_010498.15	100%	S
13290	HPV16	intact	2	P4 Sau3AI	3080 (E2)	21q21.1	NT_011512.11	98%	AS
3220	HPV16 /45	not tested	3	P2 Sau3AI	1789 (E1)	12q24.11	BLAT only	100%	AS
12282	HPV18	not tested	1	P2 Sau3AI	1551 (E1)	18q22.3	NT_025028.14	100%	S
10295	HPV18/45	not tested	1	P2 Sau3AI	1551 (E1)	Xp22.2	NT_167197.1	100%	S

Table 47 DIPS integration data for baseline samples. E2 status, histology grade, DIPS PCR primer and restriction digestion used to detect integration are shown. The site of viral disruption and integration locus are listed along with the accession number for the NCBI database sequence with the greatest similarity to human sequence.

[1] E2 PCR outcome.
[2] Histology outcome linked with the first attendance at colposcopy following first smear test. 1=CIN1, 2=CIN, 3=CIN3, 0=normal
3 Restriction enzyme and HPV specific primer used in DIPS that detected integration.
[4] Viral disruption number denotes last viral nucleotide before recombination to human sequence. Numbering of HPV sequence is according to Genebank accession numbers NC_001526.1, X05015 and X74479 for HPV16, HPV18 and HPV45, respectively.
[5] The genomic location of the site of integration chromosomal reference of all viral-cellular fusion transcripts with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al. 2002).
[6] EMBL Accession number for the NCBI database sequence with the most similarity to human sequence data.
[7] Percentage consensus of sequence data with NCBI database sequence.
[8] Orientation of HPV DNA within human sequence. S= Sense orientation, A=Antisense orientation.

6.1.3.3 Confirmation of Integration Sites

Once Beta-globin PCR, DIPS and E2 PCRs had been performed on Baseline samples there was insufficient DNA to confirm the integration sites by PCR using primers that flanked the integration site. DIPS PCRs were repeated on the ligation products for the samples in Table 47 using the primers indicated in Table 47 in an attempt to confirm the integration events detected; integration was not confirmed by replicating the DIPS assay. This indicates that the integration events detected in Baseline samples are an artefact of DIPS or they are not clonal.

6.1.3.4 Integration in Baseline Samples

Hypothesis 4 was tested by comparing the frequency of integration between varying histological grades. There was no evidence to support integration as early event in cervical neoplasia (hypothesis 4). Analyses were performed on all samples and then repeated on data excluding borderline and mild cytology data.

"Integration not detected" can be defined two ways: 1) DIPS assay performed but no integration detected or 2) DIPS assay performed but only HPV sequences detected. To allow an accurate Chi squared test, cytology was grouped into high grade cytology (moderate and severe) and low grade cytology (mild and borderline); there were no significant differences in integration detection between high grade or low grade cytology, when all samples were included (χ^2=0.00, df =1, p=0.99).

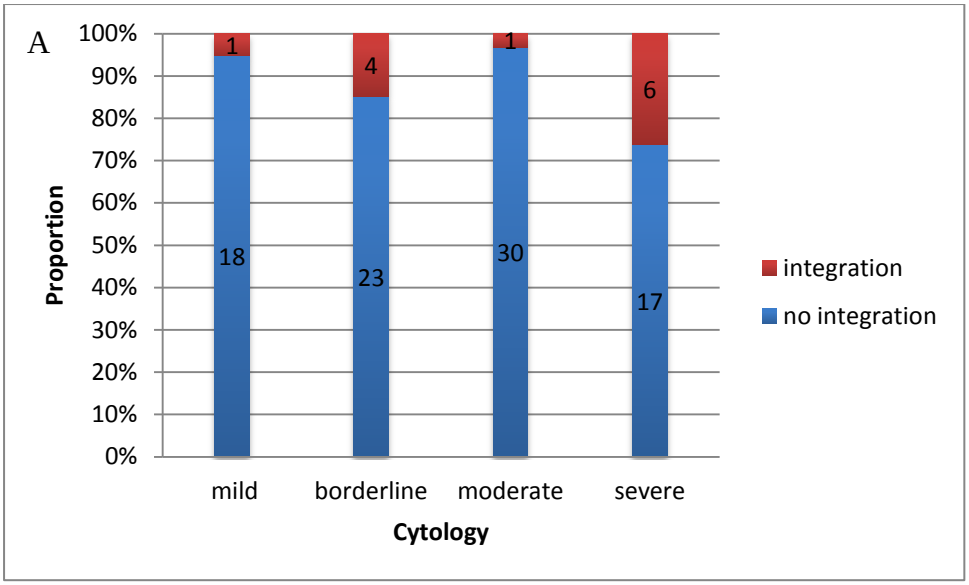


Figure 43 Integration detection by cytology groups. Frequency of integration detected versus no integration detected is shown for borderline, mild, moderate and severe dyskaryosis.

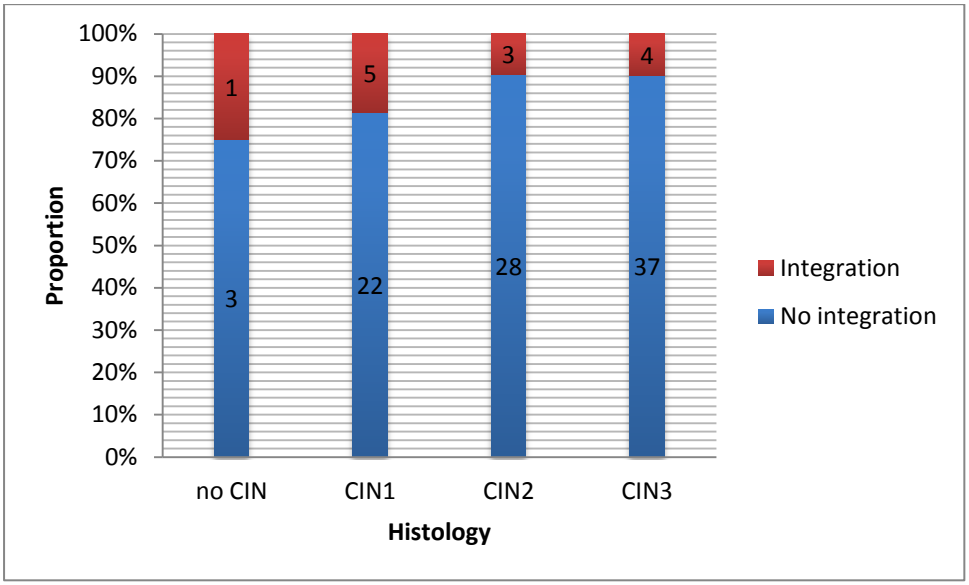


Figure 44 Integration detection by histology group: Frequency of integration detected versus no integration detected is shown for no CIN, CIN1, CIN2 and CIN3.

To test hypothesis 4, detection of integration was compared with histology result of a biopsy taken at colposcopy following first smear test. The proportion of integration events were higher in CIN1 compared to the other groups but this was not statistically significant (χ^2=1.96, df=3, p=0.58) (Figure 44A). To allow a more accurate comparison with Chi square, histology was grouped into low grade (no CIN and CIN1) and high grade (CIN2 and CIN3):

when including all samples no significant difference was detected (χ^2=1.05, df=1, p=0.30, Yates' continuity correction applied; data not shown); when including only samples positive for HPV no significant difference was detected (χ^2= 0.71, df=1, p=0.40, Yates' continuity correction applied; data not shown). These data do not support hypothesis 4 and integration is not an early event present in CIN1.

6.1.3.5 Alternative DNA extraction method

The Baseline samples were previously extracted using Proteinase K method. As mentioned in section 3.2.3 Proteinase K allows purification of crude, un-purified DNA and it remained possible that presence of protein in the extracted DNA may inhibit the DIPS assay. To ascertain whether an alternative extraction method could improve the DNA quality and improve DIPS outcome, 3 samples that were positive for HPV with DIPS, and 3 samples that failed DIPS and E6 PCR for HPV16, HPV18 and HPV45 were extracted using QIAamp® (QIAGEN). The 3 samples that were HPV positive for DIPS and E6 were still E6 positive and produced amplicons for DIPS; the three samples that had failed DIPS and E6 PCR previously failed both assays. The QIAamp® DNA was electrophoresed on a fine pore 0.7% gel and there were no bands visible. An alternative method of extraction did not improve the outcome.

6.1.4 Baseline: Summary

There was no evidence in this sample set to support hypothesis 3 or hypothesis 4. Integration was not more prevalent in CIN1 nor was presence of episomal HPV, represented by intact E2, associated with CIN1. The major limitation of this sample set was DNA quality. In future, E2 PCRs and DIPS require samples that have not been stored in SurePath™ fluid.

6.2 HPV Integration as a Biomarker of Cervical Disease: Results

The aim of this section was to pilot whether HPV integration can predict high-grade cervical disease, due to presence of integration and loss of episomal HPV, in women with cytological abnormalities (hypothesis 4 and 5). Cervical smear samples were collected from women that were referred to colposcopy following detection of cellular dyskaryosis by cytological

screening. Samples were collected in RNAProtect® (QIAGEN) to provide the best possible quality RNA and DNA. Initially, 45 samples were intended to be collected but sample collection was slower than anticipated and study initiation was delayed thus 28 samples were collected. The samples were HPV typed and E2 PCRs, DIPS, RS-PCR and APOT for HPV16, HPV18 and HPV45 applied; HPV16 E2, E6 and E7 mRNA were quantified by qPCR. The data from integration assays and qPCR were correlated with cytology and biopsy results to determine if integration is present in CIN and loss of episomes is associated with CIN3 and leads to high E6 and E7 expression (hypothesis 3).

6.2.1 Samples

Twenty eight samples were collected (Table 48 and Figure 45). In samples taken from women referred to colposcopy with severe dyskaryosis, CIN3 was the most frequent diagnosis following histological examination. The majority of samples taken from women referred to colposcopy with mild dyskaryosis, either had no CIN confirmed by histology or did not have a biopsy taken. There were 2 women (HIBCD10 and HIBCD23) referred to colposcopy with mild cellular dyskaryosis that had CIN2 or CIN3 confirmed by histological examination. HIBCD5 was from a woman did not have CIN confirmed by histological examination and was referred with sever dyskarosis. HIBCD14 was from a woman referred to colposcopy with severe dyskaryosis, who had both cervical cancer and CIN3 confirmed on histological examination. HIBCD11 and HIBCD19 were samples for which histological grade could not be confirmed; these were taken from women referred with severe and moderate dyskaryosis.

ID	Age[1]	Cytology[2]	Histology[3]	Infection	HPV[4]
HIBCD1	36	severe	CIN 3	no	
HIBCD2	52	mod	No CIN	single	45
HIBCD3	27	severe	CIN3	mixed	33,39,58
HIBCD4	31	severe	CIN 2-3	mixed	33,39,51,58
HIBCD5	35	severe	No CIN	single	16
HIBCD6	37	severe	CIN 3	single	16
HIBCD7	21	severe	CIN 3	mixed	16,35,52
HIBCD8	24	severe	CIN3	mixed	16,18,31,59,66
HIBCD9	30	severe	CIN 3	mixed	16,18,39
HIBCD10	33	mild/mod	CIN 2	mixed	33,35,39,52
HIBCD11	26	severe	CIN ungraded	single	16
HIBCD12	27	severe	CIN 3	single	33,39,58
HIBCD13	27	severe	CIN 2	single	16
HIBCD14	25	severe	SCC + CIN 3	mixed	16,59,66
HIBCD15	35	severe	CIN 3	single	58,59
HIBCD16	21	severe	CIN 3	mixed	31,35,39,51,66,68
HIBCD17	25	severe	CIN 3	mixed	33,39,55
HIBCD18	28	severe	CIN 3	mixed	33,39,58
HIBCD19	37	mod	CIN ungraded	mixed	18,58,66
HIBCD20	31	mod	CIN3	single	16
HIBCD21	22	mild	No CIN	mixed	31,35,39,59,68
HIBCD22	22	mild	No CIN	single	39
HIBCD23	34	mild	CIN2/3	single	33
HIBCD24	40	mild	No CIN	no	
HIBCD25	25	mild	No CIN	no	
HIBCD26	21	mild	No CIN	no	
HIBCD27	53	mild	No biopsy	no	
HIBCD28	54	mild	No CIN	mixed	51,66

Table 48 Summary of HI-BCD samples. Sample ID, age, cytology grade (severe, mild or moderate (mod)), histology outcome (CIN= cervical intraepithelial neoplasia, CaCx= cervical cancer), infection type (mixed, single or no infection) and HPV type.

[1] Age of woman at time of sampling.
[2] Cytology referral result; mod=moderate.
[3] Histology of biopsy taken at the time of sampling.
[4] HPV type by Papillocheck® and GP5+/6+ PCR-EIA methods.

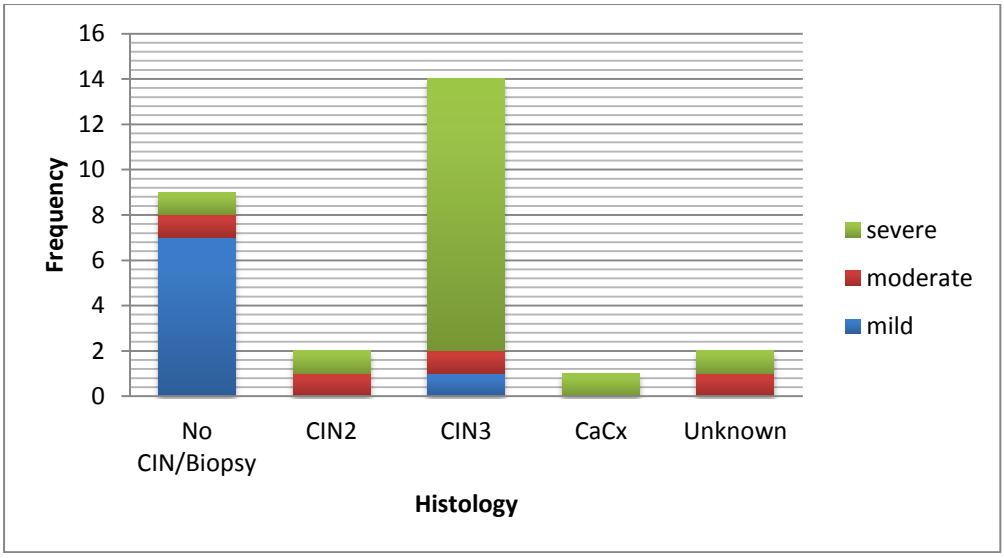


Figure 45 Histology outcome (No CIN or no biopsy, CIN1, CIN2 CIN3 and CIN unknown) in mild, moderate and severe dyskaryosis groups.

6.2.1.1 DNA and RNA quality

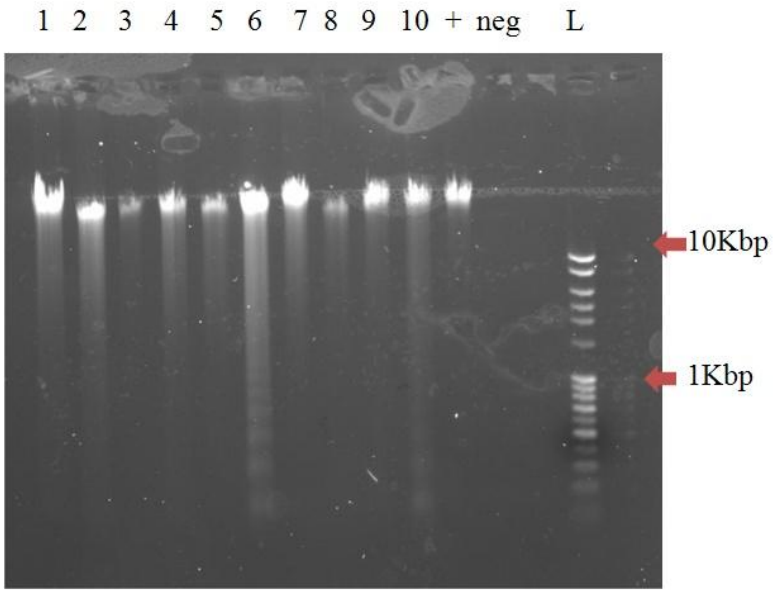


Figure 46 HI-BCD DNA Integrity. Electrophoresis of 5uL of extracted HIBCD sample DNA on a 0.8% fine-pore gel. Lanes 1 to 10 contain DNA for HIBCD samples HIBCD19 to HIBCD28; + is extraction positive control SiHa DNA and – is extraction negative control (water).

DNA was electrophoresed on a 0.8% fine pore gel and little DNA degradation was observed (

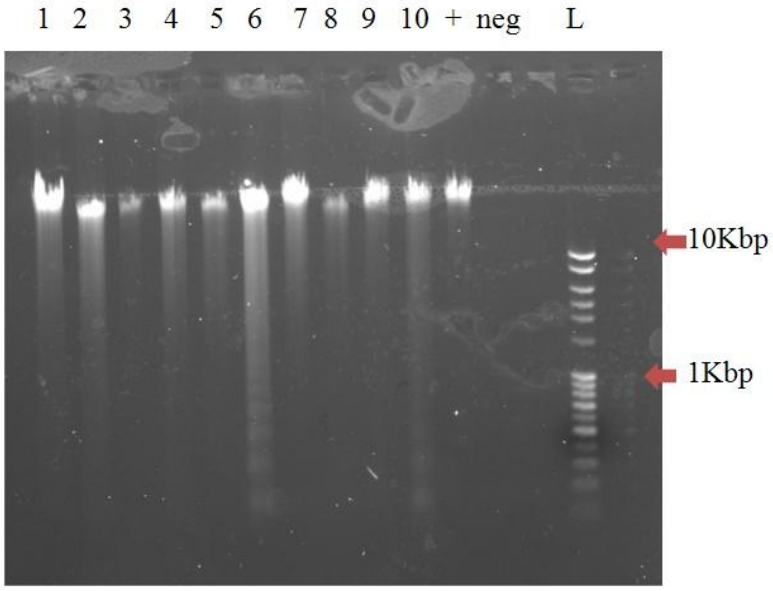


Figure 46). Beta-globin PCR was successful on all HI-BCD samples except HIBCD26 indicating 96% of samples had DNA of suitable quality for PCR. RNA integrity values were low indicating that a high level of RNA degradation was present. SiHa cells, that were stored in RNAProtect® in the same conditions as the smear samples, were used as a positive control each time a batch of samples was extracted; the RNA integrity value was 10 every time; RNA degradation did not occur in the extraction process or in the storage of the samples. Jane McRea, a senior cytologist in Cervical Screening Wales, confirmed that the likely reason for RNA degrading is due to the natural lifecycle of the epithelial cells: the cells collected by a smear are ready to be shed from the topmost layer of epithelia, are already dead and RNA has started to degrade naturally.

6.2.1.2 HPV Typing

SiHa was used as a positive control and HPV16 was detected by PapilloCheck® or by GP5+/6+PCR-EIA. Mixed infections and single infections were present across all cytology and histology groups (Table 48). HPV16 was detected most frequently in samples from women referred with severe dyskaryosis. HIBCD1 was taken from a women referred to colposcopy with severe dyskaryosis and had CIN3 diagnosed by histology; this sample had no HPV detected by PapilloCheck® or by GP5+/6+PCR-EIA. HIBCD2 was from a woman with moderate cytology who had no CIN detected when a biopsy was taken; HPV45 was detected in this sample.

6.2.1.3 E6 PCR

Integration disrupts the HPV genome and it is plausible that integration can result in failure to detect HPV by PapilloCheck® or GP5+/6+PCR-EIA typing. To overcome this, all samples were subjected to E6 PCR specific for HPV16, HPV18 and HPV45 (Table 49). Plasmids encoding HPV16, HPV18 and HPV45 DNA were used as positive controls and HPV was detected in each. All samples that were HPV negative by PapilloCheck® and GP5+/6+ PCR-EIA typing, did not produce amplicons for HPV16, HPV18 or HPV45 E6 PCR. As mentioned above HIBCD1, was diagnosed with severe cellular dyskaryosis and CIN3 but was HPV negative by PapilloCheck® and GP5+/6+ PCR-EIA typing; this sample did not produce an amplicon for HPV16, HPV45 or HPV18 E6 PCR. All samples positive for HPV16 by PapilloCheck® and GP5+/6+ PCR-EIA produced amplicons with HPV16 E6 PCR. HIBCD2 was positive for HPV45 by E6 PCR. There was 100% consensus between the detection of HPV16 and HPV45 by PapilloCheck®, GP5+/6+ PCR-EIA and E6 PCR. PapilloCheck® and GP5+/6+ PCR-EIA detected HPV18 in 4 samples: HIBCD5, HIBCD9, HIBCD8 and HIBCD19; two samples produced amplicons for HPV18 E6 PCR: HIBCD19 and HIBCD8; there was therefore 50% consensus for HPV18 between GP5+/6+ PCR-EIA, E6 PCR and PapilloCheck®.

Sample	**age**[1]	**infection**[2]	**Cytology**[3]	**CIN**[4]	**E6**[5]	**E2**[6]
HIBCD2	52	single	mod	No CIN	HPV45	disrupted
HIBCD5	35	single	severe	No CIN	HPV16	intact
HIBCD6	37	single	severe	CIN 3	HPV16	intact
HIBCD7	21	mixed	severe	CIN 3	HPV16	intact
HIBCD8	24	mixed	severe	CIN3	HPV16	intact
HIBCD8	24	mixed	severe	CIN3	HPV18	disrupted
HIBCD9	30	mixed	severe	CIN 3	HPV16	intact
HIBCD11	26	single	severe	CIN Ungraded	HPV16	intact
HIBCD13	27	single	severe	CIN 2	HPV16	intact
HIBCD14	35	mixed	severe	CaCx, CIN 3	HPV16	intact
HIBCD19	37	mixed	mod	CIN ungraded	HPV18	disrupted
HIBCD20	31	single	mod	CIN3	HPV16	intact

[1] Age of woman at time of sampling.
[2] Infection type, mixed or single infection
[3] Cytology referral result: mild, moderate (mod) or severe dyskaryosis.
[4] Histology results of biopsy taken at colposcopy.
[5] E6 PCR result
[6] E2 PCR result

Table 49 Summary of HI-BCD E2 PCR and E6 PCR results. Age, infection type, HPV type, cytology and histology outcome are shown.

6.2.2 HPV Integration

To test hypotheses that integration is an early event present in CIN1 and that episomal loss is a late event associated with CIN3 (hypotheses 4 and 5) DIPS, APOT and E2 PCRs were performed on HIBCD samples that were HPV16, HPV18 and HPV45 positive by E6 PCR, PapilloCheck® and GP5+/6+ PCR-EIA.

6.2.2.1 E2 PCRs

E2 PCR was done on all the samples that were HPV16, HPV18 and HPV45 E6 positive (Table 49). Samples that were HPV18 positive by PapilloCheck® and GP5+/6+ PCR-EIA, but not by HPV18 E6 PCR, failed the E2 PCR. All the samples that were HPV16 positive had intact E2 (Table 49). HIBCD2 had disrupted E2 for HPV45, with primer set 3 failing to amplify an amplicon; HIBCD8 had disrupted E2 for HPV18, with primers sets 1, 2 and 3 failing to produce an amplicon; HIBCD19 had disrupted E2 with primer set 3 failing to amplify. There was insufficient sample numbers to perform accurate statistical analysis but there was no pattern in the data. There were no samples that had CIN1 confirmed by a biopsy however, E2 disruption was not more frequent in CIN3 or CaCx compared to samples with no CIN. E2 PCR was used as a marker of episomal presence and these data indicate that loss of episomes is not more frequent in CIN3; hypothesis 5 is not supported.

6.2.2.2 DIPS

DIPS, with a 2 hour incubation restriction digest was performed on HI-BCD samples (Table 50). DIPS detected HPV in all the samples where E6 PCR had detected HPV16, HPV18 or HPV45. DIPS control PCR was 100% successful indicating that efficient digestion and ligation had taken place and that the DNA was of sufficient quality for DIPS. HPV16, HPV18 and HPV45 plasmid DNAs were used in addition to SiHa, HeLa and HTB-34™ as type specific positive controls. All positive controls produced amplicons consistent with predicted sizes. Primer sets that detected integration, in SiHa, HeLa and HTB-34™ produced the same amplicon sizes as previously detected and DIPS assay was therefore performing

optimally. Furthermore, in samples where multiple types were detected by PapilloCheck® and GP5+/6+ PCR-EIA, DIPS detected HPV; this shows that DIPS performance is not affected by the presence of multiple HPV types.

DIPS detected integration in 3 samples: HIBCD8, HIBCD9 and HIBCD19 (Table 50). HIBCD8 had 1 integration event detected for HPV16 on 12q23.2, with disruption to E1 ORF, and 3 integration events for HPV18 on 8p11.21, Xp11.4 and 11p12, with disruption to E1/E2 and L2 ORFs. HIBCD9 had 1 integration site detected on 6q15, with disruption to the E1 ORF and HIBCD19 had integration detected on 9q21.3, with disruption to the L2 ORF. Integration was detected in 3/16 (19%) women with severe dyskaryosis and detected in 3/14 (21%) of women with CIN3. Samples HIBCD7, HIBCD14 and HIBCD19 had other HPV types present; it is possible that in these samples another HPV type could be integrated and contributing to disease. HIBCD6 and HIBCD20 were samples taken from women that had CIN3; both of these samples had single infections and another HPV type is unlikely to be driving disease in these samples. Where integration was detected, it could not be confirmed by PCR using primers that flanked the integration site and repeating DIPS PCR could not replicate the detection of integration; thus integration is unlikely to be clonal and may not be contributing to CIN3 in these samples.

As mentioned in section 6.2.2.1, HPV18 E2 was disrupted in HIBCD8 and HIBCD19 and HPV45 E2 was disrupted in sample HIBCD2. The sequence data from DIPS was used to verify the E2 status in these samples:

For HIBCD2, E2 PCR primer set 3 failed. DIPS with P4 produced sequence data that covered the region of E2 that E2 PCR primers set 3 would amplify: E2 was intact in this sample. When the sequence for primer set 3 were compared to the sequence data, the reverse primer lacked specificity to HPV45 sequence detected in HIBCD2; this would explain why primer set 3 failed to produce an amplicon.

For HIBCD8, E2 PCR with primers sets 1, 2 and 3 failed to produce an amplicon. DIPS with HPV18 F4 produced sequence data that the region of E2 that E2PCR primers sets 1, 2 and 3 would amplify thus E2 was intact in this sample. When the sequence of the HPV18 E2 primer sets 1, 2 and 3 were compared to HIBCD8 DIPS sequence data, there was no difference in the primer sequences to the sequence data; lack of specificity cannot explain why primer sets 1, 2 and 3 failed the E2 PCR. HIBCD8 had mixed HPV types present (HPV16, HPV18, HPV31, HPV59, HPV66), when comparing the HPV18 E2 primers to HPV16 sequence (NC_001526.1), primer sets 1, 2 and 3 have high consensus (BLAST E values less than 0.073) with HPV16 sequence; primer set 4 had low consensus to HPV16 (forward primer had BLAST E value of 1 and reverse primer had E value 0.01). It is likely that a background of mixed HPV types interfere with E2 PCR specificity.

For HIBCD19, HPV18 E2 PCR with primer set 3 did not generate an amplicon. DIPS with HPV18 F5 produced sequence data that aligned to the region of E2 that E2 PCR primers set 3 would amplify: E2 was intact in this sample. When sequence of the HPV18 E2 primer set 3 was compared to HIBCD19 DIPS sequence data, the reverse primer lacked specificity to the HIBCD19 sequence data; lack of specificity can explain the failure of E2 primer set 3 to produce an amplicon in this sample.

6.2.2.3 APOT

SiHa, HeLa and HTB-34 RNA were used as positive controls and APOT detected integrated transcripts in these samples consistent with section 4.1.4. Although the RNA integrity of the samples assayed was poor, GAPDH PCR was used to detect cDNA after RT-PCR and cDNA was detected in all of the samples subjected to APOT. APOT did not detect any transcripts in samples HIBCD5, HIBCD6, HIBCD11, HIBCD13, HIBCD19 and HIBCD20 (Table 50). APOT detected episomal transcripts in samples HIBCD2, HIBCD7, HIBCD8, HIBCD9 and HIBCD14: type 3 (E7,E1^E4) and type 4 (E7,E1^E4,E5) transcripts were detected (Table 50). Integrated transcripts were not detected nor were any full length episomal transcripts with E2 detected. RNA integrity cannot explain why HPV transcripts were not detected in samples HIBCD5, HIBCD6, HIBCD19 and HIBCD20 since transcripts were detected in HIBCD14 and HIBCD8 that failed RNA integrity tests. For samples where DIPS detected

integration, integrated transcripts were not detected; the conclusion from this data is although integration was detected, it is not clonal, not being transcribed, and not driving CIN3 in these samples.

ID	HPV[1]	Infection	CIN[2]	Map[3]	Accession[4]	HPV [5]	APOT[6]
HIBCD2	45	Single	no CIN	HPV only	na	na	Episomal Type 3 and 4
HIBCD5	16	Single	no CIN	HPV only	na	na	Not detected
HIBCD6	16	Single	CIN3	HPV only	na	na	Not detected
HIBCD7	16	Mixed	CIN3	HPV only	na	na	Episomal Type 4
HIBCD8	16	Mixed	CIN3	11p12	NT_009237.18	E1 (2431bp)	Not detected
	18			Xp11.4	NT_079573.4	E2/E1 (2877bp)	Episomal Type 3 and 4
	18			8p11.21	NT_167187.1	L2 (4540bp)	
	18			12q23.2	NT_029419.12	L2 (4541bp)	
HIBCD9	16	Mixed	CIN3	6q15	NT_007299.13	E1 (2877bp)	Episomal Type 4
HIBCD11	16	Single	ungraded	HPV only			Not detected
HIBCD13	16	Single	CIN2	HPV only			Not detected
HIBCD14	16	Mixed	CIN3 Cacx	HPV only			Episomal Type 4
HIBCD19	18	Mixed	ungraded	9q21.3	NT_008470.19	L2 (4479bp)	Not detected
HIBCD20	16	Single	CIN3	HPV only			Not detected

Table 50 DIPS and APOT results. Site of integration, accession number, site of HPV disruption and transcripts detected by APOT are shown.

[1] HPV type
[2] Histology of biopsy taken at colposcopy. Ungraded= neoplasia was observed but it was not possible to grade CIN1, CIN2 or CIN3.
[3] The genomic location of the site of integration chromosomal reference of all viral-cellular fusion transcripts with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al. 2002).
[4] EMBL Accession number for the NCBI database sequence with the most likeness to human sequence data.
[5] Viral disruption number denotes last viral nucleotide before recombination to human sequence. Numbering of HPV sequence is according to Genebank accession numbers NC_001526.1, X05015 and X74479 for HPV16, HPV18 and HPV45, respectively.
[6] Transcripts detected by APOT. Type 3 is E7,E1^E4; type 4 is E7,E1^E4,E5.

6.2.2.4 HPV gene expression

To test hypothesis 3, HPV16 E2, E6 and E7 mRNA were quantified relative to housekeeping gene expression TBP2 and HPRT in all HPV16 positive HIBCD samples (Figure 47 and Table 51). Assays to quantify HPV16 E2, E6 and E7 were not available for HPV18 and HPV45 and were not developed due to limited time and cost implications. Data is not shown for samples HIBCD8, HIBCD13, HIBCD11 and HIBCD20 because their gene stability values and coefficients of variance were outside quality control parameters levels, mentioned in section 3.3.3.8.3.

Sample	infection	Cytology[1]	E2	age[2]	CIN[3]	Integ[4]	E2	E6	E7
HIBCD5	single	severe	intact	35	No CIN	no	0.89	ND	0.98
HIBCD6	single	severe	intact	37	CIN 3	no	0.56	2.13	1.76
HIBCD7	mixed	severe	intact	21	CIN 3	no	0.34	ND	0.94
HIBCD8	mixed	severe	intact	24	CIN3	yes	fail	fail	fail
HIBCD9	mixed	severe	intact	30	CIN 3	yes	0.27	0.47	0.62
HIBCD11	single	severe	intact	26	CIN Ungraded	no	fail	fail	fail
HIBCD13	single	severe	intact	27	CIN 2	no	fail	fail	fail
HIBCD14	mixed	severe	intact	35	CaCx + CIN 3	no	21.94	ND	ND
HIBCD20	single	mod	intact	31	CIN3	no	fail	fail	fail

Table 51 Calibrated Normalised Relative Quantities (CNRQs) of E2, E6 and E7 in HPV16 positive HI-BCD samples. Infection status, cytology grade, E2 status, age, histology grade and integration status are shown. CNRQ values of 1 or more indicate HPV E2, E6 or E7 expression greater than or equal to housekeeping genes. CNRQ values less than 1 indicate HPV E2, E6 or E7 expression lower than housekeeping genes. ND= not detected.

[1] Referral cytology result: mild, moderate (mod) or severe dyskaryosis.
[2] Age of woman at time of sampling.
[3] Histology of biopsy taken at the time of sampling.
[4] Integration status. Yes=integration detected; no=no integration detected

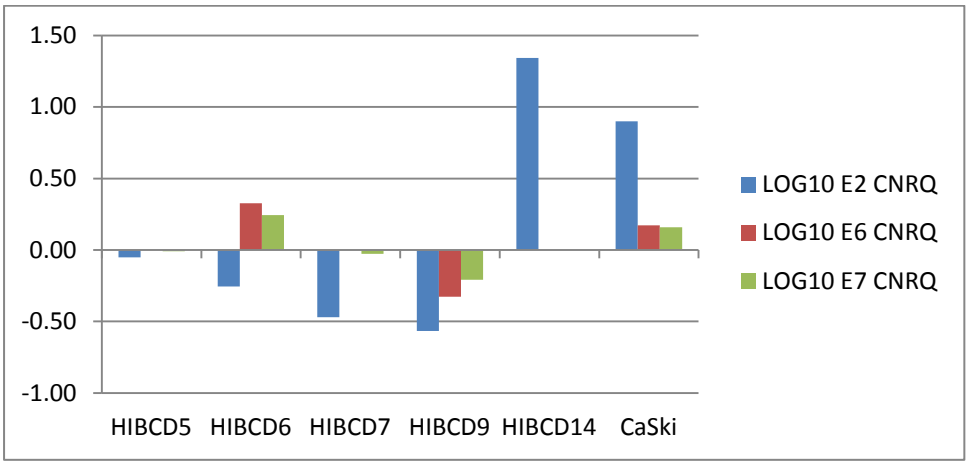


Figure 47 Log_{10} Calibrated Normalised Relative Quantities (CNRQs) of E2, E6 and E7 in HPV16 positive HIBCD samples. Values equal to zero represent HPVE2, E6 and E7 expression quantities equal to house-keeping genes; values less than zero represent expression levels lower than housekeeping genes; values greater than zero represent expression levels higher than housekeeping genes.

Theoretically, in a productive infection, E2 expression would be high and E6 and E7 expression would be low; in a high grade lesion, E2 expression would be low, or absent, and E6 and E7 expression would be high. HIBCD6, HIBCD7 and HIBCD9, were from women with CIN3, and showed lower E2 expression relative to housekeeping genes despite having intact E2. E7 and E6 mRNA expression was higher than housekeeping gene expression for HIBCD6; it is likely that E7 and E6 are driving CIN3 in HIBCD6. HIBCD14 has unusual levels of E2 expression: E2 is expressed over 21 times higher than housekeeping genes (Table 51), with no E6 or E7 mRNA detected; the assays were repeated on HIBCD14 and the same outcome was achieved. HIBCD14 had mixed HPV types and it is possible that another HPV type was contributing to CIN3 and CaCx in this sample. HIBCD5 was taken from a woman with no CIN by histology; very low levels of E2 and E7 mRNA were detected. These data do not support hypothesis 3: there was no evidence in to suggest, where integration was detected, that integration causes loss of E2 that results in increased E6 and E7 expression.

6.2.3 HI-BCD: Summary

There was no evidence to support hypotheses 3, 4 or 5: HPV integration or episomal loss are not contributing to cervical neoplasia in this sample set and the data suggests that integration cannot be used as a biomarker of cervical disease when applied to smear samples. The average age, at sampling, of the women in this sample set is 31, with the average age of

women diagnosed with CIN3 being 29. The DNA quality of the HIBCD samples was excellent and DIPS was performing optimally therefore failure to detect integration cannot be explained by poor DNA quality. In section 6.1.3.2, very low frequencies of integration were detected in women aged 20-22. The data from this section and 6.1.3.2 suggest that HPV integration detection and episomal loss do not confer high grade cervical neoplasia in women under the age of 30. Many samples had mixed HPV types and a number of samples were not explored by DIPS, E2-PCR and APOT because HPV16, HPV18 and HPV45 were not present. Types other than HPV16, HPV18 and HPV45 must be explored for integration but this will be future research not part of this PhD.

Published studies of integration have focussed on cervical cancer biopsies and integration has been reported for cancer biopsies. It is possible that failure to detect integration could be due to the sample type: performing the assays on smear samples may reduce the chances of detecting integration. Furthermore the presence of episomes in precancerous lesions may prevent the detection of integration transcripts. To test the hypothesis that integration and integrated transcripts can be detected in cancer (hypothesis 6), DIPS, E2 PCRs and RS-PCR were performed on 6 cervical cancer biopsies.

6.3 Application of Assays to Cervical Cancer Biopsies: Results

To test the hypothesis that integrated transcripts are detected in cervical cancer when episomal HPV is lost, E2 PCRs, APOT, DIPS and RS-PCR were performed on DNA and RNA from sections taken from six cervical cancer biopsies. Assay reproducibility was also compared using the biopsies due to availability of DNA and RNA. Integration was detected at low frequency in the samples tested in sections 6.1.3.2 and 6.2.2.2 and age was hypothesised as a possible explanation: integration is not detected in women under the age of 30. To test the hypothesis that age may be a factor in detection of integration, integration detection was observed in relation to age in the cancer biopsy samples.

6.3.1 HPV typing

PapilloCheck® was applied to DNA from all six biopsies. Samples 1W, 2A, 3O, and 5B were HPV16 only. For sample 6N and 4T no HPV was detected. Samples 6N and 4T were then subjected to GP5+/6+ PCR-EIA: HPV16 was detected in 6N and HPV45 was detected in sample 4T.

6.3.2 DNA and RNA quality

Electrophoresis of the DNA samples on a 0.8% finepore gel showed that the DNA had little degradation for 1W, 2A, 3O, 5B, and 6N; sample 4T was badly degraded. RNA integrity values were between 5.3 and 8.1 for samples 1W, 2A, 3O, 5B, and 6N; samples 4T had an RNA integrity value of 4 indicating the RNA had degraded.

6.3.3 Integration and E2

A different pattern of integration and E2 status was observed in each biopsy (Figure 48, Table 52, Table 53 and Table 54).

For sample 1W, from a woman aged 23, E2 was intact and RS-PCR and DIPS did not detect integration. Three non-contiguous sequence events were detected by RS-PCR (Table 53): and confirmed by PCR using primers that flanked the HPV sites of disruption. APOT did not

detect integration and detected episomal transcript types 3 and 4 (E7,E1^E4 and E7,E1^E4,E5, respectively). These data support the hypothesis that integration is not detected in women under the age of 30.

For sample 2A taken from a woman 33 years old, E2 was disrupted; RS-PCR and DIPS detected no integration. APOT did not detect any integration or episomal transcripts.

Sample 3O was taken from a woman 46 years old. E2 PCR indicated E2 was intact; RS-PCR, DIPS and APOT detected the same integration event on 2q22.3 with disruption at 944bp within the E1 ORF. PCR using primers flanking the integration site confirmed the detection of integration. These data suggest that the integration event in 3O is contributing to cancer as it is clonal and being actively transcribed.

Sample 4T was taken from a woman 63 years old. Unfortunately DNA and RNA from this sample were poor in quality and repeated attempts at integration assays failed.

Sample 5B was taken from a woman 30 years old. E2 PCR did not detect E2, however, both DIPS and RS-PCR detected HPV16. RS-PCR did not detect integration but DIPS detected integration on 8p21.3, with disruption to the E1 ORF. APOT detected no integrated transcripts suggesting this integration event was not transcribed. PCR using flanking primers did not confirm this integration event nor did repeating the DIPS assay. These data imply that integration in sample 5B is not clonal and contributing to the cancer in this sample.

Sample 6N, taken from a woman 45 years old, had no HPV detected by PapilloCheck®. When HPV16 E6 PCR was performed, E6 was detected and HPV typing by GP5+/6+ PCR-EIA confirmed that HPV16 was present as a single infection. E2 PCR showed that E2 was disrupted in this sample and both DIPS, APOT and RS-PCR detected integration of 4q28.3 with disruption to the E1 ORF. PapilloCheck® uses E1 sequence to identify different high risk types of HPV. The exact locus of HPV that is amplified as part of the PapilloCheck®

assay is not disclosed by Greiner, but it is highly likely that integration prevented the detection of HPV16 in this sample by PapilloCheck®. DIPS detected integration on Xq22.1; APOT did not detect an integrated transcript form this site and PCR or repeating the DIPS assay could not confirm the detection of integration on Xq22.1; it is likely that integration on Xq22.1 is not contributing to cancer. Integration on 4q28.3 was detected by APOT, RS-PCR and DIPS; PCR using primers flanking the integration site confirmed the detection of integration thus it is likely that this integration event is clonal and contributing to cancer. Integration on 4q28 was within an exon of solute carrier family 7 gene (SLC7A11) that encodes an amino acid transporter protein; the integrated transcript detected by APOT showed integration within an intron of SLC7A11 implying that transcription from this site does not encode a fusion protein consisting of both HPV and SLC7A11. Further work is needed to determine if integration within SCL7A11 disrupts the function or expression of the protein and contributes to cancer in sample 6N.

Sample	Age	HPV	E2 PCR[1]	RSO[2]	Viral Disruption[3]	Accession number[4]	Map[5] %	Match[6]
1W	23	HPV16	Intact	P1 BamHI	discontig seq 1116 (E1)- 6634 (L1)	NC_001526.1	na	99%
				P6 NdeI	discontig seq 5562 (L1)-707 (E6)	NC_001526.1	na	98%
				P6 EcooRI	discontig seq 5452 (L1) -7547 (L1)	NC_001526.1	na	97%
2A	33	HPV16	disrupted	not detected				
3O	46	HPV16	intact	P1 BamHI	944bp (E1)	NT_022135.16	2q22.3	97%
4T	63	HPV45	fail		not detected			
5B	30	HPV16	disrupted		not detected			
6N	45	HPV16	disrupted	P2 NdeI	2345 (E1)	NT_016354.19	4q28.3	98%

Table 52. E2 PCR and RS-PCR results for six cervical cancer biopsies. Integration sites detected are shown for each biopsy; where a match to human sequence was detected, the accession number is given with percent consensus and direction in which the HPV sequence is integrated.

[1] E2 PCR outcome.
[2] Restriction site oligo and HPV specific primer used in RS-PCR that detected integration.
[3] Viral disruption number denotes last viral nucleotide before recombination to human sequence. Numbering of HPV16 sequence is according to Genebank accession numbers NC_001526.1.
[4] EMBL Accession number for the NCBI database sequence with the most likeness to human sequence data.
[5] The genomic location of the site of integration chromosomal reference of all viral-cellular fusion transcripts with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al. 2002).
[6] Percentage consensus of sequence data with NCBI database sequence.

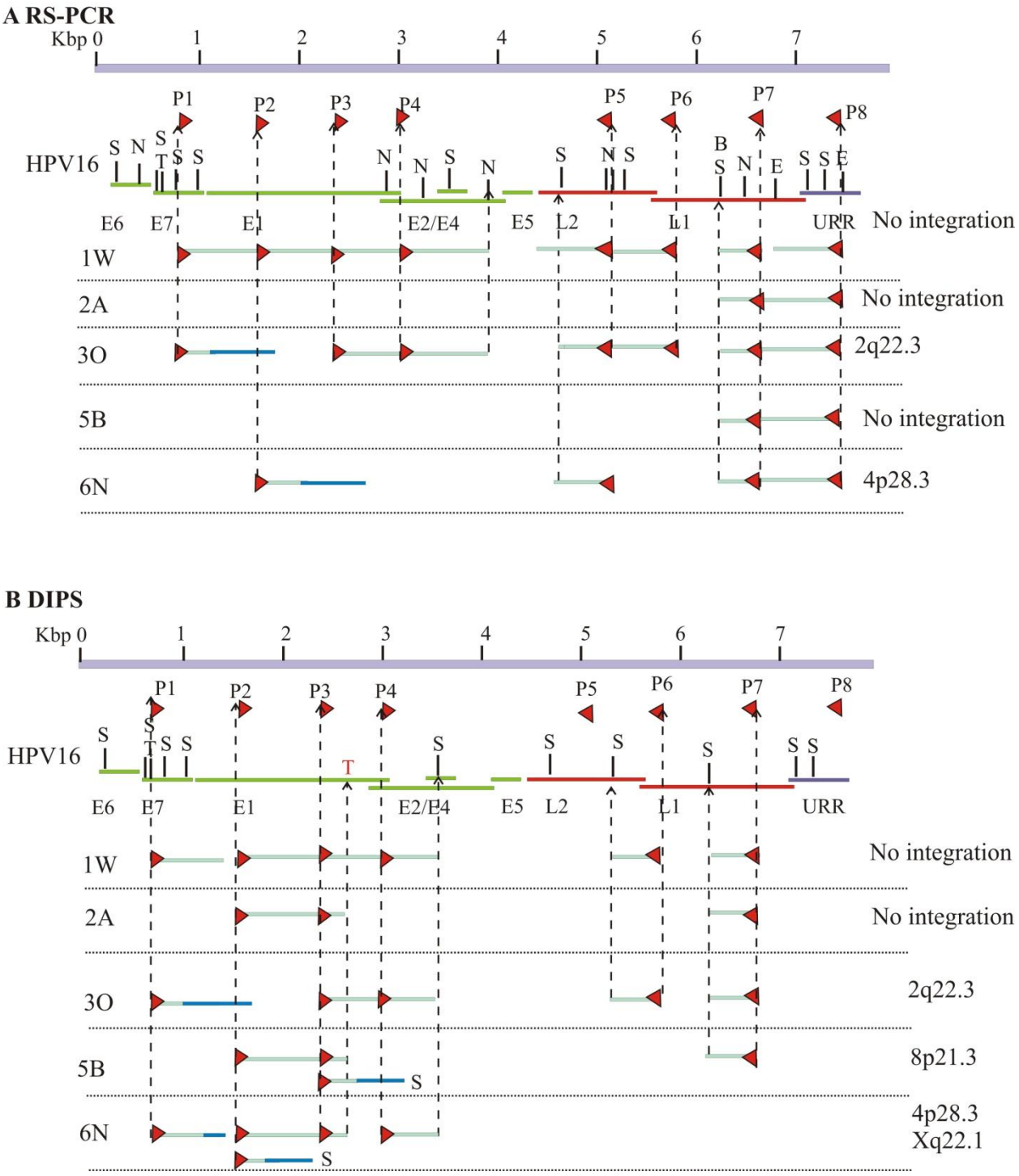


Figure 48 Schematic alignment of biopsy A) RS-PCR and B) DIPS data to HPV16 (NC_001526.1). Purple bar shows an approximate scale of HPV16 in Kbp with the point of HPV disruption for each cell line, early promoter (PE), late promoter (PL), early polyadenylation (AE) and late polyadenylation (AL) sites marked. Red triangles show approximate location of DIPS primers in relation to HPV genome in green (early genes), red (late genes), dark purple (URR) bars, with BamHI (B), EcoRI (E), NdeI(N), SauAI (S), TaqI (T) and XbaI (X) cut sites indicated. Turquoise bars show HPV DNA sequence data alignment with integrated, human DNA shown by blue bars. The site of integration is given on right hand side, "no integration" denotes no integration detected.

Sample	Age	E2 PCR[1]	Primer [2]	Viral Disruption[3]	Accession number[4]	Map[5]	% Match[6]	Orientation
1W	23	intact		Not detected				
2A	33	disrupted		Not detected				
3O	46	intact	P1 TaqI	944bp (E1)	NT_022135.16	2q22.3	90%	S
4T	63	fail		No HPV45 detected				
5B	30	disrupted	P3 Sau3AI	2673 (E1)	NT_167187.1	8p21.3	100%	S
6N	45	disrupted	P1 TaqI	1368 (E1)	NT_011651.17	Xq22.1	94%	S
			P2 Sau3AI	2345 (E1)	NT_016354.19	4q28.3	100%	S

Table 53 DIPS Biopsy Results. Integration sites detected are shown for each biopsy; where a match to human sequence was detected, the accession number is given with percent consensus and direction in which the HPV sequence is integrated**.**

[1] HPV specific primer set used in DIPS that detected integration.
2 Restriction enzyme used in DIPS that detected integration.
3 Viral disruption number denotes last viral nucleotide before recombination to human sequence or viral sequence. Numbering of HPV16 sequence is according to Genbank accession number NC_001526.1.
4 Genbank accession number for the NCBI database sequence with the most likeness to human sequence data.
5 The genomic location of the site of integration chromosomal reference of all viral-cellular fusion transcripts with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al. 2002).
6 Percentage consensus of sequence data with NCBI database sequence.

Sample	HPV	Transcript Size[1]	HPV splice/fusion site[2]	Accession number[3]	Map[4]	% Match[5]	Orientation[6]
1W	16	800bp	Episomal E7/E1^E4; Type 3	NC_001526.1	na		
		1200bp	Episomal E7/E1^E4,E5; Type 4	NC_001526.1			
2A	16	No transcripts					
3O	16	500bp	880bp (E1); Type 1	NT_022135.16	2q22.3	95%	S
		700bp	880bp (E1); Type 1	No BLAST match	2q22.3		
4T	45	No transcripts					
5B	16	No transcripts					
6N	16	600bp	880bp (E1); Type 1	NT_016354.19	4q28.3	100%	S

Table 54 APOT Biopsy results. Accession number of human sequence and orientation of HPV transcripts is shown for each transcript detected by APOT for each of the cell lines

[1] Transcript size approximate according to DNA ladder in base pairs (bp).

2 HPV splice site denotes last viral nucleotide before splicing to human sequence or viral sequence. Numbering of HPV sequence is according to Genebank accession numbers NC_001526.1, X05015 and X74479 for HPV16, HPV18 and HPV45, respectively. Transcript type is according to Figure 57, section 5.1.3.

3 EMBL Accession number for the NCBI database sequence with the most likeness to human sequence data.

4 The genomic location of the site of integration chromosomal reference of all viral-cellular fusion transcripts with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al. 2002).

5 Percentage consensus of sequence data with NCBI database sequence.

6 S=sense orientation, AS= Antisense orientation.

6.3.4 HPV Gene Expression

HPV16 E2, E7 and E6 CNRQs relative to housekeeping genes expression were calculated for HPV16 positive biopsies (Table 55 and Figure 49). Sample 5B was removed from analysis because gene stability value and coefficient of variance was outside quality control parameters levels (mentioned in section 3.3.3.8.3).

Sample	E2	Integration	Age	E2	E6	E7
1W	intact	no	23	1.55	0.48	ND
2A	disrupted	no	33	ND	ND	ND
3O	intact	yes	46	0.08	1.40	0.69
5B	disrupted	yes	30	fail	fail	fail
6N	disrupted	yes	45	ND	ND	ND

Table 55. Calibrated Normalised Relative Quantities (CNRQs) of E2, E6 and E7 in biopsy samples. CNRQ values of 1 or more indicate HPV E2, E6 or E7 expression greater than or equal to housekeeping genes. CNRQ values less than 1 indicate HPV E2, E6 or E7 expression lower than housekeeping genes. ND= not detected.

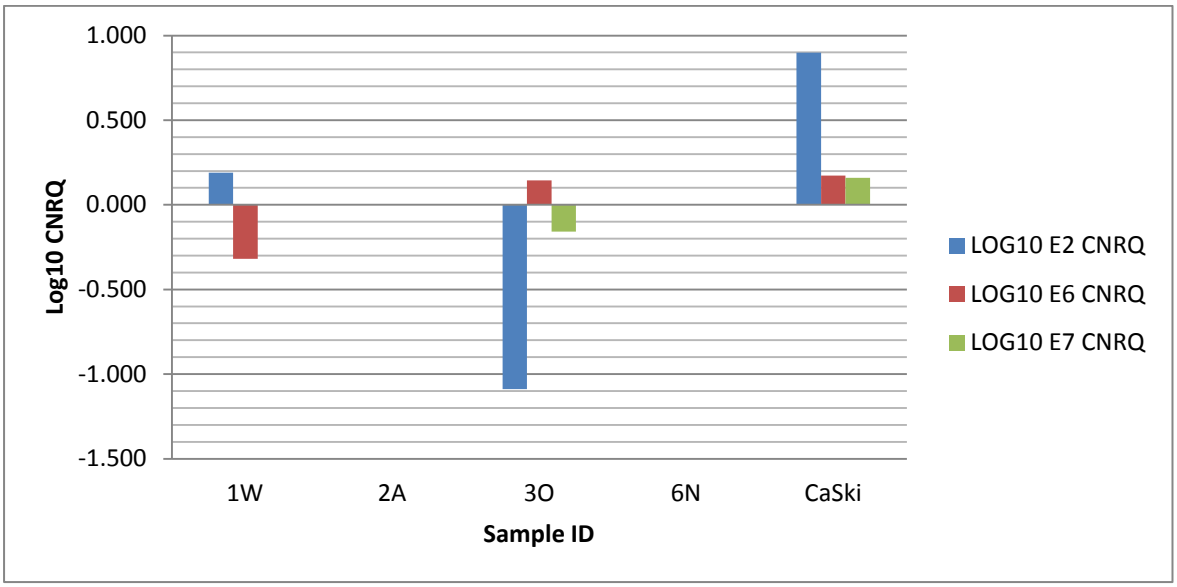


Figure 49 Log_{10} Calibrated Normalised Relative Quantities (CNRQs) of E2, E6 and E7 in HPV16 positive cervical cancer biopsies.

For sample 1W, E2 was approximately 1.5 times higher than housekeeping gene expression. E6 mRNA quantities were approximately 0.5 times the housekeeping gene mRNA quantities

and E7 was not detected. This illustrates that E6 and E7 are not required to maintain malignant phenotype in 1W.

E7, E6 and E2 were at very low levels and classed as “not detected” following qPCR data analysis in samples 2A and 6N. These data imply that although HPV16 is detected by E6 PCR, RS-PCR and DIPS, there is no evidence that HPV16 is driving the cancer.

The highest levels of HPV16 E6 and E7 expression were observed in sample 3O and E2 quantity was low relative to housekeeping genes; this is consistent with HPV E6 contributing to cancer.

6.3.5 Biopsies: Summary

These data illustrate HPV integration may not contribute to cervical cancer in women younger than 30 years of age. Integration was not detected by DIPS, RS-PCR or APOT in the biopsy from a woman aged 23 and integration was not clonal in the biopsy from the woman aged 30. Furthermore, intact E2 was detected in sample 3O indicating episomal DNA is present in this sample. These data do not support hypotheses 3 or 6: there may be another mechanism of carcinogenesis other than integration and episomal loss in this sample.

6.4 Integration in Clinical Samples: Discussion

Application of integration assays to clinical samples intended to determine if HPV integration is an early event in cervical carcinogenesis (hypothesis 4), and to determine if integration is a plausible biomarker for use in cervical screening. Furthermore, application of assays to clinical samples was done to test the hypothesis that episomal loss is associated with increasing severity of disease (hypothesis 5) and integrated transcripts combined with episomal loss would be observed more frequently in cervical cancer (hypothesis 6) . Three samples sets were used: smear samples collected from women attending their first smear prior to implementation of HPV vaccination (Baseline); smears taken from women attending colposcopy following an abnormal cervical smear (HI-BCD); cervical cancer biopsy samples.

6.4.1 E2 PCRs

E2 PCRs were applied to clinical samples to determine if episomal DNA is present at high frequency in CIN1. The data presented in this thesis illustrated that E2 PCR may not be reliable to use as a marker of E2 disruption or integration. In the study by Collins et al. (2009) E2 disruption was a marker of integration, and defined as lack of one or more E2 PCR amplicons. As mentioned in section 4.1.1, E2 PCRs were developed on cell lines with good DNA quality and known integration status; E2 PCRs produced amplicons consistent with reported integration in cell lines. When E2 PCRs were performed on clinical samples, there was evidence to suggest lack of one or more E2 PCR amplicons does not confer integration: lack of amplicon production could not be explained by poor DNA quality but could be explained by primer specificity issues or cross reactivity of primers between HPV types:

Initially, it was hypothesised that lack of E2 PCR amplicons could have been due to poor DNA quality, which would increase the rate of PCR failure and indicate disrupted E2. As part of a separate study, and data not presented in this thesis, E2 PCRs were performed on DNA from head and neck tumours that had been formalin fixed and paraffin embedded; the DNA was badly degraded but E2 was intact in all samples. This emphasises that the E2 PCR assay is successful with degraded DNA and disrupted E2, detected in the baseline samples, was not likely to be due to poor DNA quality. E2 PCR with disruption at primer 4 was most commonly observed and sensitivity tests of E2 PCRs (section 4.1.1.2) showed that primer set 4 was one of the most sensitive primer sets and able to produce an amplicon with as little as 800 copies of HPV; thus PCR sensitivity cannot explain why E2 disruption was detected in the absence of integration.

As mentioned above, E2 PCR with primer 4 most frequently failed to produce an amplicon. Variability in E2 sequence between HPV16 in clinical samples and primer sequences could explain why E2 PCR with primer 4 failed to produce an amplicon. Graham et al. (2000) reported variation within the E2 gene and three single nucleotide polymorphisms reported reside within E2 forward and reverse primer for PCR 3 (AG at 3362bp and TG at 3566bp, respectively) and reverse primer for PCR 4 (G/T at 3778bp). Thus it is possible that variability in HPV sequence at the annealing location of E2 PCR primers can result in failure

to produce an amplicon. In the HI-BCD samples, E2 PCR assay did imply that E2 disruption was present in three samples positive for HPV18 and HPV45 because E2 PCR failed to produce one or more amplicons. However, follow-up data from DIPS or RS-PCR showed that E2 was intact. The explanation for E2 failing to produce an amplicon in 2 HI-BCD samples was due to variation within HPV sequence that resulted in lack of specificity of the primer sequence to HPV. For the remaining HI-BCD sample, there was evidence to imply cross reactivity of E2 PCR primers between HPV16 and HPV18 since primer sets 1, 2 and 3 had a high consensus with HPV16 sequence.

These data strengthen the observation that E2 PCR amplicon failure does not always confer integration, but may also occur due to variation in HPV sequence or due to cross- reactivity between types. E2 PCR primer sequences used in this PhD were obtained from Collins et al. (2006) who performed E2 PCR on samples that contained HPV16, HPV18 or both; Collins et al. (2006) did not show evidence that cross reactivity between primer sets could affect the E2 PCR assay; therefore the data in this PhD are novel and indicate caution should be taken upon interpretation of E2 PCR data. In future assay design for E2 PCR, primers should be situated in regions of HPV that harbour no or minimal polymorphisms and primers should be checked for possible cross reactivity with other HPV types by checking primer specificity for other HPV types.

In conclusion, E2 PCR is not reliable to detect E2 disruption as a marker of integration. E2 PCR can be used as tool to detect intact E2, as a marker of episomal HPV. Where integration is detected, within a background of HPV in episomal form, it is likely that episomal E2 regulates HPV E6 and E7 expression. Hypothetically low-grade lesions would possess episomal HPV and intact E2 would be associated with low-grade disease. With this in mind E2 PCRs were used to detect intact E2 as a marker of episomal HPV and where disrupted E2 was detected, DIPS was performed to confirm integration.

6.4.2 Loss of episomal HPV is not associated with high grade neoplasia

Collins et al. (2009) reported integration as an early event in cervical disease progression. In the Baseline sample, intact E2 was observed in the majority of samples and there was no

evidence to support intact E2 being associated with grade of neoplasia or grade of cellular dyskaryosis. Disrupted E2 was detected most frequently in CIN2. When CIN2 and CIN3 were grouped together and defined as high-grade squamous intraepithelial neoplasia, disrupted E2 was detected more frequently in high-grade disease compared to low-grade disease. This was statistically significant but when correcting for multiple testing, this value was no longer statistically significant. Furthermore, power calculations indicated the sample size was insufficient to draw a definitive conclusion and a larger sample would be required to achieve a statistically significant outcome with a false positive rate of 20%. The maximum number of samples that met the study criteria were tested thus it was not possible to perform E2 PCR on more samples. In future work, a larger sample set (n=196) should be assayed.

For the HI-BCD samples E2 was intact in all the samples where HPV16, HPV18 or HPV45 was detected. All samples that had CIN3 diagnosed by histology and were positive for HPV16, HPV18 and HPV45 had intact E2.

The data in this PhD are not directly comparable with the Collins et al (2009) study. Collins et al (2009) did not observe the histological outcome and did not present any cytological data in relation to E2 disruption; instead, in the study of Collins et al. (2009), the population investigated comprised "the subset of women who were cytologically normal and HPV DNA negative at study entry and who first tested positive during follow-up for HPV16 or HPV18, or both". E2 disruption as an "early event" is defined as time between incident HPV16 or HPV18 detection and disrupted E2 detection. The age of the women analysed by Collins et al. (2009) was between 15 to 19 years old and they were younger than the women in the Baseline cohort, nonetheless the time between detecting E2 disruption after incident infection was 5.7 months and since it was detected in such a young age group, it was considered to be an early event.

Finally, as mentioned above, E2 disruption detected by E2 PCR requires confirmation by an alternative method such as DIPS, which was not done by Collins et al. (2009). One of the strengths in this PhD was that DIPS was used to determine if integration were present and causing E2 disruption.

6.4.3 Integration is not an early event in cervical neoplasia

DIPS was performed on a selection of HPV16, HPV18 and HPV45 Baseline samples with CIN1, CIN2 or CIN3 to determine if integration is an early event. Also, DIPS was performed on HPV16 samples where E2 disruption had been implicated by E2 PCR. Where integration was detected, intact E2 was present. Thus episomal HPV is likely to exist with integrated HPV in the Baseline samples. For samples that indicated disrupted E2 by E2 PCR, no integration was detected; this supports the observation that E2 PCR data should not be used as a marker of integration.

Overall there was no evidence in the Baseline samples that integration was an early event present in CIN1. Integration was detected but in a very small proportion of the samples and any integration detected could not be replicated using PCR or by repeating the DIPS assay. Section 4.1.3.5 illustrates that DIPS is a sensitive assay and can detect integration in as little as 12 copies of HPV, in a clonal population of SiHa cells. Therefore, the fact that integration could not be replicated is not a weakness of the PhD methodology, instead it highlights that integration may not be a clonal event since if it were, integration detection would have been replicated. With the exception of one study, which used PCR primers designed to flank the host-viral junction, to confirm integration detection (Dall et al. 2008), confirmation of integration is not reported in published data; thus attempts to confirm integration is a strength of this PhD.

The DNA quality of Baseline samples was poor because the samples had been collected in SurePath™ preservative fluid and the DNA had degraded substantially. Nonetheless Beta-globin PCR success was high and when DIPS was performed on DNA from SiHa that had been stored in SurePath™, integration was detected; thus DIPS was capable of detecting integration in degraded DNA.

As mentioned above, DIPS is sensitive enough to detect integration in degraded DNA thus failure to detect integration in the Baseline samples cannot be attributed to lack of sensitivity of DIPS. Control PCR had a high failure rate and this was due to poor DNA quality. DIPS detected integration in SiHa stored in Surepath™, however, there was still an element of

uncertainty about whether DNA quality had an effect on integration detection and ability to replicate integration. Collection of samples in SurePath™ was unsuitable for RNA stabilisation too, thus RNA based assays such as APOT and qPCR could not be performed on Baseline samples.

Integration assays were too expensive to perform on the entire Baseline sample set (n=14,128) and sample numbers were selected that would give at least 80% power (a maximum chance of false positive detection at a rate of 20%) according to the frequency of integration in CIN3 and cervical cancer taken from Klaes et al. (1999). Samples were selected according to type, single infection status and approximately equal numbers of each histology grade were chosen. HPV16 samples were all from women with single infections but the numbers of HPV18 and HPV45 samples with single infections were low. For HPV16 samples, it was highly unlikely that integration was occurring for another HPV type but for HPV18 and HPV45 samples with mixed infections, it remained possible that another type could have been integrated and contributing to the disease. Investigating HPV types other than HPV16, HPV18 and HPV45 would involve developing assays for each type and this was beyond the scope and time frame of this project. Rachel Houghton, a PhD student within the HPV research group is investigating integration for HPV31, HPV51, HPV33 and HPV35; this is currently work in progress.

A weakness of using the Baseline sample set is that the histological outcome is not directly linked to the cervical smear sample taken. Firstly, women with moderate or severe dyskaryosis have a shorter time period before attending colposcopy. This allows less time for changes in disease before being seen by a clinician. In this case, the biopsy taken in colposcopy is likely to represent disease contained in the cytological sample used for the integration assays. Women with mild or borderline dyskaryosis have a longer period of time before attending colposcopy; allowing more time for disease to regress or progress. Therefore, smear samples from women with borderline and mild dyskaryosis, assayed in this PhD, may not represent the final outcome of histology. To overcome this, separate analyses were performed on the samples taken from women who had moderate or severe dyskaryosis because the samples taken at their first smear were most likely to represent the nature of the

cervical lesion when a biopsy was taken at colposcopy; there was still no evidence that integration was an early event present in CIN1.

Women who had attended colposcopy were selected for this study; this was to enable a direct comparison of integration and cervical neoplasia. Selecting women that had attended colposcopy can bias a study towards a higher severity of disease, since women without dyskaryosis are excluded. Due to integration not being detected at a high rate in this sample set, bias was not considered to be an issue; however, it would be advantageous in future studies to include women not attending colposcopy.

To overcome the problem of poor DNA quality and lack of RNA, a pilot study was proposed by the author and smear samples were collected in a reagent designed to preserve RNA and DNA.

6.4.3.1 Smear samples: HI-BCD

The idea for collection of samples was instigated by the author. The purpose of collecting HI-BCD samples in addition to the Baseline samples was to gather a sample collection with good DNA and RNA integrity to determine whether HPV Integration is a Biomarker of Cervical Disease (HI-BCD) and to compare DIPS, APOT and RS-PCR for detection of integration.

Overall, there was no evidence to suggest integration, combined with loss of episomal HPV, can be used as a biomarker of cervical disease since integration was not detected in CIN3 and all samples, where HPV16, HPV18 or HPV45 were present, possessed intact E2. Furthermore, a comparison of the integration assays was not warranted due to the lack of integration detection.

DIPS detected integration in 3 HI-BCD samples but integration could not be confirmed by PCR using primers that flanked integration sites and integration detection could not be replicated by DIPS; these data mirror the observations made in the Baseline sample.

Integration was not associated with CIN3 or with severe cellular dyskaryosis thus there is no evidence in this sample set that integration, combined with episomal loss, is a useful biomarker of cervical disease. The DNA quality of HI-BCD sample was excellent, there was little DNA degradation detected and protein and co-purified contaminants were low; DIPS control PCR produced amplicons of expected size and DIPS detected integration in positive controls; therefore, low frequency of integration detection cannot be explained by poor DNA quality. Additionally, as mentioned above, DIPS is a sensitive assay and low frequency of integration detection cannot be explained by lack of assay sensitivity.

APOT did not detect integrated transcripts in any of the samples. Therefore any integration, detected by DIPS, was not being actively transcribed; further supporting integration not being clonal and contributing to cervical disease in this sample set. HI-BCD RNA was degraded; this was most likely due to the nature of the sample: a smear takes the top layer of cells from epithelium where RNA is likely to be degraded because the cells are dead and due to be shed as part of the natural epithelial life cycle. To determine whether RNA degradation inferred poor cDNA quality, GAPDH PCR was performed to determine cDNA integrity; GAPDH PCR produced an amplicon for all the samples. Although GAPDH PCR implied adequate cDNA, APOT data should be interpreted with caution because degraded RNA may explain why APOT did not detect integrated transcripts.

Episomal transcripts were detected and consisted of E7, E1^E4 or E7,E1^E4,E5; no episomal transcripts contained E2. In a productive infection, in the upper regions of epithelial layers, E2 is not typically expressed but E4 is required to break down the cytokeratin network to allow release of replicated virions. The pattern of low E2 expression detected by qPCR, combined with E4 transcripts, detected by APOT suggests that HPV16 may be present as a productive infection in HIBCD2, HIBCD7, HIBCD8 and HIBCD9. HIBCD7, HIBCD8 and HIBCD9 had mixed HPV infections thus it remains plausible that another HPV type is contributing to disease in these samples.

HIBCD14 was from a woman with cervical cancer and CIN3 with multiple HPV infections. No integration was detected in this sample and very high levels of E2 expression were

detected. The data for HIBCD14 suggests that HPV16 integration does not play a part in carcinogenesis. It remains possible that HPV59 or HPV66 is causing cancer in this sample.

HIBCD1 was taken from a woman referred to colposcopy with severe cellular dyskaryosis and CIN3 was confirmed by histological examination. HPV was not detected in this sample by E6 PCR, by PapilloCheck® or GP5+/6+PCR-EIA. It is possible that a type other than HPV16, HPV18 and HPV45 is integrated and HPV disrupted at the points where primers for PapilloCheck® and GP5+/6+PCR-EIA anneal to; this would result in failure to detect HPV in this sample. It would be advantageous to perform E6 PCR, specific for HR-HPV types other than HPV16, HPV18 and HPV45, to determine if HPV is present in this sample.

The main drawback of the HI-BCD study was the number of patients recruited such that numbers were below those required for statistical analyses. A limitation in using smear samples is that many HPV types are present in a sample producing a "needle in the haystack" search for the culprit HPV type causing disease progression. Additionally, a smear takes the top layer of epithelia from the entire cervix and it is possible that the presence of non-diseased cells alongside diseased cells has a diluting effect and reduces the chance of integration detection. Biopsies allow the entire diseased area to be assessed and because the majority of the biopsy is composed of diseased cells, detection of integration is hypothetically increased.

6.4.4 Assay Comparison in Biopsy Samples

To allow an assessment of the integration assay ability to detect integration in biopsies, APOT, DIPS, E2 PCR and RS-PCR were performed on 6 cervical cancer biopsies. Five samples were HPV16 positive, with single infections. One sample was HPV45 positive but failed all assays due to poor DNA and RNA quality. Although insignificant numbers of samples were tested to make a definitive conclusion, the ability of an assay to detect integration in biopsy samples depended on the nature of integration. If integration was clonal and being transcribed then APOT, RS-PCR and DIPS detected integration. If integration was not clonal then DIPS detected it by chance but it was not replicated by E2 PCR or RS-PCR or by repeating DIPS. DIPS detected a higher proportion of integration events in biopsy

samples, highlighting the importance of the input sample for optimal integration assay performance and reproducibility

Overall, HPV16 integration was detected in 3 samples and integration was confirmed by PCR for integration sites in 2 samples both of which were also detected by APOT. It is likely that integration on 2q22.3 in sample 3O and on 4q28 in sample 6N is clonal because integration in DNA was confirmed by DIPS, RS-PCR and by PCR; the fact that APOT detected integrated transcripts in 3O and 6N show that the sites are transcribed, giving the cell selective growth advantage. Integration on 8p21.3 in sample 5B and on Xq22.1 in sample 6N was not detected by RS-PCR or by APOT and was not confirmed by PCR thus these integration events are not likely to be clonal. For samples 1W and 2A, no integration was detected by DIPS, RS-PCR or APOT thus integration is not likely to be contributing to disease in these samples. Non-contiguous sequence was detected by RS-PCR in sample 1W but was not detected by DIPS however, PCR confirmed the non-contiguous junctions. In theory, DIPS with primers 6 and 7 should have detected non contiguous sequence but wild-type sequence was amplified by DIPS; this implies that intact wild-type DNA is present alongside non-contiguous sequence but it remains unknown whether HPV is integrated or in episomal form.

E2 PCR detected intact E2 in 1W and 3O; this is consistent with DIPS and RS-PCR sequence data however, it is unclear whether episomal HPV is present in 1W or 3O. Non-contiguous sequence was detected in 1W; this is similar to the detection of non-contiguous HPV in CaSki, where E2 is intact and HPV is integrated in concatenated form. Integration may be in a concatenated form in 1W, with E2 remaining intact. APOT detected only episomal transcripts in 1W if integration were present in 1W, it is not being transcribed. For sample 3O, integration disrupted E1 but no disruption to E2 was detected, thus it remains possible that E2 remained intact in the presence of integration.

Integration was detected and E2 PCR detected disrupted E2 in 5B and 6N; this is consistent with presence of integration without episomal HPV. As mentioned above, disruption to E2 requires confirmation by DIPS or RS-PCR because failure to produce an E2 PCR amplicon may be due to variation in HPV16 sequence leading to reduced primer specificity. Data for

2A is inconclusive since integration was not detected by DIPS, RS-PCR or APOT although E2 was disrupted.

6.4.5 Integrated transcripts are not present in all cervical cancers and are not associated with episomal loss.

The above data do not support the hypothesis that integration results in loss of E2 expression and increased E6 and E7 expression (hypothesis 3). Each of the biopsies presented with differing data and there was no consistent pattern that would imply integration or episomal loss in cervical cancer. Initially it was planned to sample over 100 biopsies but access to biopsies was limited thus 6 samples were selected. To replicate the data presented in this thesis DIPS, E2-PCR, DIPS and APOT would need to be performed on a larger sample.

Biopsy 1W had no integration and only episomal transcripts, consisting of E7/E1^E4 and E7/E1^E4, E5 with E2 being intact. qPCR detected mRNA quantities of E2/E4 in excess of the house keeping gene expression and low levels of E6 mRNA was detected. Although no integration was detected in 1W, the expression data follows a pattern of episomal transformation similar to that of Gray et al (2010) who reported a series of W12 ($W12_{ser4(EPI)}$) that possessed no integrants yet had transforming abilities with high E6 and E2 protein levels and low levels of E7. Gray et al (2010) also reported that high copy numbers of episomes were maintained in $W12_{ser4(EPI)}$. E2 is vital for HPV episome maintenance and this would explain a role for E2 expression in transformation. HPV16 E6 acts through ubiquitination of p53 to suppress apoptosis; increased E6 in $W12_{ser4(EPI)}$ would lead to decreased p53 and suppress apoptosis. Clonal selection in 1W may result from episomal maintenance and decreased apoptosis through E2 and E6 expression. HPV copy numbers were not estimated in 1W to investigate how HPV copy numbers relate to HPV gene expression. In future studies of cervical cancer with episomal HPV, it would be advantageous to estimate HPV copy number in relation to HPV gene expression, HPV protein and human protein quantities, to allow further understanding of episomal transformation. No E7 expression was detected in 1W; this is inconsistent with APOT data that detected transcripts containing E7. It is possible that APOT is more sensitive than qPCR, due to a nested PCR reaction being employed by APOT.

For biopsy 2A, it is unlikely that HPV16 is maintaining the malignant phenotype in this sample because integration was not detected by RS-PCR, DIPS or APOT and qPCR did not detect E2, E6 or E7 mRNA. HPV16 was the only type present in biopsy 2A and it is unlikely that another type is contributing to cancer in this sample. It is possible that before transformation occurred in this sample, high levels of HPV E6 and E7 expression damaged the cellular DNA; this potentially would lead to chromosomal abnormalities that lead to selection of a cell without the need for HPV oncogene expression to maintain the transformed phenotype. This hypothesis requires further exploration by a method such as whole genome sequencing that could detect differences in chromosome structure between cancer tissue and stromal tissue taken from biopsy 2A. This observation would be consistent with Gray et al. (2010) who reported a rapid decrease in the expression of HPV E6, E7 and E2 when $W12_{ser4(EPI)}$ grown in raft culture became invasive and would highlight that E6 and E7 is not required to maintain malignant phenotype.

Biopsy 3O had a different mechanism of selection to 1W and 2A. It is unclear if episomes were present in 3O but integration was detected and integration was not within a human gene; it is unlikely that the integration event contributed to selection through insertional mutagenesis in this sample. Integration from this site was clonal, because DIPS, PCR and APOT detected integrated transcripts. There was no evidence to confirm disruption of HPV URR thus it is likely that integrated transcripts are being generated from the HPV early viral promoter. qPCR detected very small quantites of E2 mRNA expression and quantities of E6 were higher than housekeeping genes; thus E6 expression is required to maintain the transformed phenotype in this sample. Expression of E2 may be suppressed due to epigenetic factors such as methylation (see section 8.4.1); this would explain why intact E2 is detected by E2-PCR and little E2 expression was detected by qPCR. Increased E6 and E7 expression would confer a strong selection advantage through suppression of apoptosis though hTERT, p53 and pRB pathways described in the introduction. Furthermore, as mentioned in the introduction, integrated transcripts have a longer half life compared to mRNA of episomal transcripts and stability of mRNA would enhance the selection of a cell harbouring an integrant (Jeon and Lambert 1995).

In biopsy 5B DNA integration was detected on 8p21, but there was no detection of integrated transcripts from this site; PCR did not confirm the site of integration and it is likely that integration is not clonal in this sample and contributing to cancer. Unfortunately, qPCR data for 5B did not fall within the quality control parameters and the data was not included in this study; qPCR was not repeated on this sample due to limited time and financial constraints; it would be worthwhile repeating qPCR on 5B to determine if HPV oncogene expression is likely to be contributing to cancer in this sample.

In biopsy 6N, no E2, or E7 expression was detected and E6 expression was below that of housekeeping genes thus there was no evidence that HPV gene expression was contributing to cancer in this sample. Integration was within an exon of solute carrier family 7 gene (SLC7A11) that encodes an amino acid transporter protein that controls the transportation of glutathionine (GSH) across the cellular membrane. SLC7A11 has been implicated in chemo resistance and maintains cellular GSH as a chemo resistance mediator to certain drugs; reduced SLC7A11 would lead to increased GSH and lead to chemoresistance (Huang 2005). Disruption to the SLC7A11 gene due to integration would confer a selective growth advantage especially if chemotherapy was used; unfortunately, the clinical history of patient 6N was unknown. Further work is needed to determine if integration within SCL7A11 disrupts the function or expression of the protein and contributes to cancer in sample 6N.

6.5 Integration in Clinical Samples: Conclusion

There was no evidence to support that integration is an early event, present in CIN1 (hypothesis 4) nor was there evidence to support episomal loss is a late event associated with CIN3 (hypothesis 5). There was an interesting link made in this study between integration detection and age of women. It is possible that integration is an event that is detected more frequently in women over the age of 30. This hypothesis requires testing in a large sample set comprised of women of varying ages.

7 Mechanism of Integration

HPV integration occurs at a unique site each time and there are consistent reports that HPV integration occurs at chromosome fragile sites. Fragility at chromosome fragile sites may be attributable to tandem repeats that slow DNA repair. In this section, integration sites in relation to DNA fragile sites and repeat elements were explored to determine if integration is a random event or if DNA regions such as repeat elements and CFS are more prone to HPV integration. Non-homologous end-joining DNA repair is the mechanism reported to be involved incorporating the HPV DNA into the human DNA; DNA sequence, at the site of integration, was examined to determine whether there was identical or differing sequence between human and HPV DNA to gain insight into the mechanism behind HPV integration. Finally, it is reported that HPV integration commonly disrupts the HPV genome within the E2 and E1 ORF; the site at which integration disrupted the HPV genome was examined to test this theory.

7.1 Mechanism of Integration: Results

To test the hypothesises that integration disrupts the HPV genome at E1 and E2 ORFs (hypothesis 1); integration occurs at DNA fragile sites (hypothesis 8) and NHEJ is involved in integration (hypothesis 9), DNA sequence analysis was performed.

Sample	HPV [1]	Map[2]	FS[3]	RPT[4]	Details (I/OL)[5]
CaSki	E1;1455	Xq27.3	FRAXA (R)	SINE	TAC (O)
		6p21.1	none	na[6]	na[5]
SiHa	E2;3132	13q22.1	none	LTR	TAGC (I)
W12[7]	E1;1756	8q24.21	none	DNA	CATTATATTT (I)
	E2;3197	12q14.3	none	LINE	45bp (I)
W12p12	E1;2749	2q35	none	LINE	A (O)
W12p32	E2:3726	2q31.1	FRA2G	LINE	GACAT (O)
HeLa	E1;2497	8q24.1	FRA8C/E(C/R)	LINE, SINE	ATGTTA (O)
	E2;3100	8q24.1		SINE	AAACA (I)
	L2;5736	8q24.1		SINE, LTA	ATAA (O)
C4-I	E2;2952	8q21.3	none	SINE	CAGGGTTCTGTTCTCACT (I)
HTB-34	E1;1878	18q11.2	none	ALU	ACC (I)
3O	E1;944	2q22.3	none	AT rich	TGCCC (I)
5B	E1;2673	8p21.3	none	SINE	TATGAGC (I)
6N	E1;1368	Xq22.1	FRAXC (C)	LINE	TGCA (O)
6N	E1;2345	4q28.3	none	DNA	G (O)
PCO9,D	L2;5003	18p11.31	none	AT rich	GATGTACAAGTTTGCACCTGTCT (I)
PCO9,D	E2;3303	18p11.31	none	SINE	AAAATAAAGTATGGGAAGTTCATGTC (I)
PCO9,H	E1;2490	11p15.3	none	LINE	Flush
PCO9,H	E1;1848	22q12.3	FRA22B (C)	SINE	Flush
PCO9, Het	E2;3632	2q36.1	none	SINE	AAGCACT (O)
PCO9, Het	L1;6033	2q36.1	none	LINE	AAG (O)
PCO8,M,	E1;1194	3q28	FRA3C (C)	LTR	TG (I)
PCO8,Y,	E1;2116	3p21.31	none	ALU	CCC(I)
	E2;3167	3p21.31	none	DNA	ATGTGGCTC (I)
265 (BL)	E2;3376	5q22.5	none	LINE	TATCTCTC (O)
265(BL)	E2;3420	15q15.1	none	SINE	GACC (O)
696(BL)	E2;3206	unknown	unknown	LINE	GGGTC(O)
913(BL)	E2;3098	Xp11.1	none	none	GATA (O)
913(BL)	L1;6543	9p13.3	none	none	GCAT (O)
1513(BL)	L1;6478	8q21.13	none	LTR	C (O)
1571(BL)	E1;1781	10p12	none	SINE	GATG (O)
1571(BL)	E2;3205	17q21.33	none	DNA, SINE	GGTC (O)
1571(BL)	E2;3080	8q11.23	none	LTR	GATG (O)
3577(BL)	E2;3206	3p14.1	FRA3B (C)	LTR	GGTCA (O)
13290(BL)	E2;3080	16q11.2	none	AT rich	GATG (O)
13290(BL)	E2;3080	21q21.1	none	(T)n repeat	GATG (O)
3220(BL)	E1;1789	12q24.11	FRA12E (C)	SINE	GAGCC (O)
12282(BL)	E1;1551	18q22.3	none	LINE	CAGC (O)
10295(BL)	E1;1551	Xp22.2	none	SINE, LINE	GAGC (O)

Table 56 Summary of integration sites detected. Details of sample ID (BL=Baseline), HPV disruption, fragile site, repeat elements and sequence at site of integration is shown.

1 EMBL Accession number for the NCBI database sequence with the most likeness to human sequence data.
2 The genomic location of the site of integration chromosomal reference of sequence data with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al. 2002).
3 Fragile site within 5Mbp of integration site. C=common, R=rare, none=no fragile site
4 Repeat elements within 2Kbp of the integration site. SINE= short interspersed nuclear element; LINE= long interspersed nuclear element; DNA= DNA transposon; LTR= long transposed region.
5 Details at the site of integration: DNA overlap (O) or inserted (I); sequence given.
6 Not applicable due to transcript integration site not DNA integration site.
7 Data for both W12p12 and W12p32.

Sample	HPV [1]	Map[2]	FS[3]	RPT[4]	Details (I/OL)[5]
HIBCD8	E1; 2431	11p12	FRA11E (C)	SINE	TAGATGA (O)
HIBCD8	E1; 1903	Xp11.4	none	SINE	ATAG (O)
HIBCD8	L2; 4540	8p11.21	none	LTR	TGGGT (O)
HIBCD8	L2; 4541	12q23.2	none	SINE	GGTC (O)
HIBCD9	E1/2;2877	6q15	FRA6G (C)	SINE	flush
HIBCD19	L2; 4479	9q21.3	none	SINE	CTAT (O)

Table 56 continued. Summary of integration sites detected. Details of HPV disruption, fragile sites, repeat elements and sequence at site of integration is shown.

1 EMBL Accession number for the NCBI database sequence with the most likeness to human sequence data.
2 The genomic location of the site of integration chromosomal reference of sequence data with respect to Giemsa-stained bands was taken from the UCSC database (Kent et al. 2002).
3 Fragile site within 1Mbp of integration site. C=common, R=rare, none=no fragile site
4 Repeat elements within 2Kbp of the integration site. SINE= short interspersed nuclear element; LINE= long interspersed nuclear element; DNA= DNA transposon; LTR= long transposed region.
5 Details at the site of integration: DNA overlap (O) or inserted (I); sequence given.

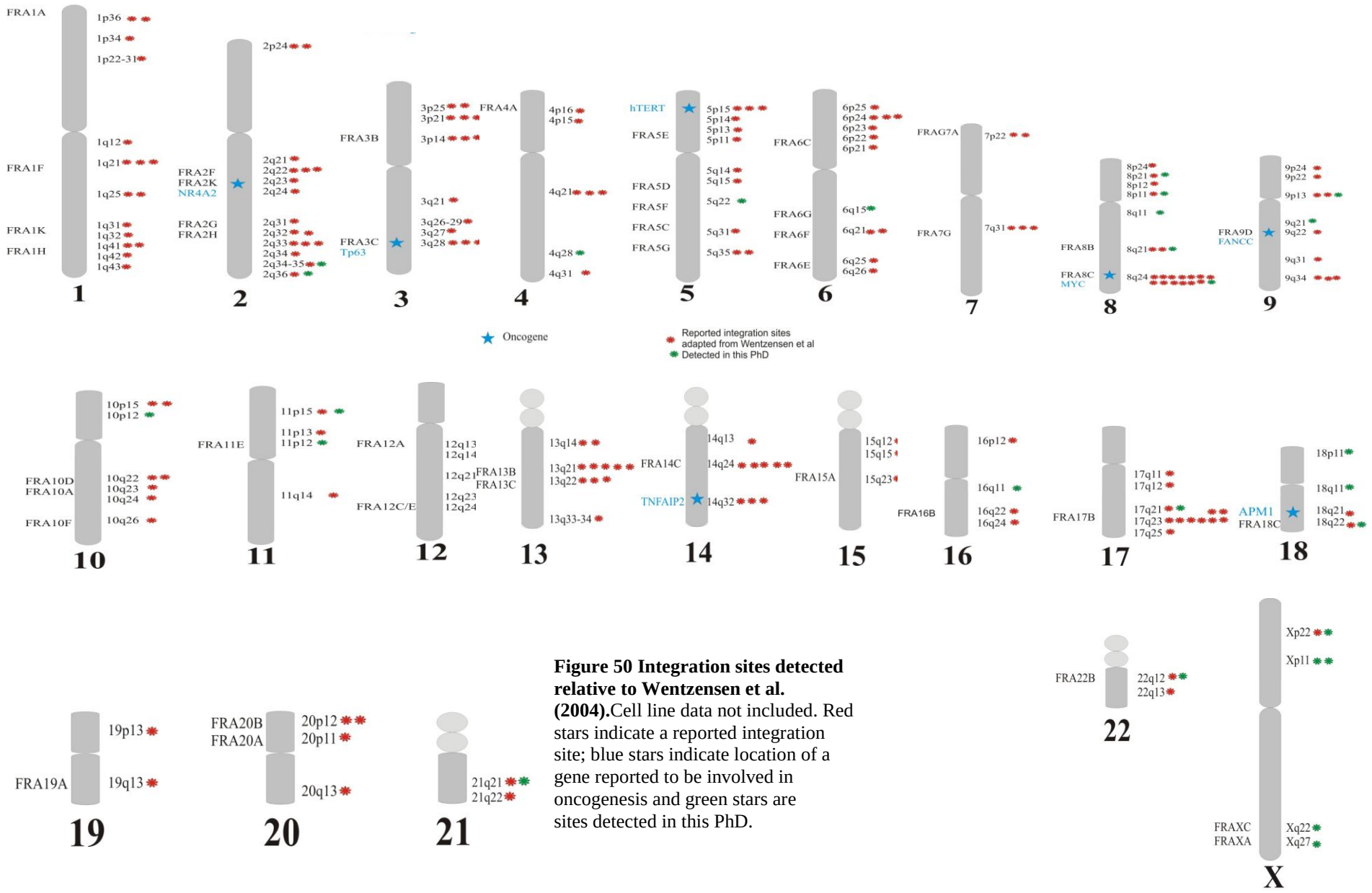


Figure 50 Integration sites detected relative to Wentzensen et al. (2004). Cell line data not included. Red stars indicate a reported integration site; blue stars indicate location of a gene reported to be involved in oncogenesis and green stars are sites detected in this PhD.

7.1.1 Sites of HPV integration

In total, 40 different integration sites were detected (Figure 51 and Table 56). Integration occurred at a unique site each time: no integration event was the same and there were no hot-spots where integration occurred within the same band. As mentioned in the introduction, Wentzensen et al. (2004) constructed a review of HPV integration sites in cell lines, vulval, head and neck and cervical cancer. HeLa, SiHa, C4I and CaSki integration sites had already been included in the analysis of Wentzensen et al. (2004) and further integration data for HeLa, SiHa, C4I and CaSki from this PhD were not included in a comparison to Wentzensen et al. (2004). When data were compared to the study of Wentzensen et al. (2004) (Figure 51), 20/35 (60%) integration sites were in same bands as previously reported. There are a total of 315 entire bands on UCSC database (for example counting 22q11.1, 22q11.2 and 22q11.3 as one band on 22q11) and Wentzensen et al. (2004) detected integration in 118 of them. Twenty integration events were detected in the same bands as those reviewed by Wentzensen et al. (2004), this is statistically different from the mean ($Z=2.44$, $p=0.0073$; mean=13.11; 95% CI=12.16-14.15). The power to detect a false positive result in a sample of this size is 0% (assuming mean=13.11; 95%CI=12.16-14.15 and $\alpha=0.05$). These data imply that although no hot-spots for integration were detected in this PhD, there are hotspots for integration when combining data with published data.

7.1.2 Integration at Fragile sites

Integration was detected at three sites on 8q24 in HeLa, at 2 sites on 18p11.31 in PC09 clone D and at 2 sites on 3p21.31 for PC08 clone Y; for fragile site analysis purposes these were classed as one event for each sample; 40 samples were included in the overall analysis. Ten integration sites occurred within known fragile sites; 9 were within a common fragile site, 1 was within a rare fragile site; this is statistically different to the mean ($Z=3.96$, $p=0.0001$; mean=3.2; 95%CI=2.97-4.02). There would be 0% chance of a false positive result in a sample this size (assuming mean=3.2; 95%CI=2.97-4.02). These data suggest there is a strong correlation between fragile sites and integration; hypothesis 8 is supported.

7.1.3 Integration at DNA repeat elements

The majority of integration sites detected fell within repeat elements. These are novel data and there are no published reports of association of integration at repeat elements. To determine whether integration was random or associated with repeat element a "goodness of fit" Chi squared analysis was performed.

CaSki integration site on 6p21.1 was detected by APOT and the exact integration site in DNA was not known; CaSki was not included in DNA repeat element analysis. Under the assumption that 70% of the human genome consists of repeat elements, 32/45 integration sites would be expected to be detected within a repeat element if integration were random. Forty-three (95%) integration sites were within a repeat element; this is a statistically significant difference to the expected (Goodness of fit, with Yates' continuity correction, χ^2 =8, df=1, p=0.005). In a sample of this size (n=45, ϕ_c=0.42, α=0.05, df=1), the chance of detecting a false positive result is 20%. Additionally, when correcting for multiple testing, with Bonferroni correction (number of tests performed in section 3.4 was 5), this p value remained statistically significant (p=0.02). These data show that repeat elements play a role in HPV integration.

Forty-five percent of the human genome is composed of transposon elements (Lander et al. 2001). Thirty seven (82%) of the integration events were within transposons: short interspersed repeat elements (SINE), long interspersed repeat elements (LINE), DNA transposons or long transposable regions (LTR); this is statistically significant from expected (Goodness of fit, with Yates' continuity correction, χ^2 =12.24, df=1, p=0.0005). In a sample of this size (n=45, ϕ_c=0.52, α=0.05,df=1), the chance of detecting a false positive result is 7%. Additionally, when correcting for multiple testing, with Bonferroni correction (number of tests performed in section 3.4 was 5), this p value remained statistically significant (p=0.0025). These data show that transposon repeat elements play a role in HPV integration.

7.1.4 DNA sequence at site of integration

At 63% of integration sites there was overlapping sequence of DNA that was the same between HPV and human. Thirty percent of integration sites had DNA inserted, this DNA was not HPV or human DNA; in 3 of these sites, a long insertion of DNA was detected; there was no similarity or consensus sequence for long DNA insertions at the site of integration. At 3 of the integration sites the sequence was flush: HPV sequence adjoined human sequence with no overlap or insertions. These data imply that integration is common where there is similarity between human and HPV sequence; this is a trait of non-homologous end-joining DNA repair. These data support hypothesis 9.

7.1.5 HPV disruption

To examine hotspots for HPV disruption, and to test hypothesis 1, sites of HPV disruption were compared.

The integration site on 6p21.1 in CaSki was detected by APOT, the point of disruption in DNA was not known thus CaSki was not included in HPV disruption analysis. The majority of integration sites involved disruption of E1 ORF (45%); E2 was disrupted at 36% of the integration sites and L1 and L2 were disrupted in 7% and 11% of integration sites, respectively. Assuming 37% of the HPV genome is either E1 or E2 ORF the detection of 36 integration events with disruption to E1 or E2 is statistically significant from the mean ($Z=6.16$, $p<0.0001$, mean=16.28, 95%CI=15.33-17.22). Assuming 36% of the HPV genome constitutes either L1 or L2 ORF: the detection of 8 integration sites with disruption to the L2 ORF or L1 ORF is statistically significant ($Z=-2.48$, $p=0.006$, mean=15.94, 95%CI=14.99-16.88).

When the precise location of HPV disruption in base pairs was examined by dividing the E1 and E2 ORF into 10x 300bp sections (Figure 51), the region between 2966bp and 3265bp was more prone to disruption and less likely to occur randomly ($Z=4.11$, $p<0.00001$, mean=3.6, 95%CI=3.01-4.08) (Figure 51). Seven of the integration events detected in the baseline samples had disruption between 2966-3265. As mentioned in section 6.1.3.2, star

activity and poor DNA quality were observed in the baseline samples and disruption seen in these may not be a result of true integration but due to star activity of Sau3AI restriction digest and ligation reaction. When the baseline samples, with disruption between 2966-3265, were excluded from the analysis, HPV disruption was randomly distributed across the E1 and E2 ORF ($-1.96<Z<1.96$; $p<0.05$; mean=3.6, 95%CI=3.01-4.08).

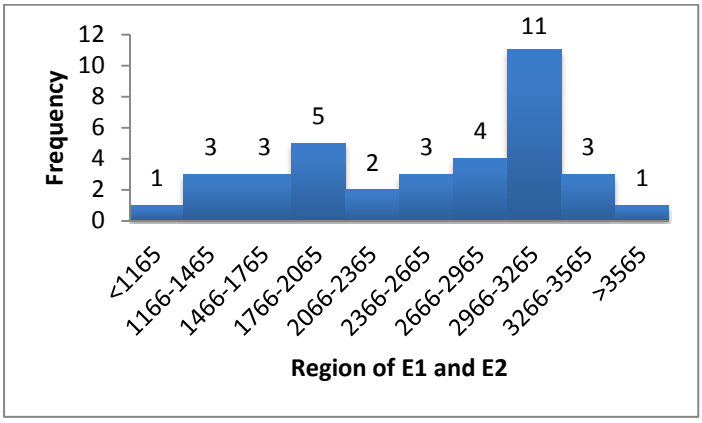


Figure 51 Frequency of HPV disruption across E1 to E2 ORFs.

7.2 Mechanism of Integration and disease progression: Discussion

7.2.1 Integration was detected at fragile sites.

Integration detected in this PhD was at a unique locus every time. When integration in relation to fragile sites was investigated, to test hypothesis 8, integration was not a random process but had affinity for chromosome fragile sites. These data support the previous studies of integration reporting integration at fragile sites (Cannizzaro et al. 1988; Wilke et al. 1996; Thorland et al. 2000; Wentzensen et al. 2002; Thorland et al. 2003; Wentzensen et al. 2004; Yu et al. 2005; Dall et al. 2008; Kraus et al. 2008). However, not all integration sites detected were located within a fragile site: 62% were in chromosomal regions that did not contain a fragile site. It is possible that there are fragile sites at other loci throughout the human genome not yet detected and that further fragile sites may be identified as hotspots for integration. It does remain possible that integration occurs at fragile sites by chance. When conducting a Z test, as a rule of thumb, the mean should not be less than 5. This rule is important when the probability of an event occurring is close to zero because the data is less likely to be normally distributed. When probability is close to zero a sufficient number of samples need to be tested to achieve a mean greater than 5. The mean rate of integration detection at a fragile site in this PhD was 3.2. Although there was 0% chance of detecting a false positive outcome the data presented here need to be interpreted with caution. A sample size of 63 would be required to achieve a mean of 5. In both of the studies by Thorland et al, (Thorland et al. 2000; Thorland et al. 2003) insufficient sample sizes were tested (n=20 and 43, respectively) thus the data present by Thorland et al. may not be accurate. Normally distributed data takes the shape of a symmetrical bell shaped histogram and data with a mean below 5 will not follow this pattern. Future analyses should include greater than 63 samples to achieve reliable data that is normally distributed.

7.2.2 Hot-spots for integration are present.

Wentzensen et al. (2004) compiled a review of HPV integration in the genome, summarised in Figure 11. The integration sites detected in this PhD were not detected more than once but when comparing them to the data from Wentzensen et al. (2004), hotspots for integration, where integration is reported at least 3 times on the same band, are observed; 65% of hotspots reside at a chromosome fragile site. These data support the suggestion that integration could be used to identify further fragile sites and it would be advantageous in future work to

perform an another review, similar to that of Wentzensen et al. (2004), that includes all integration studies published since 2004.

7.2.3 Integration is frequent at repeat elements

Integration at DNA repeat elements was common in this PhD. Repeat elements are various DNA sequences present in multiple copies; repeat elements can be defined as tandem arrays or interspersed repeats. DNA repeats and AT rich elements were detected at integration sites; they fall under the tandem repeat category. As mentioned in 1.4.1, slowed DNA replication at replication forks may contribute to fragility of DNA. One of the reasons that DNA replication could be delayed is the tandem repeat elements that form secondary DNA structures that block DNA replication (Ozeri-Galai et al. 2011). It is therefore possible that integration occurs at tandem repeats due to slowed DNA replication.

Integration was most frequently detected at interspersed repeats, otherwise known as retroelements; this is supported by a recent publication that reported a high incidence of integration at LINE and SINE repeat elements (Li et al. 2013). Retroelements, otherwise known as transposons, are regions of DNA that are able to amplify and relocate to another region in the genome through an RNA intermediate. Long Terminal Repeats (LTR), Short Interspersed Elements (SINE), Long Interspersed Elements (LINE) and ALU repeats were detected at the sites of integration. The mechanism of DNA relocation involves formation of a transcript from the retroelement, reverse transcription of the retroelement mRNA to form cDNA; the retro-element cDNA is then integrated into the genome. The detection of integration at retroelements highlights that there may be another mechanism of integration that has not yet been investigated; it may be possible that HPV integration occurs at the same time as a retroelement relocates or it may be possible that retroelements are more sensitive to integration of foreign DNA; this warrants further investigation. Furthermore, retroelements have been implicated in a number of cancers (Florl et al. 1999; Takai et al. 2000; Wolff et al. 2010). Typically, retroelements are methylated in the human genome preventing further translocation. In cancer, hypo-methylation activates the retroelement promoter and translocation of the retroelement induces genomic instability. In future it would be interesting to examine DNA sequence at the location of integration to determine if integration occurred

as part of translocation. Hypothetically, if integration occurs at the same time as translocation, repeat elements may be intact; if integration disrupts a repeat element then integration may not be a result of translocation but could be due to the nature of the DNA sequence at the repeat element. Additionally, it would be advantageous to examine methylation of promoter regions of repeat elements at the sites of integration because hypo-methylation at the promoter region may result in translocation of repeat elements and HPV; this may explain why CaSki has concatenated HPV at many different sites in the human genome.

7.2.4 Integration commonly disrupts E1 and E2 ORFs

Integration commonly disrupted either the E1 or E2 ORFs of HPV; this observation supports hypothesis 1. Integration did not disrupt the same point in E1 or E2, indicating that there is not an exact break point that has more susceptibility to breakage. In the Baseline samples there were a number of integration sites where HPV disruption occurred at the same point with identical overlapping sequence between HPV and human DNA. It may be that disruption of E1 and E2 gives the cell a selective advantage through decreased E2 expression and increased E6 and E7 expression but this was not observed in any of the samples with integration. This pattern of HPV E2, E6 and E7 expression was observed in two samples: HIBCD6, where no integration was detected and in sample 3O, where integration was detected but E2 was intact. For the remaining clinical samples or primary cultures there was no evidence that E1 or E2 disruption influenced selection. It does remain possible that E6 and E7 expression was increased initially in the samples with integration and high-grade disease, before the samples were taken.

7.2.5 DNA sequence is similar between HPV and human DNA at the sites of integration

NHEJ has been implicated in HPV integration (Durst et al. 1987; el Awady et al. 1987; Ziegert et al. 2003) because, at the site of integration, there is either overlapping sequence of homology between HPV and human sequence or an insert of DNA that bears no homology to human or HPV. Sequence similarity at the site of HPV integration was common and this indicates that integration is favoured where similarities exist between HPV and human sequences.

7.3 Mechanism of Integration: Conclusion

These data support hypotheses 1, 8 and 9 and illustrate that although integration occurs at a unique site in the genome each time, there are hotspots for integration. Integration is not a random event: integration has affinity for common fragile sites and transposable repeat elements. Integration frequently disrupts the HPV genome within E1 and E2 ORFs but there is not a specific site within E1 or E2 that is more prone to disruption. Integration commonly occurs at sites within the human DNA where there is sequence similarity between HPV and human sequence thus NHEJ is likely to be involved in incorporation of HPV DNA into the human genome.

8 General Discussion

8.1 HPV Integration and Cervical Neoplasia

This is the first study to investigate integration at the nucleotide level in cervical neoplasia using DIPS and RS-PCR. Previous studies have looked at integration using DIPS and RS-PCR in cancer rather than precancers and have reported integration in at least 80% of cancers (Klaes et al. 1999; Thorland et al. 2000; Luft et al. 2001; Ferber et al. 2003a; Ferber et al. 2003b; Thorland et al. 2003). Integration was not detected frequently in CIN1 in this PhD; this implies integration is not an early event. Integration was not detected at a high rate in CIN3 either and this was surprising because CIN3 has a significant rate of progression to cervical cancer and less chance of regressing to normal. Hypothetically, the changes that are required to allow a cell to progress to cancer, such as integration, would have already occurred in CIN3. There could be a number of explanations for not detecting integration in CIN3:

Integration is an event that occurs once invasive disease develops thus may not be detected in CIN3. In HI-BCD samples, integration was detected in 13% of CIN3; these data are consistent with Klaes et al. (1999) who reported integrated transcripts in 15% of CIN3 and 85% of cervical cancers; however it remains possible that integrated transcripts were not detected in CIN3 in the HI-BCD samples due to poor RNA quality. On the other hand, these data contradict the confirmed finding of integration in the W12p12 cell line representative of a low-grade cervical lesion. W12 was established from a non-tumourigenic, CIN1 lesion of the cervix of a 22 year old woman with a history of mild cellular dyskaryosis; once the lesion was removed, the woman did not have further abnormal smears (Stanley et al. 1989). Continued culture of W12 leads to the selection of clones with integrated HPV that differ between culture series (Jeon et al. 1995; Dall et al. 2008). Culturing W12 may have induced integration at a very early stage and the *in-vitro* integration status of HPV16 may not represent the integration state of the lesion *in-vivo*. Ultimately it is possible that any observations made in W12p12 in this PhD are an artefact of cell culture and do not represent events that would develop *in-vivo*.

Presence of episomal HPV at levels above the sensitivity threshold of DIPS could mask the detection of integration in these sample sets. In both the Baseline and HI-BCD sample sets, intact E2 was detected in the majority of samples where HPV16, HPV18 and HPV45 were present; this is potentially a marker of episomal HPV. Typically, CIN3 is not a productive infection and one would not expect to observe HPV DNA replication but it is possible that episomes are maintained in high grade cervical lesions at high copy numbers as described in $W12_{ser4(EPI)}$ by Gray et al (2010).

Integration is in concatenated form. DIPS could not detect integration in CaSki because HPV is integrated in concatenated form. It is possible that integration is in concatenated form in the Baseline and HI-BCD samples and biopsy sample 1W since integration was not detected by DIPS or RS-PCR. Although integration is in concatenated form in CaSki, APOT detected integrated transcripts because integration is being transcribed. The fact that APOT did not detect integration in the HI-BCD samples or in biopsy 1W strengthens the hypothesis that integration is not present. Southern blot is ideal to detect integration in concatenated form but requires more DNA than was available in the clinical samples used in this study.

Age may influence whether integrated HPV or episomal plays a role in cervical carcinogenesis. The ages of the women included in the Baseline study were 20-22 years as they were attending their first smear test in Wales, whilst the average age of women sampled in HI-BCD study was 31. The age of women from whom the biopsies were taken, ranged from 23 to 63 years, with an average age of 40. As mentioned in the introduction, cervical cancer is prevalent from age 25 upwards and it takes approximately ten years to progress from CIN1 to CIN3 (Figure 7). The average age of first intercourse in the UK is 16 years of age (Wellings et al. 2001) and this is the age at which first possible exposure to HPV occurs. The women who had CIN3 diagnosed by histology in the Baseline sample set were likely to have disease that progressed quickly with the time from exposure to HPV infection through to CIN3 being a maximum of 6 years. Biopsy 1W was from a cervical cancer that displayed characteristics of episomal transformation and was taken from a woman aged 23 years. It remains plausible that cervical disease progression in young women does not involve integration but episomal HPV, as observed in 1W. Integration could be involved in cervical disease progression in older women as seen in 3O, aged 46 and in 6N, aged 45. Biopsy 5B

was taken from a 30 year old woman and although integration was detected, it was not clonal and not likely to be contributing to disease. Overall, the data presented in this PhD suggests that integration of HPV may not be involved in cervical oncogenesis in women under 30 years. This hypothesis is supported by studies of integration in women from a similar age group: Firstly, Evans et al. (2008) used Southern blot to investigate integration in CIN in a sample set with an average age of 30 (Table 3); this group reported very low incidence of integration in CIN2/CIN3 and benign lesions with integration being detected alongside episomal HPV. Ramanakumar et al. used RS-PCR to investigate integration in persistent and transient infections in women aged less than 25 years (2010); they did not detect integration but reported high HPV viral loads. It is therefore possible that high viral loads, which are reported to be associated with women under the age of 25 years (Flores et al. 2006), may play a role in carcinogenesis in young women. A high viral load may confer increased E2, which allows episome maintenance, as well as high E6 and E7 expression resulting in transformation. Furthermore, previous reports of HPV integration in cervical cancer (Klaes et al. 1999; Thorland et al. 2000; Luft et al. 2001; Ferber et al. 2003a; Ferber et al. 2003b; Thorland et al. 2003) did not specify the age of the samples used in the study and it is not possible to draw a conclusion as to why integration was detected at such a high rate in these studies. Vinokurova et al. (2008) examined age at diagnosis in relation to integration status by comparing the age at diagnosis between patients with and without integration; the mean age of the entire sample set was 40 years; the median ages at diagnosis of cervical cancer in the Vinokurova et al. (2008) study was between 43–44 years for HPV16, HPV18 and HPV45 and there were no women under the age of 25 years with cervical cancer in the study. The data from the Vinokurova et al. (2008) study support the hypothesis that integration predominates in cervical cancers of older women. No women younger than 25 years old were included thus there is no evidence to support integration being absent in younger women in the Vinokurova et al. (2008) study. The link between age and integration is novel and interesting but due to the small sample numbers investigated in this PhD, it is not possible to draw a definitive conclusion. A hypothetical explanation for more integration in women over 30 years of age could be environmental factors. Continual exposure to factors such as smoking, contraceptive drugs, alcohol, poor diet combined with persistent HPV infection can damage DNA over time and this would increase the risk of HPV integration.

Types other than HPV16, HPV18 and HPV45 are contributing to disease in these samples sets. For example, in the HI-BCD samples mixed infections were detected in the majority of samples and although no integration was detected, it is possible that another HPV type is integrated and causing disease. HIBCD14 had cervical cancer and qPCR implied that HPV16 expression was similar to that expected in a productive infection. A limitation of this PhD is that assays were only developed for HPV16, HPV18 and HPV45 and other HPV types were not examined for integration; it would be advantageous to explore HI-BCD samples, with mixed HPV types, for integration with assays for other HPV types.

8.2 Integration and E2, E6 and E7 expression

The data presented in this thesis are consistent with the hypothesis that integration disrupts E1 and E2 ORFs (hypothesis 1). E1 and E2 ORFs can repress immortalisation and hypothetically, over-expression of E6 and E7 genes, due to the disruption of E1 or E2 ORFs upon integration, may provide a selective growth advantage (Romanczuk and Howley 1992) through interaction with hTERT, p53 and pRb. Furthermore, E6 and E7 expression has been demonstrated to be vital for ongoing cell proliferation of primary cervical carcinoma cell lines (Magaldi et al. 2012) and removal of E6 and E7 expression in culture induces senescence. There was no evidence in this PhD to support the hypothesis that integration disrupts E1 and E2 resulting in increased expression of E6 and E7. One of the limitations of the samples used in this study is that they are taken at a single time point. PC09 clone D initially showed unstable mRNA quantities of E2, E6 and E7 that increased between passage 5 and 10. After passage 10 there was a dramatic decrease in E2, E6 and E7 mRNA; this mimics an observation made in the $W12_{Ser4(EPI)}$ cell line (Gray et al. 2010) where HPV E6, E7 and E2 expression decreased at the point of invasiveness. E6 and E7 have the potential, independently of each other, to cause genomic instability through centrosome duplication errors and chromosome miss-alignment during mitosis (Duensing and Munger 2002; Duensing and Münger 2003). It is plausible that sufficient DNA damage by E6 or E7 was already inflicted in the clinical samples prior to sampling. DNA damage may have been sufficient to drive transformation with maintained cell growth favouring cells lacking HPV expression. There was evidence in biopsy 3O where integration was detected and HPV oncogene expression was higher than housekeeping genes Therefore, integration and E6 and

E7 were likely to be contributing to cancer in 3O; this was an exception rather than a rule of transformation.

8.3 Integration in Human Genes

An alternative mechanism for transformation and selection other than HPV oncogene expression is disruption to human genes. There was evidence to support integration into human genes in the PC0 cultures and in the biopsy samples. It is possible that a cell with human gene disruption does not require HPV oncogene expression and although integrated HPV is detected, HPV oncogene expression is not required for cell survival. Evidence for this was observed in biopsy sample 6W where integration was in gene SLC7A11 but there was no detectable E6 or E7. To conclusively determine if integration disrupts SLC7A11 gene expression, further work would be required to determine if protein levels are altered by integration at this site.

8.4 Alternative mechanisms

8.4.1 Methylation

HPVE2 expression was lower and HPV oncogene expression was higher than housekeeping genes in 3O and HIBCD6. No integration was detected in HIBCD6, and 3O had intact E2, even though integration was detected; this means that suppression of E2 expression was due to a mechanism other than integration and episomal loss. As mentioned in section chapter 4, CaSki possesses many integration events in concatenated form but only 1 copy is actively transcribed. The mechanism of suppression of integrated HPV in CaSki is methylation (Badal et al. 2003). It is possible, in CaSki, that one copy of HPV escaped methylation; this in turn provided the cell with an un-methylated HPV copy, a selective advantage and allowed malignant progression. As mentioned in 1.2.3.3.5, the HPV URR possesses CpG rich regions that have been reported to be hypomethylated or hypermethylated, depending on the stage of the HPV lifecycle (Kim et al. 2003; Kalantari et al. 2008). Changes in methylation at the CpG regions in HPV URR can alter the HPV lifecycle, increase E6 and E7 expression, and potentially lead to cancer. Dean Bryant, a PhD student in the HPV research group at Cardiff University investigated integration in relation to HPV URR methylation in vulval neoplasia (Figure 62 and Table 57, in appendix). There was a higher percentage of methylation of the

L2/L1 CpG region in 50% of the samples. This may lead to decreased expression of capsid proteins needed to complete the HPV life cycle and disruption of the HPV life cycle may lead to transformation. Also, where integration was detected and disruption to the E1 ORF occurred, hyper-methylation of the E2 CpG was observed. Methylation of the E2 promoter region did not lead to decreased E2 expression or increased E6 or E7 expression but the link between integration and methylation is an interesting one. There is increasing evidence that the host methylates foreign DNA integrated into the genome, but the reason and mechanism is not fully understood. It is possible that methylation is a defence mechanism to protect the host cell and may be dependent on the site at which integration occurred and chromosomal interactions (Doerfler 2007). Furthermore, methylation of foreign DNA in a host genome may influence the methylation of DNA surrounding the integration site; this potentially could have an impact on host cell gene expression. Methylation of genes involved in oncogenic pathways could increase the risk of oncogenesis and provide a selective advantage without the need for HPV E6 and E7 expression as observed in Dean Bryant's data. In future research of HPV methylation, it would be interesting to observe methylation patterns of the genome surrounding integration sites and changes in human gene expression.

8.4.2 DNA instability by Replication of Integrated HPV

Replication of integrated HPV DNA may offer an alternative mechanism of genomic instability other than E6 and E7 expression. It has been reported that where integrants are transcriptionally active and where episomes are present, HPV DNA replication can still occur from the origin of replication in the integrated HPV DNA (Kadaja et al. 2009b). E1 and E2 are required for HPV genome replication and where episomal HPV and integrated HPV occur together, replication of integrated HPV is warranted (Kadaja et al. 2007). In a review of HPV replication and genomic instability, Kadaja et al. (2009b) proposed a model for genomic instability that resulted from replication of integrated HPV, adapted in Figure 52. Non-contiguous sequences have been a common theme in this PhD: firstly, non-contiguous sequence was detected in CaSki in the assay development stage and has been previously reported in CaSki (Meissner 1999); also non-contiguous sequence was detected in a biopsy sample (1W) and in PC08 (clone P). Although episomes are no longer present in CaSki, it is possible that non-contiguous sequence arose due to replication of integrated HPV DNA, when episomes were present. It is also possible that concatenated HPV at the site of

integration is a result of DNA replication of integrated HPV. Structural abnormalities have been reported at the sites of integration in cervical cancers (Peter et al. 2010). The majority of structural abnormalities have developed due to local amplification of DNA by replication of integrated viral sequences. Furthermore, Peter et al. (2010) reported that HPV copy numbers were higher in cancers with genomic alterations at the site of integration; this suggests that integration precedes genomic alteration in these samples and episomal HPV allows replication of integrated HPV. Genomic abnormalities are important to understand in each individual cancer because chromosomal abnormalities can give a cell a selective advantage; this is important in treatment of cancer because genomic changes may enhance a cell's survival and result in persistent disease. Chromosome abnormalities at the sites of integration were not studied in this PhD but should be explored in future work to improve understanding of the role of integration in cancer cell evolution. It is important to note that integration was not detected in sample 1W or in the P clone. If the model of Kadaja et al. (2009b) only applies to integrated HPV, then in theory integration must exist in 1W or in the P clone. It is possible that integration occurs in concatenated form in 1W and P clone and this may hinder detection of integration by the methods used; whole genome sequencing, Southern blot, or FISH would be useful to detect integration in these samples.

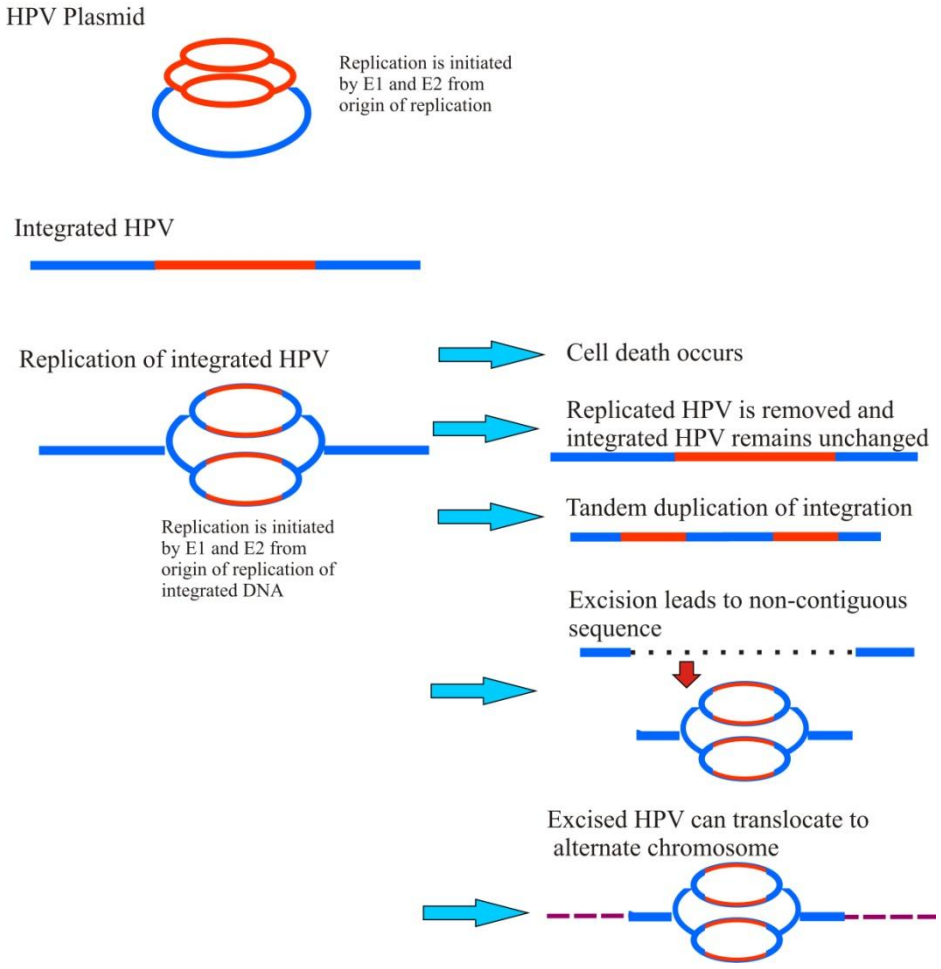


Figure 52 Model of replication of integrated HPV, adapted from Kadaja et al. (2009). Where episomal HPV is present, replication of integrated HPV can be initiated from origin of replication. In most cases cell death occurs or HPV is removed by homologous recombination repair. Replication of integrated HPV may lead to duplications and concatenated HPV. Excision of replicated HPV may lead to non-contiguous sequence. Excised HPV can be translocated to alternate chromosomes.

8.5 Clinical Implications

8.5.1 Integration in Vulval and Vaginal Neoplasia: Clinical Implications

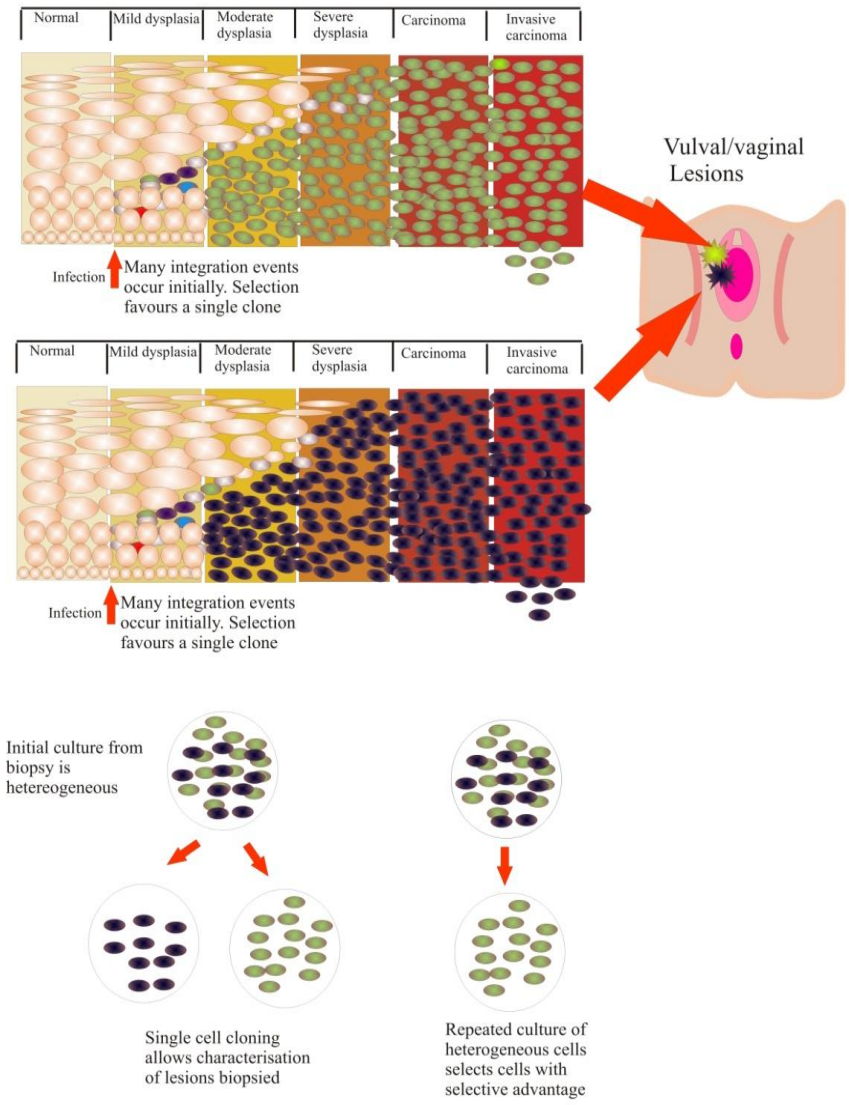


Figure 53 Multifocal clonal integration in vulval and vaginal neoplasia. Multifocal lesions arise from separate infections and have different integration characteristics that can be observed through single cell cloning; culture of heterogeneous cells results in selection of cells with strongest selective advantage, which do not retain original characteristics of the lesion.

Data presented here, from single cell cloning of cells, taken from vaginal and vulval intraepithelial neoplasia grade 3 shows that there is variability in integration status between clones from the same biopsy. If a single integrant were contributing to grade 3 intraepithelial neoplasia then, when single cell cloning were performed, all clones would be expected to be

the same; this was not the case for PC08 or PC09. Multifocality is a risk factor for recurrence of vulval and vaginal neoplasia (Dodge et al. 2001) and multifocality would explain the detection of differences in integration status in each clone with several separate infections developing different integration events (Figure 53). In the pilot study of PC08 and PC09 development, PC08 was confirmed as a multifocal infection but PC09 was not; it is possible that for PC09 a number of separate lesions developed in close proximity to each other appearing as one lesion upon clinical examination. The presence of cellular differences within a lesion could have implications for treatment: each cell that differs could potentially have different viral and host characteristics that respond differently to treatment. In addition, cells resistant to treatment would survive and clonal evolution would select these resistant cells leading to disease progression or recurrence. The techniques developed in this PhD could aid understanding of integration in vulval and vaginal neoplasia and give an insight into how treatments can be developed according to an individual patient's needs, with decisions on treatment based upon knowledge of a patient's HPV integration status. For example, in PC08, integration was within two important candidate genes for cancer; PC08 displayed no evidence that HPV E6 or E7 expression contributed to cell survival and anti-viral treatment may not be suitable for this patient. Therapies that target human proteins such as LEPREL1 and CDCP1 pathways may be a better choice of treatment. In PC09, there was evidence to imply both viral and human factors playing a role in cell survival. Antiviral drugs may be more effective in this patient but if integration remains within a cell, disruption to human genes will remain; a combination of anti-viral and novel therapeutics to target human protein pathways may be required.

8.5.2 Integration in Cervical Neoplasia: Clinical Implications

One of the aims of this PhD was to determine if HPV integration could be used as biomarker of high-grade cervical disease to improve the management of women with mild cellular dyskaryosis. Selecting the Baseline samples of young women may have been a weakness of the PhD, but there was no evidence in the literature to suggest that integration occurs over time and is present in older women. Cervical cancer is becoming an increasing problem in young women and women who want to have a family need to decide whether to have treatment if they have low-grade abnormalities. It was important to determine if integration played a role in disease progression in this age group to improve current screening methods.

Performing the assays on this age group has highlighted that a different pattern of disease progression may exist between younger women and older women; this is a strength rather than a limitation of this PhD.

The data from HI-BCD study shows that integration was not prevalent in CIN3 in the samples assayed thus may not be suitable as a biomarker of cervical disease. To be confident that integration does not play a role in CIN3, it would need to be confirmed in a larger, adequately powered sample with a wider range of ages included. Also, there are additional explanations, other than integration not being detected in CIN3 that would justify that detection of integration by DIPS, APOT or RS-PCR is unsuitable as a diagnostic biomarker for use in a clinical setting to predict high-grade disease:

The assays used in this study required good quality DNA and RNA, where nucleic acids are intact without protein and co-purified contaminants. Current methods for the collection of liquid based cytology (LBC) samples utilise a fixative agent to preserve the structure of the cells for cytological investigation. Baseline smear samples were collected in SurePath™ preservative fluid, the ingredients of which are not disclosed by the manufacturer. It was clear that contact of cytological smear samples or SiHa cells with SurePath™ resulted in rapid degradation of DNA. Thus LBC samples, collected in Surepath™ are not ideal. Furthermore, the need for good quality DNA and RNA for reliable data implied that the assays are not robust enough for routine clinical practice. Ideally, an assay that performs optimally on both intact and degraded DNA or RNA is required for clinical practice as it would reduce the number of specimens needing to be repeated. Collection of samples in RNAProtect® would improve DNA quality but these samples could not be used for cytological investigation since RNAProtect® lyses cells, rendering the sample useless for cytological examination. To overcome this 2 smear samples would have to be collected, each in a different reagent. Also, RNA was degraded in smear samples as observed in the HI-BCD study although every precaution was taken to ensure good RNA quality; thus smears do not provide RNA of good quality.

An ideal diagnostic biomarker would be used on smear samples to determine which women require biopsies, since smear sampling is a less invasive process. Integration assays are not suitable for this purpose because smear samples contain a mixture of HPV types as a result of transient infections. As described above, a single assay detects integration for a single type but where many types are present, it is a tedious search for integration, using a different assay for each type. Biopsy samples commonly possess a single HPV type contributing to disease (van der Marel et al. 2012) and this reduces the number of integration assays that need to be performed but biopsy is a more invasive process. Integration may be a useful as a prognostic biomarker, rather than a diagnostic biomarker: a recent study reported that detection of episomal transcripts by APOT conferred disease free survival whereas women with integrated transcripts had more recurrence of disease (Das et al. 2012). Another recent study reported that integration has potential as a prognostic biomarker to predict relapse by detecting integrated HPV in the blood of cervical cancer patients (Campitelli et al. 2012). DIPS method was used to identify integration sites in cervical tumours. Circulating tumour DNA (ctDNA) that possessed the same integration site as the tumour, was quantified from blood; in patients that responded to treatment there was a decrease in ctDNA whereas in patients that relapsed an increase in ctDNA was detected. In future research of HPV integration it would be interesting to replicate the study of Campitelli et al. (2012) to enhance understanding of integration in cervical cancer relapse because ctDNA may be useful in management of cervical cancer by improving follow-up of patients after treatment. In future studies of integration in cervical neoplasia, DIPS and APOT should be applied to biopsies to determine if integration is present in CIN3 and to determine if integration can be used as prognostic biomarker.

One fascinating observation made in this study was in sample HI-BCD14, taken from a woman with cervical cancer: no integration was detected but HPV16 E2, E6 and E7 expression followed a pattern expected from a productive infection. Currently women with cervical cancer are advised by clinicians that cervical cancer carries no risk of sexually transmitted HPV infection since a cervical cancer is not producing infectious virions. The possible detection of a productive infection alongside cervical cancer could change the way patients are advised by clinicians: clinicians would advise that even though cervical cancer is present, it is still possible to infect partners with HPV if a productive infection is present too.

HIBCD14 also contained HPV59 and HPV66 and these types require investigation for integration to test the hypothesis of a productive infection alongside cervical cancer.

Another observation of clinical importance arose in cervical cancer biopsy sample 6N: PapilloCheck® was initially used to test 6N but no HPV was detected. When GP5+/6+PCR-EIA was used to type 6N, HPV16 was detected; integration analysis showed HPV was disrupted within E1 ORF. PapilloCheck® genotypes HPV according to the E1 ORF and disruption to E1 is very common in integration. Biopsy 6N had integration that disrupted E1 and although the region of E1 that is amplified by PapilloCheck® is not disclosed by Greiner Bio-One, it is very likely that integration prevented the detection of HPV in this sample. Integration may result in false negative detection of HPV in high-grade disease, where episomal HPV is absent thus any HPV negative outcome needs to be interpreted with caution especially if there is evidence of disease.

8.6 Unanswered Questions and Future Work

8.6.1 Integration and Vulval and Vaginal Neoplasia

Antiviral response is very important to consider during the treatment of HPV infections as rapid loss of episomes from integrant containing cells may speed up malignant progression as was reported in a study of Interferon treated W12 cells (Herdman et al. 2006). The action of Cidofovir in vulval and vaginal neoplasia is poorly understood and there is much conflicting data on the mechanism of action. Currently, Cidofovir treatment of PC08 and PC09 is being studied by the Cardiff University HPV research group to determine if antiviral treatment is suitable for infections containing only episomal HPV and to further the understanding of mechanism of action of Cidofovir in vulval and vaginal neoplasia. Additionally, a multicentre UK clinical trial (RT3VIN) of Cidofovir and Imiquimod in the topical treatment of vulval neoplasia is being conducted, with Cardiff as the lead centre. Integration techniques developed in this PhD are being applied to biopsies taken from women in the trial before and after treatment to gain insight into how factors such as integration affect treatment outcomes.

8.6.2 Integration and Cervical Neoplasia

Viral load was not explored in this PhD. At the outset of the PhD, it was hypothesised that viral load decreases as transformation progresses due to episomal loss and was therefore not useful to investigate. Integration was not detected at a high frequency in Baseline or HI-BCD where the average age was lower than studies that reported integration (Vinokurova et al. 2008). Integration may not be involved in cervical neoplasia in the age group investigated in this PhD and viral load may be more important in transformation. In future studies, viral load will be examined in cervical neoplasia. High viral loads are common in women younger than 25 years but cervical cancer is rare in this age group (Figure 2) and hypothetically high viral load alone will not lead to cancer. It may be possible that women with early onset cervical cancers have a genetic predisposition or other factors that when combined with high viral load, trigger transformation. In a study of oral cancer, HPV and variants within p53 related genes (Wang et al. 2012) reported an increased risk of oral cancer in patients infected with HPV16 who possessed risk variants of p53 related genes. In future studies of cervical neoplasia it would be interesting to investigate factors such as human genetic variants of genes involved in the HPV lifecycle such as p53, hTERT and pRB.

An interesting observation made in this PhD is that not all integration is clonal and contributing to disease; integration may be detected by chance and not be replicated by repeating the assay or by PCR. The data from this PhD have allowed the HPV research group to evolve DIPS to differentiate non-clonal integration from integration events that are clonal. The DIPS assay is performed in duplicate: 2 digestions, ligations and PCRs per sample. Electrophoresis of the duplicates side by side allows comparison of amplicons to each other. Amplicons that are identical between duplicates and differ in size from amplicons from HPV plasmid controls are likely to be clonal integration events. Amplicons that differ between duplicates and are different in size to amplicons from HPV plasmid controls are likely to be non-clonal integration events. Non-clonal and clonal amplicons are selected for DNA sequence analysis. Future studies would ideally examine clonal and non-clonal integration events in an adequately powered sample to determine the role of clonal and non-clonal integration in HPV related neoplasia.

Finally, the samples assayed in this study were taken from a single time point. Taking a sample at a single time point cannot give insight into physiological events prior or subsequent to the sample being taken. A longitudinal study of samples taken at time points throughout a woman's attendance at clinic would allow understanding of HPV disease progression and regression.

9 Conclusion

Cervical malignant progression is a complex process and there is no single mechanism that explains why transformation occurs after HPV integration. There appears to be a pattern emerging in the data presented here: integration may give a cell a selective advantage either due to disruption of HPV oncogene expression or due to disruption of human genes. Additionally, there was evidence to imply that E6 and E7 contribute to genome instability but are not required to maintain the malignant phenotype. The clinical samples used in this study are taken at a fixed time point. To enhance understanding of cervical neoplasia, future work should include samples taken over a period of time to observe changes in E6 and E7 expression and human genome instability in relation to integration and episomal loss.

Integration was not detected abundantly in CIN3 and there was no evidence to suggest that episomal loss drives cervical disease progression; this implies that integration is not suitable as a biomarker of cervical disease. Also, it is possible that presence of episomes alongside integrated HPV can induce human genome instability and contributes to disease progression; this requires further investigation.

The samples examined in this PhD were taken from young women and it does remain possible that age can determine whether a woman has integrated or episomal related oncogenesis. An alternative mechanism of oncogenesis could be related to presence of episomal HPV in high copy numbers. The sample sizes used in this PhD were too small to make a definitive conclusion about the relationship between age and integration. Further investigation using an adequately powered collection of samples from women with a range of ages would allow further understanding of the role of age in cervical neoplasia.

Integration was detected in fragile sites in approximately a quarter of the sample. This suggests that fragile sites may play a role in integration in some cases, but not all and that a mechanism other than DNA damage and repair to incorporate the HPV into human DNA may exist. Transposon repeat elements were observed at the majority of integration sites. This

highlights the possibility of an alternative mechanism of integration of HPV into the human genome.

Finally, the importance of clonal evolution has been highlighted in this PhD. An understanding of how integration contributes to oncogenesis in each patient is vital for selection of optimal management of cervical cancer. The techniques developed in this PhD could also aid understanding of integration in vulval and vaginal neoplasia and give an insight into how treatments can be developed according to an individual patient's needs, with decisions on treatment based upon knowledge of a patient's HPV integration status.

10 Appendices

10.1 Appendix 1: HI-BCD Study Protocol and Consent Form

Human Papillomavirus Integration analysis in liquid based cytology samples as a Biomarker for Cervical Disease

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

Chris Holmes, Consultant, Cardiff University

Helen Beer, Senior Information Manager and Research Specialist

Rachel Raybould, PhD student in Gynaecological Oncology, Cardiff University

Study Advisors

Shantini Paranjothy, Senior Lecturer and MPH Programme Director, Cardiff University

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 and Public Health Wales, Screening Division.

Aim

To pilot whether:

HPV viral integration can predict high-grade cervical abnormality in women with cytological abnormalities.

A novel repertoire of monoclonal antibodies can be used as tools to objectively assess the quality, adequacy and pathological status of smears.

Number of Subjects

45

Study Design

Pilot, cohort, observational study.

Inclusion Criteria

Women referred to colposcopy following an abnormal cytology

Exclusion Criteria

Women unable to give informed written consent.

Study Interventions

HPV testing and biomarker assays for viral integration performed on an additional smear taken during colposcopy.

Primary endpoint

Histologically proven high-grade disease at first colposcopy visit.

Secondary endpoint

Histologically proven high-grade disease during subsequent visits (up to 4 years).

Contents

1.0 Lay Summary

2.0 Introduction

2.1 Cervical Screening

2.2 Human Papillomavirus

2.3 Viral Integration

2.4 Novel Monoclonal Antibodies

2.5 Aims and Objectives

2.6 Rationale

3.0 Investigational Plan

3.1 Overall Study Design

3.2 Discussion of Design

3.3 Study Population

3.4 Inclusion and Exclusion Criteria

3.5 Recruitment

3.6 Interventions

3.7 Study Plan

3.8 Endpoints

4.0 Sample Management

4.1 Cytology

4.2 HPV Testing

4.3 HPV Viral Integration

4.4 Novel Monoclonal Antibodies

5.0 Data Management

5.1 Statistical Methods

5.2 Sample Size and Power Calculations

5.3 Data Analysis and Dissemination

6.0 Project Management

7.0 Administrative Procedures

7.1 Changes to the Protocol

7.2 Recording of Data and Retention of Documents

7.3 Publication of Results

8.0 Ethical Considerations and Consent

9.0 References

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. The result will not go in the woman's hospital or screening records. The women will be invited to take part if they have been referred to colposcopy clinic following an abnormal smear. If they consent to being included in the study, a smear will be taken and tested first for human Papillomavirus (HPV).

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 with tests for viral integration. 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 Screening Services Wales, Public Health Wales with the support of a data analyst.

2.0 Introduction

2.1 Cervical Screening

The UK has national cervical screening programmes that have been estimated to prevent 80% of deaths from cervical cancer (Peto et al. 2004). 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+) can then be 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
- Three tests reported as abnormal at any grade in a 10 year period

Whilst there is a high incidence of CIN2+ following a moderate or severe cytology result (74-77% and 80-90% respectively), approximately only 20% of women with persistent low-grade cytology will have 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 who 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 reduce the number of women that are unnecessarily concerned with the possibility of having cervical cancer, in addition to focusing resources on those women that need to be seen.

In this study we aim to investigate viral integration as a potential prognostic marker in women referred to colposcopy. Whilst there are many women with low-grade cytological abnormalities referred for colposcopy who do not have disease; samples from women with all grades of cytological abnormality will be taken to evaluate integration status in relation to grade of abnormality. It may then be possible to identify biomarkers in low-grade smears that will predict high-grade disease.

2.2 Human Papillomavirus

High risk Human Papilloma Virus (HPV) plays a central aetiological role in anogenital 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 replace cervical cytology with a test for the virus (Cuzick et al. 2003; Cuzick et al. 2006). 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) (Solomon et al. 2001; Kulasingam et al. 2006; Moss et al. 2006).

However, HPV is common with an 80% lifetime risk of infection. The majority of infections are transient and asymptomatic with less than 5% of those 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.

1989). Within HPV episomes the expression of viral oncogenes is highly regulated and 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 to not function as 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 these cells contain truncated viral genomes integrated into the host genome. *In vitro* viral integration increases cell proliferation (Jeon et al. 1995; Wentzensen et al. 2004) 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 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 that the biomarkers might be used if introduced into clinical practice.

2.4 Novel Monoclonal Antibodies

The assessment of cytological smears underpinning the cervical screening programme remains a subjective, error-prone and labour intensive process. A novel repertoire of monoclonal antibodies has been developed against human cervical tissue that unequivocally identifies and discriminates between the complex epithelial cell subpopulations present in the cervix. When used on cytological smears, the antibodies provide a clear indication of the origin of the component cells from within the cervix in vivo.

The research aims to use these reagents as tools to objectively assess the quality, adequacy and pathological status of smears, thereby reducing errors during sample taking in the clinic and providing information on the pathological status of smears. Use of antibodies in this way renders the approach eminently suitable for automated analysis. This stage of the research programme requires detailed analysis of the performance of antibodies against endocervical and tranformation zone cells in cervical smear preparations from LBC (liquid based cytology) samples.

2.5 Aims and Objectives

The aim of the study is to perform a pilot study to investigate the potential of viral integration as a biomarker in predicting high-grade cervical abnormality in women referred to colposcopy. A novel repertoire of monoclonal antibodies will also be investigated to see if they can be used as tools to objectively assess the quality, adequacy and pathological status of smears.

Primary Outcomes

1. Comparison of three viral integration methods for identification and prediction of high-grade disease.
2. 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).

3. Functionality of novel monoclonal antibodies as tools to objectively assess the quality, adequacy and pathological status of smears.

2.6 Rationale

This pilot, cohort, observational study has been designed to examine the benefit of viral integration following a HPV positive result in a clinical cohort of patients attending colposcopy clinic. A novel set of monoclonal antibodies will also be assessed to further investigate the pathological status of the smears.

3.0 Investigational Plan

3.1 Overall Study Design

This is a cohort study to investigate viral integration as a biomarker and novel monoclonal antibodies targeting human cervical tissue 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. The functionality of the novel monoclonal antibodies as tools to objectively assess the quality, adequacy and pathological status of smears will also be investigated.

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 thought 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 and 15 with severe dyskaryosis. Approximately 20-40 women per month are expected to be eligible for the study within the single site. Recruitment is expected to be completed within three months.

3.6 Interventions

This study is designed to evaluate viral integration and the potential functionality of novel monoclonal antibodies specific for cervical tissue in a real life setting and 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 and x monoclonal antibodies will be assessed.

3.7 Study Plan

1. Women will be informed of their smear test result and colposcopy referral as per Cervical Screening Wales (CSW) standard operating procedures (CSAD Quality Manual Feb 2005).

2. 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 and 15 with severe dyskaryosis.

3. 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.

4. 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) after entry to be made available for the study. These data will be retrieved by Cervical Screening Wales.

5. A ThinPrep liquid based cervical sample will be taken, according to standard operating procedures, from all women who give written consent. The LBC pots will be labelled with the women's date of birth and hospital number for transportation to the HPV Laboratory at Cardiff University. Once received in the lab the samples will be pseudo-anonymised and all patient identifiers will be removed and a unique study code allocated. An excel worksheet linking study ID and patient identifiers will be forwarded to CSAD.

6. Colposcopy will be undertaken according to normal practice as set out in the current version of CSW's Quality Manual Version

7. HPV testing and analysis of viral integration will be performed by the department of HPV Research Laboratories, Obstetrics and Gynaecology at Cardiff University.

8. Monoclonal antibody testing will be performed in the School of Medicine, Cardiff University.

3.9 Endpoints

- Primary endpoint will be high-grade disease on histology at first colposcopy visit.

- Secondary endpoint will be correlation of biomarker results and antibody results with histology at subsequent visits over next screening round (up to 4 years).

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

ThinPrep samples will be taken from women with written informed consent within the Cardiff and Vale NHS Trust Colposcopy Clinics following standard operating procedures. The LBC pots 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, pseudo anonomysed 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 + samples will be sub-typed using type-specific probes.

- The commercial 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.

4.4 Novel Monoclonal Antibodies

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 45 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=16) (Jones et al. 2009). All samples will be tested with the novel repertoire of monoclonal antibodies.

5.3 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, viral integration and monoclonal antibody results 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 support further grant proposals to increase sample size with prospective cohort studies and randomised control trials that aim to reduce cervical cancer mortality through decreased incidence with early detection of CIN2+.

6.0 Project Management

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 1 and 11 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 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.

Dr Chris Holmes

Dr Shantini Paranjothy

Dr Paranjothy is a Clinical Senior Lecturer in Public Health Medicine, Cardiff University and Public Health Wales and has research expertise in screening.

Mr Bryan Rose

Mr Rose is the All Wales Programme Manager for Cervical Screening Wales.

Mr Dave Nuttall

Mr Nuttall is the Head of Laboratory Services for Cervical Screening Wales.

Miss Helen Beer

Miss Beer is a Senior Information Analyst / Manager for Screening Services and has expertise is 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.

Mrs Rachel Raybould

Mrs Raybould is a PhD student with the HPV Research group at Cardiff University and has expertise in molecular biology and virology.

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 for presentation at national and international conferences.

8.0 Ethical Considerations and Consent

Women participating in this study will provide one extra samples 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 three years. The women will be sent information in advance and will be permitted 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 and monoclonal antibody 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

1. Peto, J., et al., *The cervical cancer epidemic that screening has prevented in the UK.* Lancet, 2004. **364**(9430): p. 249-56.

2. Wales, C.S., *KC53/61/65 Statistical Report 2008/2009.* 2007.

3. 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.

4. Walboomers, J.M., et al., *Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.* J Pathol, 1999. **189**(1): p. 12-9.

5. 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.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. Stanley, M.A., et al., *Properties of a non-tumorigenic human cervical keratinocyte cell line.* Int J Cancer, 1989. **43**(4): p. 672-6.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

10 . Consent form follows on next page

Human Papillomavirus Integration analysis in liquid based cytology samples as a Biomarker for Cervical Disease (HI BCD)

(A new study for looking at cervical smear tests in the laboratory)

We would like to invite you to take part in a research study. Before you decide you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. Ask us if there is anything that is not clear or if you would like more information. Talk to friends and family about the study if you wish.

What is the study about?

This is a study to see if carrying out new tests on cervical smear samples could safely cut down the number of women who have to go to hospital following an abnormal smear result.

Cervical Screening (regular smear tests) has greatly reduced the number of cervical cancers in the UK, but a lot of women are seen in the hospital who do not need to have treatment. The purpose of the screening programme is to pick out women who would most benefit from having treatment to prevent cervical cancer developing in years to come. The smear taken at your doctors can identify whether you have an increased chance of having abnormal cells in the cervix (the neck of the womb). You can then be referred to hospital, to see if you need to have treatment. Not all women picked up by the current cervical smear test and referred to hospital will actually have abnormal cells. Even fewer women will need to have treatment.

We would like to do some new tests on a smear sample and compare the results of the new tests with what is found when women are seen in the hospital. This is to see if carrying out the new tests will help us know who needs treatment and who does not.

Why have I been chosen?

You have had an abnormal smear result and are being referred to hospital for a colposcopy (closer look at your cervix) to see if you have abnormal cells. Most women in your situation will not need to have treatment. We are trying to find out if new tests can help identify which women need to be seen and which women do not.

What are the tests being used for?

Current cervical smear testing involves looking at the shapes of the cells in your cervical smear to see if they look abnormal. The new tests will concentrate more on what is actually happening inside the cells. Normally, cells divide but this is tightly controlled by within the cells. Cancer develops when this control is lost and cells can then divide randomly and rapidly. We want to see if there are any early changes within the cells that can determine which cells could eventually lead to cancer.

The first test we will do will be a test for Human Papillomavirus (HPV). HPV is a very common viral infection and most women will be infected at some point in their lives. HPV usually clears from the body in about a year, without any treatment. We know that HPV causes abnormal smears and cervical cancer, especially HPV types 16 and 18 (these are the ones we are currently vaccinating school girls against) however, having a HPV infection very rarely leads to cancer. As HPV is so common, just testing for the virus would mean too many women would have to go to hospital to be checked.

Occasionally, HPV can cause abnormal cell growth by interfering with the normal cell controls. The tests will start to look for possible changes in the virus that could be the cause of this interference in the normal cell controls.

What Are The Possible Benefits of Taking Part?

Taking part in this study will not change the way you are treated. The study will not help you now, but the information we get from this study will help improve the way cervical screening is carried out in the future.

What are the possible disadvantages of taking part?

Taking part in the study will not change the way you are treated.

Do I have to take part?

It is up to you to decide whether you want to take part in the study or not. If, after reading this information sheet, you decide you would like to take part we will ask you to sign a consent form when you attend the colposcopy clinic. If you decide not to take part or you decide to withdraw at any time, it will not affect the standard of care you receive, now or in the future.

What would I have to do?

When you come for your colposcopy appointment we will answer any questions you may have and ask you if you are happy to take part in the study. It is up to you to decide if you are happy to take part or not. If you are happy to take part we will ask you to sign a consent form and give you a copy of the form to take home. Before we have a look at your cervix we will take a smear sample for the study. This smear will be taken in exactly the same way as your other smears. Although we would not normally take a smear during this visit, the sample taken will not affect your usual care. This smear sample will be sent to our Laboratory in the University Hospital of Wales so that the new tests can be done. The people testing your sample will not know who you are. The results of these tests will not change any treatment you are offered. We will not give you the result of these tests.

What extra tests will I need if I take part?

You will have one extra smear sample taken from you during your colposcopy visit if you agree to take part in the study.

What if something goes wrong?

If you are unhappy with any aspect of the study we would like you to tell us. Regardless of this, if you wished to complain about any aspect of the way in which you have been approached or treated, the normal NHS complaints mechanism is available to you.

Would my taking part in this study be kept confidential?

All information which is collected about you during the course of the research will be kept strictly confidential. The fact you are included in the study will be documented in your hospital notes. You would be assigned a study number which will be used on information or samples that leave the hospital. Your medical records might be inspected by regulatory authorities to check the study is properly carried out. Signing the consent form to take part in the study means you agree to this access.

What happens to the results of the study?

The results may be published in a medical journal and/or presented at a scientific meeting. It would not be possible to identify you from any of the information published or presented.

Who has reviewed the study?

All research in the NHS is looked at by independent group of people, called a Research Ethics Committee to protect your safety, rights, wellbeing and dignity. This study has been reviewed and given favourable opinion by South East Wales Research Ethics Committee.

Who is organising and funding the research?

The research is being coordinated by the Academic Department of Obstetrics and Gynaecology at Wales College of Medicine, Cardiff University. It is being supported by Emma Jayne Demery Bequest Fund, Cardiff University.

What happens to my sample at the end of the study?

On completion of the study all material will be disposed of in line with the Human Tissue Act 2004.

Further Information

Further information about cervical smear tests, results and treatment is available on the Cervical Screening Wales website: www.screeningservices.org.uk/csw

If you have any questions at any time during the study please contact:

Dr Sam Hibbitts 02920 745005

Dr Amanda Tristram 02920 742337

Research Nurse Sue Ashman 02920 745365

Thank you for taking the time to read this information and considering whether or not you would like to take part in this study of a new test for cervical screening.

GIG CYMRU NHS WALES
Bwrdd Iechyd Prifysgol Caerdydd a'r Fro
Cardiff and Vale University Health Board

Ysbyty Athrofaol Llandochau
University Hospital Llandough

Penlan Road, Llandough Heol Penlan, Llandochau

Consent Form

HPV Integration analysis in LBC samples as a Biomarker for Cervical Disease (HI BCD)

(A study of a new test for cervical screening)

Centre Number:	
Study Number:	
Referral Smear Laboratory Number:	

Name of Researcher: Dr Sam Hibbitts

Please initial each box

1. I confirm that I have read and understand the information sheet dated 14.10.10 version 1.0 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected.

3. I understand that relevant sections of my medical notes and data collected during the study, may be looked at by individuals from Cardiff University, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐

4. I agree to take part in the above study. ☐

Name of Patient	Date	Signature
Name of Person taking consent	Date	Signature

When completed, 1 for patient; 1 to Cytology Laboratory; 1 (original) to be kept in medical notes ☐

10.2 Appendix 2: E2 supplementary data

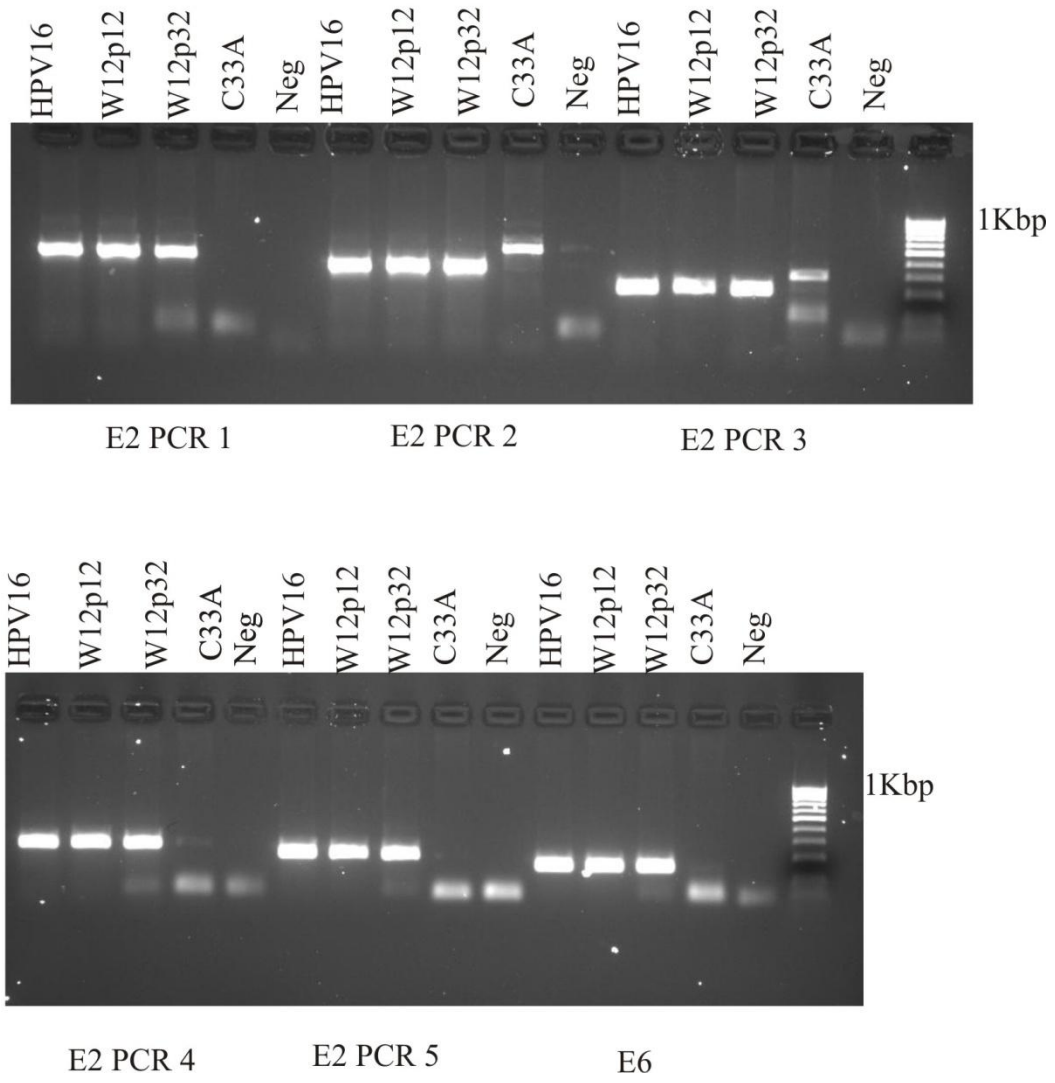


Figure 54 E2 PCR gel electrophoresis of W12p12, W12p32 and C33A. HPV16 plasmid DNA was used as a positive control and water as a negative control. W12p12 and W12p32 had intact E2, indicating the presence of episomal HPV16.

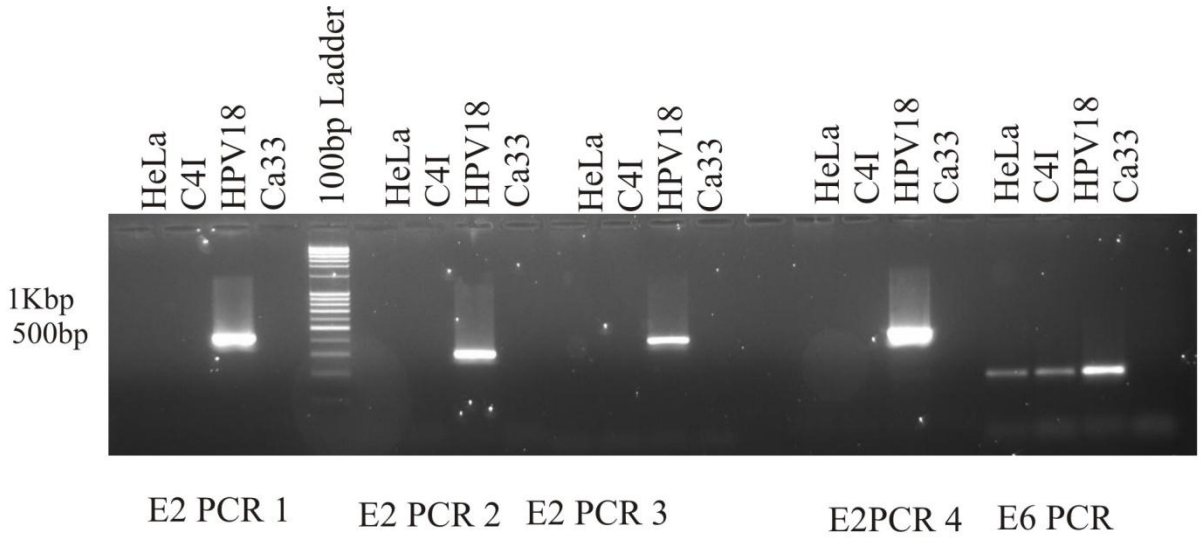


Figure 55 E2 PCR gel electrophoresis of Hela, C4I and C33A. HPV18 plasmid DNA was used as a positive control and water as a negative control (not shown). C4I and HeLa had no E2, indicating the absence of episomal HPV18.

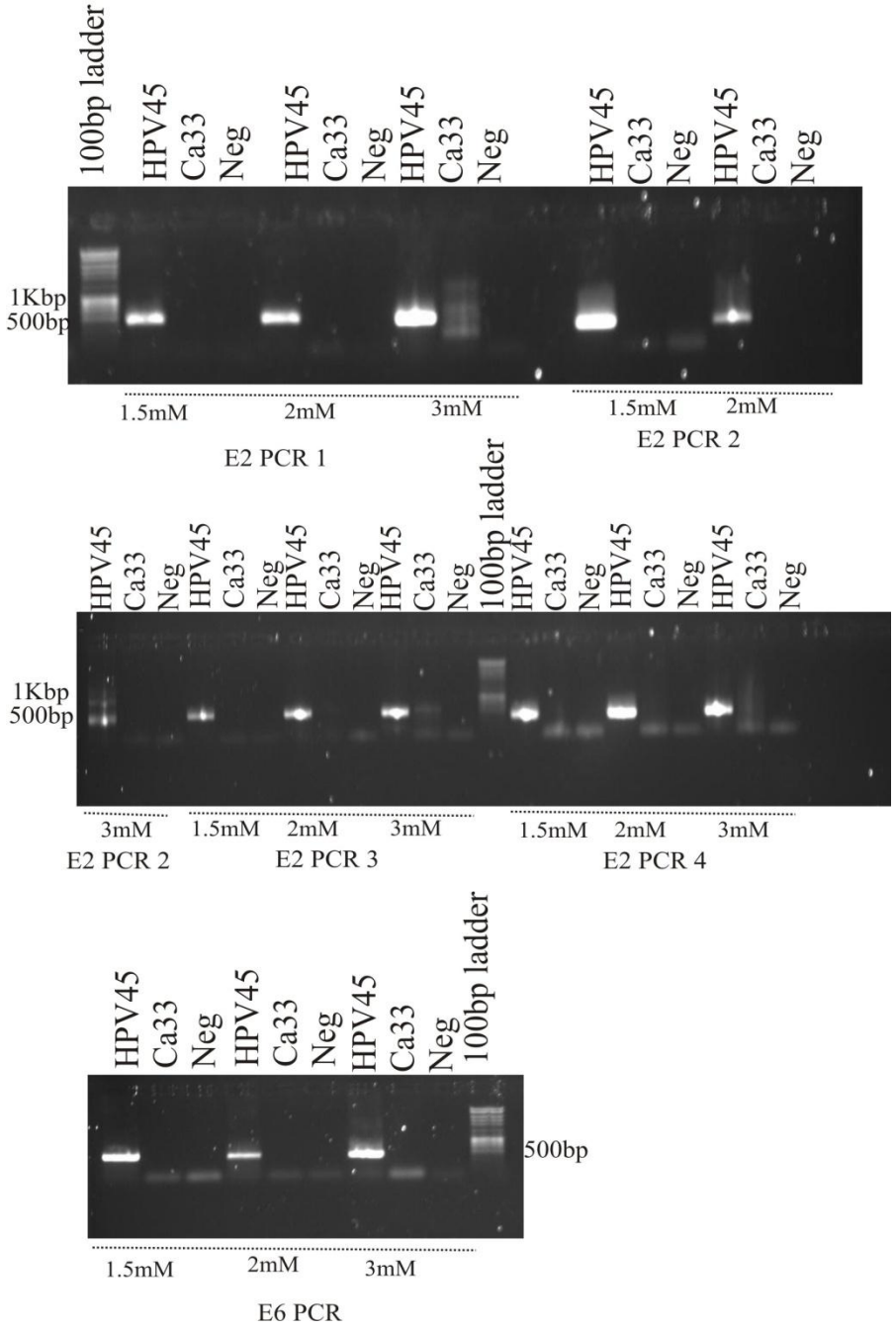


Figure 56 Optimisation of HPV45 E2 PCRs at differing $MgCl_2$ concentrations. HTB34™ has no E2 (not shown) thus HPV45 plasmid DNA was used as a positive control for optimisation and produced an amplicon for each E2 or E6 PCR. C33A was used as a HPV negative control and produced no amplicons. Water was used as a negative control.

10.3 Appendix 3:Tiffany Onions' PC08 and PC09 Data

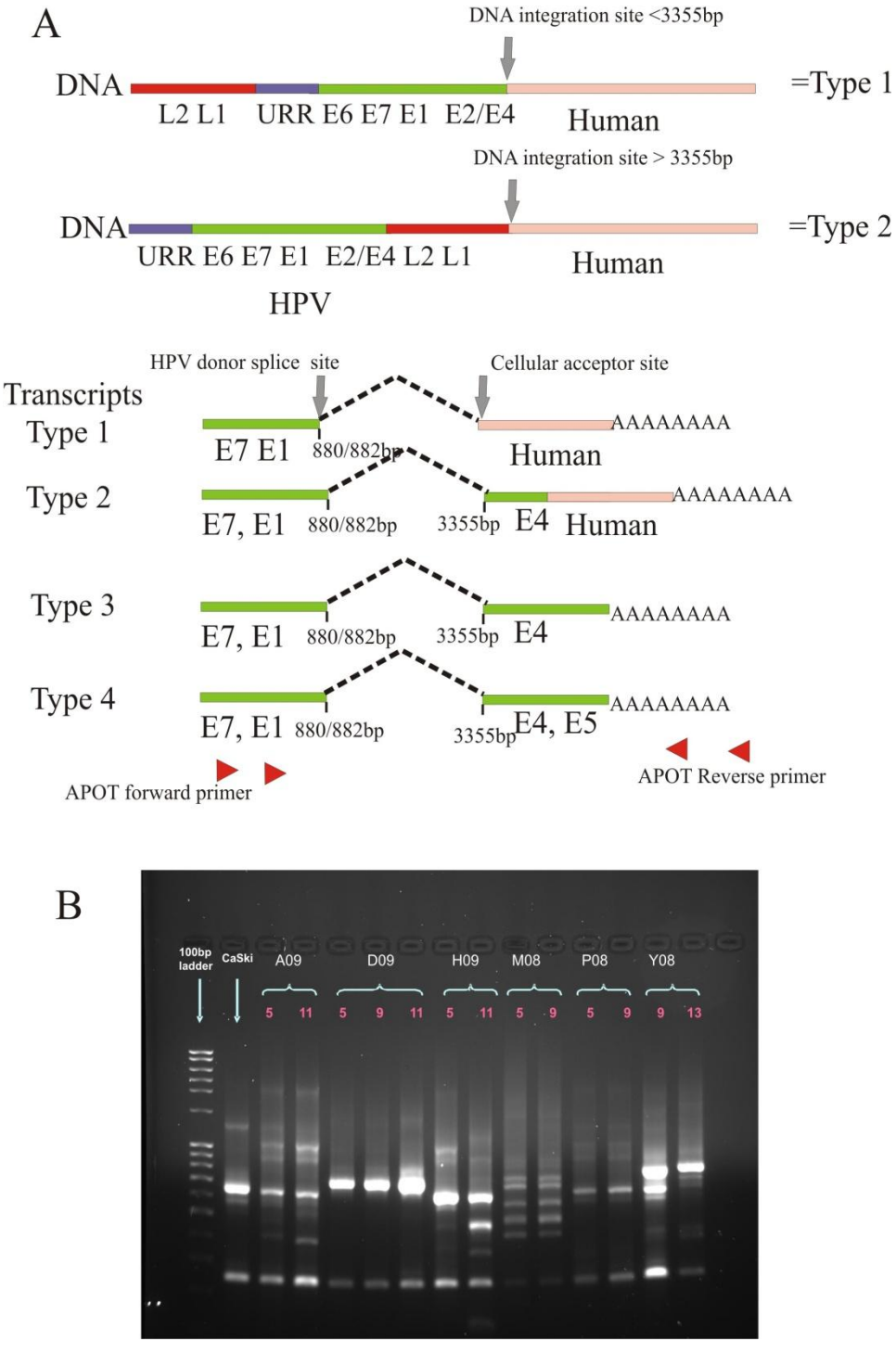


Figure 57 Transcripts detected in PC08 and PC09 by APOT. A) Schematic diagram of transcripts detected and how they relate to integration in DNA. Integration in DNA, disrupting HPV before 3355bp, will produce type 1 transcripts, integration in DNA, disrupting HPV after 3355bp, will result in a type 2 transcript. B) Electrophoretic separation of APOT amplicons for early and late passage clones A, D, H, M, P, and Y ; image used with permission from Tiffany Onions.

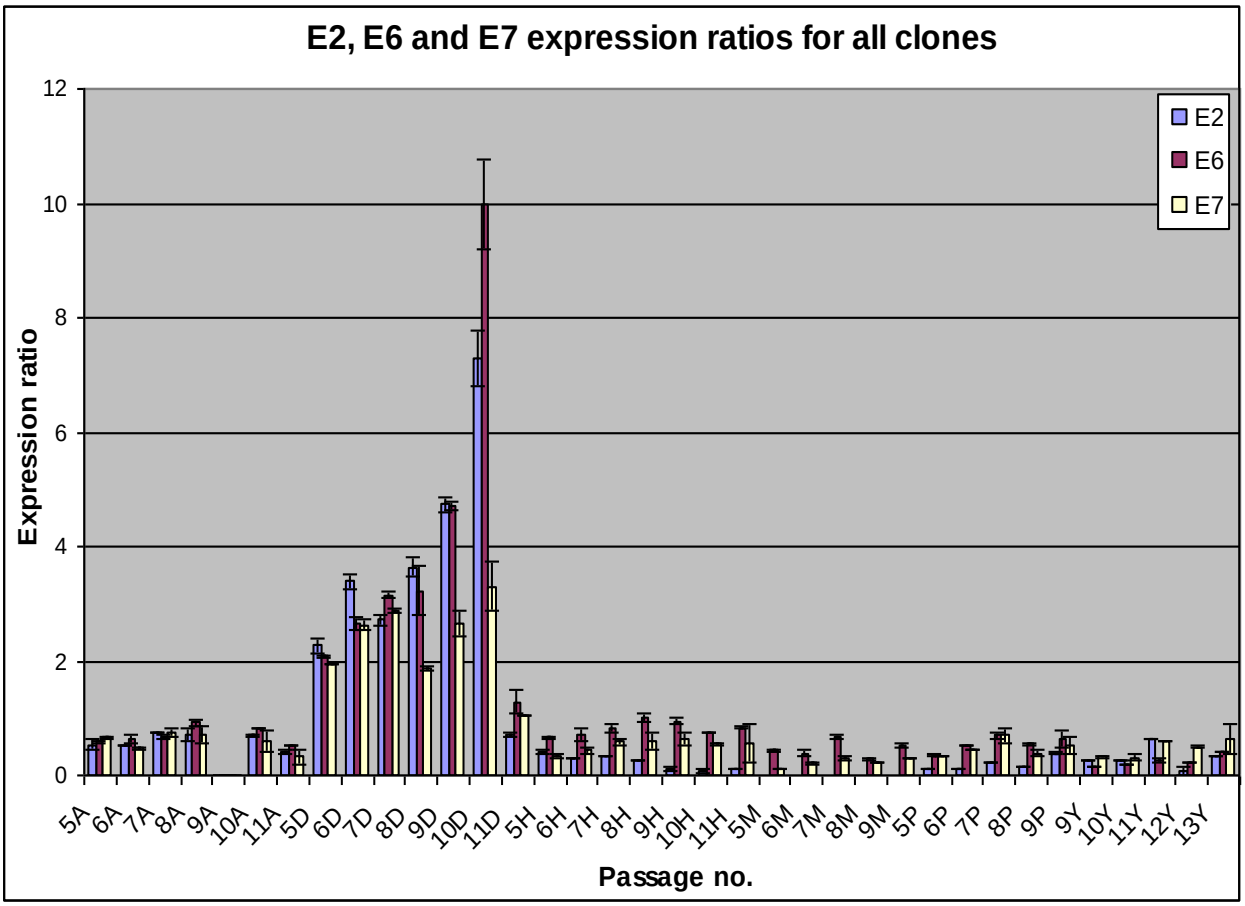
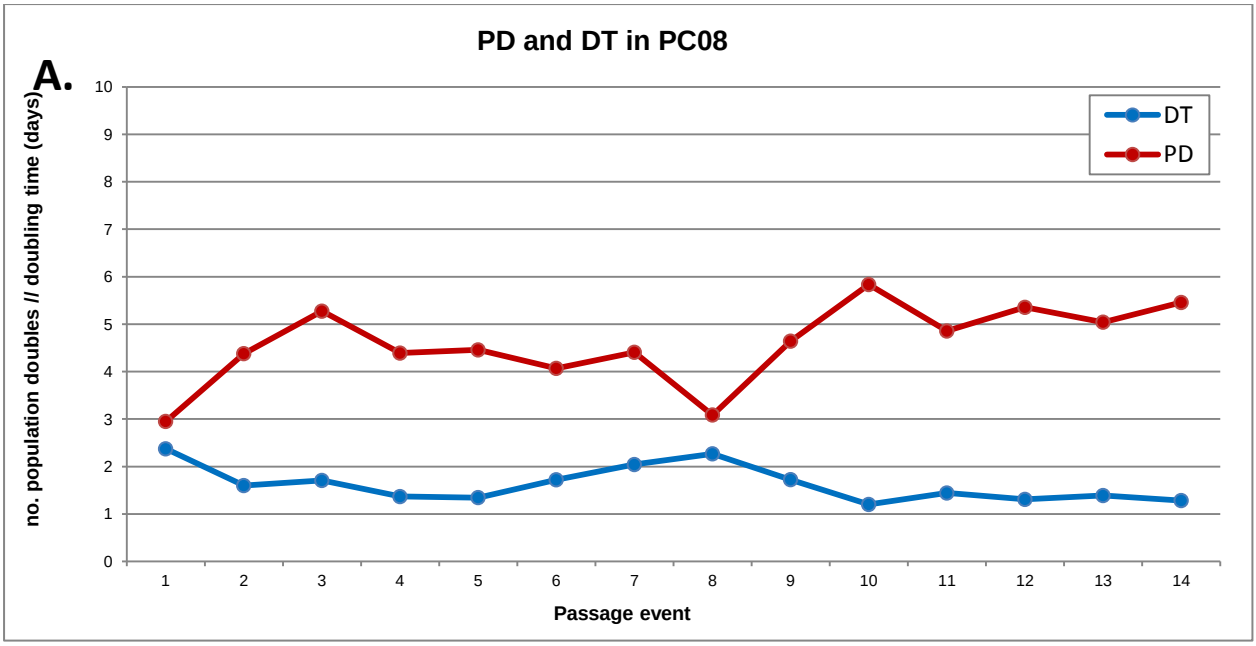


Figure 58 HPV16 E2, E6 and E7 expression ratios, relative to CaSki and house-keeping genes TBP2 and HPRT for all PC08 clones (M, P and Y) and PC09 (A, D and H) clones at each passage. Note that an expression ratio of 1 is equal to CaSki, more than 1 is more than CaSki; an expression ratio less than 1 denotes expression levels lower than CaSki.



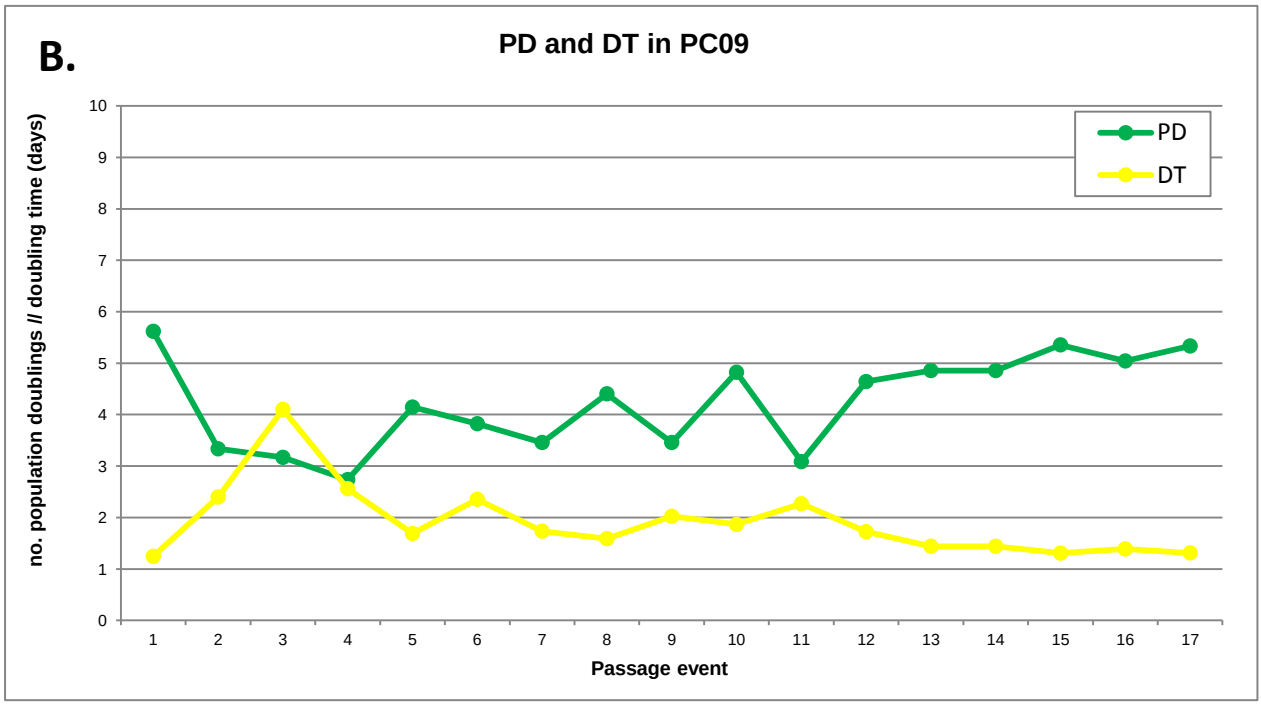


Figure 59 Population doubling (PD) and doubling time (DT) for heterogeneous PC08 (A) and PC09 (B) cell lines. Population doubling and doubling time fluctuates slightly over all passages for both cell lines. However, there appears to be a trend with increasing passage; population doubling is inversely proportional to doubling time.

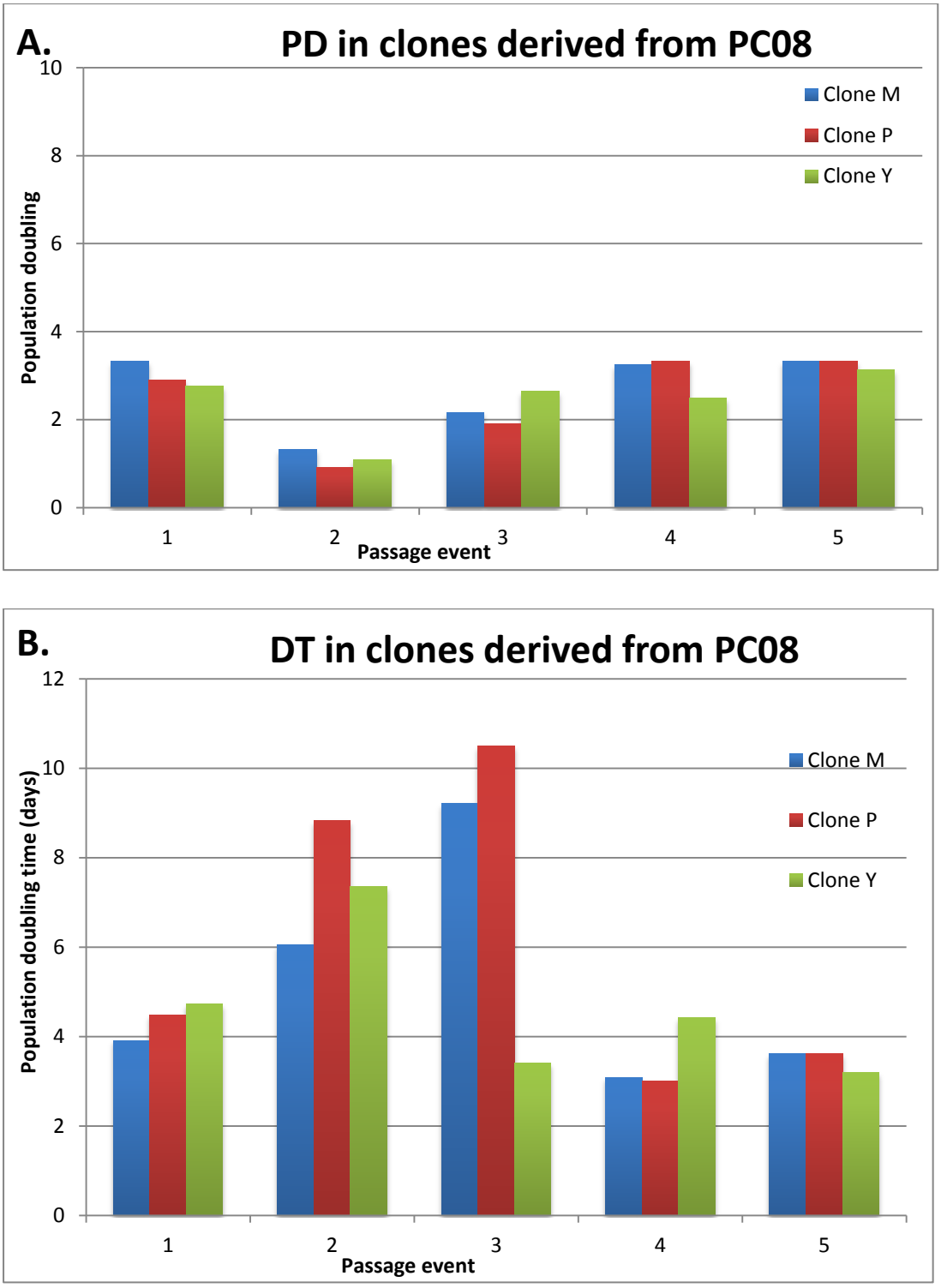
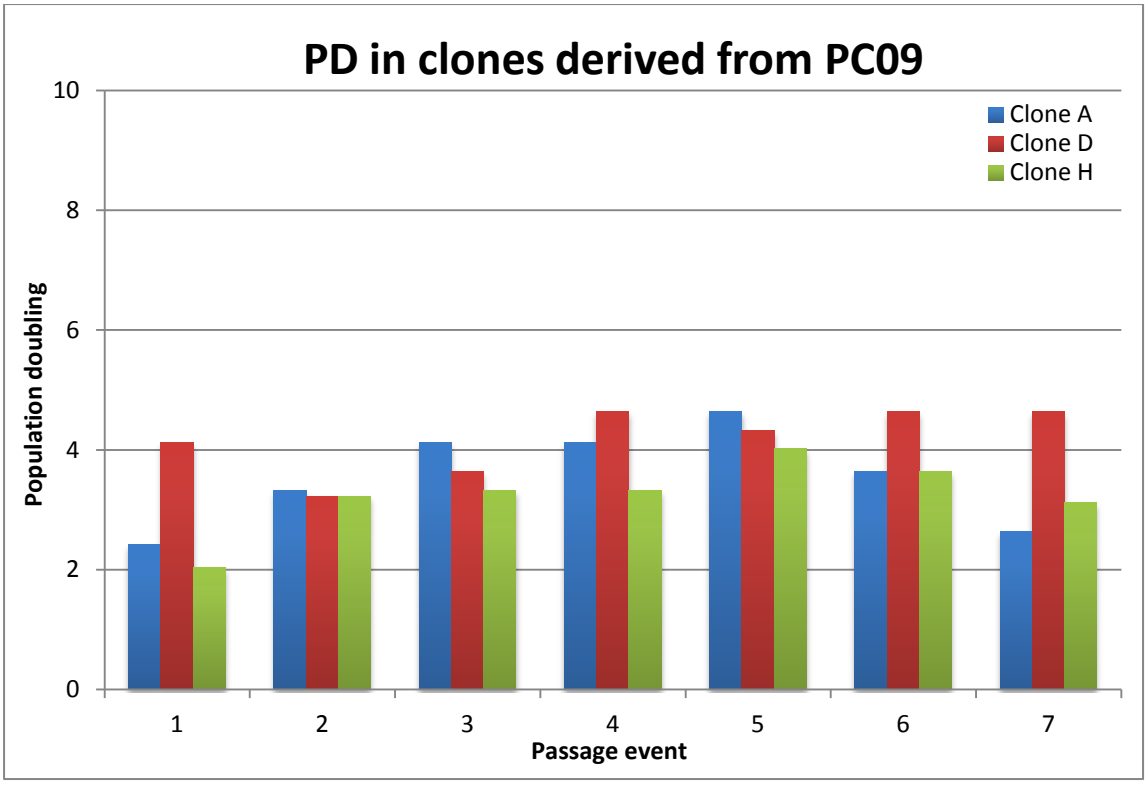


Figure 60 Population doubling (A) and doubling time (B) in clones derived from PC08.



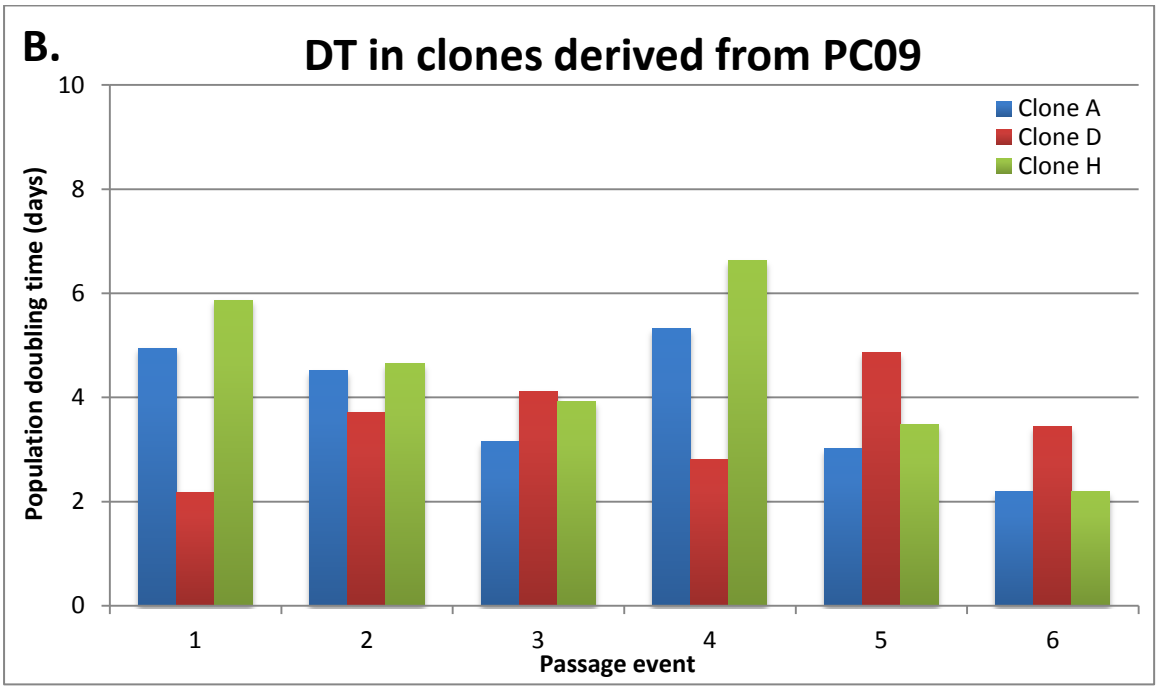


Figure 61 Population doubling (A) and doubling time (B) in clones derived from PC09.

10.4 Appendix 3: Dean Bryant's Methylation Data

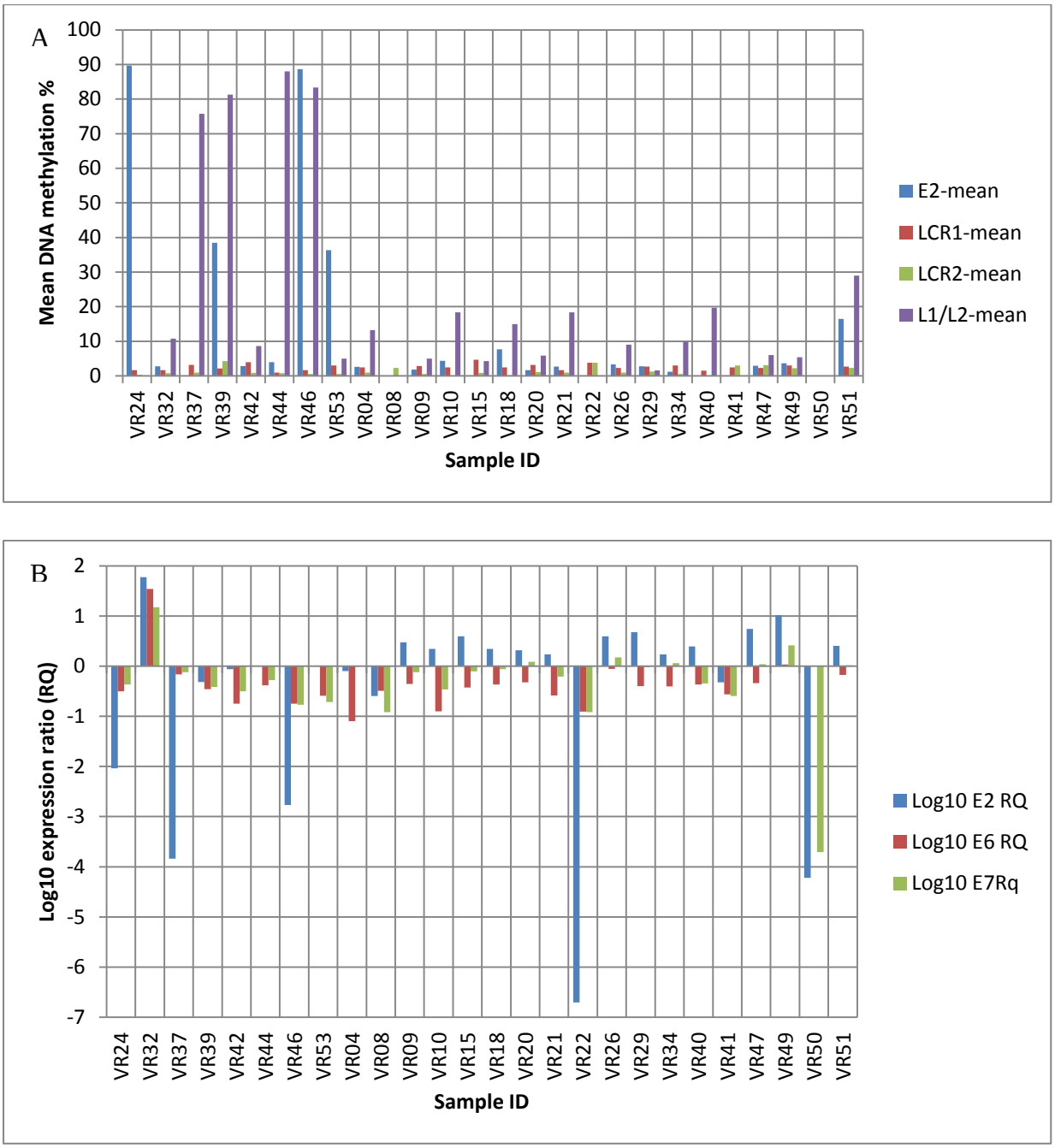


Figure 62 HPV16 URR methylation and HPV E2, E6 and E7 expression in VIN biopsies. A) Methylation at E2, L1, L2 promoters and Long Control Region (LCR also referred to as URR). B) Log_{10} expression ratios/ relative quantities (RQ) of E2, E6 and E7 relative to CaSki E2, E6 and E7 expression, normalised to housekeeping genes TBP2 and HPRT.

Sample	HPV disruption	Position (bp)	Human genome region
VR24	*E1*	1,540	Chr8 (q24.21)
VR32	*E2*	3,348	Chr2 (q35)
VR37	*E2* & *L2*	3,161 & 4,917	Chr1 (q32.1)
VR39	*E1*	1,786	Chr1 (p36.22)
VR42	*L2*	4,910	Satellite DNA*
VR44	*E2*	3,166	Chr17 (q24.3)
VR46	*E1*	2,508	Chr9 (q21.31)
VR53	*L2*	4,989	Chr3 (q26.31)

Table 57 Summary of integration events detected by DIPS by Dean Bryant in VIN biopsy samples. HPV disruption refers to the HPV16 ORF where the integration event disrupts HPV16. Position (bp) gives the integration site into the HPV genome. Human genome region gives the region in which integration occured. *Homology was found to HSATII satellite DNA in Chr7, Chr22, Chr2, Chr16, Chr10 and ChrY. VR37 had two detected integration sites, showing disruption of the *E2* and *L2* genes. Both events were within the same region of the human genome and were orientated in opposite directions and so are likely to represent the opposite ends of a single integration event rather than representing independent integration events.

10.5 Appendix 4: Associated Publications

This section shows the manuscript , written by the author, that was submitted to the Open Clincal Cancer Journal in November 2010:

Raybould, R. et al. 2011. Human Papillomavirus Integration and its Role in Cervical Malignant Progression. *Open Clinical Cancer Journal* 5, pp. 1-7.

Human Papillomavirus Integration And Its Role In Cervical Malignant Progression

Rachel Raybould, Alison Fiander and Sam Hibbitts

Obstetrics and Gynaecology, School of Medicine, Cardiff University, CF14 4XN

Telephone: 029 2074 4742

Email: raybouldre1@cf.ac.uk, fianderan@cf.ac.uk, hibbittssj@Cardiff.ac.uk

Abstract

High risk Human Papilloma Virus (HR-HPV) DNA integration into the human genome is one of the key stages in the progression of cervical neoplasia. This mini-review highlights the importance of HPV integration as a potential biomarker for cervical screening and briefly describes the main methods used to detect HPV integration: Amplification of Papillomavirus Oncogene Transcripts (APOT), Restriction Site PCR (RS-PCR), Southern blot and Detection of Integrated Papillomavirus Sequences (DIPS). The potential mechanisms of HPV integration are discussed with a focus on DNA instability; site of integration; and transcriptional regulation of integrants. This article provides an overview of the role HPV integration plays in malignant progression.

Keywords: Human Papillomavirus; integration; cervical cancer; E6 and E7

Introduction

Cervical cancer is the second most common cancer among women worldwide. Globally there are around 500,000 new cases and 274,000 deaths due to cervical cancer annually (WHO 2008). For centuries, cervical carcinoma has been recognised to behave as a sexually transmitted disease and in the mid 1970's it was proposed that there was an aetiological link with Human Papillomavirus (HPV) (zur Hausen et al. 1974; Meisels and Fortin 1976; zur Hausen 1976). Infection with HPV is now recognised as an essential, but not sufficient, factor for the development of cervical cancer with many HPV lesions regressing without a need for treatment. Current cervical screening serves to detect and treat precancerous cells (Cervical Intraepithelial Neoplasia) before invasive disease develops. A biomarker able to predict malignant progression in women at risk would enhance screening algorithms. High risk HPV (HR-HPV) DNA integration into the human genome is one of the key stages in malignant progression and is therefore a potential biomarker that precedes invasive disease.

HPV Genome

HPV is strictly epitheliotropic and infects epithelial cells either of the skin or mucous membranes, particularly of the anogenital tract and oropharynx. Viral transcription is tightly regulated and linked spatially and temporally with epithelial differentiation. The HPV genome consists of double stranded DNA (dsDNA) approximately 8kbp long contained within a capsid (Fig 1). Only one coding strand of the dsDNA acts as a template for transcription. The HPV genome is divided into three regions: an Early region containing genes encoding non-structural proteins (E), a Late region containing genes encoding capsid proteins (L), and an Upstream Regulatory Region (URR) (or Long Control Region), which contains a DNA replication origin, transcription regulatory sequences and one or more promoters which control expression of the viral oncoproteins E6 and E7.

HPV Proteins

The late proteins, L1 and L2, are expressed late in the HPV lifecycle in the upper, granular layer of the epidermis. The HPV late proteins make up the virion shell and play an important

role in mediating virus infectivity. To achieve a successful lifecycle HPV must be able to bind to cell surface receptors. Heparan sulphate proteoglycans on the cell surface are considered the primary binding sites for L1 and L2 from certain HPV evolutionary groups however different secondary receptors are involved for other HPV types.

The Early proteins are predominantly expressed in the basal and suprabasal layers of the epidermis. E5, E6 and E7 are oncogenic proteins (Table 1). E5 expression enhances oncogenic potential (Stoppler et al. 1996; Maufort et al. 2010) but the exact function of E5 remains poorly understood. E6 and E7 expression is essential for maintenance of the transformed state and malignant progression (von Knebel Doeberitz et al. 1988; Bosch et al. 1990) and the roles E6 and E7 play in carcinogenesis are well documented. HPV integration increases the production of E6 and E7 proteins and influences cancer progression through interaction with hTERT, p53 and Retinoblastoma protein (pRB). hTERT is a catalytic subunit of Telomerase that acts to synthesise telomere ends of linear chromosomes during DNA replication. p53 is a transcription factor that regulates cell cycle arrest, apoptosis, senescence, DNA repair and cell metabolism; p53 activity is inhibited by ubiquitin ligase which also ubiquitinates p53 to initiate p53 degradation. pRB is a tumour suppressor protein and interacts with transcription factor E2F to repress the transcription of genes required for the S phase of the cell cycle. In the normal HPV life-cycle expression of E5, E6 and E7 is tightly regulated within cells that are destined to be lost from the surface epithelial layers, such that they do not pose a carcinogenic threat (Fig 2).

HPV Integration

There are more than 100 types of HPV (WHO 2008), which can be sub-divided by their oncogenic risk. HPV16, HPV18, HPV31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV56 and HPV58 are the predominant HR types detected in cervical cancer cases and HR-HPV DNA is often integrated into the human genome. As disease progresses, the risk of integration is higher (Pirami et al. 1997; Klaes et al. 1999). Many studies of HPV16 have demonstrated that upon viral integration, variable parts of the HPV genome are disrupted; fragments containing E2, E4 and E5 ORFs are missing whereas the entire E1, E6 and E7 ORFs are integrated and retained (Matsukura et al. 1986; Pater et al. 1986).

Integration is not part of the normal HPV lifecycle, unlike the retroviruses that encode the protein integrase, which facilitates their integration into the host genome. The mechanisms by which HPV integrates its DNA into the human genome are not fully understood although there are a number of hypotheses:

DNA instability

There is considerable debate over whether integration precedes genetic instability or integration arises due to genetic instability. A number of studies have observed that DNA integration occurs in the presence of Double Strand Breaks (DSBs). DSBs occur in regions of DNA in which the DNA repair process has failed. Regions that harbour DNA instability, known as Chromosome Fragile Sites (CFSs), are distributed throughout the genome. Studies have reported an increased frequency of HR-HPV integration in regions of DNA that contain CFSs (Thorland et al. 2003; Wentzensen et al. 2004; Dall et al. 2008). However, it remains unclear as to whether DNA integration is more common in CFSs because DNA is more prone to breakage, as a result of factors such as increased expression of E6 and E7, or because the host DNA contains sequences that increase the likelihood of integration occurring.

Matzner et al. (Matzner et al. 2003) investigated the integration frequency of exogenous DNA into a breast cancer cell line containing many fragile sites within the genome. This study reported preferential integration of exogenous DNA into fragile sites of cellular DNA, supporting the hypothesis that integration occurs due to chromosome instability.

There is also speculation that HPV oncogenes E6 and E7 induce DNA damage (Duensing and Munger 2002) and increase the risk of HPV DNA integration (Kessis et al. 1996). Kessis et al (Kessis et al. 1996) assessed the integration frequency of a reporter plasmid in a cell line expressing individual E6 or E7 genes from HR and low-risk (LR) HPV types and proposed that the difference in oncogenic potential observed between HR and LR HPV types may be due to the increased ability of HR-HPV to integrate into host DNA.

Site of integration

Integration is random throughout the genome and much research has been carried out investigating sites of HR-HPV integration in cell lines and clinical samples of HPV related cancers at various sites of the body. A systematic review of HR-HPV integration in epithelial dysplasia and infections of the female lower genital tract assessed over 190 HR-HPV integration loci all of which were randomly located throughout the genome with a predilection of integration in CFSs (Wentzensen et al. 2004). Integration of HPV DNA into genomic regions harbouring oncogenes may also contribute to the malignant phenotype. A number of studies have observed disruption and/or over-expression of oncogenes as a consequence of integration at specific sites (Durst et al. 1987; Couturier et al. 1991; Einstein and Goldberg 2002; Ferber et al. 2003a; Thorland et al. 2003; Peter et al. 2006); however this data is not conclusive with contradicting reports from different research groups (Wentzensen et al. 2004; Dall et al. 2008).

Transcriptional regulation of integrants

As mentioned above, HR-HPV DNA integration occurs randomly throughout the human genome but not all integrants will be involved in malignant progression of the host cell. Transcription of integrated HR-HPV is regulated and both host cell and HPV regulatory elements have been implicated in this event. Disruption of regulatory elements may contribute to the fate of the cell.

HPV gene expression is tightly associated with the epithelial cell life cycle. Differentiated and quiescent epithelial cells are the natural host of HPV. When epithelial cells have undergone differentiation, the cells stop expressing genes involved in cell replication (Baker and Kligman 1967) and HPV must reactivate the host cell machinery in order to replicate (Fig 2). In low-grade cervical lesions, expression of HPV E6 and E7 in undifferentiated basal cells is repressed; this is partly due to host trans-acting repressors of the HPV URR, and as the cell differentiates E7 protein inactivates host proteins that repress transcription and viral

gene expression increases. In contrast, high grade lesions that contain integrated HPV have high levels of viral transcription throughout the epithelium (Durst et al. 1992). A study that injected an immortalised cell line into nude mice reported that viral transcription was more evident in basal cells where the HPV DNA was mainly integrated (Durst et al. 1991). It is probable that a combination of HPV DNA integration and host cell de-repression of viral gene expression contributes to the malignant phenotype.

The physical state of HPV DNA (integrated or episomal) plays a role in influencing viral gene regulation. Episomal HPV within the host cell contributes to regulation of integrant transcription, with E1 and E2 ORFs repressing immortalisation and over-expression of E6 and E7 genes. The loss of E1 or E2 ORFs following integration may therefore provide a selective growth advantage (Romanczuk and Howley 1992). Loss of episomes from integrant containing cells enables expression of E6 and E7, highlighting the importance of both episomal loss and integration in malignant progression (Pett et al. 2006). HPV infected cells can be treated to promote episome loss and recent studies have observed that this may speed up malignant progression (Herdman et al. 2006). Detection of HPV integration may potentially be a useful biomarker, which could be used to make more informed decisions about management of HPV infection and cervical disease.

Protein	Function/s	Molecular Activity
E1	Maintains viral genome. Contributes to viral genome replication.	E1 interacts with cellular DNA polymerase and replication proteins to initiate viral DNA replication.
E2	Maintains viral genome. Anchors episomes to mitotic chromosomes in basal cells. Controls transcription.	E2 engages E1 to viral origin and binds E1 to cellular proteins essential for DNA replication. Low levels of E2 activate transcription and high levels of E2 inhibit transcription through interaction with viral early promoters. E2 represses hTERT promoter activity.
E4	Function unclear. E4 has potential roles in the late stage of viral lifecycle, apoptosis and cell cycle arrest.	E4 may bind to keratins and lead to breakdown of the cytokeratin network. E4 may bind to cdk/cyclin complexes leading to arrest of the cell cycle in G2 phase. E4 may interact with E2 and mediate relocation of E2 from nucleus to cytoplasm.
E5	Transformation	HPV16 E5 interacts with E6 and E7 to stimulate cell proliferation. E5 may inhibit apoptosis and cell to cell communication.
E6	Transformation	HR-HPV E6 mediates ubiquitination and degradation of p53. HR-HPV E6 activates expression of hTERT.
E7	Transformation	E7 disrupts association of pRb with E2F transcription factors, activates cellular proteins involved in viral DNA replication and pushes the cell from G_1 phase to S phase. Stimulates cell proliferation through interaction with histone deacetylases, AP1 transcription complex, p21 and p27 cyclin-dependant kinase inhibitors.

Table 1, A summary of the molecular activity of HPV Early proteins.

Cervical Screening

Cervical screening serves to detect and treat precancerous cells before invasive disease develops however, cytological testing may lack sensitivity and an increasing number of women develop invasive cancer following a routine normal smear result. HPV testing can detect high grade lesions but also detects transient infections that are not associated with high grade lesion development; thus HPV testing has a good negative predictive value but a poor positive predictive value. High risk HPV (HR-HPV) DNA integration into the human genome is one of the key stages in the onset of malignant progression and so it makes a very plausible positive predictive biomarker of invasive disease. Not only would development of assays to detect integration improve cervical screening and improve management of women with

cervical HPV infections, the assays would be applicable to HPV related cancers in other parts of the body, in men and women e.g. head and neck cancers, vulval cancer and anogenital cancers. Thus understanding the biology and mechanisms behind HPV integration will aid the prevention and/or detection of many HPV related cancers.

HPV Vaccination

HPV16 and HPV18 are responsible for approximately 70% of cervical cancer cases globally (Clifford et al. 2003) and most studies have focused on investigating integration for these 2 genotypes. Current vaccination programs serve to protect women against HR HPV16 and HPV18 infection. Vaccination protects against HPV16 and HPV18 infections throughout the body; this in turn will reduce the prevalence of HPV related cancers including cancers of the anogenital region, vulva and head and neck. There are two vaccines currently available: Gardasil®, a quadrivalent HPV6/11/16/18 vaccine (Merck) and Cervarix™, a bivalent HPV16/18 vaccine (GSK). Vaccine trials have been conducted for both Cervarix™ (GSK001/007 (Harper et al. 2004) and PATRICIA (Paavonen et al. 2009)) and Gardasil® (Merck 007 (Villa et al. 2006), FUTURE I (Garland et al. 2007) and FUTURE II (2007)). The vaccine trials followed women from Asia, Australia, Europe, Latin America and North America for up to four years; see table 2 for an overview of trial designs. The vaccines reduce persistent HPV infection, of the targeted HPV types and the HPV6/11/16/18 vaccine eliminated genital warts in the vaccinated group compared to the placebo (Saleem et al. 2009). In the PATRICIA trial, Cervarix™ was at least 90% effective against persistent infection with HPV16/18; with a 100% reduction of high grade lesions of the cervix associated with HPV16/18 in the vaccine recipient group compared to the placebo group. The PATRICIA trial also indicated that the benefits of the vaccine would be greater in women that have not had previous exposure to HPV16/18 infections. Clinical trials of the vaccine have also suggested the possibility of cross-protection against other HR-types and there are some concerns that type-replacement may occur with one or more of the other HR types becoming more predominant in cancer cases in the absence of HPV16 and HPV18 (Saleem et al. 2009). It will therefore be advantageous to develop novel methods to investigate HPV integration for less prevalent HR-HPV types.

Clinical Trial	Vaccine	Endpoint	Outcome
GSK001/007	Cervarix™	N=1,113, AR=15-25 years, SF= 6 months, Mean FU= 48 months. Objective: to examine HPV16/18 infection incidence, infection persistence, CIN1+ and adverse events.	Efficacy was 91.6% against HPV16/18 infection and 100% against persistent infection with HPV16/18.
PATRICIA	Cervarix™	N=18,644, AR=15-25 years, SF= 12 months, Mean FU= 15 months. Objective: to examine HPV16/18 CIN2+, infection persistence, CIN1+ and adverse events.	There was 100% reduction against HPV16/18 CIN2/3. Efficacy against 6 month and 12 month HPV16/18 persistent infection was 93.8% and 91.2%, respectively.
Merck 007	Gardasil®	N=552, AR=16-23 years, SF= 6 months, Mean FU= 60 months. Objective: to examine HPV6/11/16/18 infection persistence, cervical and external genital disease and adverse events.	Persistent infection with HPV6/11/16/18 was reduced by 89%.
FUTURE I	Gardasil®	N=5,455, AR=16-24 years, SF= 6 months, Mean FU= 36 months. Objective: to examine HPV6/11/16/18 CIN1+, external genital lesions and adverse events.	HPV associated genital warts was reduced by 100%. Vaccination reduced vulvar, vaginal and perianal lesions by 34% and reduced cervical lesions by 20% regardless of HPV type..
FUTURE II	Gardasil®	N=12,167, AR=15-26 years, SF= 12 months, Mean FU= 36 months. Objective: to examine HPV16/18 CIN2+ and adverse events.	HPV associated genital warts was reduced by 98%. Estimated vaccine efficacy in cervical lesions regardless of HPV type was 17%.

Table **2**, A summary of HPV vaccine clinical trials. N= number of women, AR= age range, SF= screening frequency, Mean FU= mean follow-up, CIN=Cervical Intraepithelial Neoplasia, VIN= Vulval Intraepithelial Neoplasia, VAIN= Vaginal Intraepithelial Neoplasia.

Detection of HPV Integration

There are a number of methods currently employed to detect HR-HPV integrants in the human genome (summary in Table 3). 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.

The majority of studies have investigated integration in clinical samples whilst a number of studies have chosen to investigate integration mechanisms in carcinoma cell line cultures such as CaSki, HeLa and SiHa. The W12 cell line has been used as a model to study HPV16 integration. This cell line was derived from a low-grade cervical lesion infected with HPV16 (Stanley et al. 1989). Early passages of W12 contain HPV16 in an episomal form (ca 100 episomes per cell) and long-term passage produces cells containing fully integrated HPV16 DNA. The properties of the W12 cell line are rare and although cell lines enable practical approaches to investigate mechanisms of HPV integration they are limited and not available for types other than HPV16 and HPV18. Investigation of HPV integration, in large and well defined clinical cohorts, may give insight into natural integration events and determine whether HPV integration can be used as a reliable biomarker to predict malignant progression.

Conclusion

HR-HPV integration is vital, but not the only contributing factor to the development of invasive cervical cancer. Integration induces changes in the viral DNA structure that favours an immortalized host cell, although this event is disruptive to the viral lifecycle. Further studies of HPV integration are required in large clinical samples to improve our understanding of HPV integration and its role in the pathogenesis of cervical carcinoma.

Table **3**, A summary of the most common DNA and RNA based HR-HPV integration detection methods: RS-PCR= Restriction Site PCR; APOT=Amplification of Oncogene Transcripts; DIPS=Detection of Integrated Papillomavirus.

Method	Description	Strengths	Limitations
RS-PCR	Host/viral genomic regions are amplified by PCR using HPV specific primers and a primer designed to bind to restriction enzyme sites.	Can obtain DNA sequence of host/viral junctions.	Large concentrations of DNA required and labour intensive.
APOT	RT-PCR followed by PCR using HPV specific primer discriminates HPV mRNAs derived from integrated and episomal viral genomes.	Can obtain DNA sequence of host/viral junctions.	Labour intensive and expensive.
DIPS	Single-side-specific ligation-mediated PCR. Involves vectorette PCR and suppression PCR to detect integrated HPV DNA.	Can obtain DNA sequence of host/viral junctions.	Labour intensive.
Southern blot	Cellular DNA digestion and electrophoresis followed by hybridisation of labelled HPV DNA probes to determine the physical state (integrated or episomal) of HPV.	Can reliably distinguish episomal from integrated HPV DNA.	Uses large concentrations of DNA and labour intensive. The use of radio-labelled probes has health and safety implications.
Real-time PCR	Physical state of HPV is estimated by calculating HPV E2:E6/E7 ratio by real-time PCR amplification of HPV E2 and E6/E7.	Uses small concentrations of DNA and is less labour intensive.	HPV E2:E6/E7 ratio may not reliably distinguish integrated DNA in a background of episomal DNA. Consumables expensive.

Figures

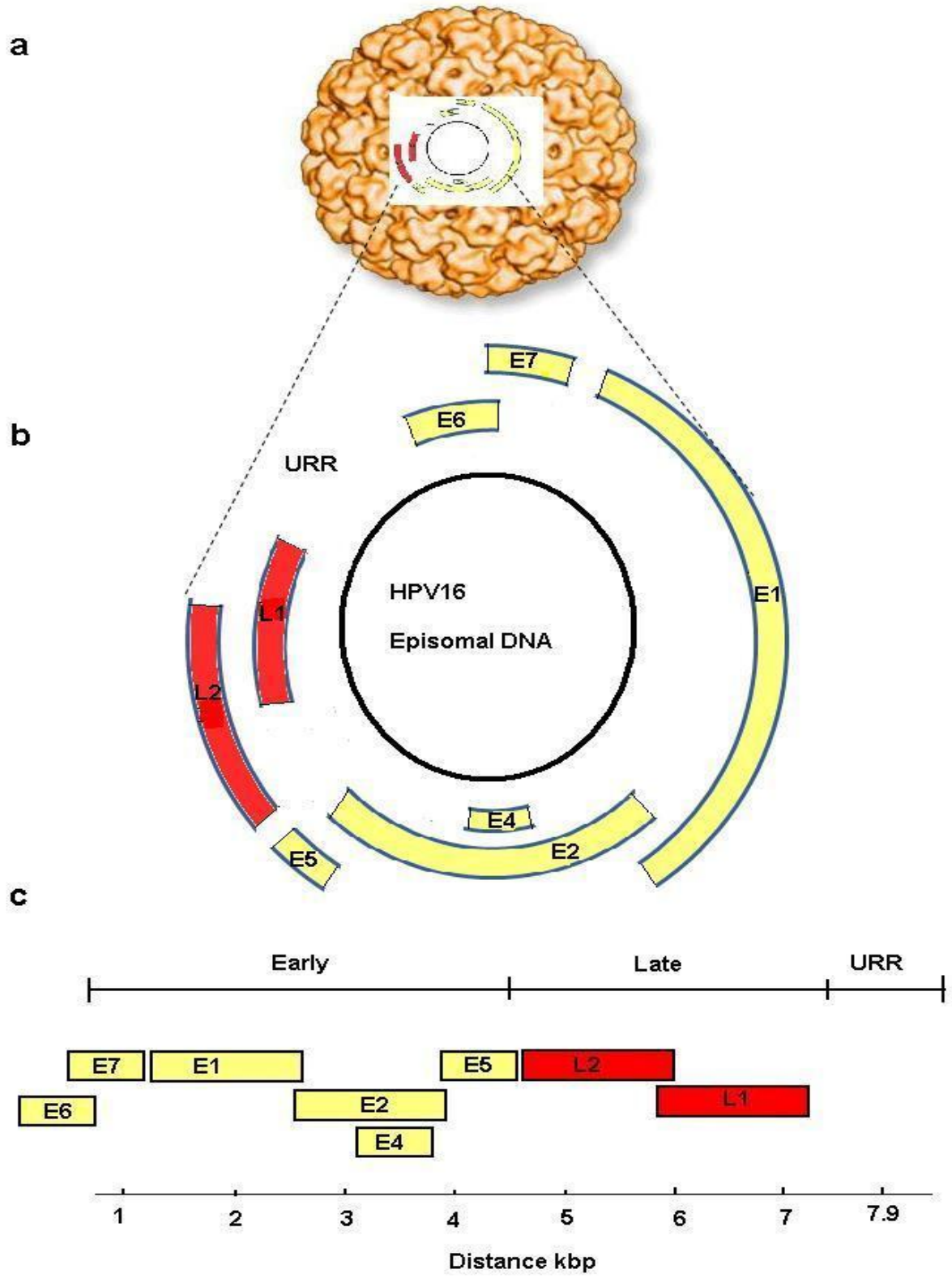
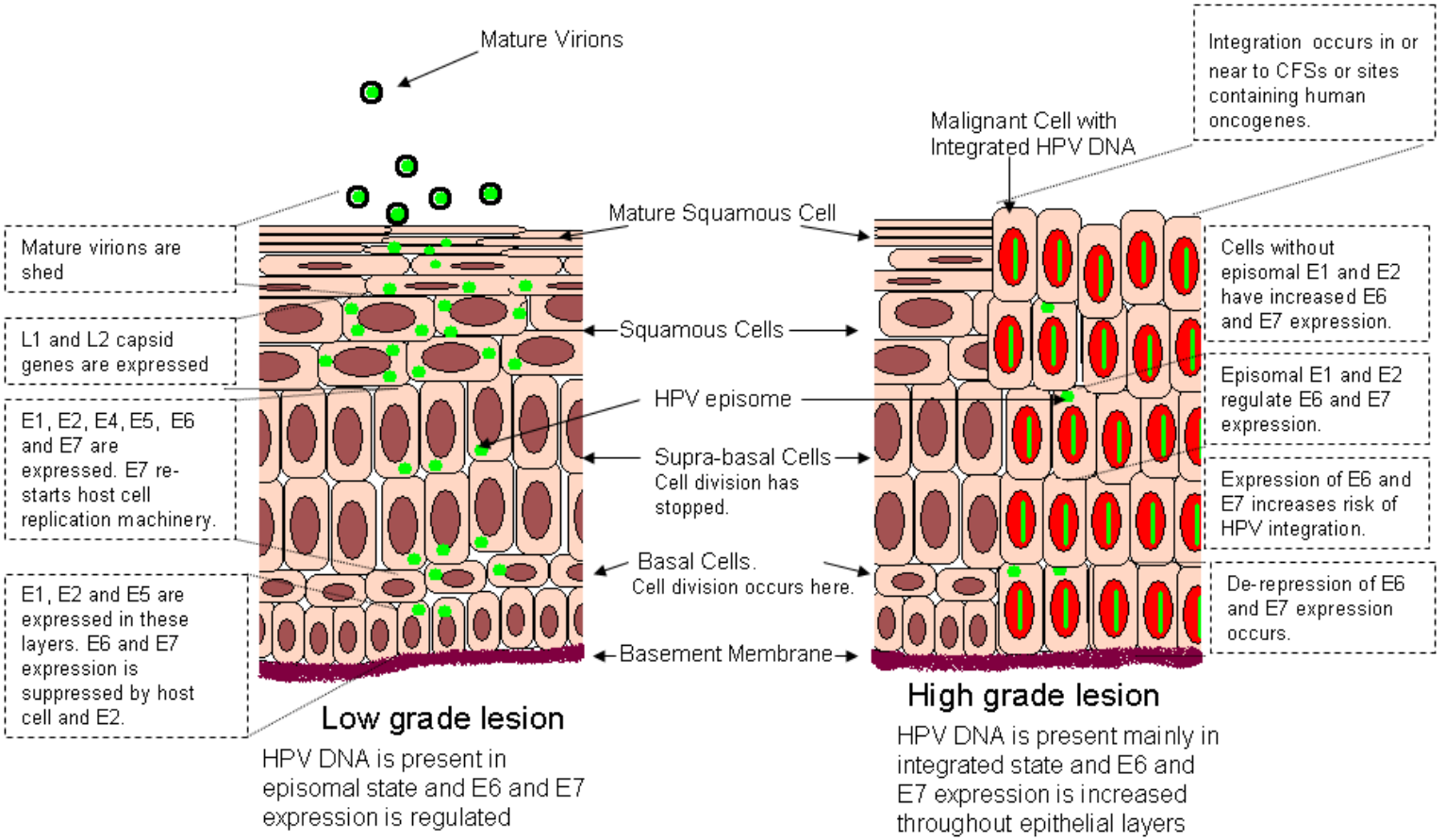


Fig 1 A schematic representation of HPV16: a) A representation of HPV16 episomal DNA contained within the HPV capsid, b) A schematic representation of HPV16 in episomal form, c) A schematic representation of linear HPV16

Fig 2 A schematic comparison of HPV physical state and HPV gene expression for low grade and high-grade cervical lesions. CFS= chromosome fragile site.



References:

[1] **WHO, Department of Reproductive Health and Research. Cervical cancer, human papillomavirus (HPV), and HPV vaccines - Key points for policy-makers and health professionals. 2008. Available from: http://whqlibdoc.who.int/hq/2008/WHO_RHR_08.14_eng.pdf**

[2] **Meisels A and Fortin R. Condylomatous lesions of the cervix and vagina. I. Cytologic patterns. Acta Cytol 1976; 20: 505-509.**

[3] **zur Hausen H, Meinhof W, Scheiber W and Bornkamm GW. Attempts to detect virus-secific DNA in human tumors. I. Nucleic acid hybridizations with complementary RNA of human wart virus. Int J Cancer 1974; 13: 650-656.**

[4] **zur Hausen H. Condylomata acuminata and human genital cancer. Cancer Res 1976; 36: 794.**

[5] **Stoppler MC, Straight SW, Tsao G, Schlegel R and McCance DJ. The E5 gene of HPV-16 enhances keratinocyte immortalization by full-length DNA. Virology 1996; 223: 251-254.**

[6] **Maufort JP, Shai A, Pitot HC and Lambert PF. A role for HPV16 E5 in cervical carcinogenesis. Cancer Res 2010; 70: 2924-2931.**

[7] **Bosch FX, Schwarz E, Boukamp P, Fusenig NE, Bartsch D and zur Hausen H. Suppression in vivo of human papillomavirus type 18 E6-E7 gene expression in nontumorigenic HeLa X fibroblast hybrid cells. J Virol 1990; 64: 4743-4754.**

[8] **von Knebel Doeberitz M, Oltersdorf T, Schwarz E and Gissmann L. Correlation of modified human papilloma virus early gene expression with altered growth properties in C4-1 cervical carcinoma cells. Cancer Res 1988; 48: 3780-3786.**

[9] **Klaes R, Woerner SM, Ridder R, Wentzensen N, Duerst M, Schneider A, Lotz B, Melsheimer P and von Knebel Doeberitz M. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. Cancer Res 1999; 59: 6132-6136.**

[10] **Pirami L, Giache V and Becciolini A. Analysis of HPV16, 18, 31, and 35 DNA in pre-invasive and invasive lesions of the uterine cervix. J Clin Pathol 1997; 50: 600-604.**

[11] **Matsukura T, Kanda T, Furuno A, Yoshikawa H, Kawana T and Yoshiike K. Cloning of monomeric human papillomavirus type 16 DNA integrated within cell DNA from a cervical carcinoma. J Virol 1986; 58: 979-982.**

[12] **Pater MM, Dunne J, Hogan G, Ghatage P and Pater A. Human papillomavirus types 16 and 18 sequences in early cervical neoplasia. Virology 1986; 155: 13-18.**

[13] **Dall KL, Scarpini CG, Roberts I, Winder DM, Stanley MA, Muralidhar B, Herdman MT, Pett MR and Coleman N. Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions. Cancer Res 2008; 68: 8249-8259.**

[14] **Wentzensen N, Vinokurova S and von Knebel Doeberitz M. 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: 3878-3884.**

[15] **Thorland EC, Myers SL, Gostout BS and Smith DI. Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. Oncogene 2003; 22: 1225-1237.**

[16] Matzner I, Savelyeva L and Schwab M. Preferential integration of a transfected marker gene into spontaneously expressed fragile sites of a breast cancer cell line. Cancer Lett 2003; 189: 207-219.

[17] Duensing S and Munger K. The human papillomavirus type 16 E6 and E7 oncoproteins independently induce numerical and structural chromosome instability. Cancer Res 2002; 62: 7075-7082.

[18] Kessis TD, Connolly DC, Hedrick L and Cho KR. Expression of HPV16 E6 or E7 increases integration of foreign DNA. Oncogene 1996; 13: 427-431.

[19] Couturier J, Sastre-Garau X, Schneider-Maunoury S, Labib A and Orth G. Integration of papillomavirus DNA near myc genes in genital carcinomas and its consequences for proto-oncogene expression. J Virol 1991; 65: 4534-4538.

[20] Durst M, Croce CM, Gissmann L, Schwarz E and Huebner K. Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas. Proc Natl Acad Sci U S A 1987; 84: 1070-1074.

[21] Einstein MH and Goldberg GL. Human papillomavirus and cervical neoplasia. Cancer Invest 2002; 20: 1080-1085.

[22] Ferber MJ, Montoya DP, Yu C, Aderca I, McGee A, Thorland EC, Nagorney DM, Gostout BS, Burgart LJ, Boix L, Bruix J, McMahon BJ, Cheung TH, Chung TK, Wong YF, Smith DI and Roberts LR. Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (hTERT) gene in liver and cervical cancers. Oncogene 2003; 22: 3813-3820.

[23] Peter M, Rosty C, Couturier J, Radvanyi F, Teshima H and Sastre-Garau X. MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors. Oncogene 2006; 25: 5985-5993.

[24] Baker H and Kligman AM. Technique for estimating turnover time of human stratum corneum. Archives of dermatology 1967; 95: 408-411.

[25] Durst M, Glitz D, Schneider A and zur Hausen H. Human papillomavirus type 16 (HPV 16) gene expression and DNA replication in cervical neoplasia: analysis by in situ hybridization. Virology 1992; 189: 132-140.

[26] Durst M, Bosch FX, Glitz D, Schneider A and zur Hausen H. Inverse relationship between human papillomavirus (HPV) type 16 early gene expression and cell differentiation in nude mouse epithelial cysts and tumors induced by HPV-positive human cell lines. J Virol 1991; 65: 796-804.

[27] Romanczuk H and Howley PM. Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. Proc Natl Acad Sci U S A 1992; 89: 3159-3163.

[28] Pett MR, Herdman MT, Palmer RD, Yeo GS, Shivji MK, Stanley MA and Coleman N. Selection of cervical keratinocytes containing integrated HPV16 associates with episome loss and an endogenous antiviral response. Proc Natl Acad Sci U S A 2006; 103: 3822-3827.

[29] Herdman MT, Pett MR, Roberts I, Alazawi WO, Teschendorff AE, Zhang XY, Stanley MA and Coleman N. Interferon-beta treatment of cervical keratinocytes naturally infected with human papillomavirus 16 episomes promotes rapid reduction in episome numbers and emergence of latent integrants. Carcinogenesis 2006; 27: 2341-2353.

[30] Clifford GM, Smith JS, Aguado T and Franceschi S. Comparison of HPV type distribution in high-grade cervical lesions and cervical cancer: a meta-analysis. Br J Cancer 2003; 89: 101-105.

[31] **Harper DM, Franco EL, Wheeler C, Ferris DG, Jenkins D, Schuind A, Zahaf T, Innis B, Naud P, De Carvalho NS, Roteli-Martins CM, Teixeira J, Blatter MM, Korn AP, Quint W and Dubin G. Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. Lancet 2004; 364: 1757-1765.**

[32] **Paavonen J, Naud P, Salmeron J, Wheeler CM, Chow SN, Apter D, Kitchener H, Castellsague X, Teixeira JC, Skinner SR, Hedrick J, Jaisamrarn U, Limson G, Garland S, Szarewski A, Romanowski B, Aoki FY, Schwarz TF, Poppe WA, Bosch FX, Jenkins D, Hardt K, Zahaf T, Descamps D, Struyf F, Lehtinen M, Dubin G and Greenacre M. Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. Lancet 2009; 374: 301-314.**

[33] **Villa LL, Costa RL, Petta CA, Andrade RP, Paavonen J, Iversen OE, Olsson SE, Hoye J, Steinwall M, Riis-Johannessen G, Andersson-Ellstrom A, Elfgren K, Krogh G, Lehtinen M, Malm C, Tamms GM, Giacoletti K, Lupinacci L, Railkar R, Taddeo FJ, Bryan J, Esser MT, Sings HL, Saah AJ and Barr E. High sustained efficacy of a prophylactic quadrivalent human papillomavirus types 6/11/16/18 L1 virus-like particle vaccine through 5 years of follow-up. British journal of cancer 2006; 95: 1459-1466.**

[34] **Garland SM, Hernandez-Avila M, Wheeler CM, Perez G, Harper DM, Leodolter S, Tang GW, Ferris DG, Steben M, Bryan J, Taddeo FJ, Railkar R, Esser MT, Sings HL, Nelson M, Boslego J, Sattler C, Barr E and Koutsky LA. Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. The New England journal of medicine 2007; 356: 1928-1943.**

[35] **Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. The New England journal of medicine 2007; 356: 1915-1927.**

[36] **Saleem A, Tristram A, Fiander A and Hibbitts S. Prophylactic HPV vaccination: a major breakthrough in the fight against cervical cancer? Minerva Med 2009; 100: 503-523.**

[37] **Stanley MA, Browne HM, Appleby M and Minson AC. Properties of a non-tumorigenic human cervical keratinocyte cell line. Int J Cancer 1989; 43: 672-676.**

11 Bibliography

2007. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. *The New England journal of medicine* 356(19), pp. 1915-1927.

Alazawi, W. et al. 2002. Changes in cervical keratinocyte gene expression associated with integration of human papillomavirus 16. *Cancer Res* 62(23), pp. 6959-6965.

Allen, D. et al. 2008. Reference gene selection for real-time rtPCR in human epidermal keratinocytes. *J Dermatol Sci* 49(3), pp. 217-225.

Andrei, G. et al. 1998. Antiproliferative effects of acyclic nucleoside phosphonates on human papillomavirus (HPV)-harboring cell lines compared with HPV-negative cell lines. *Oncol Res* 10(10), pp. 523-531.

Arbyn, M. et al. 2008. Perinatal mortality and other severe adverse pregnancy outcomes associated with treatment of cervical intraepithelial neoplasia: meta-analysis. *BMJ* 337, p. a1284.

Arbyn, M. et al. 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(3 Suppl 1), pp. S7-11.

Arbyn, M. et al. 2009. How to evaluate emerging technologies in cervical cancer screening? *International Journal of Cancer* 125(11).

Arnold, C. and Hodgson, I. J. 1991. Vectorette PCR: a novel approach to genomic walking. *PCR Methods Appl* 1(1), pp. 39-42.

Ashida, S. et al. 2006. Expression of novel molecules, MICAL2-PV (MICAL2 prostate cancer variants), increases with high Gleason score and prostate cancer progression. *Clin Cancer Res* 12(9), pp. 2767-2773.

Audet-Lapointe, P. et al. 1990. Vaginal intraepithelial neoplasia. *Gynecol Oncol* 36(2), pp. 232-239.

Auersperg, N. and Hawryluk, A. P. 1962. Chromosome observations on three epithelial-cell cultures derived from carcinomas of the human cervix. *J Natl Cancer Inst* 28, pp. 605-627.

Badal, V. et al. 2003. CpG methylation of human papillomavirus type 16 DNA in cervical cancer cell lines and in clinical specimens: genomic hypomethylation correlates with carcinogenic progression. *J Virol* 77(11), pp. 6227-6234.

Baker, H. and Kligman, A. M. 1967. Technique for estimating turnover time of human stratum corneum. *Archives of dermatology* 95(4), pp. 408-411.

Begum, S. et al. 2005. Tissue distribution of human papillomavirus 16 DNA integration in patients with tonsillar carcinoma. *Clinical Cancer Research* 11(16), pp. 5694-5699.

Bosch, F. X. et al. 1990. Suppression in vivo of human papillomavirus type 18 E6-E7 gene expression in nontumorigenic HeLa X fibroblast hybrid cells. *J Virol* 64(10), pp. 4743-4754.

Bustin, S. A. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 25(2), pp. 169-193.

Campitelli, M. et al. 2012. Human Papillomavirus Mutational Insertion: Specific Marker of Circulating Tumor DNA in Cervical Cancer Patients. *Plos One* 7(8).

Canfell, K. et al. 2004. The predicted effect of changes in cervical screening practice in the UK: results from a modelling study. *British journal of cancer* 91(3), pp. 530-536.

Cannizzaro, L. A. et al. 1988. Regional chromosome localization of human papillomavirus integration sites near fragile sites, oncogenes, and cancer chromosome breakpoints. *Cancer Genet Cytogenet* 33(1), pp. 93-98.

Castellsague, X. et al. 2007. HPV and Cervical Cancer in the World: 2007 Report. *Vaccine* 25(Supplement), pp. C27-C219.

Castle, P. E. et al. 2007. A cross-sectional study of a prototype carcinogenic human papillomavirus E6/E7 messenger RNA assay for detection of cervical precancer and cancer. *Clin Cancer Res* 13(9), pp. 2599-2605.

Chan, A. M. et al. 1994. Expression cDNA cloning of a novel oncogene with sequence similarity to regulators of small GTP-binding proteins. *Oncogene* 9(4), pp. 1057-1063.

Cheng, S. et al. 1995. Differentiation-dependent up-regulation of the human papillomavirus E7 gene reactivates cellular DNA replication in suprabasal differentiated keratinocytes. *Genes Dev* 9(19), pp. 2335-2349.

Cheung, J. L. et al. 2006. Viral load, E2 gene disruption status, and lineage of human papillomavirus type 16 infection in cervical neoplasia. *The Journal of infectious diseases* 194(12), pp. 1706-1712.

Clifford, G. M. et al. 2003. Comparison of HPV type distribution in high-grade cervical lesions and cervical cancer: a meta-analysis. *Br J Cancer* 89(1), pp. 101-105.

Collins, S. I. et al. 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(9), pp. 3828-3832.

Couturier, J. et al. 1991. Integration of papillomavirus DNA near myc genes in genital carcinomas and its consequences for proto-oncogene expression. *J Virol* 65(8), pp. 4534-4538.

Cuzick, J. et al. 2006. Overview of the European and North American studies on HPV testing in primary cervical cancer screening. *Int J Cancer* 119(5), pp. 1095-1101.

Cuzick, J. et al. 2003. Management of women who test positive for high-risk types of human papillomavirus: the HART study. *Lancet* 362(9399), pp. 1871-1876.

Dall, K. L. et al. 2008. Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions. *Cancer Res* 68(20), pp. 8249-8259.

Das, P. et al. 2012. HPV Genotyping and Site of Viral Integration in Cervical Cancers in Indian Women. *Plos One* 7(7), p. 9.

Davies, P. et al. 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(4), pp. 791-796.

Day, P. M. et al. 2004. Establishment of papillomavirus infection is enhanced by promyelocytic leukemia protein (PML) expression. *Proc Natl Acad Sci U S A* 101(39), pp. 14252-14257.

Day, P. M. et al. 1998. The papillomavirus minor capsid protein, L2, induces localization of the major capsid protein, L1, and the viral transcription/replication protein, E2, to PML oncogenic domains. *J Virol* 72(1), pp. 142-150.

de Koning, A. P. et al. 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 7(12), p. e1002384.

de Villiers, E. M. et al. 2004. Classification of papillomaviruses. *Virology* 324(1), pp. 17-27.

Diakomanolis, E. et al. 2002. Vaginal intraepithelial neoplasia: report of 102 cases. *Eur J Gynaecol Oncol* 23(5), pp. 457-459.

Dodge, J. A. et al. 2001. Clinical features and risk of recurrence among patients with vaginal intraepithelial neoplasia. *Gynecol Oncol* 83(2), pp. 363-369.

Doorbar, J. et al. 1991. Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. *Nature* 352(6338), pp. 824-827.

Dray, M. et al. 2005. p16(INK4a) as a complementary marker of high-grade intraepithelial lesions of the uterine cervix. I: Experience with squamous lesions in 189 consecutive cervical biopsies. *Pathology* 37(2), pp. 112-124.

Duensing, S. and Munger, K. 2002. The human papillomavirus type 16 E6 and E7 oncoproteins independently induce numerical and structural chromosome instability. *Cancer Res* 62(23), pp. 7075-7082.

Duensing, S. and Münger, K. 2003. Centrosome abnormalities and genomic instability induced by human papillomavirus oncoproteins. *Prog Cell Cycle Res* 5, pp. 383-391.

Durkin, S. G. and Glover, T. W. 2007. Chromosome fragile sites. *Annu Rev Genet* 41, pp. 169-192.

Durst, M. et al. 1991. Inverse relationship between human papillomavirus (HPV) type 16 early gene expression and cell differentiation in nude mouse epithelial cysts and tumors induced by HPV-positive human cell lines. *J Virol* 65(2), pp. 796-804.

Durst, M. et al. 1987. Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas. *Proc Natl Acad Sci U S A* 84(4), pp. 1070-1074.

Durst, M. et al. 1983. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proc Natl Acad Sci U S A* 80(12), pp. 3812-3815.

Durst, M. et al. 1992. Human papillomavirus type 16 (HPV 16) gene expression and DNA replication in cervical neoplasia: analysis by in situ hybridization. *Virology* 189(1), pp. 132-140.

Duvvuri, U. and Myers, J. N. 2009. Cancer of the head and neck is the sixth most common cancer worldwide. *Curr Probl Surg* 46(2), pp. 114-117.

Einstein, M. H. et al. 2002. Utilization of the human genome sequence localizes human papillomavirus type 16 DNA integrated into the TNFAIP2 gene in a fatal cervical cancer from a 39-year-old woman. *Clin Cancer Res* 8(2), pp. 549-554.

Einstein, M. H. and Goldberg, G. L. 2002. Human papillomavirus and cervical neoplasia. *Cancer Invest* 20(7-8), pp. 1080-1085.

el Awady, M. K. et al. 1987. Molecular analysis of integrated human papillomavirus 16 sequences in the cervical cancer cell line SiHa. *Virology* 159(2), pp. 389-398.

Evans, M. F. et al. 2008. Evidence of HPV16 integration in low- and high-grade cervical lesions that regress demonstrated by multiple displacement amplification and Southern blot hybridisation. *Journal of Clinical Pathology* 61(4), pp. 541-543.

Fakhry, C. et al. 2008. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst* 100(4), pp. 261-269.

Ferber, M. J. et al. 2003a. Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (hTERT) gene in liver and cervical cancers. *Oncogene* 22(24), pp. 3813-3820.

Ferber, M. J. et al. 2003b. Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. *Oncogene* 22(46), pp. 7233-7242.

Ferlay, J. et al. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International journal of cancer. Journal international du cancer* 127(12), pp. 2893-2917.

Flores, R. et al. 2006. Cross-sectional analysis of oncogenic HPV viral load and cervical intraepithelial neoplasia. *Int J Cancer* 118(5), pp. 1187-1193.

Florl, A. R. et al. 1999. DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas. *Br J Cancer* 80(9), pp. 1312-1321.

Fontaine, J. et al. 2005a. High level of correlation of human papillomavirus-16 DNA viral load estimates generated by three real-time PCR assays applied on genital specimens. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for*

Cancer Research, cosponsored by the American Society of Preventive Oncology 14(9), pp. 2200-2207.

Fontaine, J. et al. 2005b. High levels of HPV-16 DNA are associated with high-grade cervical lesions in women at risk or infected with HIV. *AIDS* 19(8), pp. 785-794.

Frattini, M. G. et al. 1996. In vitro synthesis of oncogenic human papillomaviruses requires episomal genomes for differentiation-dependent late expression. *Proc Natl Acad Sci U S A* 93(7), pp. 3062-3067.

Friedl, F. et al. 1970. Studies on a new human cell line (SiHa) derived from carcinoma of uterus. I. Its establishment and morphology. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine* 135(2), pp. 543-545.

Frohman, M. A. et al. 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A* 85(23), pp. 8998-9002.

Funk, J. O. et al. 1997. Inhibition of CDK activity and PCNA-dependent DNA replication by p21 is blocked by interaction with the HPV-16 E7 oncoprotein. *Genes Dev* 11(16), pp. 2090-2100.

Garland, S. M. et al. 2007. Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. *The New England journal of medicine* 356(19), pp. 1928-1943.

Gecz, J. et al. 1997. Gene structure and subcellular localization of FMR2, a member of a new family of putative transcription activators. *Genomics* 44(2), pp. 201-213.

Geisbill, J. et al. 1997. Detection and characterization of human papillomavirus type 45 DNA in the cervical carcinoma cell line MS751. *J Gen Virol* 78 (Pt 3), pp. 655-658.

Gey, G. O. et al. 1952. TISSUE CULTURE STUDIES OF THE PROLIFERATIVE CAPACITY OF CERVICAL CARCINOMA AND NORMAL EPITHELIUM. *Cancer Research* 12(4).

Graham, D. A. and 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. *Mol Pathol* 53(4), pp. 201-206.

Gray, E. et al. 2010. In vitro progression of human papillomavirus 16 episome-associated cervical neoplasia displays fundamental similarities to integrant-associated carcinogenesis. *Cancer research* 70(10), pp. 4081-4091.

Hamid, N. A. et al. 2009. The regulation of cell proliferation by the papillomavirus early proteins. *Cell Mol Life Sci* 66(10), pp. 1700-1717.

Hansen, R. S. et al. 1997. A variable domain of delayed replication in FRAXA fragile X chromosomes: X inactivation-like spread of late replication. *Proceedings of the National Academy of Sciences of the United States of America* 94(9), pp. 4587-4592.

Harper, D. M. et al. 2004. Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. *Lancet* 364(9447), pp. 1757-1765.

Harwood, C. A. et al. 2004. Increased risk of skin cancer associated with the presence of epidermodysplasia verruciformis human papillomavirus types in normal skin. *Br J Dermatol* 150(5), pp. 949-957.

Hecht, F. 1988. Fragile sites, cancer chromosome breakpoints, and oncogenes all cluster in light G bands. *Cancer Genet Cytogenet* 31(1), pp. 17-24.

Hellemans, J. et al. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8(2), p. R19.

Herdman, M. T. et al. 2006. Interferon-beta treatment of cervical keratinocytes naturally infected with human papillomavirus 16 episomes promotes rapid reduction in episome numbers and emergence of latent integrants. *Carcinogenesis* 27(11), pp. 2341-2353.

Heselmeyer, K. et al. 1997. Advanced-stage cervical carcinomas are defined by a recurrent pattern of chromosomal aberrations revealing high genetic instability and a consistent gain of chromosome arm 3q. *Genes Chromosomes Cancer* 19(4), pp. 233-240.

Heselmeyer-Haddad, K. et al. 2005. Genomic amplification of the human telomerase gene (TERC) in pap smears predicts the development of cervical cancer. *Am J Pathol* 166(4), pp. 1229-1238.

Hines, C. S. et al. 1998. DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins. *J Mol Biol* 276(4), pp. 809-818.

Ho, C. M. et al. 2011. Integration of human papillomavirus correlates with high levels of viral oncogene transcripts in cervical carcinogenesis. *Virus Res* 161(2), pp. 124-130.

Howley, P. M. and Lowy, D. R. 2007. Papillomaviruses.*Fields Virology Editors 5th Edition*. Vol. Volume 2. 2007 Lippincott Williams and Wilkins, a Wolters Kluwer Business, pp. 2300-2338.

Huang, L. W. et al. 2008. Integration of human papillomavirus type-16 and type-18 is a very early event in cervical carcinogenesis. *J Clin Pathol* 61(5), pp. 627-631.

Huang, S. Z. et al. 1989. [Detection of sickle cell gene by analysis of amplified DNA sequences]. *Yi Chuan Xue Bao* 16(6), pp. 475-482.

Hudson, J. B. et al. 1990. Immortalization and altered differentiation of human keratinocytes in vitro by the E6 and E7 open reading frames of human papillomavirus type 18. *J Virol* 64(2), pp. 519-526.

Ikeda, J. et al. 2009. Expression of CUB domain containing protein (CDCP1) is correlated with prognosis and survival of patients with adenocarcinoma of lung. *Cancer Sci* 100(3), pp. 429-433.

Ikeda, K. et al. 2006. Loss of expression of type IV collagen alpha5 and alpha6 chains in colorectal cancer associated with the hypermethylation of their promoter region. *Am J Pathol* 168(3), pp. 856-865.

Jacobs, M. V. et al. 1995. Group-specific differentiation between high- and low-risk human papillomavirus genotypes by general primer-mediated PCR and two cocktails of oligonucleotide probes. *J Clin Microbiol* 33(4), pp. 901-905.

Jacobs, M. V. et al. 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(3), pp. 791-795.

Jeon, S. et al. 1995. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J Virol* 69(5), pp. 2989-2997.

Jeon, S. and Lambert, P. F. 1995. Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. *Proc Natl Acad Sci U S A* 92(5), pp. 1654-1658.

Jones, C. et al. 1995. Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene CBL2. *Nature* 376(6536), pp. 145-149.

Jones, D. L. et al. 1997. The human papillomavirus E7 oncoprotein can uncouple cellular differentiation and proliferation in human keratinocytes by abrogating p21Cip1-mediated inhibition of cdk2. *Genes Dev* 11(16), pp. 2101-2111.

Jones, H. W. et al. 1971. George Otto Gey. (1899-1970). The HeLa cell and a reappraisal of its origin. *Obstet Gynecol* 38(6), pp. 945-949.

Jones, J. et al. 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(2), pp. 100-104.

Jones, M. H. et al. 1996. The mildly abnormal cervical smear: patient anxiety and choice of management. *J R Soc Med* 89(5), pp. 257-260.

Joyce, J. G. et al. 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(9), pp. 5810-5822.

Kadaja, M. et al. 2009a. Mechanism of genomic instability in cells infected with the high-risk human papillomaviruses. *PLoS pathogens* 5(4), p. e1000397.

Kadaja, M. et al. 2009b. Papillomavirus DNA replication - from initiation to genomic instability. *Virology* 384(2), pp. 360-368.

Kadaja, M. et al. 2007. Genomic instability of the host cell induced by the human papillomavirus replication machinery. *The EMBO journal* 26(8), pp. 2180-2191.

Kalantari, M. et al. 2008. Effects of cellular differentiation, chromosomal integration and 5-aza-2'-deoxycytidine treatment on human papillomavirus-16 DNA methylation in cultured cell lines. *Virology* 374(2), pp. 292-303.

KC53/61/65. 2012. Cervical Screening Wales. In: Helen Beer et al. eds. *CERVICAL SCREENING PROGRAMME*.

Kelly, R. S. et al. 2011. HPV testing as a triage for borderline or mild dyskaryosis on cervical cytology: results from the Sentinel Sites study. *Br J Cancer* 105(7), pp. 983-988.

Kent, W. J. et al. 2002. The human genome browser at UCSC. *Genome Res* 12(6), pp. 996-1006.

Kessis, T. D. et al. 1996. Expression of HPV16 E6 or E7 increases integration of foreign DNA. *Oncogene* 13(2), pp. 427-431.

Kim, K. et al. 2003. Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection. *J Virol* 77(23), pp. 12450-12459.

Kitchener, H. C. et al. 2006. Chapter 7: Achievements and limitations of cervical cytology screening. *Vaccine* 24 Suppl 3, pp. S3/63-70.

Klaes, R. et al. 1999. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. *Cancer Res* 59(24), pp. 6132-6136.

Klingelhutz, A. J. et al. 1996. Telomerase activation by the E6 gene product of human papillomavirus type 16. *Nature* 380(6569), pp. 79-82.

Kraus, I. et al. 2008. The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes. *Cancer Res* 68(7), pp. 2514-2522.

Kraus, I. et al. 2006. Presence of E6 and E7 mRNA from human papillomavirus types 16, 18, 31, 33, and 45 in the majority of cervical carcinomas. *J Clin Microbiol* 44(4), pp. 1310-1317.

Kremer, E. J. et al. 1991. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* 252(5013), pp. 1711-1714.

Kulasingam, S. L. et al. 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(2), pp. 92-100.

Kulmala, S. M. A. et al. 2006. Early integration of high copy HPV16 detectable in women with normal and low grade cervical cytology and histology. *Journal of Clinical Pathology* 59(5), pp. 513-517.

Kyrgiou, M. et al. 2006. Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis. *Lancet* 367(9509), pp. 489-498.

Lacey, C. J. 2005. Therapy for genital human papillomavirus-related disease. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology* 32 Suppl 1, pp. S82-90.

Lander, E. S. et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409(6822), pp. 860-921.

Li, H. et al. 2013. Preferential sites for the integration and disruption of human papillomavirus 16 in cervical lesions. *J Clin Virol*.

Li, W. et al. 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(12), pp. 1355-1361.

Lie, A. K. et al. 2005. DNA- versus RNA-based methods for human papillomavirus detection in cervical neoplasia. *Gynecol Oncol* 97(3), pp. 908-915.

Luft, F. et al. 2001. Detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells. *Int J Cancer* 92(1), pp. 9-17.

Lukusa, T. and Fryns, J. P. 2008. Human chromosome fragility. *Biochim Biophys Acta* 1779(1), pp. 3-16.

Magaldi, T. G. et al. 2012. Primary human cervical carcinoma cells require human papillomavirus E6 and E7 expression for ongoing proliferation. *Virology* 422(1), pp. 114-124.

Maglennon, G. A. et al. 2011. Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression. *Virology* 414(2), pp. 153-163.

Matsukura, T. et al. 1986. Cloning of monomeric human papillomavirus type 16 DNA integrated within cell DNA from a cervical carcinoma. *J Virol* 58(3), pp. 979-982.

Matzner, I. et al. 2003. Preferential integration of a transfected marker gene into spontaneously expressed fragile sites of a breast cancer cell line. *Cancer Lett* 189(2), pp. 207-219.

Maufort, J. P. et al. 2010. A role for HPV16 E5 in cervical carcinogenesis. *Cancer Res* 70(7), pp. 2924-2931.

Mayrand, M. H. et al. 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(3), pp. 615-623.

Meisels, A. and Fortin, R. 1976. Condylomatous lesions of the cervix and vagina. I. Cytologic patterns. *Acta Cytol* 20(6), pp. 505-509.

Meissner, J. D. 1999. Nucleotide sequences and further characterization of human papillomavirus DNA present in the CaSki, SiHa and HeLa cervical carcinoma cell lines. *J Gen Virol* 80 (Pt 7), pp. 1725-1733.

Melsheimer, P. et al. 2004. DNA aneuploidy and integration of human papillomavirus type 16 e6/e7 oncogenes in intraepithelial neoplasia and invasive squamous cell carcinoma of the cervix uteri. *Clin Cancer Res* 10(9), pp. 3059-3063.

Milligan, S. G. et al. 2007. Analysis of novel human papillomavirus type 16 late mRNAs in differentiated W12 cervical epithelial cells. *Virology* 360(1), pp. 172-181.

Mincheva, A. et al. 1987. Chromosomal integration sites of human papillomavirus DNA in three cervical cancer cell lines mapped by in situ hybridization. *Med Microbiol Immunol* 176(5), pp. 245-256.

Minner, F. and Poumay, Y. 2009. Candidate housekeeping genes require evaluation before their selection for studies of human epidermal keratinocytes. *J Invest Dermatol* 129(3), pp. 770-773.

Moberg, M. et al. 2004. Type-specific associations of human papillomavirus load with risk of developing cervical carcinoma in situ. *Int J Cancer* 112(5), pp. 854-859.

Molden, T. et al. 2006. Human papillomavirus E6/E7 mRNA expression in women younger than 30 years of age. *Gynecol Oncol* 100(1), pp. 95-100.

Molden, T. et al. 2005. Comparison of human papillomavirus messenger RNA and DNA detection: a cross-sectional study of 4,136 women >30 years of age with a 2-year follow-up of high-grade squamous intraepithelial lesion. *Cancer Epidemiol Biomarkers Prev* 14(2), pp. 367-372.

Moore, P. S. and Chang, Y. 2010. Why do viruses cause cancer? Highlights of the first century of human tumour virology. *Nature reviews. Cancer* 10(12), pp. 878-889.

Morrison, T. B. et al. 1998. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques* 24(6), pp. 954-958, 960, 962.

Moscufo, N. et al. 1999. Two distinct regions of the BPV1 E1 replication protein interact with the activation domain of E2. *Virus Res* 65(2), pp. 141-154.

Moss, S. et al. 2006. Effect of testing for human papillomavirus as a triage during screening for cervical cancer: observational before and after study. *Bmj* 332(7533), pp. 83-85.

Munger, K. et al. 2004. Mechanisms of human papillomavirus-induced oncogenesis. *J Virol* 78(21), pp. 11451-11460.

Munoz, N. et al. 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *The New England journal of medicine* 348(6), pp. 518-527.

Münger, K. et al. 2001. Biological activities and molecular targets of the human papillomavirus E7 oncoprotein. *Oncogene* 20(54), pp. 7888-7898.

Nakahara, T. et al. 2005. Human papillomavirus type 16 E1circumflexE4 contributes to multiple facets of the papillomavirus life cycle. *J Virol* 79(20), pp. 13150-13165.

NHSCSP, P. N. 2010. Colposcopy and Programme

Management. In: David Luesley, S.L. ed. *Guidelines for the NHS Cervical Screening Programme*. NHS Cancer Screening Programmes.

ONS. 2010. *Office for National Statistics Cancer Statistics Registrations, England (Series MB1), No. 41, 2010*.

Ozeri-Galai, E. et al. 2011. Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. *Mol Cell* 43(1), pp. 122-131.

Paavonen, J. et al. 2009. Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. *Lancet* 374(9686), pp. 301-314.

Pater, M. M. et al. 1986. Human papillomavirus types 16 and 18 sequences in early cervical neoplasia. *Virology* 155(1), pp. 13-18.

Pater, M. M. and Pater, A. 1985. Human papillomavirus types 16 and 18 sequences in carcinoma cell lines of the cervix. *Virology* 145(2), pp. 313-318.

Pattillo, R. A. et al. 1977. Tumor antigen and human chorionic gonadotropin in CaSki cells: a new epidermoid cervical cancer cell line. *Science* 196(4297), pp. 1456-1458.

Peitsaro, P. et al. 2002. Integrated human papillomavirus type 16 is frequently found in cervical cancer precursors as demonstrated by a novel quantitative real-time PCR technique. *J Clin Microbiol* 40(3), pp. 886-891.

Peter, M. et al. 2006. MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors. *Oncogene* 25(44), pp. 5985-5993.

Peter, M. et al. 2010. Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma. *J Pathol* 221(3), pp. 320-330.

Peto, J. et al. 2004. The cervical cancer epidemic that screening has prevented in the UK. *Lancet* 364(9430), pp. 249-256.

Pett, M. and Coleman, N. 2007. Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? *J Pathol* 212(4), pp. 356-367.

Pett, M. R. et al. 2004. Acquisition of high-level chromosomal instability is associated with integration of human papillomavirus type 16 in cervical keratinocytes. *Cancer Res* 64(4), pp. 1359-1368.

Pett, M. R. et al. 2006. Selection of cervical keratinocytes containing integrated HPV16 associates with episome loss and an endogenous antiviral response. *Proc Natl Acad Sci U S A* 103(10), pp. 3822-3827.

Pirami, L. et al. 1997. Analysis of HPV16, 18, 31, and 35 DNA in pre-invasive and invasive lesions of the uterine cervix. *J Clin Pathol* 50(7), pp. 600-604.

Powell, N. G. et al. 2011. The risk of cervical cancer associated with specific types of human papillomavirus: a case-control study in a UK population. *International journal of cancer. Journal international du cancer* 128(7), pp. 1676-1682.

Ramanakumar, A. V. et al. 2010. Human papillomavirus (HPV) types 16, 18, 31, 45 DNA loads and HPV-16 integration in persistent and transient infections in young women. *Bmc Infectious Diseases* 10.

Raybould, R. et al. 2011. Human Papillomavirus Integration and its Role in Cervical Malignant Progression. *Open Clinical Cancer Journal* 5, pp. 1-7.

Richart, R. M. 1973. Cervical intraepithelial neoplasia. *Pathol Annu* 8, pp. 301-328.

Rieck, G. C. et al. 2006. Cervical screening in 20-24-year olds. *Journal of medical screening* 13(2), pp. 64-71; discussion 62-63.

Rigoni, S. 1987. Statistical facts about cancers on which Doctor Rigoni-Stern based his contribution to the Surgeons' Subgroup of the IV Congress of the Italian Scientists on 23 September 1842. (translation). *Statistics in medicine* 6(8), pp. 881-884.

Roberts, I. et al. 2008. Critical evaluation of HPV16 gene copy number quantification by SYBR green PCR. *BMC Biotechnol* 8, p. 57.

Romanczuk, H. and Howley, P. M. 1992. Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. *Proc Natl Acad Sci U S A* 89(7), pp. 3159-3163.

Rome, R. M. and England, P. G. 2000. Management of vaginal intraepithelial neoplasia: A series of 132 cases with long-term follow-up. *Int J Gynecol Cancer* 10(5), pp. 382-390.

Rozen, S. and Skaletsky, H. J. 2000. *Primer3 on the WWW for general users and for biologist programmers.* Humana Press, Totowa, NJ,.

Saiki, R. K. et al. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239(4839), pp. 487-491.

Saleem, A. et al. 2009. Prophylactic HPV vaccination: a major breakthrough in the fight against cervical cancer? *Minerva Med* 100(6), pp. 503-523.

Sanger, F. et al. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74(12), pp. 5463-5467.

Sarafi, T. R. and McBride, A. A. 1995. Domains of the BPV-1 E1 replication protein required for origin-specific DNA binding and interaction with the E2 transactivator. *Virology* 211(2), pp. 385-396.

Sarkar, G. et al. 1993. Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. *PCR Methods Appl* 2(4), pp. 318-322.

Sasieni, P. et al. 2003. Benefit of cervical screening at different ages: evidence from the UK audit of screening histories. *British journal of cancer* 89(1), pp. 88-93.

Sasieni, P. et al. 2009. Effectiveness of cervical screening with age: population based case-control study of prospectively recorded data. *BMJ* 339, p. b2968.

Sasieni, P. et al. 2010. The Impact Of Cervical Screening On Young Women: A Critical Review Of The Literature 2002–2009.*NHSCSP Publication No 31*. NHS Cancer Screening Programmes.

Scheffner, M. 1998. Ubiquitin, E6-AP, and their role in p53 inactivation. *Pharmacol Ther* 78(3), pp. 129-139.

Scheffner, M. et al. 1993. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* 75(3), pp. 495-505.

Scheffner, M. et al. 1990. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* 63(6), pp. 1129-1136.

Scherl-Mostageer, M. et al. 2001. Identification of a novel gene, CDCP1, overexpressed in human colorectal cancer. *Oncogene* 20(32), pp. 4402-4408.

Schiffman, M. et al. 2007. Human papillomavirus and cervical cancer. *Lancet* 370(9590), pp. 890-907.

Schiller, J. T. et al. 2012. A review of clinical trials of human papillomavirus prophylactic vaccines. *Vaccine* 30 Suppl 5, pp. F123-138.

Schmitz, M. et al. 2012. Loss of gene function as a consequence of human papillomavirus DNA integration. *International Journal of Cancer* 131(5), pp. E593-E602.

Schwartz, S. 2000. Regulation of human papillomavirus late gene expression. *Ups J Med Sci* 105(3), pp. 171-192.

Seo, S. R. et al. 2004. The novel E3 ubiquitin ligase Tiul1 associates with TGIF to target Smad2 for degradation. *EMBO J* 23(19), pp. 3780-3792.

Shah, R. et al. 2009. The prolyl 3-hydroxylases P3H2 and P3H3 are novel targets for epigenetic silencing in breast cancer. *Br J Cancer* 100(10), pp. 1687-1696.

Shrivastav, M. et al. 2008. Regulation of DNA double-strand break repair pathway choice. *Cell research* 18(1), pp. 134-147.

Siebert, P. D. et al. 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res* 23(6), pp. 1087-1088.

Solomon, D. et al. 2002. The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA* 287(16), pp. 2114-2119.

Solomon, D. et al. 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(4), pp. 293-299.

Stanley, M. A. et al. 1989. Properties of a non-tumorigenic human cervical keratinocyte cell line. *Int J Cancer* 43(4), pp. 672-676.

Statistics, O. f. N. 2000. *Cancer Statistics Registrations, England (Series MB1) - No. 31, 2000*.

Steger, G. and Corbach, S. 1997. Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein. *J Virol* 71(1), pp. 50-58.

Stoppler, M. C. et al. 1996. The E5 gene of HPV-16 enhances keratinocyte immortalization by full-length DNA. *Virology* 223(1), pp. 251-254.

Stubenrauch, F. and Laimins, L. A. 1999. Human papillomavirus life cycle: active and latent phases. *Seminars in cancer biology* 9(6), pp. 379-386.

Stubenrauch, F. et al. 1998. Differential requirements for conserved E2 binding sites in the life cycle of oncogenic human papillomavirus type 31. *J Virol* 72(2), pp. 1071-1077.

Stunkel, W. and Bernard, H. U. 1999. The chromatin structure of the long control region of human papillomavirus type 16 represses viral oncoprotein expression. *J Virol* 73(3), pp. 1918-1930.

Sutcliffe, J. G. 1979. COMPLETE NUCLEOTIDE-SEQUENCE OF THE ESCHERICHIA-COLI PLASMID-PBR322. *Cold Spring Harbor Symposia on Quantitative Biology* 43, pp. 77-90.

Sutherland, G. R. et al. 1998. Fragile sites still breaking. *Trends in genetics : TIG* 14(12), pp. 501-506.

Sutherland, G. R. and Richards, R. I. 1999. Fragile sites-cytogenetic similarity with molecular diversity. *Am J Hum Genet* 64(2), pp. 354-359.

Sykes, J. A. et al. 1970. Some properties of a new epithelial cell line of human origin. *J Natl Cancer Inst* 45(1), pp. 107-122.

Takai, D. et al. 2000. Hypomethylation of LINE1 retrotransposon in human hepatocellular carcinomas, but not in surrounding liver cirrhosis. *Jpn J Clin Oncol* 30(7), pp. 306-309.

Tan, S. H. et al. 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(10), pp. 6411-6420.

Thorland, E. C. et al. 2003. Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. *Oncogene* 22(8), pp. 1225-1237.

Thorland, E. C. et al. 2000. Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. *Cancer Res* 60(21), pp. 5916-5921.

Tiainen, P. et al. 2008. Characterization of recombinant human prolyl 3-hydroxylase isoenzyme 2, an enzyme modifying the basement membrane collagen IV. *J Biol Chem* 283(28), pp. 19432-19439.

Tristram, A. and Fiander, A. 2005. Clinical responses to Cidofovir applied topically to women with high grade vulval intraepithelial neoplasia. *Gynecol Oncol* 99(3), pp. 652-655.

Twigg, A. J. and Sherratt, D. 1980. Trans-complementable copy-number mutants of plasmid ColE1. *Nature* 283(5743), pp. 216-218.

Valerie, K. and Povirk, L. F. 2003. Regulation and mechanisms of mammalian double-strand break repair. *Oncogene* 22(37), pp. 5792-5812.

van der Marel, J. et al. 2012. Molecular mapping of high-grade cervical intraepithelial neoplasia shows etiological dominance of HPV16. *Int J Cancer* 131(6), pp. E946-953.

Van Tine, B. A. et al. 2004. Clonal selection for transcriptionally active viral oncogenes during progression to cancer. *Journal of virology* 78(20), pp. 11172-11186.

Venuti, A. et al. 2011. Papillomavirus E5: the smallest oncoprotein with many functions. *Mol Cancer* 10, p. 140.

Villa, L. L. et al. 2006. High sustained efficacy of a prophylactic quadrivalent human papillomavirus types 6/11/16/18 L1 virus-like particle vaccine through 5 years of follow-up. *British journal of cancer* 95(11), pp. 1459-1466.

Vinokurova, S. et al. 2008. Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. *Cancer Res* 68(1), pp. 307-313.

von Knebel Doeberitz, M. et al. 1988. Correlation of modified human papilloma virus early gene expression with altered growth properties in C4-1 cervical carcinoma cells. *Cancer Res* 48(13), pp. 3780-3786.

Walboomers, J. M. et al. 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* 189(1), pp. 12-19.

Wales, C. S. 2007. KC53/61/65 Statistical Report 2008/2009.

Wang, Q. et al. 2004. Functional analysis of the human papillomavirus type 16 E1=E4 protein provides a mechanism for in vivo and in vitro keratin filament reorganization. *J Virol* 78(2), pp. 821-833.

Wang, Z. et al. 2012. Combined p53-related genetic variants together with HPV infection increase oral cancer risk. *Int J Cancer* 131(3), pp. E251-258.

Wang-Johanning, F. et al. 2002. Quantitation of human papillomavirus 16 E6 and E7 DNA and RNA in residual material from ThinPrep Papanicolaou tests using real-time polymerase chain reaction analysis. *Cancer* 94(8), pp. 2199-2210.

Wellings, K. et al. 2001. Sexual behaviour in Britain: early heterosexual experience. *Lancet* 358(9296), pp. 1843-1850.

Wentzensen, N. et al. 2005. Evaluation of a nuclear score for p16INK4a-stained cervical squamous cells in liquid-based cytology samples. *Cancer* 105(6), pp. 461-467.

Wentzensen, N. et al. 2002. Characterization of viral-cellular fusion transcripts in a large series of HPV16 and 18 positive anogenital lesions. *Oncogene* 21(3), pp. 419-426.

Wentzensen, N. et al. 2009. Utility of methylation markers in cervical cancer early detection: Appraisal of the state-of-the-science. *Gynecologic Oncology* 112(2).

Wentzensen, N. et al. 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(11), pp. 3878-3884.

WHO. 2008. Cervical cancer, human papillomavirus (HPV), and HPV vaccines - Key points for policy-makers and health professionals. In: Research, D.o.R.H.a. ed.

Wilke, C. M. et al. 1996. FRA3B extends over a broad region and contains a spontaneous HPV16 integration site: direct evidence for the coincidence of viral integration sites and fragile sites. *Human molecular genetics* 5(2), pp. 187-195.

Wolff, E. M. et al. 2010. Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. *PLoS Genet* 6(4), p. e1000917.

Xie, W. et al. 2003. Loss of Smad signaling in human colorectal cancer is associated with advanced disease and poor prognosis. *Cancer J* 9(4), pp. 302-312.

Yee, C. et al. 1985. Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. *The American journal of pathology* 119(3), pp. 361-366.

Yu, T. et al. 2005. The role of viral integration in the development of cervical cancer. *Cancer Genet Cytogenet* 158(1), pp. 27-34.

Yunis, J. J. and Soreng, A. L. 1984. Constitutive fragile sites and cancer. *Science* 226(4679), pp. 1199-1204.

Zerfass-Thome, K. et al. 1996. Inactivation of the cdk inhibitor p27KIP1 by the human papillomavirus type 16 E7 oncoprotein. *Oncogene* 13(11), pp. 2323-2330.

Zheng, Z. M. and Baker, C. C. 2006. Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front Biosci* 11, pp. 2286-2302.

Ziegert, C. et al. 2003. A comprehensive analysis of HPV integration loci in anogenital lesions combining transcript and genome-based amplification techniques. *Oncogene* 22(25), pp. 3977-3984.

zur Hausen, H. 1976. Condylomata acuminata and human genital cancer. *Cancer Res* 36(2 pt 2), p. 794.

zur Hausen, H. 1986. Intracellular surveillance of persisting viral infections. Human genital cancer results from deficient cellular control of papillomavirus gene expression. *Lancet* 2(8505), pp. 489-491.

zur Hausen, H. et al. 1974. Attempts to detect virus-secific DNA in human tumors. I. Nucleic acid hybridizations with complementary RNA of human wart virus. *Int J Cancer* 13(5), pp. 650-656.