HPV RESEARCH GROUP

INSTITUTE OF CANCER & GENETICS

SCHOOL OF MEDICINE

CARDIFF UNIVERSITY

Methylation of Human Papillomavirus DNA: Biological Significance and Clinical Utility

By Dean Bryant

2012

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Acknowledgements

It would not have been possible to write this thesis without the support of the generous and inspiring people around me, to only some of whom it is possible to give particular mention here.

Above all, I dedicate this thesis to my loving fiancée Emily who has unequivocally supported me throughout. My parents and brothers have each supported me in their own unique way and for this, I extend my gratitude.

I would also like to thank Dr Ned Powell, Dr Amanda Tristram and Professor Alison Fiander for their invaluable assistance and support both as project supervisors and as friends. I would like to thank each member of the HPV Research Team (University of Cardiff, School of Medicine) for their assistance and support throughout the duration of this project.

I would like to extend further gratitude to the HPV Research Team PhD students. At some point, each has contributed academically and personally and for this I am grateful.

I would also like to thank Dr T Liloglou, Amanda Gilkes and Michelle Lazenby for their technical advice and guidance.

I am especially grateful to Cancer Research Wales for funding this project and finally, I would like to thank my institution for the opportunity to perform this research.

Summary

DNA methylation helps regulate transcriptional activity and is widely studied in cancer biology. This investigation aimed to establish the significance of Human Papillomavirus (HPV) DNA methylation in HPV-associated disease both in terms of basic biology and as a potential biomarker.

Assays to assess DNA methylation and gene expression were developed and evaluated. Pyrosequencing was used to assess DNA methylation of four regions of the HPV16 genome (*E2*, *L1/L2*, enhancer, promoter). Gene expression was assessed using quantitative PCR with assays for *E2*, *E6* and *E7*. HPV integration was assessed using Detection of Integrated Papillomavirus Sequences (DIPS).

The relationship between HPV methylation, gene expression and integration was explored *in vitro* and *in vivo* using cell cultures and clinical cohorts. A variety of sample materials were used including short term and immortal cell lines, cervical cancer biopsies, cytology samples and Vulval Intraepithelial Neoplasia (VIN) biopsies.

In general, hypermethylation of the HPV genome was associated with low HPV gene expression and the presence of integrated HPV genomes.

To better understand the potential clinical utility of HPV DNA methylation, the relationship between HPV DNA methylation and various stages of cervical disease was determined. The HPV genome was progressively hypermethylated with increasing severity of cervical disease and certain regions of the HPV genome were more affected than others. A longitudinal study was also performed in order to determine a relationship between HPV methylation and clinical outcome. Differences in HPV methylation among patients who had persistent HPV infection and low grade disease, persistent infection and high grade disease and patients that cleared HPV infections were observed. Throughout the study the potential application of a HPV biomarker was considered and the correct biomarker design procedures were referred to. Several of the early biomarker development steps were successfully achieved.

Abbreviations

Abbreviation	Definition
AIN	Anal intraepithelial neoplasia
ANOVA	Analysis of variance
APOT	Amplification of papillomavirus oncogene transcripts
BS	Bisulfite (i.e. Sodium bisulfite)
CIN	Cervical intraepithelial neoplasia
CpG	Cytosine-guanine dinucleotide
CRUK	Cancer Research UK
Ct	Cycle threshold number
DIM	Diindolylmethane
DIPS	Detection of integrated papillomavirus sequences
DNMT	DNA methyltransferase
E1-E8	The HPV early genes
E2-(1-8)	The 8 *E2* CpGs tested
E2BS$_{(1-4)}$	E2 binding sites 1-4
EBV	Epstein Barr virus
GLM	General linear model
HBV	Hepatitis B virus
HDAC	Histone deacetylase
HPRT	Hypoxanthine guanine phosphoribosyl transferase
HPV	Human papillomavirus
HrHPV	High risk HPV
HSIL	High grade squamous intraepithelial lesion
L1 and *L2*	The HPV late genes
L1/L2	The overlap between *L1* and *L2*
L1L2-(1-4)	The 4 CpGs of the *L1/L2* regions tested
LBC	Liquid based cytology
LCR	Long control region
LCR1	The LCR1 assay (enhancer)
LCR1-(1-3)	The 3 LCR1 CpGs tested
LCR2	The LCR2 assay (promoter)
LCR2-(1-5)	The 5 LCR2 CpGs tested
LrHPV	Low risk HPV
LS	Lichen sclerosus
LSIL	Low grade squamous intraepithelial lesion
meCpG	Methylcytosine-quanine dinucleotide
MS-PCR	Methylation specific PCR
ORF	Open reading frame
P97	The HPV16 early promoter
Pap	Papanicolaou (stain)
PC08/09	Primary culture 08/09 - short term cell lines
qMS-PCR	Quantitative methylation specific PCR
qPCR	Quantitative PCR (also qRTRTPCR)
qRTRTPCR	Quantitative reverse transcriptase real time PCR
RQ	Relative quantification
RT	Reverse transcription/transcriptase
SCC	Squamous cell carcinoma
TBP	TATA-binding protein
VAIN	Vaginal intraepithelial neoplasia
VIN	Vulval intraepithelial neoplasia
VLP	Virus like particle
W12Ser7	Series 7 of the W12 cell line

Opening Notes

This project was an investigation into the biological and clinical significance of HPV DNA methylation. DNA methylation at several sites of the HPV genome was assessed and the relationship between DNA methylation, gene expression and HPV integration investigated.

A vital component of this investigation was the development of suitable assays with which to assess DNA methylation and gene expression. Method development formed a large part of the study and is described in a separate chapter. The experiments described in the four other results chapters were designed to test hypotheses concerning the relationship between HPV DNA methylation and other characteristics.

Several components of the various studies were common to, or had broader relevance to more than one study. As such, discussion of these topics was grouped together into the general discussion in order to avoid repetition.

This investigation is organised in the following manner:

- Chapter 1. Introduction: HPV and HPV-Associated Cancer
- Chapter 2. Introduction: DNA Methylation and Epigenetics
- Chapter 3. Methods and Materials
- Results and Discussion Chapters
    - Chapter 4. Method Development
    - Chapter 5. DNA Methylation and Disease Grade
    - Chapter 6. DNA Methylation and Clinical Outcome
    - Chapter 7. DNA Methylation, Integration and Gene Expression *in vitro*
    - Chapter 8. DNA Methylation, Integration and Gene Expression *in vivo*
- Chapter 9. General Discussion
- Chapter 10. General Conclusions

Table of Contents

Chapter 1. Introduction: HPV and HPV-Associated Cancer

1.1. Human Papillomaviruses

Papillomaviruses (PV) are a diverse range of viruses that infect squamous and glandular epithelial tissues in a range of species. At least 120 types of Human Papillomavirus (HPV) are capable of infecting humans (Bernard *et al.*, 2010) with around 40 HPV types infecting the anogenital tract (Sanders and Taira, 2003, Stanley, 2006a). HPV infection is very common and most individuals contract a genital HPV at some point in their lives, however relatively few people experience symptoms or HPV-associated anogenital neoplasia (Woodman *et al.*, 2001, Ho *et al.*, 1998). Although the majority of infections are subclinical and transient, HPV infection occasionally results in benign or malignant tumours of the cutaneous and mucosal epithelia; most notably of the cervix where HPV is associated with >99% of cancers (Duensing and Münger, 2004, McIndoe *et al.*, 1984, Stanley, 2007).

Mucosal HPV are further classified according to their oncogenicity. Low risk HPV (LrHPV) types rarely or never lead to cancers; types 6 and 11 for example are implicated in the formation of benign genital warts and are rarely associated with cancers. Other HPV types are denoted as High risk HPV (HrHPV) due to their common association with the development of cancers (Sanders and Taira, 2003, Braaten and Laufer, 2008, Doorbar, 2006). Persistent infection with HrHPV is considered a risk factor for the development of anogenital neoplasia (Ho *et al.*, 1995). The HrHPV types most commonly associated with anogenital cancer are HPV16 and HPV18 which are present in around 70% of cervical cancers worldwide (Muñoz *et al.*, 2004) and in varying proportions of other anogenital carcinomas.

1.1.1. Taxonomy of Human Papillomaviruses

There are 12 genera of Papillomaviruses (PV), defined as having at least 60% sequence identity of the *L1* gene. Human papillomaviruses are divided into alpha, beta, gamma, mu and nu (Bernard *et al.*, 2010). The two main genera of HPV are the Alpha and Beta papillomaviruses, approximately 90% of all characterised HPV fit into one of these two groups and have a general tendency to infect either cutaneous or mucosal epithelia. Beta papillomaviruses are typically associated with unapparent cutaneous infections, and only tend to cause clinically relevant disease if the patient is immunocompromised or suffering from epidermodysplasia verruciformis (Doorbar, 2006). Alpha papillomaviruses include the HPV types associated with genital/mucosal infections and several cutaneous types such as HPV2 which tend to cause common warts (Doorbar, 2006, de Villiers *et al.*, 2004). PV genera are divided by species,

representing 60-70% *L1* sequence identity to other PV of the same species; viral type (71-89% *L1* identity); viral subtype (90-98% *L1* identity) and variant (over 98% *L1* identity) (Howley and Lowy, 2006).

1.1.2. HPV and the HPV Lifecycle

The structure, genetics and lifecycle of HPV will now be considered before the functions of the HPV gene products and the association between HPV and cancer are discussed.

1.1.2.1. HPV

HPV has a double stranded, circular DNA genome of approximately 8,000 bp in length (e.g. HPV16 is 7,904 bp) and clinically important HPV typically encode 8 Open Reading Frames (ORFs). The terms ORF and gene are widely used interchangeably in the literature. There are six early genes, *E1*, *E2*, *E4*, *E5*, *E6* and *E7* that encode regulatory proteins and the two late genes *L1* and *L2* that encode structural capsid proteins (Figure 1.1). The early gene products include those involved with regulating viral DNA replication and are primarily transcribed from the early promoter (e.g. P97 for HPV16) whilst the late genes are transcribed from the late promoter (e.g. P670 in HPV16) (Abdul *et al.*, 2009, Zheng and Baker, 2006). Early promoter activity and the regulation of replication is governed by interactions with the approximately 1 kbp non-coding Long Control Region (LCR) that contains the viral origin of replication and a multitude of transcription factor binding sites and DNA regulatory elements (Chiang *et al.*, 1992, Demeret *et al.*, 1997, Dollard *et al.*, 1993, Dong *et al.*, 1994, Tan *et al.*, 1994, Berg and Stenlund, 1997).

HPV virions are comprised of a DNA genome encapsulated in an icosahedral capsid approximately 55 nm in diameter (Scheurer *et al.*, 2005). Capsids are constructed from 72 capsomers and each capsomer is a pentamer of the HPV L1 protein. The HPV L2 minor capsid protein occurs at a frequency of about 12 proteins per capsid (Modis *et al.*, 2002).
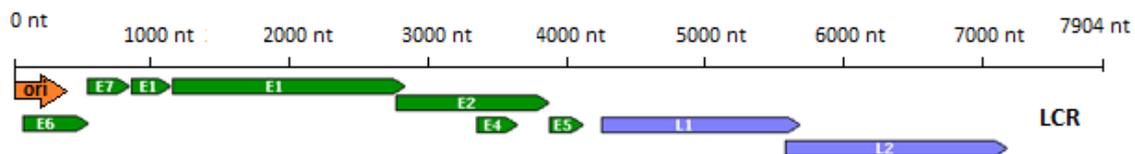
Figure 1.1: Linear representation of the HPV16 genome.

Adapted from (NCBI, 1993), (GenBank ID NC001526.1). Orange arrow labelled *ori* represents the origin of replication, green arrows represent early ORFs and blue arrows represent late capsid ORFs, the direction of the arrow indicates the direction of transcription. The Long Control Region (LCR) is a non-coding region that contains binding sites for viral and cellular regulatory proteins and includes the viral origin of replication.

1.1.2.2. *The HPV Life Cycle in a Productive Infection*

1.1.2.2.1. Infection

Recently, a new model of infection initiation has been proposed. Infection with HPV requires micro-abrasion of the epithelium which exposes the basement membrane to HPV virions. HPV L1 proteins bind structures on the epithelial basement membrane which induces conformational changes that expose L2. L2 is cleaved and previously concealed parts of the L1 protein are exposed to cell surface receptors on keratinocytes that have migrated across the basement membrane as part of the wound healing process (Schiller *et al.*, 2010, Kines *et al.*, 2009).

Following attachment to keratinocytes, the HPV virions enter the cell via an endocytic pathway and are localised within an endosome where the virus eventually uncoats. An L2 dependent mechanism promotes escape from the endosome; L2 bound HPV DNA moves through the cytoplasm and enters the nucleus. In the nucleus the L2-DNA complex associates with sub-nuclear domains and RNA transcription is initiated (Schiller *et al.*, 2010).

The HPV lifecycle from initial infection onwards is intrinsically associated with the differentiation programme of the keratinocyte that it infects. As such, the study of the HPV lifecycle *in vitro* is especially difficult and very few cell culture models exist that are not derived from transformed cancer lines.

1.1.2.2.2. Genome Maintenance/Cell Proliferation

The HPV16 early promoter (P97) governs the transcription of the early genes. Located at the 3' end of the LCR, transcription from the P97 promoter is controlled by the binding of various cellular transcription factors and the E2 viral gene product (Kammer *et al.* 2000; Thain *et al.* 1996). P97 activation by the binding of cellular transcription factors allows initial low level *E2* and *E1* expression which direct host cell replication factors to the viral origin of replication, driving viral genome replication and the maintenance of infection (Kadaja *et al.*, 2009, Steger and Corbach, 1997). Viral genome replication occurs and the HPV genome is maintained as low copy number episomes (50-100 copies per cell) using host DNA replication processes (Flores *et al.*, 1999). In addition to providing an important contribution to viral replication and transcription, E2 plays a major role in ensuring long term episomal maintenance and establishment of a productive HPV infection. E2 is involved in the partitioning of episomal genomes during mitosis by anchoring viral genomes to mitotic spindles during cell division, thus ensuring efficient viral genome segregation into daughter cells (You *et al.*, 2004, Van Tine *et al.*, 2004a, Lehman and Botchan, 1998).

For DNA replication, HPV require the cellular polymerases and other cellular factors that are produced only in mitotically active cells. As infected keratinocytes begin differentiation, E6/E7 activity is up-regulated (Middleton *et al.* 2003; Schwartz 1998). E7 targets pRb (retinoblastoma protein) family proteins for degradation, disrupting the association between pRb and the E2F family of transcription factors which results in the transactivation of certain cellular factors required for viral DNA replication (Dyson *et al.*, 1989, Gage *et al.*, 1990, Scheffner *et al.*, 1991). E6 effectively targets p53 for proteasomal degradation, inhibiting cellular apoptosis that would otherwise result from E7 mediated cell cycle entry in the differentiated epithelial layers (Scheffner *et al.*, 1991, Scheffner *et al.*, 1990, Doorbar, 2006). Together, low level E6 and E7 activity allow terminally differentiating cells to avoid differentiation and continue cycling. This both maintains the availability of S-phase cellular factors required for viral genome replication and increases the number of cells producing progeny HPV virions (Middleton *et al.*, 2003, Ruesch and Laimins, 1997, Stanley *et al.*, 2007, Münger and Howley, 2002).

P97 is responsible for transcription of *E2*. Low E2 concentration positively regulates P97 activity whilst high E2 concentration negatively regulates P97 activity, as such P97 activity is essentially self regulating (Steger and Corbach, 1997). Regulated P97 activity results in a relatively constant level of the other viral gene products. E1 and E2 contribute to the maintenance of a relatively stable level of genome replication during viral maintenance in undifferentiated cells.

1.1.2.2.2.1. *Origin Identification and Initiation of DNA Replication*

HPV E1 proteins are highly conserved and recognise AT rich sequences at the HPV origin of replication. E1 alone has a weak DNA affinity. E2 binding sequences are located adjacent to E1 recognition sites and allow E2 to direct the origin specific recruitment of E1 proteins (Sedman and Stenlund, 1998). Once an E1-E2-DNA complex is formed, E2 can be displaced in an ATP-dependent step and additional E1 molecules are added to form an active E1 hexamer. E2 effectively directs the formation of a DNA specific E1 complex to the viral origin (Sanders and Stenlund, 1998).

E1 proteins exhibit both ATPase and 3'-5' helicase activities, bind DNA polymerase α/primase and aid the recruitment of other cellular replication complexes to the viral origin of replication (Longworth and Laimins, 2004, Lin *et al.*, 2002, Conger *et al.*, 1999, Hughes and Romanos, 1993). Helicase activity at the origin of replication allows the unwinding and separation of the two complementary strands of a supercoiled DNA double helix, exposing the base sequence to the relevant DNA replication machinery (Sedman and Stenlund, 1998). Purified E1 protein has been shown to elicit efficient DNA unwinding *in vitro* and requires topoisomerase I, single-

stranded-DNA-binding protein, ATP and supercoiled DNA as a substrate (Lin *et al.*, 2002, Conger *et al.*, 1999).

1.1.2.2.3. Genome Amplification

Premature keratinocytes, including those infected with HPV, progress further into the suprabasal layers of the epithelial tissue. Upon detachment from the basement membrane, epithelial cells begin to terminally differentiate and exit the cell cycle (Madison, 2003, Stanley *et al.*, 2007). In more terminally differentiated cells, activation of the E2 independent late promoter (P670 in HPV16) results in increased levels of the proteins involved in DNA replication (E1, E2, E4 and E5) (Spink and Laimins, 2005, Grassmann *et al.*, 1996, Klumpp and Laimins, 1999). The HPV early and late promoters are not exclusively involved in early/late gene expression, each promoter can fulfil some functions of the other (Spink and Laimins, 2005). As a result of increased early protein levels, vegetative viral DNA replication occurs and results in a viral copy number of at least 1,000 copies per cell (Flores and Lambert, 1997).

1.1.2.2.4. Viral Encapsidation and Release

Whilst *L1/L2* pre-mRNA can be detected in less terminally differentiated cells, L1 and L2 protein is only detectable in highly differentiated cells (Stoler *et al.*, 1989). A variety of post-transcriptional regulatory mechanisms are implicated in the apparent lack of translation of the late mRNAs.

L1 and L2 proteins accumulate following genome amplification. Viral assembly is the result of the coordinated associations between viral DNA, L1, L2 and E2 proteins and human proteins (Doorbar, 2006). HPV cannot lyse infected cells. Instead, HPV utilises desquamation of epithelial cells to facilitate the spread of infective virions (Woodman *et al.*, 2007, Flores *et al.*, 1999). HPV E4 protein may be involved in final release of HPV virions through involvement in processes which result in destabilisation of cellular keratin networks and disruption of the cornified envelope (Doorbar *et al.*, 1991, Bryan and Brown, 2000).

1.1.3. HPV Genome and Gene Products

1.1.3.1. The Long Control Region

The LCR of HPV varies between types, however most HPV LCRs contain binding sites for the same subset of factors, the most significant of which appear to be YY1, CDP, Ap1, Sp1 and TBP (Spink and Laimins, 2005, O'Connor *et al.*, 2000, Ai *et al.*, 2000). For the promoter region, YY1 and CDP binding is associated with repression of P97. The expression of *YY1* and *CDP* decreases upon differentiation alleviating these repressive effects (O'Connor *et al.*, 1996, O'Connor *et al.*,

2000). The viral enhancer contains several Ap1 binding sites. Ap1 expression and abundance varies with cellular differentiation and binding of the viral enhancer is altered as a result (Thierry *et al.*, 1992, Sen *et al.*, 2004).

Specific cell factors bind the HPV LCR and are responsible for the tissue specific tropisms exhibited by HPV (Apt *et al.*, 1993). Four palindromic E2BSs (E2 binding site; consensus sequence, 5'-ACCN$_6$GGT-3') upstream of the promoter are bound by E2, regulating P97 activity (Howley and Lowy, 2006). Three of the four HPV E2BSs; E2BS$_2$, E2BS$_3$ and E2BS$_4$ are involved in HPV16 P97 repression (Figure 1.2). E2 functions as a transcriptional repressor of P97 by competing with cellular transcription factors for binding sites (Romanczuk *et al.*, 1990). E2 bound to E2BS$_4$ for example prevents the binding to an adjacent TATA box by the TFIID complex (binding mediated by the *TBP* (TATA box binding protein) subunit) (Dostatni *et al.*, 1991) and may affect the stability of the preinitiation complex once *TBP* binding has occurred (Hou *et al.*, 2000). P97 activation is functionally dependent on Sp1 binding. The Sp1 binding site is separated by only a single nucleotide from E2BS$_3$ (Gloss and Bernard, 1990). E2 binding at E2BS$_3$ can displace Sp1 and any associated *TBP* from the adjacent Sp1 binding site (Tan *et al.*, 1994, Emili *et al.*, 1994). E2 binding at E2BS$_1$ increases early gene expression (Steger and Corbach, 1997, Dong *et al.*, 1994) by stimulating YY1 function at the adjacent YY1 binding site (Tan *et al.*, 2003).
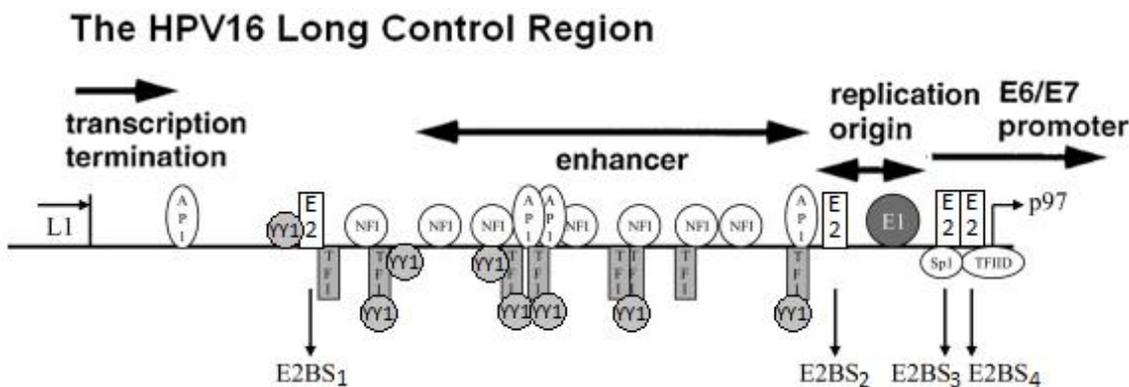
Figure 1.2: Diagram of the HPV16 LCR including binding positions of several HPV and human cellular factors.

Adapted from (Snellenberg *et al.*, 2012). There are four E2BSs within the HPV16. Binding sites for Specificity factor 1 (Sp1), Ying Yang 1 (YY1), Transcription Factor I (TFI), Nuclear Factor (NFI) Activator Protein 1 (AP1), Transcription Factor IID (TFIID) are shown.

E2BSs also function in genome replication with E2BS$_2$ playing a major role in terms of E1 origin recruitment (Badal *et al.*, 2003, Tan *et al.*, 1994). The affinity of E2 for the repressive E2BS is

much lower than the affinity of E2 for the transcriptional activator $E2BS_1$. E2 has high affinity for $E2BS_1$, $E2BS_1$ is located several hundred bp upstream of the promoter and E2 binding to $E2BS_1$ increases promoter activity (Demeret *et al.*, 1998). Thus, low concentrations of E2 will stimulate *E6* and *E7* transcription whereas high concentrations of E2 repress transcription (Steger and Corbach, 1997, Badal *et al.*, 2003).

Histone Deacetylase (HDAC) activity at the HPV genome results in histone deacetylation and specific positioning of two nucleosomes over the viral enhancer and promoter that can inhibit promoter function (O'Connor *et al.*, 2000, Stunkel and Bernard, 1999). A series of nuclear matrix attachment regions are also present in the HPV16 LCR which are thought to be responsible for altering the topology of the DNA and HPV gene expression in association with HPV integration (Stunkel *et al.*, 2000). Methylation of cytosine residues has also been shown to repress viral and cellular gene expression as methylated DNA can result in the displacement of transcription factors and the alteration of chromatin configuration via the focusing of HDAC activity (Kalantari *et al.*, 2004).

In summary, the LCR is responsible for regulating HPV early gene expression and is affected by a multitude of factors, both HPV and human in origin.

1.1.3.2. Post-Transcriptional Regulation of HPV Gene Expression

HPV transcripts originating from P97 are polycistronic and typically contain three exons and two introns. Variable splicing of transcripts leads to early mRNA transcripts with a range of coding potential (Figure 1.3) (Zheng and Baker, 2006). Papillomaviruses further utilise inefficient splice signals to differentially express early viral proteins throughout the HPV lifecycle (Zheng and Baker, 2006).

Figure 1.3: Early splice variants originating from the P97 promoter.

Adapted from (Zheng and Baker, 2006). A variety of potential splice variants exist; thin lines indicate regions spliced out, the coding potential of each spliced transcript is indicated on the right.

As well as splice variants, translational inefficiency also appears to affect HPV gene expression. HPV transcripts are polycistronic, translation of downstream ORFs requires ribosomes to bypass upstream ORFs in order to translate downstream ORFs (Zheng and Baker, 2006). Several potential mechanisms have been proposed including translation reinitiation, leaky ribosome scanning and ribosome shunting across the transcripts (Remm *et al.*, 1999, Sen *et al.*, 2004, Stacey *et al.*, 2000, Zheng *et al.*, 2004). The two most likely mechanisms appear to be leaky ribosomal scanning and translation reinitiation (Zheng *et al.*, 2004, Remm *et al.*, 1999, Longworth and Laimins, 2004). Leaky ribosome scanning means start codons towards the 5' end of an mRNA are skipped and initiation of translation occurs via ribosomal binding at downstream start codons (Longworth and Laimins 2004; Remm *et al.* 1999). By this complex

mechanism, transcriptional regulation depends upon the position of an ORF on the polycistronic mRNA, with 5' cistrons being translated with greater efficiency than 3' cistrons. The HPV early promoter produces a transcript in the order 5'- *E6*, *E7*, *E1*, *E2*, *E4*, *E5* - 3' which would explain the almost negligible amounts of E4 and E5 protein produced (Remm *et al.*, 1999). As a consequence of differentiated cell optimised codon usage, late promoter activation and alternative splicing of late mRNAs, transcription and translation of the late gene products (*L1* and *L2*) occurs (Zhou *et al.*, 1995, Zhou *et al.*, 1999, McPhillips *et al.*, 2004, Stoler *et al.*, 1989). Various additional post-transcriptional regulatory mechanisms have also been proposed. A negative regulatory element in the 3' regions of the HPV16 late (*L1* and *L2* containing) mRNA transcripts negatively influences late gene expression, reduces the stability of mRNA species and binds cellular proteins (Howley and Lowy, 2006, Koffa *et al.*, 2000). As an infected host cell undergoes cellular differentiation, the range and distribution of cellular factors fluctuates which may remove post-transcriptional repression of HPV late mRNAs permitting their translation (Koffa *et al.*, 2000, Howley and Lowy, 2006).

1.1.3.3. Viral Gene Products
Each of the HPV gene products have additional functions that were not previously discussed but are pertinent to the role of HPV in HPV-associated carcinogenesis and require further consideration.

1.1.3.3.1. E1 and E2
E1 proteins are comprised of three regions: an N-terminal region, a variable length spacer region and a large, highly conserved C-terminal region that has homology to characterised ATPases and helicases (Wilson *et al.*, 2002). E1 displaces the H1 histone protein, suggesting that E1 can promote the initiation of replication and elongation by altering the viral chromatin structure and disrupting nucleosomes at the replication fork (Sedman and Stenlund, 1998).

E2 protein function can be isolated to specific regions within the protein; for instance the C-terminus DNA binding domain is a dimeric ß-barrel structure that bends DNA and interacts with E1, whilst the N-terminus contains a transactivation domain (Hegde *et al.*, 1992, Chen and Stenlund, 2000). The *E2* ORF can give rise to multiple *E2* gene products e.g. the E2TA full length protein and the truncated E2TR polypeptides, Bovine Papillomavirus (BPV) studies also find an additional array of E8^E2TR polypeptides (Lambert *et al.*, 1990, Kim *et al.*, 2003). Truncated E2 forms appear to repress full length E2 activity via competitive exclusion at E2BS or the formation of inactive heterodimers (Lambert *et al.*, 1990). E2 is occasionally described as having a proapoptotic effect on the infected host cell. E2 expressing cells effectively repress *E6* and *E7* expression, thus p53 is stabilised, p21 is activated and the cell cycle arrests at G1

(Blachon and Demeret, 2003). The repression of *E6* and *E7* by E2 also appears to be largely responsible for the cellular senescence exhibited by these arrested cells (Goodwin *et al.*, 2000). In addition, several HrHPV E2 proteins when expressed ectopically induce a cell death with the characteristic features of apoptosis, independent of E2 DNA binding and functional transcriptional regulation of the oncogenes (Blachon and Demeret, 2003). The introduction of *in vitro* truncated E2 proteins efficiently induced apoptosis in HeLa cells (Blachon and Demeret, 2003).

1.1.3.3.2. E4 and E5

In comparison to the other HPV encoded proteins, the roles of E4 and E5 are less well defined. HPV E5 is a small hydrophobic transmembrane protein primarily associated with the endoplasmic reticulum and the Golgi apparatus (Conrad *et al.*, 1993). E5 overexpression for some, but not all, HPV inhibits the degradation of and increases the phosphorylation of Epidermal Growth Factor (EGF) receptors (Straight *et al.*, 1993, Fehrmann *et al.*, 2003). E5 forms a complex with EGF receptors (Hwang *et al.*, 1995) and binds cellular proteins in order to delay endosomal acidification (Genther *et al.*, 2003, Conrad *et al.*, 1993, Straight *et al.*, 1995). This might indirectly influence the recycling of cell surface growth factor receptors, such that there is an increase in EGF-mediated receptor signalling. As a consequence of increased sensitivity to EGF, a replication competent environment is maintained in the otherwise differentiated layers of the epithelia, permitting genome amplification (Doorbar, 2006). HPV E5 appears to have a minor role in genome amplification in the productive stages of a HPV infection. E5 mutants show fewer supraparabasal cells undergoing DNA synthesis in comparison to differentiation competent cell lines with wild type E5 (Genther *et al.*, 2003). In terms of carcinogenesis, several roles of E5 could be relevant to early carcinogenesis by modulation of integrant selection, viral persistence, evasion of the host immune system and insensitivity to homeostatic growth control signals from neighbouring cells (Venuti *et al.*, 2011)

E4 is produced from a spliced mRNA transcript as an E1^E4 fusion protein, i.e. formed from the N-terminus of the *E1* ORF product (first 5 amino acids) and most of the *E4* ORF product (Longworth and Laimins, 2004, Roberts *et al.*, 2008b). Transfection of keratinocytes with HPV16 genomes containing mutant *E4*, produces transformed cells that are incapable of completing viral amplification (Nakahara *et al.*, 2005). The exact role of E4 in regulation of viral amplification is poorly defined, however due to the organisation of the HPV late polycistronic mRNA; late gene expression and thus productive virus replication coincides with a large increase in HPV E4 protein (Longworth and Laimins, 2004, Roberts *et al.*, 2008b). Overexpressed HrHPV E1^E4 in cell lines can associate with cellular keratin networks, inducing their collapse (Doorbar *et al.*, 1991), may interfere with cornified envelope formation (Bryan

and Brown, 2000) and may incite cellular apoptosis in differentiated cells via alteration of mitochondrial function (Raj *et al.*, 2004); thus suggesting a role of E4 in viral egress (i.e. escape from desquamated keratinocytes) (Longworth and Laimins, 2004). It has also been hypothesised that the inhibition of cell cycle progression and DNA synthesis by coexpressed truncated E4 proteins serves to inhibit competing host DNA synthesis and favour viral genome amplification (Roberts *et al.*, 2008b).

1.1.3.3.3. E6 and E7

HrHPV E6 and E7 oncoproteins (so named due to their implication in HPV induced carcinogenesis) can effectively transform cells upon transfection to form a hyperproliferative, immortal cell line. Both E6 and E7 are required to effectively transform cells (Hawley-Nelson *et al.*, 1989, Giarre *et al.*, 2001).

Many of the cellular factors required by HPV for DNA synthesis are transcriptionally regulated through an E2F-dependent mechanism. In turn, the E2F family of transcription factors are regulated by pocket proteins such as pRb (retinoblastoma protein) whose regulation is determined by cyclical phosphorylation and dephosphorylation. E2F responsive genes (encoding e.g. DNA polymerase alpha, dihydrofolate reductase, thymidine kinase etc) are important for the progression through S-phase (Münger and Howley, 2002, Howley and Lowy, 2006).

Some of the key components of the following events are presented in Figure 1.4. Hypophosphorylated pRb (as found in G1 of the cell cycle) complexes with E2F in order to repress its activity as a transcription factor. Conversely, hyperphosphorylated pRb does not interact with E2F and E2F is free to transcriptionally activate the E2F responsive genes (Howley and Lowy 2006; Münger and Howley 2002; Pan and Griep 1994). Late in M-phase, hyperphosphorylated pRb is dephosphorylated and the E2F-pRb inactive transcription factor complex is reformed (Münger and Howley, 2002). An interesting and important role for HrHPV E7 is the proteolytic degradation of the retinoblastoma proteins pRb and related pRb family proteins p107 and p130 (Duensing and Münger, 2004).
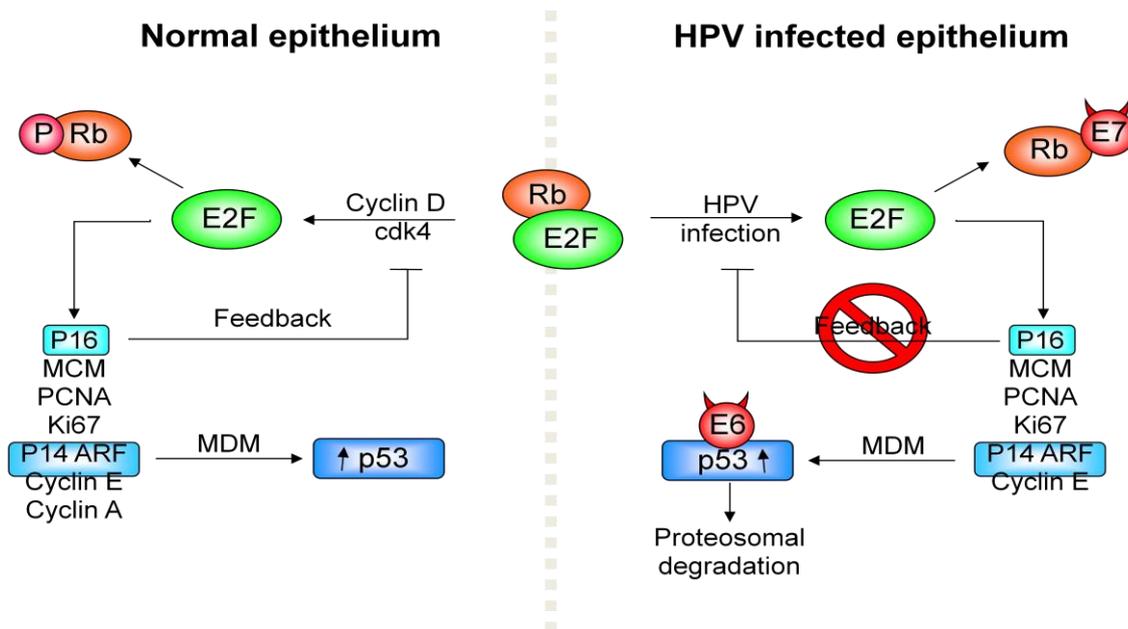
Figure 1.4: E2F regulation in normal and HPV infected epithelium.
E2F if negatively regulated by bound pRb. Phosphorylated pRb does not bind E2F and E2F are free to enhance the transcription of S-phase gene products. p16 represents a negative feedback loop and one consequence of E2F deregulation is enhanced p53 level. In HPV infected epithelia, deregulation of E2F is caused by E7 binding pRb whilst enhanced p53 level is counteracted by E6 mediated degradation of p53 protein.

The E7 proteins of HrHPV types but not LrHPV types have a high affinity for pRb and target pRb for proteolytic degradation (Boyer *et al.*, 1996, Gage *et al.*, 1990). Proteasomal degradation of pRb prevents downregulation of E2F dependent genes, independent of cyclin dependent kinase activity. This results in increased expression of cyclin E and cyclin A and aberrant Cyclin dependent kinase 2 (Cdk2) activity (Duensing and Münger, 2004). Aberrant Cdk2 activity induces cellular transformation and the transcription of the G1/S-phase enzymes required for DNA replication in terminally differentiated cells (Münger and Howley, 2002, Doorbar, 2006). E7 expression is thus responsible for expression of DNA synthesis genes required for a productive HPV infection (Flores *et al.*, 2000).

The majority of cervical carcinomas exhibit increased levels of p16^{INK4a} mRNA and protein as a result of E7 expression and E2F activation (Ivanova *et al.*, 2007). p16^{INK4a} acts as a repressor of cell cycle progression by inhibiting the phosphorylation of pRb. Most cervical carcinomas exhibit p16^{INK44a} overexpression, to the degree that the detection of overexpressed p16^{INK4a} has been widely proposed and tested as a potential biomarker of cervical disease (Ivanova *et al.*, 2007, Wentzensen *et al.*, 2005, Lesnikova *et al.*, 2009).

p21 is normally induced during keratinocyte differentiation. p21 reduces the activity of cyclin/Cdk2 complexes, negatively regulates cell cycle progression and can also bind to proliferating cell nuclear antigen. p21 may thus effectively block DNA synthesis in differentiated keratinocytes (Gartel and Radhakrishnan, 2005). A range of factors influence p21 cellular protein levels, most notably the activity of p53 tumour suppressor protein which is a major determinant of p21 transcription level. Downregulation of p53 activity results in reduced p21 activity (Gartel and Radhakrishnan, 2005). The induction of E2F by E7 mediated pRb deregulation may inadvertently stabilise p53 by compromising the action of mdm2 (Münger and Howley, 2002). Thus the result of E7 activity is increased p53 protein which could result in an apoptotic response (Ruesch and Laimins, 1997).

HrHPV E6 complexes with the cellular HECT domain E3 ubiquitin ligase E6-AP (E6 associated protein) to target p53 for ubiquitinylation. Ubiquitinylation of p53 leads to proteasomal degradation (Huibregtse *et al.* 1993; Snijders *et al.* 2006). The proteasomal degradation of p53, and inhibition of p21 by HrHPV E7 results in a decreased intracellular pool of active p53 and p21 (Gartel and Radhakrishnan, 2005, Münger and Howley, 2002). This significantly reduces the integrity of the G2 cell cycle checkpoint in response to DNA damage (Bunz *et al.*, 1998) and the ability of p21 to orchestrate p53 mediated cell cycle arrest at the G1/S checkpoint (Waldman *et al.*, 1995). The effect of HrHPV E6 and E7 activity on cell cycle controls effectively maintains a replication permissive background as required for genome amplification, late gene expression and the avoidance of apoptosis (Münger and Howley, 2002, Ruesch and Laimins, 1997).

The catalytic subunit of human Telomerase Reverse Transcriptase (hTERT) is expressed in cells such as stem cells that do not undergo replicative senescence. hTERT is a ribonucleoprotein polymerase and prevents telomere shortening with multiple cell divisions by the addition of the telomere repeat TTAGGG (NCBI, 2009). Somatic cells typically repress *hTERT* expression but *hTERT* activation is a common observation in cancerous cells (Cheng *et al.*, 2008) and is required for cellular immortalization by HPV (Kiyono *et al.*, 1998). HPV E6 indirectly promotes the activation of *hTERT* expression; *hTERT* is transcriptionally activated by the binding of c-Myc and Sp1 transcription factors to the *hTERT* promoter (Oh *et al.*, 2001). p53 both represses *MYC* (the c-Myc gene) transcription and binds Sp1 thus preventing its access to the *hTERT* promoter; degradation of p53 by E6 activity effectively circumvents *hTERT* repression (Xu *et al.* 2000). E6 also directly activates *hTERT* expression, the *hTERT* promoter repressor NFX1-91 is also targeted by E6 for proteasomal degradation and the ability of c-Myc to activate the *hTERT* promoter is increased in the presence of E6 (Xu *et al.*, 2008, Liu *et al.*, 2009, Veldman *et al.*,

2003). Further, E6 associates with telomeric DNA and physically interacts with hTERT protein enhancing telomerase activity (Liu *et al.*, 2009). Enhanced hTERT activity enables cancer cells to resist cellular senescence and undergo successive rounds of replication (Yugawa and Kiyono, 2009).

1.2. HPV and Cancer

The causal link between HPV and cervical cancer was first postulated by Harald zur Hausen in the 1970s (zur Hausen *et al.*, 1974, zur Hausen, 1976, zur Hausen, 1977). Since then, HPV DNA has been consistently identified in over 99% of cervical cancers and up to 94% of women with Cervical Intraepithelial Neoplasia (CIN) and the association between HPV and cancer is now firmly established (Scheurer *et al.*, 2005).

HPV are typically found in around 10% of normal cervical cytology samples, although the incidence of HPV infection varies with geographical location and some studies have identified HPV in up to 46% of the cytologically normal women studied (de Sanjose *et al.*, 2007). Most (over 80%) of infections in the cervix with high risk HPV types (HrHPV) are likely to be transient, without resulting in CIN (Meijer *et al.*, 2000).

In 2008, the global incidence of cervical cancer in women was an estimated 529,000 cases, leading to 274,000 deaths. More than 85% of cervical cancer deaths occurred in developing countries where it is frequently the commonest cancer to affect women, accounting for 13% of all female cancers (W.H.O., 2010). Developing countries lack the resources and infrastructure to facilitate effective screening and prevention strategies, as well as lacking medical facilities for treatment. HPV is also linked to a large percentage of carcinomas and intraepithelial neoplasias of the vulva, vagina, anus, penis, and head and neck (Table 1.1) (Stanley, 2007). The majority of these cases also occur in developing countries.

		Developed countries		Developing countries	
Site	Attributable to HPV (%)	Total cancers	Attributable to HPV	Total cancers	Attributable to HPV
Cervix	100	83,400	83,400	409,400	409,400
Penis	40	5,200	2,100	21,100	8,400
Vulva, vagina	40	18,300	7,300	21,700	8,700
Anus	90	14,500	13,100	15,900	14,300
Mouth	3	91,200	2,700	183,100	5,500
Oropharynx	12	24,400	2,900	27,700	3,300
Other	0	4,779,100	0	5,148,600	0
Total at all Sites		5,016,100	111,500	5,827,500	449,600

Table 1.1: The incidence of a range of malignancies and the number of which are attributable to an infection with HPV.

Adapted from (Stanley, 2007).The data is divided between occurrence in economically developed and developing countries.

1.2.1. Risk Factors

Several factors such as promiscuity, unprotected sexual encounters, early age of first sexual contact and relationships with high risk partners are more likely to result in infection with HPV (Braaten and Laufer, 2008, de Sanjose *et al.*, 2007). Most cervical cancers arise at the squamocolumnar junction between the columnar epithelium of the endocervix and the squamous epithelium of the ectocervix (also known as the transformation zone) (Doorbar, 2006). The transformation zone is a site of continuous metaplastic change where squamous epithelium is replaced with glandular columnar epithelium and vice versa, the greatest risk of infection correlates to periods of high metaplastic activity (Burd, 2003). Metaplastic changes of the cervix are at their greatest level at puberty and first pregnancy and the level declines after menopause. Age is an important determinant of risk of HPV infection and HPV-associated disease. HPV infection is most common in young, sexually active women between 18 and 30 years of age, however cervical cancer is most common in women older than 35 years of age, indicating that persistent infection is associated with HPV mediated cervical carcinogenesis (Burd, 2003, Ho *et al.*, 1995). A multitude of other factors such as race, low socioeconomic status, gender, tobacco use, oral contraceptive use and even a genetic determinant increase the risk that a given HPV infection will develop into a neoplasia (Braaten and Laufer, 2008, Frazer, 2009, de Sanjose *et al.*, 2007).

The host immune response is vital for protection against an HPV infection. Immunodeficient hosts show increased persistence and neoplastic progression of HPV-induced intraepithelial lesions in several instances (O'Brien and Saveria Campo, 2002). HIV mediated

15

immunosuppression for instance is associated with an increased risk of developing HSIL and carcinoma. Prior to the HIV epidemic, the incidence of anal cancer was estimated at 36.9 per 100,000 among high risk groups. This incidence is similar to the incidence of cervical cancer prior to the widespread adoption of cervical cancer screening programmes (Arain *et al.*, 2005). Since the beginning of the HIV epidemic, the incidence of anal Squamous Cell Carcinoma (SCC) and Anal Intraepithelial Neoplasia (AIN) have dramatically increased among homosexual males and anal SCC incidence in HIV positive populations is typically twice that of HIV negative populations (Arain *et al.*, 2005).

1.2.2. Detection of Cervical Intraepithelial Neoplasia

The Papanicolaou (Pap) stain is a staining procedure that permits cervical epithelial cells to be cytologically examined for dysplastic changes, an indicator of the pre-malignant condition intraepithelial neoplasia. CIN can then be treated, thus preventing the development of cervical cancer. Cervical screening is widely regarded as the most successful cancer screening program in the world and has been attributed to reducing cervical cancer deaths by 50-75% (Nephew, 2009, Etzioni *et al.*, 2003, Kurman Rj and *et al.*, 1994). Liquid Based Cytology (LBC) reduced the rate of inadequate samples and replaced the traditional Pap in England from 2003 (Albrow *et al.*, 2012). Cervical SCCs are the most common histological type of cervical cancer. The classical model of HPV induced cervical carcinogenesis is that cervical carcinoma develops from CIN of increasing severity over the course of many years. However, some CIN2/3 lesions are thought to develop rapidly after initial HPV infection i.e. within three years (Sherman and Kelly, 1992). CIN is classified histologically based upon the degree of atypia and abnormality of squamous maturation of the epithelial cells layers (Snijders *et al.*, 2006).

Worldwide, cervical cytology results are reported using a variety of classifications. In the UK, cytology results are reported according to British Society for Clinical Cytology (BSCC) guidelines based on the presence and severity of dysplastic cell changes, ranging from normal, to borderline changes, mild dyskaryosis, moderate dyskaryosis and severe dyskaryosis (Figure 1.5). Mild, moderate and severe dyskaryosis loosely correlate with CIN1, CIN2 and CIN3 on histological examination respectively (Snijders *et al.*, 2006).

| Normal cytology | Borderline changes | Mild dyskarosis | Moderate dyskaryosis | Severe dyskaryosis |
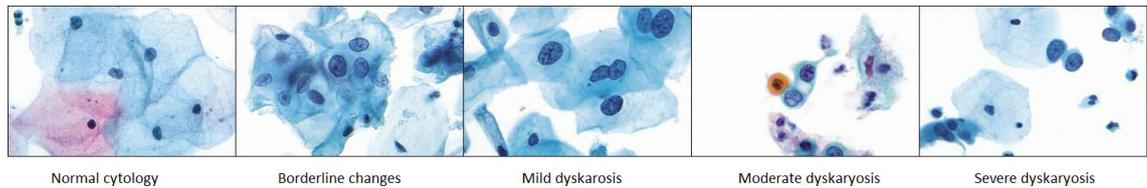
Figure 1.5: UK cervical cytological screening results in ascending order of severity.

Adapted from (NHS Cervical Screening Programme, 2012).

Whilst cytological examination determines the morphology of individual cells, biopsy of epithelial tissue allows the progression of atypical cells from basal epithelial layers to suprabasal differentiated epithelial layers to be detected and quantified by histological examination. The extent of atypical cells through the epithelium forms the basis for CIN grading; CIN1 indicates mild dysplasia (atypia of the lower third of the epithelium), CIN2 indicates moderate dysplasia (atypia of up to the lower two thirds of the epithelium) and CIN3 represents severe dysplasia (atypia of over two thirds of the epithelium) which is sometimes called carcinoma *in situ* (Figure 1.6). Other nomenclature includes Low grade Squamous Intraepithelial Lesion (LSIL) and High grade Squamous Intraepithelial Lesion (HSIL) which correspond to CIN1 and CIN2/3 respectively (Snijders *et al.*, 2006).
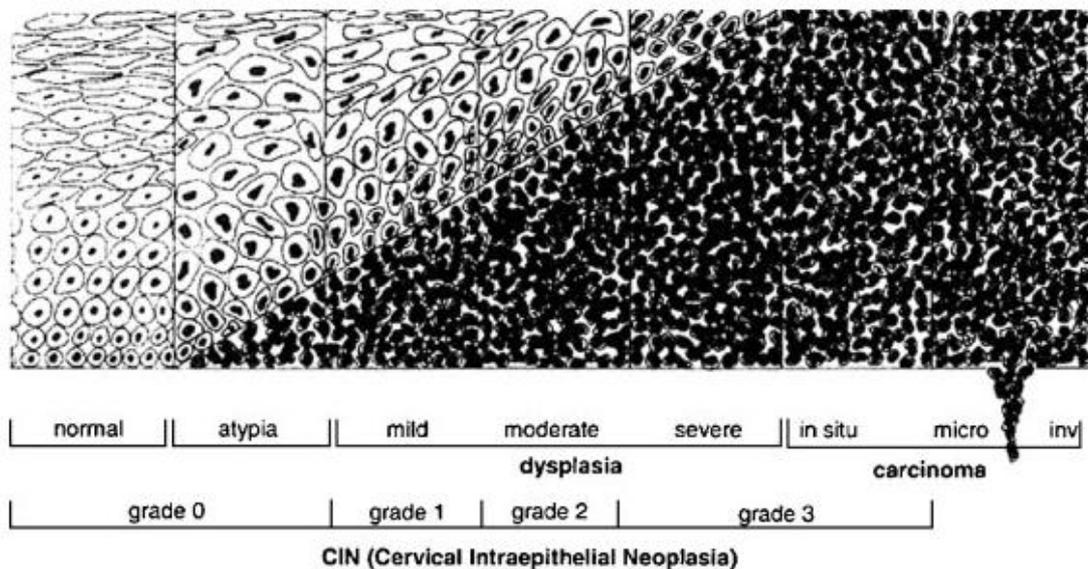
Figure 1.6: Schematic presentation of the classification of morphological alterations seen in cervical lesions.

Adapted from (Snijders *et al.*, 2006).

Whilst cytological screening has been successful, a non-subjective molecular approach that includes HPV typing has been shown more sensitive for CIN2+, whilst allowing the study of HPV

epidemiology. A molecular based approach may also facilitate risk assessment for a given HPV infection. HPV based cervical screening would also be useful in assessing the effect of the prophylactic HPV vaccination programme (Hibbitts *et al.*, 2008).

1.2.3. Other Anogenital Neoplasias

1.2.3.1. Genital Warts

Various LrHPV types are also the causative agent of warts, including genital warts. Genital warts are primarily caused by HPV6 and HPV11 and are relatively common. In 2011, there were 76,071 cases of genital warts in England alone (Health Protection Agency, 2012, Lacey *et al.*, 2006).

1.2.3.2. Vulval Intraepithelial Neoplasia

Several components of this investigation relied on the use of material from Vulval Intraepithelial Neoplasia (VIN) lesions. In the UK, VIN is typically graded histologically using a scale similar to that used for CIN (Figure 1.6). An additional means of VIN classification has been proposed to reflect the potential difference in aetiology between two types of VIN (usual and differentiated) and the inconsistent definition of VIN1 as a precursor of vulval cancer (Sideri *et al.*, 2005) however it is not universally used in the UK.

HPV-associated VIN is known as usual VIN (van de Nieuwenhof *et al.*, 2008). Many (50-80%) vulval SCCs are not HPV-associated (JCNI, 2010, van de Nieuwenhof *et al.*, 2009) and arise via a different route and premalignant phase to HPV-associated vulval SCC (van de Nieuwenhof *et al.*, 2011, van de Nieuwenhof *et al.*, 2008). Lichen sclerosus (LS) is a chronic inflammatory skin condition that affects both male and female patients, it is most common in the anogenital region of post-menopausal women (Tasker and Wojnarowska, 2003). LS is not typically associated with HPV infection (van de Nieuwenhof *et al.*, 2008) and LS patients are on average, older than HPV-associated VIN patients (Bonvicini *et al.*, 2005). A multitude of factors have been implicated in the aetiology of LS including several infectious agents, genetic predisposition, oestrogen level, vulval irritation and autoimmune disorders (van de Nieuwenhof *et al.*, 2008). LS is associated with progression to SCC and progression is observed in 2-6% of cases. For most vulval SCC patients, LS can also be found in adjacent areas (Kagie *et al.*, 1997, Leibowitch *et al.*, 1990, Carli *et al.*, 1995).

Differentiated VIN is a rare diagnosis and its existence is disputed (McCluggage, 2009). When diagnosed, differentiated VIN usually occurs in a background of LS and whilst its aetiology is uncertain, it is not associated with HPV (Bonvicini *et al.*, 2005).

1.2.4. HPV, the Immune System and the HPV Vaccine

Both the innate and adaptive immune effector mechanisms are influenced by triggering of the innate immune mechanisms; dsRNA, ssDNA and foreign antigens either free or displayed by host cells are capable of initiating an immune response. The life cycle and gene expression of HPV however occurs within differentiating epithelial cells, there is no viraemic stage, no cellular lysis and virions are released by the desquamation of keratinised epithelial cells (Stanley, 2006b, Frazer, 2009). The immune system thus has limited opportunity to detect HPV infection and relies on the identification of cells displaying stress signals due to E6 and E7 mediated cell cycle deregulation (Frazer, 2009). Hence the ability of the host to detect and subsequently clear an HPV infection is severely limited and HPV infections can be chronic and difficult to clear (Stanley, 2006b). Chronic HPV infection is associated with CIN and cancer (Ho *et al.*, 1995, Stanley *et al.*, 2007).

In August 2008 the UK initiated a nationwide prophylactic vaccination programme for adolescent females, utilising the bivalent vaccine "Cervarix" manufactured by GlaxoSmithKline UK which provides protection against HPV16 and HPV18 induced anogenital neoplasia. The vaccine utilises Virus Like Particles (VLPs). I*n vitro* L1 proteins spontaneously self assemble to form VLPs which are structurally very similar to HPV capsids (Hagensee *et al.*, 1994). Several types of VLP can be constructed; L1 and L2 chimeric VLPs, L1 external chimera VLPs and L1 internal chimera VLPs. L1 external chimeras (as used in Cervarix) and L1 and L2 mixed chimeras are structurally very similar. Internal chimeras of L1 are an interesting prospect as they could allow the incorporation and expression of polypeptides of other viral and cellular proteins into the VLPs, the polypeptides could be for instance HPV early gene products or even the gene products of another virus (Schiller and Lowy, 2000). L1 external chimeras induce a good humoral immune response, produce significant quantities of serum antibody and thus effectively confer some degree of resistance against HPV infection (Breitburd *et al.*, 1995, Kahn and Burk, 2007). Cervarix contains equal quantities of L1 protein from HPV16 and HPV18, is administered in three stages over the course of six months (GlaxoSmithKline, 2008) and functions with a high efficacy against the target HPV types. The HPV18 L1 component also appears to offer some protection against HPV45 (Kahn and Burk, 2007). A quadrivalent vaccine manufactured by Merck ("Gardasil") contains VLPs for HPV16, 18, 6 and 11 thereby conferring protection against HPV16 and 18 anogenital neoplasia and in addition, protection against LrHPV types 6 and 11 which are responsible for over 90% of genital warts (CDC, 2011).

1.2.5. Carcinogenesis

In high grade cervical lesions, viral genome amplification and *E6* and *E7* oncogene expression occur closer to the epithelial surface than in low grade cervical lesions and expression of the late proteins is retarded (Middleton *et al.*, 2003). The malignant transformation of a host cell is an accidental event that is atypical of the normal viral lifecycle, and is associated with persistent HPV infection (Ho *et al.*, 1995). Malignant cells are incapable of viral amplification and late gene expression such that no new virions can be produced and the productive life cycle terminates (Münger and Howley, 2002). Several factors may stimulate oncogene deregulation, including viral integration and the mutation/deletion of transcriptional repressors (Badal *et al.*, 2003). Deregulated and aberrant *E6* and *E7* expression in high grade disease results in the disruption of the usual cell cycle control mechanisms via the interaction of HPV proteins with a range of host factors including p53 and pRb. Essentially, p53 mediated DNA damage cell cycle controls cannot be enforced, cells resist apoptosis and the accumulation of mutations may go unchecked (Doorbar, 2006).

Genomic instability is an early event in HPV induced carcinogenesis (zur Hausen, 1991) and results in alterations of chromosome number and localised structural anomalies (Duensing and Munger 2002; Heselmeyer *et al.* 1996). In the context of HPV induced cancers, some viral lifecycle and molecular events affect genomic stability which may lead to carcinogenesis. For example, E6 and E7 mediated degradation of p53 and pRb results in reduced ability to arrest in response to DNA damage (Thorland *et al.*, 2003).

Activity of the HPV E1 and E2 proteins from integrated HPV is associated with deletion of host DNA proximal to the integration site, the presence of multiple HPV genomes in the host genome may also generate sources of homology for homology dependent recombination resulting in genomic instability (Kadaja *et al.*, 2007). The HPV E7 mediated induction of centrosome duplication results in supernumerary spindle poles (Duensing *et al.*, 2000). E6 mediated p53 degradation effectively represses DNA damage controls and cell cycle checkpoint integrity, such that aneuploid cells and cells with major genomic changes are not detected and/or apoptosed (Duensing *et al.*, 2000). The improper detection and response to DNA breakage due to the action of E6 and E7 may lead to structural aberrations of the host cell chromosomes such as anaphase bridges. Anaphase bridges may result in the development of chromosomal rearrangements; translocations, duplications and deletions (Yugawa and Kiyono, 2009). HPV E6 both directly and indirectly activates *hTERT* expression, and physically interacts with the hTERT protein enhancing its telomerase activity expression (Veldman *et al.*, 2003, Liu

et al., 2009). Enhanced hTERT activity enables cancer cells to resist cellular senescence and undergo successive rounds of replication (Yugawa and Kiyono, 2009).

The accumulation of mutations, genomic rearrangements and deletions in cells with deregulated DNA maintenance mechanisms and a lack of cellular senescence signals promotes selection of cells that exhibit deregulated cell cycle controls and thus a growth advantage. HPV integration typically leads to the retention of the *E6* and *E7* genes (Thorland *et al.*, 2003) but loss of *E1*, *E2* and other downstream sequences leading to the lack of E2 trans-acting repression of the P97 promoter (Howley and Lowy, 2006, Cricca *et al.*, 2009). HPV integration also appears to influence E6 and E7 protein levels through the stabilising effects on the polycistronic *E6* and *E7* mRNA as a result of fusion of the *E6* and *E7* ORFs to downstream cellular DNA (Howley and Lowy, 2006). Integrated HPV *E6*/*E7* transcripts often include 3' cellular untranslated sequences and polyadenylation signals, increasing the stability of the transcript (Klaes *et al.*, 1999). Increased *E6*/*E7* mRNA stability results in elevated E6 and E7 translation, and thus increased levels of E6 and E7 oncoproteins which are considered important factors in HPV induced carcinogenesis (Jeon and Lambert, 1995).

1.2.6. Viral Integration

1.2.6.1. Integration is Associated with Disease

Cell cultures with integrated HPV16 demonstrate a selective growth advantage over cells harbouring episomal HPV16 (Jeon *et al.*, 1995) and the majority of cervical cancers contain integrated forms of HPV (Choo *et al.*, 1987). HPV integration is often considered an early event in HPV induced carcinogenesis and frequently results in the loss of *E2* and/or *E1* (Arias-Pulido *et al.*, 2006). Loss of E2 trans-acting repression of the LCR may correspond to the deregulated expression of *E6* and *E7* (Jeon and Lambert, 1995, Thierry, 2009, Romanczuk and Howley, 1992, Rosl *et al.*, 1993). HPV integration initially occurs within a background of episomal HPV, and therefore functional E2 protein produced by episomal HPV. Loss of *E2* expression via episomal clearance might therefore be necessary for integration induced transformation and is a feature of certain cell lines (Pett *et al.*, 2006, Pett and Coleman, 2007). HPV integration can also affect human gene expression; integrants were found proximal to the *MYC* locus in several HPV containing tumour cell lines, and as a result *MYC* was overexpressed (Herrick *et al.*, 2005, Peter *et al.*, 2006).

HPV16 and HPV18 are more commonly integrated in more severe cervical disease; in cervical cancers, rates of integration of up to 80% and 100% are observed for HPV16 and HPV18

respectively (Corden *et al.*, 1999, Cullen *et al.*, 1991, Pirami *et al.*, 1997, Melsheimer *et al.*, 2004, Arias-Pulido *et al.*, 2006). There also appears to be a correlation between integration status and risk of CIN progression (Arias-Pulido *et al.*, 2006). Nonetheless, a proportion of HPV16 induced cervical carcinomas are purely episomal, indicating that HPV16 integration is only one of several factors associated with an increased risk of progression.

1.2.6.2. *Aetiology of HPV Integration*

There are two main types of HPV integrants described during integration studies; single insertions and concatemeric insertions. These terms represent events where HPV genomes have integrated as individual copies (possibly at multiple sites) and as repeated concatemers integrated into the host cell genome respectively (Jeon and Lambert, 1995, Jeon *et al.*, 1995). HPV integration by non-homologous end joining between the viral and host genomes is thought to be a random event, however HPV integration occurs most frequently at or close to common fragile sites (CFSs) (Dall *et al.*, 2008, Thorland *et al.*, 2003, Matovina *et al.*, 2009). Nonetheless, no consensus sequence has been observed, and integration at similar sites in different tumours occurs within disparate sequences (Pett and Coleman, 2007). CFSs are unstable, recombinogenic chromosomal loci of the genome and are preferential sites for sister chromatid exchange, translocations, deletions, intrachromosomal gene amplification and integration of plasmid and oncogenic virus DNA (Smith *et al.* 1998; Thorland *et al.* 2003).

Linear HPV genome fragments produced by defective viral replication of episomal HPV or dsDNA breaks may serve as templates for HPV integration into the host genome via non-homologous end joining (Mannik *et al.*, 2002, Kadaja *et al.*, 2007). Once integrated, HPV origins are still targeted for amplification by episomal E1 and E2 proteins and can result in replication of the integrated HPV sequences and human sequences adjacent to HPV to produce homologous linear DNA sequences (Kadaja *et al.*, 2007). Further recombination of these linear fragments with the host genome could result in DNA rearrangement, deletion and/or further integration (Kadaja *et al.*, 2007, Pett *et al.*, 2004), i.e. characteristics of cancers (Klausner, 2002, Duensing and Münger, 2004).

Whilst integration is very common in cancers, many cervical cancers and many more precancerous cervical lesions do not contain integrated HPV. Integration could be a result of oncogene deregulation and associated genomic instability, rather than a cause of deregulation and transformation. As such, an alternative means of initial deregulation may exist (Vinokurova and von Knebel Doeberitz, 2011). Epigenetics represents a series of potential alternative transforming events for a HPV infection.

1.2.7. Integration Study Techniques

Integration of HPV DNA into the host genome is a very common event in HPV induced malignancies and has been proposed as a potential molecular biomarker of an increased risk of progression to cancer (Pett and Coleman, 2007).

1.2.7.1. APOT

Amplification of Papillomavirus Oncogene Transcripts (APOT) is an assay that permits the differentiation of HPV mRNAs produced by transcriptionally active integrated and episomal HPV genomes (Klaes *et al.*, 1999). The technique amplifies HPV-human fusion transcripts that can be sequenced to allow the identification of human sequences at the 3' end of viral oncogene transcripts (Figure 1.7).
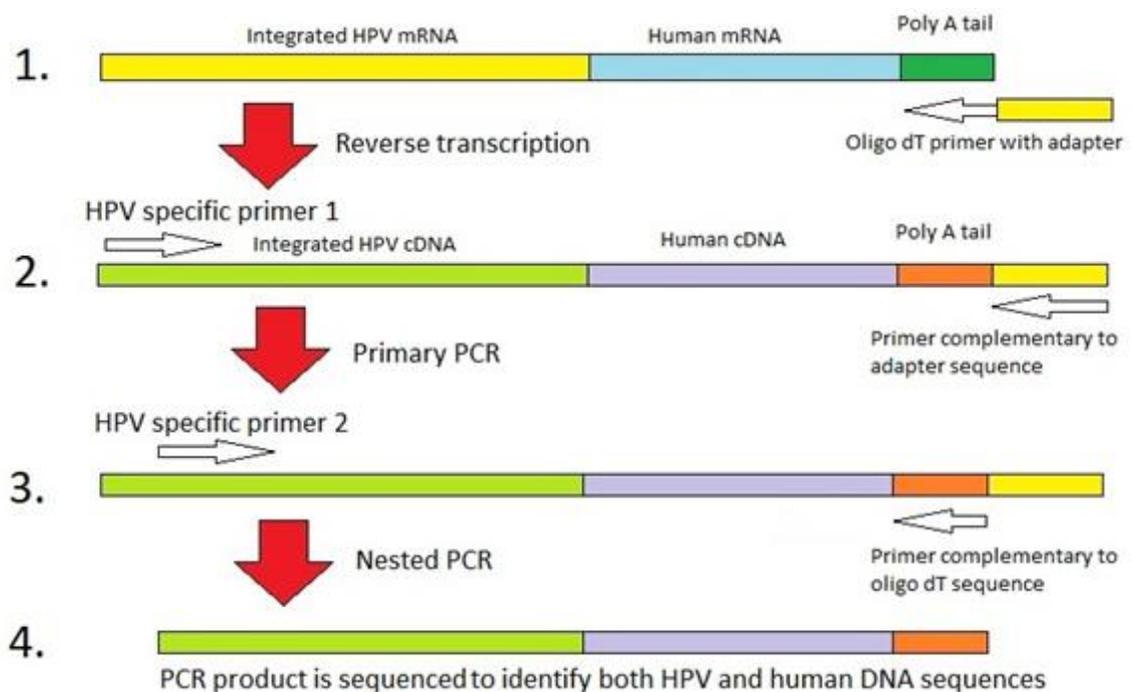
Figure 1.7: The APOT procedure.

1; mRNA transcripts are reverse transcribed with an oligo dT primer with an adapter to form cDNA. If the HPV is integrated then there is a chance that a HPV-human fusion transcript will be produced with human sequence on the 3' end of the transcript. 2; a primary PCR is performed on the cDNA using a primer specific to HPV *E7* and a primer specific to the adapter sequence. 3; A second round of PCR is required using another *E7* specific HPV primer downstream of the first and a reverse primer that is specific to the oligo dT primer sequence. 4; the resulting PCR products are separated by electrophoresis, excised and sequenced. Integrated sequences will contain both papillomavirus and human sequences and sequencing will allow the identification of the location of the integration events (Klaes *et al.*, 1999).

1.2.7.2. RS-PCR

Restriction site PCR (RS-PCR) is DNA based and uses a series of HPV specific primers and primers specific to the recognition sequence of several restriction enzymes. The restriction enzymes chosen are those without, or with only one restriction site within the HPV genome. The basis of this technique is that restriction sites are a common occurrence in the human genome and have a known sequence. As such, if one of the targeted restriction sites occurs in a human genome within close proximity of the integrated HPV it should be amplified by one of the PCR reactions (using a HPV specific and restriction site specific PCR primer) (Ragin *et al.*, 2004).

1.2.7.3. E2 PCRs

HPV16 integration frequently results in the disruption of the *E2* gene (Baker *et al.*, 1987). A simple method of detecting *E2* disruption is to use overlapping PCRs that cover the whole *E2* gene (Collins *et al.*, 2009). If *E2* is disrupted then one or more of the PCR reactions will fail. It is important to consider that this method does not detect integration events where the *E2* gene remains intact or integration events that have resulted in concatemeric HPV sequences (such as observed in CaSki) (Collins *et al.*, 2009). Knowledge of *E2* disruption is biologically relevant because of the important regulatory role of E2 (Collins *et al.*, 2009, Baker *et al.*, 1987).

1.2.7.4. DIPS

Detection of integrated papillomavirus sequences (DIPS) is DNA based. DIPS relies on cutting human DNA located 5' to HPV DNA using restriction endonucleases, then ligating an adapter of known sequence to the sticky end (Figure 1.8). For integrated HPVs, the result would be HPV DNA at the 3' end of the DNA, human DNA of unknown sequence at the 5' end, but with an adapter of known sequence positioned at the very end. The ligated constructs are first amplified in a linear manner using a HPV specific primer, and then amplified in an exponential manner using a HPV primer and an adapter specific primer. These products are then separated by electrophoresis, the bands excised, purified and sequenced in order to identify both HPV and human sequence in the amplimers (Luft *et al.*, 2001).
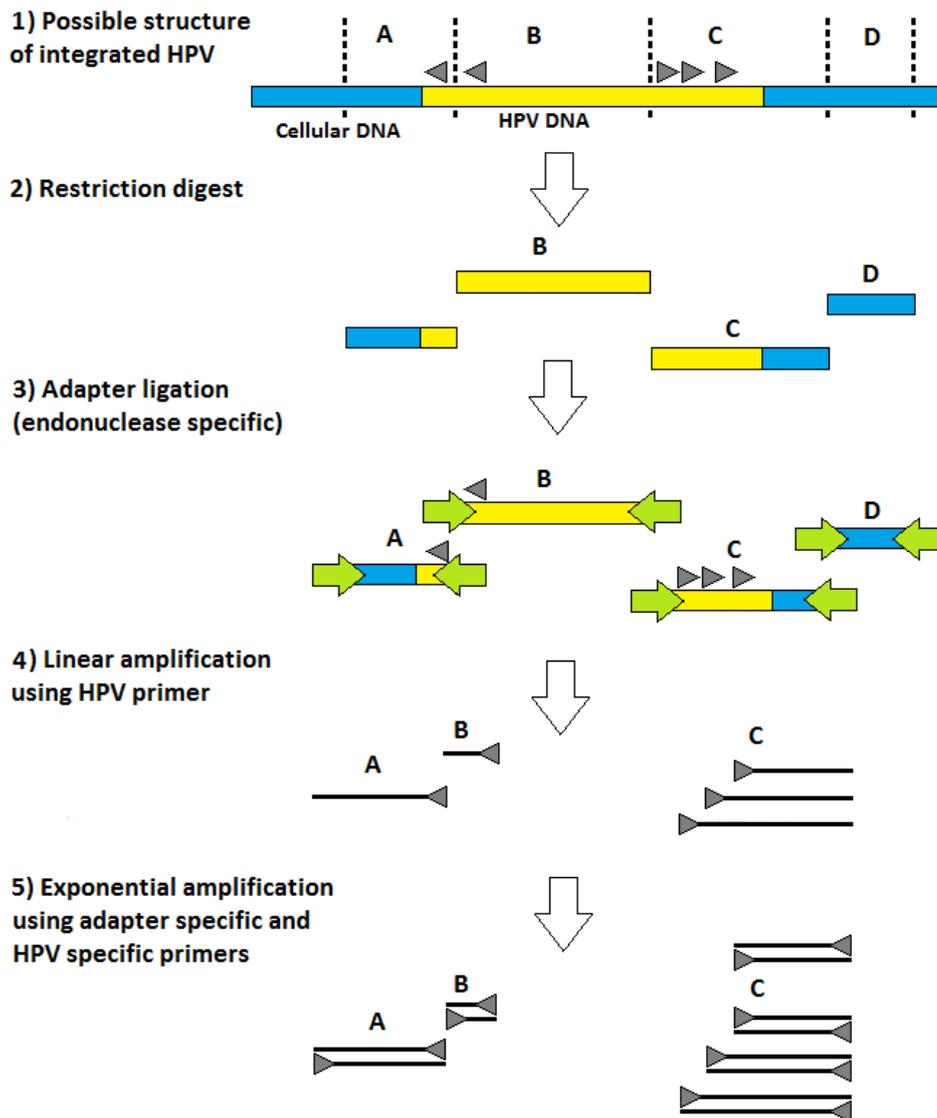
Figure 1.8: Schematic representation of the DIPS method.

Adapted from (Luft *et al.*, 2001, Ragin *et al.*, 2004). 1; Possible structure of an integrated HPV genome. Dotted lines represent endonuclease restriction sites and grey triangles represent HPV specific primer sites. 2; Restriction of genomic DNA yields four classes of fragments, HPV only (B), human only (D) and both human and HPV (A and C). Episomal HPVs produce type B only. 3; Endonuclease specific dsDNA adapters (green arrows) are ligated to the restriction fragments which introduces PCR primer binding sites. 4; Linear PCR is performed using HPV specific primers to amplify fragment types A, B and C. Single stranded PCR products are generated. 5; An exponential nested PCR is performed using HPV specific and adapter specific primers to generate double stranded PCR products for sequencing. B type products are identified based on size and are not sequenced (Luft *et al.*, 2001).

Chapter 2. Introduction: DNA Methylation and Epigenetics

Epigenetic inheritance is defined as cellular information other than DNA sequence that is heritable during cell division. This includes DNA methylation, histone modification and nucleosome positioning (Feinberg and Tycko, 2004, Esteller, 2002).

2.1. DNA Methylation

DNA methylation is the covalent addition of a methyl group to the 5' position of cytosine (Figure 2.1). In mammals, this normally occurs within cytosine-guanine dinucleotides (5'-CpG-3', also known as CpGs) and there is an overwhelming tendency for symmetrical methylation (i.e. methylation on both strands of the DNA) (Law and Jacobsen, 2010, Lister *et al.*, 2009). DNA methylation outside of the CpG context is considerably less common. Non-CpG methylation at CHH and CHG (where H is A, C or T) has been described in stem cells, primarily within gene bodies, and correlates with gene expression (and is less common in protein binding sites and enhancers) (Lister *et al.*, 2009). The role and mechanisms of non-CpG methylation within human cells are poorly understood but appear to be involved in pluripotency (Portela and Esteller, 2010).
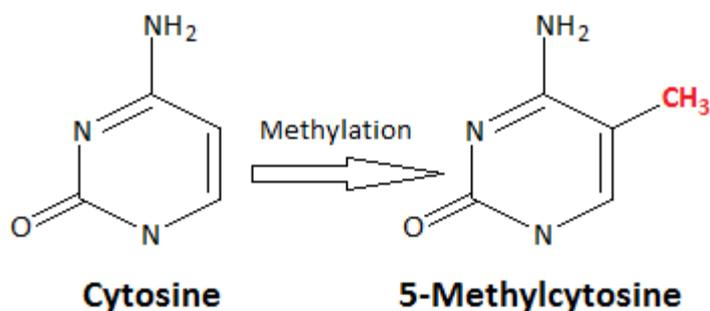
Figure 2.1: A schematic diagram of cytosine methylation.

Timely and accurate DNA methylation is essential for the development and maturation of cells and organisms; more specifically for normal embryonic development, X-chromosome inactivation, genomic imprinting and the suppression of parasitic DNA sequences (Robertson and Jones, 2000). Many human diseases, including cancers, neurological diseases and autoimmune diseases occur as the result of untimely or inaccurate DNA methylation (Chen and Li, 2004, Zilberman *et al.*, 2007). *In vitro* studies using mice with knockout mutations in genes involved in DNA methylation produced a range of defects as diverse as early developmental arrest, embryonic lethality, death at a young age, neural tube defects, abnormal behavioural patterns in adulthood, sterility, complex neurological defects, tendency to be overweight and issues surrounding X-inactivation (Li, 2002).

2.2. Histone Modification

Histone modification is another well studied example of epigenetic information that plays an important role in development and gene expression (Portela and Esteller, 2010). There are 4 core histones H2A, H2B, H3 and H4 that form a complex of two H2A-H2B dimers and one H3-H4 tetramer to form the nucleosome (Luger *et al.*, 1997). A length of 147 bp of DNA is wrapped around each nucleosome and there is a short stretch of linker DNA with an average length of 50 bp between adjacent nucleosomes (Figure 2.2). Histone H1 binds this linker DNA in order to seal off the nucleosomes where the DNA is exposed (Daujat *et al.*, 2005). All of the histones can be chemically modified post-transcriptionally, these modifications occur mainly in the N-terminal tails and include acetylation, methylation, phosphorylation, ubiquitination and others (Portela and Esteller, 2010). Histone modifications are mediated by a similarly diverse array of enzymes, for example histone acetyltransferases (HATs), deacetylases (HDACs), methyltransferases (HMTs) and demethylases (HDMs), kinases and phosphatases, ubiquitin ligases and deubiquitinases, SUMO ligases and proteases (Rodriguez-Paredes and Esteller, 2011).
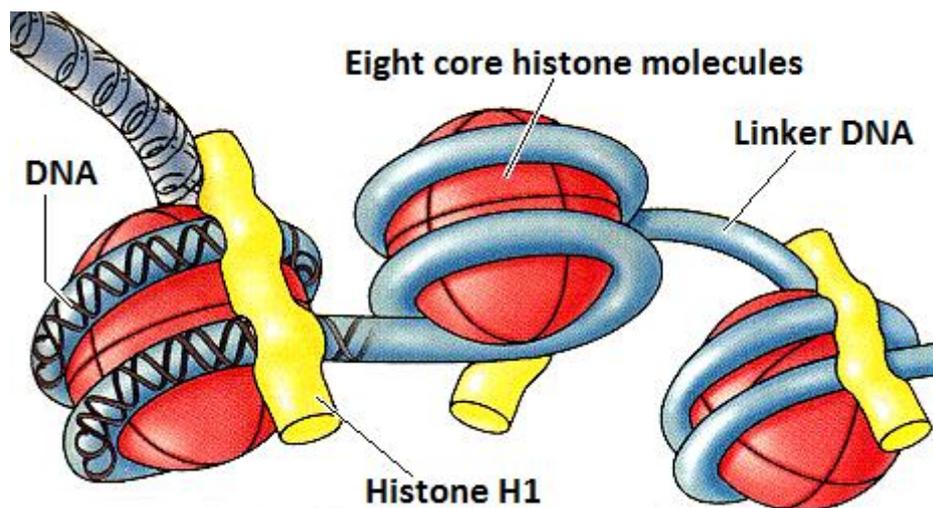
Figure 2.2: Diagram of DNA wrapped around nucleosomes.
Adapted from (Luong, 2009).

The presence, absence and interactions of histone modifications modulate chromatin conformation to influence the transcriptional activity of a DNA region. Transcriptionally active euchromatin and silenced heterochromatin are each characterised by a different subset of histone modifications (Rodriguez-Paredes and Esteller, 2011, de Ruijter *et al.*, 2003). Similarly, gene expression microarray studies observe specific promoter histone modification profiles (especially modifications of H3 histones) that were predictive of transcription level (Karlic *et*

al., 2010). Abnormal histone modification profiles are also a hallmark of certain cancers (e.g. skin, prostate, and colon cancers) (Fraga *et al.*, 2005, Esteller, 2007a).

Many combinations of histone modifications have been defined with a variety of potential biological outcomes in a wide range of processes (Ernst and Kellis, 2010). To add to the complexity, there is also significant interplay between different epigenetic modifications, for example the DNA Methyltransferase (DNMT) DNMT3L specifically interacts with H3 histones, recruits other DNMTs and induces DNA methylation, this interaction is inhibited by methylation of lysine 4 of the H3 histone (Ooi *et al.*, 2007). There are also several histone methyltransferases that recruit DNMTs and direct DNA methylation to specific regions of the genome, and several histone methyltransferases and demethylases can modulate the stability of DNMT proteins, regulating DNA methylation level (Portela and Esteller, 2010). The converse is also true in that DNA methylation can direct histone modifying enzymes (Portela and Esteller, 2010). In summary, histone modifications play a role in a diverse range of biological processes and abnormal histone modification profiles are a hallmark of cancer (Rodriguez-Paredes and Esteller, 2011, Arico *et al.*, 2011).

2.3. The Role of DNA Methylation

2.3.1. DNA Methylation and Early Development

DNA methylation plays a crucial role in early development. Post-fertilisation, the maternal and paternal gamete genomes are epigenetically different and functionally non-equivalent (Surani *et al.*, 1984, McGrath and Solter, 1984). This phenomenon is known as genomic imprinting and results in maternal/paternal specific allele expression for many genes (Reik and Walter, 2001).

Genomic imprints are erased upon the formation of germ cells and a new imprinting profile is established. In somatic cells, genomic imprints are maintained and later modified during development (Reik and Walter, 2001). Abnormal genomic imprinting is implicated in several syndromes that result in developmental, sexual and mental deficiencies (i.e. Beckwith-Wiedemann Syndrome, Prader-Willi syndrome and Angelman syndrome) and also numerous human cancers such as neuroblastoma, Acute Myeloblastic Leukemia (AML), Wilms' tumour, rhabdomyosarcoma, sporadic osteosarcoma and familial glomus tumours (Falls *et al.*, 1999).

2.3.2. DNA Methylation and Cellular Differentiation

Many genes involved in differentiation following cellular specialisation appear to be affected by DNA methylation. For example the *Oct4* and *Nanog* genes, which regulate pluripotency are silenced by promoter hypermethylation in differentiated cells but are hypomethylated and expressed in germline cells (Hattori *et al.*, 2004, Berdasco and Esteller, 2010, Hattori *et al.*, 2007). Indeed, several hundred genes have been identified in mouse and human studies that are differentially methylated in undifferentiated somatic tissues compared to differentiated germline tissues (Mohn *et al.*, 2008, Schilling and Rehli, 2007). Nonetheless, the role of DNA methylation in the establishment and maintenance of cell-type-specific expression of developmentally regulated genes is not completely understood. Methylation of these genes in many instances was not negatively correlated with gene expression as might be expected (Eckhardt *et al.*, 2006, Berdasco and Esteller, 2010).

2.3.3. Mechanisms of DNA Methylation

Methylation of mammalian CpG dinucleotides is mediated by a family of DNA-[cytosine-5] methyltransferases that catalyse the transfer of a methyl group from S-adenosyl methionine to cytosine residues in DNA. There are five DNMTs currently reported in humans, these are DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L however only DNMT1, DNMT3A and DNMT3B possess methyltransferase activity (Portela and Esteller, 2010). DNMT1 is thought to function as the primary maintenance methyltransferase, and has a strong preference for hemimethylated CpG sites produced during DNA replication (Pradhan *et al.*, 1999). This allows for replication of the existing methylation pattern onto newly synthesised DNA strands (Portela and Esteller, 2010, Chen and Li, 2004). The two *de novo* methyltransferases DNMT3A and DNMT3B (Chen and Riggs, 2011) act to establish the methylation patterns of the genome during gametogenesis, embryogenesis and somatic tissue development (Rodriguez-Osorio *et al.*, 2010).

Several diseases are caused by genetic defects in the DNA methylation machinery genes. The extremely rare condition Immunodeficiency, Centromere instability and Facial abnormality (ICF) syndrome and some cases of AML for example have been found to possess *DNMT3B* and *DNMT3A* mutations respectively (Ley *et al.*, 2010, Jiang *et al.*, 2005). Methylation is also important for early embryonic development; mammalian embryos deficient in DNMT3A/B do not survive (Okano *et al.*, 1999). DNMT3L does not possess methyltransferase activity, but acts as a stimulatory factor for DNMT3A and DNMT3B (Zilberman *et al.*, 2007). DNMT2 has very

little DNMT activity and does not appear to methylate DNA, but does methylate tRNA[Asp] (Goll *et al.*, 2006).

The distinction between the roles of DNMT3A/B and DNMT1 may not be as clear as once thought. A new model suggests a role for DNMT3A and DNMT3B in maintenance methylation and a role for DNMT1 in *de novo* methylation (Jones and Liang, 2009). In this model, DNMT1 is bound to Proliferating Cell Nuclear Antigen (PCNA) during DNA synthesis and is therefore directed towards the replication fork. In this position DNMT1 methylates the vast majority of newly synthesized hemimethylated daughter DNA strands. Rather than being located at the replication fork, DNMT3A and DNMT3B are bound to nucleosomes in areas of DNA methylation, in this position DNMT3A and DNMT3B act to maintain the methylation state in these areas and methylate hemimethylated CpGs missed by DNMT1 (Jones and Liang, 2009).

DNA regions rich in CpGs are susceptible to a range of negative consequences including silencing by methylation and increased mutation rates (Bird and Wolffe, 1999, Jones, 1996, Jones *et al.*, 2001). As such, most CpGs have been lost during evolution and only occur at 20% of their statistically expected frequency (Schorderet and Gartler, 1992, Badal *et al.*, 2003). The evolutionary loss of CpGs is not however universal throughout the genome, some regions retain CpGs at a higher density. These regions, mostly 0.5 to several kb in length, are termed CpG islands (Robertson and Jones, 2000, Bird, 1986). CpG islands are defined as regions longer than 200 bp with a G+C content of over 50% and a ratio of observed to statistically expected CpG frequencies of at least 0.6 (Portela and Esteller, 2010).

The increased rate of mutation for methylated DNA sequences is a result of the biochemistry of cytosine and methylcytosine nucleotides. Cytosine nucleotides will spontaneously undergo hydrolytic deamination to form uracil nucleotides. Because uracil is only normally found in RNA, it is easily detected and removed by the highly abundant and efficient Base Excision Repair (BER) enzyme uracil DNA glycosylase (UDG) (Robertson and Jones, 2000, Jones, 1996). Methylated cytosine (5-methylcytosine) however spontaneously deaminates to form thymine, a naturally occurring DNA base that is considerably more difficult to detect and repair via the thymine DNA glycosylase pathway (TDG). TDG mismatch repair is less efficient and up to several hundred times slower than UDG repair (Jiricny, 1996, Jones, 1996). The elevated rate of cytosine to thymine transitions for methylated residues is a source of increased mutability and over time has resulted in a decreased frequency of CpG dinucleotides (Jiricny, 1996).

2.4. Methylation of the Human Genome

Most (up to 80%) of the CpGs in the human genome are highly methylated however those within CpG islands are typically less methylated (Takai and Jones, 2002). The human genome contains up to 45,000 CpG islands, most are found within or proximal to 5' untranslated promoter regions or the first exons of genes (Antequera and Bird, 1993). CpGs are distributed such that up to 50% of the gene promoters in the human genome contain CpG islands (Bird, 1986) and over 70% of gene promoters have been described as having a CpG density that is higher than the genome average (Saxonov *et al.*, 2006). The maintenance of CpG islands might be either a requirement of high sequence fidelity (a large proportion of CpG islands are associated with housekeeping genes (Bird, 1986)) or a requirement for the region to be regulated by methylation. A variety of genes are under the control of promoters located within CpG islands, which importantly in the context of carcinogenesis, includes numerous genes involved in cell cycle control (e.g. *ARF*, *INK4a*, *INK4b*) (Esteller, 2002). CpG islands occur infrequently within the coding sequence of a gene and the 3' region of genes; these CpG islands are more likely to become methylated than typical CpG islands and were assumed not to have an effect on the progression of the relevant RNA transcript (Bird, 1986, Jones, 1999). However more recent studies suggest that in some instances intragenic, or gene body, methylation can have a demonstrable effect on gene expression/protein translation (Lorincz *et al.*, 2004).

Provided that the relevant transcription factors are available, the unmethylated state of a typical CpG island permits an uncondensed chromatin structure that is favourable to transcription (Esteller, 2007b). A notable feature of CpG islands is the presence of acetylated H3 and H4 histones (forming the nucleosomes) and a distinct reduction of H1 linker histone resulting in the relaxed chromatin state (Tazi and Bird, 1990). The majority of CpG dinucleotides within the genome occur within mobile genetic elements and the ancestral remnants of a range of parasitic DNA sequences that comprise upwards of 40% of the entire human genome (Robertson and Wolffe, 2000). DNA methylation has been proposed as a mechanism evolved to silence the expression, mobility and thus the multiplication of such genetic elements. Silencing of these regions of DNA would be essential to prevent retrotransposon activity, and to maintain genomic structural integrity by preventing recombination between homologous sequences (Robertson and Wolffe, 2000).

2.4.1. DNA Methylation and Transcriptional Regulation

DNA methylation plays a significant role in transcriptional regulation. The promoters of transcriptionally active genes are often located within or proximal to unmethylated CpG islands which when methylated can silence the expression of these genes (Eden and Cedar, 1994). Typical CpG islands lack methylation at any stage of development and allow the expression of associated genes if the appropriate transcription factors are present and the chromatin structure is accessible (Esteller, 2002). A range of genes containing CpG islands are found to be hypermethylated in tumours, these include a number of cell cycle control genes and tumour suppressor genes (Robertson, 2001, Badal *et al.*, 2003).

Transcriptional repression due to DNA methylation may occur through several mechanisms, these include the inability of several transcription factors to bind methylated target sites (Bird and Wolffe 1999; Thain *et al.* 1996) and the alteration of chromatin conformation around binding sites or even entire gene promoters (Jones *et al.*, 2001). The most direct method by which DNA methylation can affect transcription is the inhibition of binding of DNA binding proteins to their target sites (Kuroda *et al.*, 2009). These proteins include the transcriptional machinery and transcription factors that require contact with cytosine in the major groove of the DNA double helix. Many mammalian transcription factors contain GC-rich binding sites and often contain CpGs within their DNA recognition elements (Bird and Wolffe, 1999). CpG methylation at transcription factor binding sites may impede the ability of a transcription factor to bind DNA however not all transcription factors are affected by methylation (i.e. Sp1) and there are some instances of highly methylated genes that are still transcribed (Bird and Wolffe, 1999).

It is thought that transcriptional regulation through the inhibition of binding of transcription factors to methylated binding sites is a relatively rare event in the human genome (Weber *et al.*, 2005, Li, 2002). For example the glial fibrillary acidic protein is activated during astrocyte differentiation by the demethylation of a CpG in a STAT3 (signal transducer and activator of transcription 3) binding element in its promoter region (Takizawa *et al.*, 2001). Conversely, DNA methylation of a silencer element of the *Igf2* (insulin like growth factor 2) gene blocks the binding of repressor proteins such that transcription from the promoter increases (Murrell *et al.*, 2001).

In order to transcribe a given region of DNA, the supercoiled tertiary structure must be relaxed to allow the necessary cellular enzymes access to the DNA. In the supercoiled state the genes within the DNA region are transcriptionally silenced. This supercoiled chromatin state is generated by the action of HDAC proteins on the nucleosomal histones of newly synthesised

DNAs (Jones *et al.*, 2001). Histone deacetylation involves the removal of acetyl groups from lysine residues at the N-terminus of histone proteins; acetylated histones show a markedly reduced affinity for DNA (de Ruijter *et al.*, 2003) permitting a more relaxed and accessible structure. Histone deacetylation therefore encourages the maintenance of a supercoiled structure by promoting the association between histone proteins and DNA, permitting DNA supercoiling. Repression by DNA methylation works by effectively targeting this HDAC activity towards specific areas of the genome which results in histone deacetylation, supercoiling and silencing of methylated regions (Burgers *et al.*, 2002). DNMT1 (the maintenance methyl transferase active during DNA synthesis) is found at the replication forks, and the association between DNMT1 and HDAC2 (HDAC2 associates with the DNMT1 non-catalytic N-terminus) acts to direct HDAC activity towards newly synthesised DNA (Burgers *et al.*, 2002). DNMT3A and DNMT3B also recruit HDACs to silence genes independent of their methyltransferase activity (Burgers *et al.*, 2002).

In addition to the association between the methyltransferases and HDACs, the HDAC complex can be directed to methylated DNA via proteins such as methyl-CpG binding proteins (de Ruijter *et al.*, 2003). Two methyl-CpG binding proteins (MeCP1 and MeCP2) recruit HDAC complexes to areas containing methylcytosine in order to mediate transcriptional repression. MeCP2 is comprised of a methyl-CpG-binding domain which recognises methylated CpG dinucleotides and a transcriptional repression domain which represses transcription through an interaction with the cellular transcriptional protein TFIIB (Jones *et al.*, 2001, Robertson and Wolffe, 2000). MeCP2 can bind to histone bound DNA and in some instances is responsible for H1 displacement. MeCP2 also recruits HDAC containing transcriptional repression complexes via its interaction with the Sin3A corepressor (Jones *et al.*, 2001, Robertson and Wolffe, 2000). In summary, the recruitment of HDACs to methylated DNA sequences through various mechanisms results in histone deacetylation and chromatin condensation, a state of DNA that is inaccessible to the cellular transcription machinery.

2.4.1.1.   *Intragenic Methylation*

Intragenic or gene body methylation was traditionally assumed to not affect transcription. Whilst this is still a poorly understood and considerably understudied area, more recent studies have shown otherwise. Intragenic methylation can have a significant effect on transcription elongation of mammalian transgenes, possibly as a result of an association with a more condensed chromatin state that reduces elongation efficiency (Lorincz *et al.*, 2004). Intragenic methylation of *Arabidopsis thaliana* genes was also shown to inhibit transcriptional elongation. Gene body methylation in plant protoplasts was also seen to interfere with

elongation, especially if the methylation was close to the 5' end (Hohn *et al.*, 1996); this inhibition might be dependent on both the length and activity of the gene (Zilberman *et al.*, 2007). It is hypothesised that intragenic DNA methylation might prevent spurious initiation of transcription by preventing binding of transcription factors to gene body binding sites which might actually improve the transcription rate in some instances (Maunakea *et al.*, 2010, Zilberman *et al.*, 2007).

The role of intragenic methylation in humans is poorly understood. Human intragenic methylation might also be associated with repression or enhancement of transcription depending on the precise nature of the genes in question. Gene body methylation and promoter hypomethylation are positively correlated with gene expression during X-inactivation in early development. A similar scenario was observed in human B cells where it was also suggested that intragenic methylation might regulate the use of alternative human promoters (Hellman and Chess, 2007, Rauch *et al.*, 2009). In a range of human tissues, promoter hypomethylation and gene body hypermethylation are associated with a higher rate of cell division (Aran *et al.*, 2011). However data are contradictory regarding whether intragenic methylation is a cause or a consequence of high gene transcription (Shenker and Flanagan, 2012). Despite this, it is well established that aberrant DNA methylation of the human genome is associated with cancer and various roles (e.g. transcriptional regulation, suppression of non-coding RNA expression, cellular defence etc) could be assigned to aberrant intergenic, intragenic and promoter hypermethylation (Shenker and Flanagan, 2012).

2.4.2. DNA Demethylation

Aside from cancer cells, the genomic DNA methylation pattern of mature cells is usually considered to be a fixed state, recently however this has been questioned. DNA demethylation has been observed in several contexts and appears to be achieved by both an active and passive means (Wu and Zhang, 2010). Passive DNA demethylation occurs when the maintenance DNMT, DNMT1 is inhibited or absent during DNA replication. The newly synthesised DNA strands are not methylated which leads to gradual demethylation with cell division (Wu and Zhang, 2010). Passive demethylation is also observed due to exclusion of maternal DNMT1 from the nucleus during blastocyte development in a cell division dependent manner, however imprinted genes resist demethylation (Monk *et al.*, 1987, Rougier *et al.*, 1998). The alternative to passive demethylation is the active demethylation of DNA as a result of enzymatic activity. The significance of active demethylation has been controversial (Ooi and Bestor, 2008).

Ramchandani *et al* proposed that DNA methylation is a dynamic state that is the result of the balance between DNA methylation and enzymatic demethylation (Ramchandani *et al.*, 1999, Szyf *et al.*, 2004). Active DNA demethylation is observed as a genome wide effect at specific times during early development and in more specific loci in somatic cells in response to various stimuli (Wu and Zhang, 2010).

During fertilisation, genome wide demethylation is seen in the paternal pronuclei shortly after fertilisation, prior to the first cell division and is not affected by replication inhibitors, implying that an active demethylation event is responsible (Wu and Zhang, 2010, Mayer *et al.*, 2000). As per passive demethylation, several regions of the genome are resistant to active demethylation (Santos *et al.*, 2002, Rougier *et al.*, 1998, Olek and Walter, 1997, Lane *et al.*, 2003). Experiments involving injecting sperm nuclei from one species into the eggs of others revealed that demethylation was variably species specific, but consistently paternal genome specific (Beaujean *et al.*, 2004, Rougier *et al.*, 1998, Barton *et al.*, 2001, Santos *et al.*, 2002). As well as global active demethylation, there are reports of active demethylation of specific loci, each of which were triggered by various stimuli and in the absence of DNA replication, thereby excluding passive demethylation (Szyf *et al.*, 1985, Metivier *et al.*, 2008, Kim *et al.*, 2009, Gjerset and Martin, 1982, Hu *et al.*, 2000).

There are several putative demethylases, the most significant of which appears to be Methyl-CpG Binding Domain protein 2 (MBD2). MBD2 has a methylcytosine binding domain with a propensity towards CpGs, MBD2 also has DNA demethylase activity towards 5-methyl cytosine (Hamm *et al.*, 2008, Bhattacharya *et al.*, 1999), and thus demethylase activity could be directed towards sites of CpG methylation. Other proposed mechanisms of active demethylation include BER, nucleotide excision repair, oxidative demethylation and radical S-adenosylmethionine based demethylation.

2.4.2.1.   DNA Methylation as a Result of Chromatin Structure

An extension of the theory that DNA methylation is a dynamic state is that DNA methylation may not be the cause of gene repression, but rather a consequence of other changes that lead to repression. Several observations support this theory. It is normally assumed that areas rich in DNA methylation are associated with protein complexes that result in histone modifications, repressing the DNA region. However histone modifications and the subsequent silencing of the tumour suppressor p16[INK4a] has been observed prior to and independently of DNA hypermethylation; suggesting that repressive chromatin structures could promote DNA methylation (Bachman *et al.*, 2003).

Additionally, active chromatin formations may actively demethylate DNA; treating certain mammalian cell lines or *Neurospora* with histone deacetylase inhibitors (acetylated histones are associated with condensed chromatin structures) results in global genomic demethylation (Selker, 1998, Szyf *et al.*, 1985). In summary, a bilateral relationship between DNA methylation and chromatin structure might exist, i.e. the presence of DNA methylation could affect chromatin structure however the chromatin structure could in turn have an effect on DNA methylation (Szyf *et al.*, 2004).

2.5. DNA Methylation and Cancer

It is widely recognised that aberrant patterns of methylation and histone modification are a characteristic of many tumours. Cancer cells typically display a massive global reduction in methylation and have approximately 20-60% less methylcytosine than non-cancerous cells (Hellman and Chess, 2007, Rodriguez-Paredes and Esteller, 2011). Typically, tumour cells exhibit hypomethylation for most of the genome and hypermethylation of specific regions such as gene promoters (Robertson, 2001, Weber *et al.*, 2005).

A range of gene promoters, including promoters of cell cycle control, tumour suppressor and DNA repair genes (e.g. *INK4a*, *INK4b*, *ARF*, *p73*, A*PC*, *BRCA1*, *hMLH1*, *GSTP1*, *MGMT*, *CDH1*, *TIMP3* and *DAPK*) are found to be frequently hypermethylated in cancers, often in a conserved and characteristic manner for that cancer type (Esteller *et al.*, 2001). Approx 100-400 hypermethylated promoter CpG islands are found in any given tumour (Esteller, 2007a). The epigenetic repression of cell cycle control and DNA repair genes by DNA methylation may also predispose a cell to the accumulation of genetic mutations. For example, DNA hypermethylation associated inactivation of the DNA repair gene O6-MGMT (O^6-methylguanine-DNA methyltransferase) is a hallmark of some cancers (e.g. lymphomas, colon, head and neck and non-small-cell lung cancer) that results in the enhanced accumulation of G-A point mutations (Esteller *et al.*, 1999, Esteller and Herman, 2004).

As well as promoter hypermethylation, abnormal hypomethylation has been linked to the activation of certain protooncogenes (Watt *et al.*, 2000) and is also associated with various tumours (Wilson *et al.*, 2007). Additionally, whilst tandemly repetitive DNA sequences in normal cells are silenced by hypermethylation (Robertson, 2001, Badal *et al.*, 2003), in tumours these regions tend to be hypomethylated which may be associated with genomic instability (Deng *et al.*, 2006, Rodriguez-Paredes and Esteller, 2011).

Epigenetic features typical of cancers are commonly seen in benign neoplasias and early stage tumours. Epigenetic deregulation therefore may precede the typical transforming events (tumour suppressor and/or protooncogene mutation and genomic instability) (Rodriguez-Paredes and Esteller, 2011). Following mutation of one allele of a tumour suppressor gene, methylation might then be responsible for silencing the remaining functional allele and providing the second hit for cancer initiation (Rodriguez-Paredes and Esteller, 2011). *BRCA1* (Breast Cancer 1) for instance is frequently hypermethylated in ovarian and breast cancer (Esteller, 2002).

2.5.1.1. *Therapeutic Demethylation*

The DNA methylation profile of several cancers includes hypermethylation of specific CpG islands, many of which are known to be associated with tumour suppressor genes and other cell cycle control genes (Esteller, 2007a). The idea behind therapeutic demethylation is to relieve the repression of these regions by reducing the amount of DNA methylation, thereby elevating the levels of tumour suppressor gene products. Most demethylating agents are nucleotide analogues. A benefit of nucleotide analogues is that they are specifically targeted to cells undergoing DNA replication, i.e. cycling cells such as cancer cells. The most highly studied demethylating agents are analogues of cytidine and deoxycytidine, five of which are especially well characterised (Figure 2.3); 5-aza-2'-deoxycytidine, 5-azacytidine, 5'-fluoro-2'-deoxycytidine and 2-Pyrimidone-1-β-D-riboside (Zebularine), all of which appear to have varying potential as anticancer agents (Cai *et al.*, 2011). 5-azacytidine for instance is approved in several countries for the treatment of myelodysplastic syndrome (Cai *et al.*, 2011), but has also been trialled against various solid tumours (Chik and Szyf, 2011).
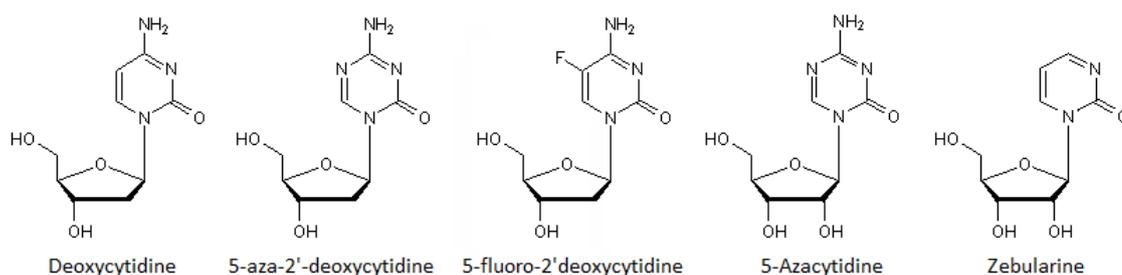
Figure 2.3: The chemical structure of deoxycytidine and several methylation inhibiting nucleoside analogues.

Most nucleoside analogues appear to demethylate DNA via a similar mechanism (Cai *et al.*, 2011). 5-aza-2'-deoxycytidine is an analogue of 2'-deoxycytidine in which the 5-carbon of the pyrimidine ring has been replaced by nitrogen (Momparler, 2005). 5-aza-2'-deoxycytidine is rapidly taken into cultured cells and incorporated into the DNA, at sub-toxic doses this leads to

incorporation into newly synthesised DNAs (Momparler, 2005). When these new DNAs are bound by DNMT1 the 5-azacytosine ring in the 5-aza-2'-deoxycytidine residues become covalently and irreversibly bound to the DNMT1 enzyme, leading to a depleted cellular level of DNMT1 and passive DNA demethylation (Jüttermann *et al.*, 1994). This demethylating effect is widely utilised to study the relationship between DNA methylation and gene suppression (Christman, 2002). 5-aza-2'-deoxycytidine also inhibits DNMT3B and leads to enhanced acetylation of histones H3 and H4 at promoter regions (Cai *et al.*, 2011), treatment with 5-aza-2'-deoxycytidine leads to demethylation and subsequent reactivation of silenced tumour suppressor genes and prometastatic genes (Chik and Szyf, 2011). In breast cancer cells, tumourigenicity is suppressed by the addition of 5-aza-2'-deoxycytidine and the alleviation of methylation of certain tumour suppressor genes (e.g. *PDLIM2*) (Qu *et al.*, 2010). Breast cancer cell line work also suggests that 5-aza-2'-deoxycytidine could have a role in enhancing chemosensitivity of breast cancers to other anticancer agents (Mirza *et al.*, 2010)

In addition, a growing list of non-nucleoside based DNA methylation inhibitors are becoming available that inhibit DNA methylation by binding directly to the catalytic region of the DNMT, without first incorporating into DNA (Schuebel and Baylin, 2005, Chuang *et al.*, 2005). However, very few non-nucleoside inhibitors have yet made it into clinical trials (Cai *et al.*, 2011).

2.6. HPV and DNA Methylation

HPV16 E7 has been shown to bind directly to DNMT1 (Burgers *et al.*, 2007). This is significant as in addition to the transcriptional modulation associated with DNA methylation, DNMT1 activity may also be intrinsically associated with carcinogenesis. Elevated DNMT1 activity effectively induces the transformation of mouse fibroblasts (Wu *et al.*, 1993). Elevated DNMT1 activity and mRNA levels have also been reported in various cancers (e.g. intestinal, colon and lung cancer) and inhibition of elevated DNMT1 is associated with suppression of tumourigenesis (Robertson, 2001, el-Deiry *et al.*, 1991, Laird *et al.*, 1995, Belinsky *et al.*, 1996). Given the role of DNMT1 in the methylation of newly replicated DNA, its association with the maintenance of chromatin structure and its implication in carcinogenesis, it is interesting to note the observed interactions between cellular DNMT1 and HPV E7. HPV16 E7 binds directly to DNMT1 which is associated with an increased methyltransferase activity *in vitro* (Burgers *et al.*, 2007). The action of E7 on DNMT1 is a potential mechanism of cellular transformation in addition to E7's role in pRb degradation and may be implicated in the aberrant cellular CpG

methylation patterns observed for HPV induced cancers (Burgers *et al.*, 2007, Robertson, 2001).

As per the human genome, the HPV genome also presents a low CpG frequency; the HPV18 genome for instance is 7,857 bp in length and the expected CpG frequency is about 400 per genome, however the actual occurrence is only 172 (Badal *et al.*, 2004). Similarly, the HPV16 genome at 7,904 bp contains only 111 CpG dinucleotides (GenBank ID NC001526.1) suggesting that the evolutionary consequences of a genome susceptible to methylation are similar for both human and HPV genomes.

Transcriptional regulation by methylation is an interesting concept for HPV biology as it provides another potential mechanism of HPV gene regulation. Both integrated and episomal HPV16 genomes are known to be methylated by the cellular methylation machinery and genomes that have been methylated *in vitro* are transcriptionally silenced, suggesting a potential role of methylation in viral gene regulation and/or a host defence mechanism acting to silence foreign DNAs (Badal *et al.*, 2003, Rosl *et al.*, 1993). In keeping with *de novo* methylation as a defence against foreign DNAs, recombination of HPV DNA into cellular chromosomes appears to incite DNA methylation whilst episomal forms of HPV and repetitive DNA sequences are also regarded as sites for DNA methylation (Yoder *et al.*, 1997, Badal *et al.*, 2003).

The methylation status of the HPV16 genome has been assessed in several studies and is seen to vary with several characteristics of the host cell/sample, including cellular differentiation, HPV integration and disease grade. The relationship between HPV methylation and a variety of disease characteristics has also been briefly considered from the perspective of biomarkers. Various techniques have been used to assess DNA methylation, including methylation specific PCR, bisulfite sequencing, methylation sensitive restriction enzymes, pyrosequencing and several others.

2.6.1. HPV Methylation and Cellular Differentiation

Investigations utilising laser capture microdissection showed that the methylation status of HPV16 correlated with the differentiation status of the host cell. HPV16 genomes (including the LCR) were typically hypomethylated in well differentiated cells and hypermethylated in less differentiated cells. The LCR in undifferentiated cells also commonly contained methylated E2BSs (Kalantari *et al.*, 2008a, Vinokurova and von Knebel Doeberitz, 2011, Kim *et al.*, 2003).

2.6.2. HPV Methylation and Disease Grade

Several studies have investigated the relationship between HPV methylation and cervical disease, and have established that patterns and levels of HPV16 DNA methylation vary with disease grade. HPV16 in cervical cancer and precancerous CIN lesions have a different methylation state than HPV16 in lower grade disease, this is especially apparent in the *L1* ORF and the LCR (Sun *et al.*, 2011, Fernandez and Esteller, 2010). The HPV16 genome is typically observed to become more methylated with progression from low grade disease, to precancerous cervical lesions through to cervical carcinomas (Kalantari *et al.*, 2004, Fernandez *et al.*, 2009, Sun *et al.*, 2011, Kalantari *et al.*, 2009) although one study observed progressive hypomethylation (Badal *et al.*, 2003).

The changes in methylation observed are specific to regions of the HPV genome. The HPV LCR may remain hypomethylated with disease progression, however this is contentious. Several studies have reported that the HPV16 LCR remained, or became hypomethylated during progression to high grade disease (Badal *et al.*, 2003, Hublarova *et al.*, 2009, Patel *et al.*, 2012, Piyathilake *et al.*, 2011) whilst others reported LCR hypermethylation with high grade disease (Kalantari *et al.*, 2004, Bhattacharjee and Sengupta, 2006a, Ding *et al.*, 2009, Brandsma *et al.*, 2009, Hong *et al.*, 2008). Similarly varied is the reported methylation status of the HPV16 LCR in HPV16 positive, but very low grade disease (i.e. LSIL/normal cytology/no CIN samples) where both LCR hypermethylation and hypomethylation have been reported (Hong *et al.*, 2008, Ding *et al.*, 2009, Fernandez *et al.*, 2009, Bhattacharjee and Sengupta, 2006a, Badal *et al.*, 2003, Patel *et al.*, 2012, Kalantari *et al.*, 2004). Despite the variety of results, studies agree that HPV methylation is associated with disease grade and thus may be clinically useful.

Each of the studies considers DNA methylation and disease grade from a different perspective, targeting different CpGs and using a variety of methylation assessment methods. Bisulfite sequencing can be performed non-quantitatively or semi-quantitatively by the cloning of PCR products prior to sequencing. The presence/absence of a cloning step might have a profound effect on the results (Badal *et al.*, 2003, Fernandez *et al.*, 2009, Kalantari *et al.*, 2004). Recently, fewer bisulfite sequencing studies have been published without the cloning steps and the use of quantitative methods such as pyrosequencing has become more common.

2.6.3. Integration and Methylation

Whilst the methylation and integration status of several HPV-associated cell lines are quite widely reported, there have been relatively few studies that have assessed both DNA methylation and HPV integration *in vivo*. Methylation of the HPV genome was typically higher for clinical samples containing integrated HPV than for non-integrated samples (Fernandez *et*

al., 2009, Kalantari *et al.*, 2008a, Kalantari *et al.*, 2010). Fernandez *et al* considered integration and methylation *in vitro* and observed hypermethylation of most of the HPV genome (LCR, *E6*, *E7*, *E5* and *L1* excluded) to be associated with HPV integrating into the host genome of primary foreskin keratinocytes (Fernandez *et al.*, 2009).

Newly integrated HPV DNAs could be targeted by host methylation mechanisms, in order to maintain a transformed phenotype the continued expression of *E6* and *E7* is required. As such many regions (including the LCR) of single insertion integrants such as those observed in the SiHa cell line often lack methylated CpGs (Kalantari *et al.*, 2008a, Kalantari *et al.*, 2004, Badal *et al.*, 2003, Fernandez *et al.*, 2009). HPV genomes inserted as concatemers however are frequently heavily methylated. The cell line CaSki for example contains several hundred HPV16 genomes inserted as concatemers at various host genome loci, and yet all but one is repressed by hypermethylation of the LCR which may reflect properties of the flanking human DNA (Kalantari *et al.*, 2008a, Van Tine *et al.*, 2004b, Kalantari *et al.*, 2004, Fernandez *et al.*, 2009).

In an early *in vitro* study using the precancerous cervical disease cell line W12, Kalantari *et al* observed a relationship between HPV methylation and integration. High copy number integrants were typically hypermethylated in the regions tested whilst single copy integrants were hypomethylated (Kalantari *et al.*, 2008a). This data formed the basis of a later study to identify a relationship between *L1* methylation, integration and disease grade in a clinical cohort. A relationship was observed however there was considerable variation among the samples (Kalantari *et al.*, 2010).

There are not many studies that have considered both HPV integration and methylation, however *de novo* methylation of integrated DNAs other than HPV are well known. Upon integration, foreign DNAs including adenoviral, plasmid and bacteriophage DNAs are hypermethylated which may be a cellular defence mechanism to repress foreign DNA elements (Sutter *et al.*, 1978, Orend *et al.*, 1995, Heller *et al.*, 1995, Doerfler, 2008). If a more robust relationship between HPV methylation and integration is established, human cells actively targeting foreign DNA elements for silencing by DNA methylation might be a suitable explanation (Doerfler *et al.*, 2001, Doerfler, 2011). The cell differentiation status also appears to have minimal affect on the methylation status of the *L1* and LCR of integrated HPV16, indicating a different methylation strategy for integrated HPV genomes (Kalantari *et al.*, 2008a).

2.6.4. Methylation of Other HPV Types and Disease Sites

As well as HPV16 methylation in cervical samples, a small number of studies have considered methylation of other HPVs and HPVs in other anogenital sites. There was good agreement between the methylation studies in cervical disease, and those concerning penile carcinoma (Kalantari *et al.*, 2008b), AIN and oral SCC (Balderas-Loaeza *et al.*, 2007, Kalantari *et al.*, 2008b, Wiley *et al.*, 2005). There was also similarity between the results of HPV16 methylation in cervical lesions, and those for HPV18 in cervical lesions (Badal *et al.*, 2004, Kalantari *et al.*, 2008b, Turan *et al.*, 2007, Turan *et al.*, 2006, Fernandez *et al.*, 2009).

The relationship between disease severity and hypermethylation is not specific to HPV. The dsDNA tumour causing viruses HBV (Hepatitis B Virus) and EBV (Epstein Barr Virus) also undergo progressive hypermethylation with increasing disease severity (Fernandez *et al.*, 2009). Both EBV and HBV have a latency period, and whilst HPV has no known latency stage, long term persistence is a feature of HPV that is associated with cervical disease. Progressive methylation of dsDNA tumour virus genomes indicates that foreign DNA is efficiently targeted for *de novo* methylation, or that methylation is involved in the latency mechanisms (Tao and Robertson, 2003, Fernandez *et al.*, 2009, Vinokurova and von Knebel Doeberitz, 2011).

2.6.5. Consequences of HPV Methylation

The methylation state of the long control region is of particular interest due to its role in regulation of HPV gene expression and the known effects of DNA methylation on gene expression. Methylation of the LCR has been determined in multiple studies and it is usually found to be generally hypomethylated but with hypermethylation of the E2 binding sites in high grade disease/cancer (Fernandez *et al.*, 2009, Brandsma *et al.*, 2009, Sun *et al.*, 2011, Kalantari *et al.*, 2004, Kalantari *et al.*, 2009).

The LCR contains four E2BS and each E2BS contains two CpGs (Figure 1.2). Methylation of these E2BS CpGs prevents E2 binding, inhibition of E2 mediated P97 repression could potentially result in *E6* and *E7* deregulation, as required for cellular transformation and carcinogenesis (Fernandez *et al.*, 2009, Thain *et al.*, 1996). Whilst E2BS hypermethylation is likely associated with increased P97 activity, general LCR hypermethylation is likely associated with decreased P97 activity. The topological structure of the DNA may become altered through HDAC recruitment to methylated sequences, resulting in enhanced chromatin condensation (Burgers *et al.*, 2002). A compact and inaccessible chromatin structure at the LCR could deny access to the multitude of cellular transcription factors that have binding sites in the LCR, negatively affecting HPV gene expression (Spink and Laimins, 2005, Stunkel *et al.*, 2000, Ai *et al.*, 2000).

The methylation status of a series of HPV16 positive primary cervical cancers was analysed. Of the samples without *E2* deletion, the majority (6 of 8) were shown to contain hypermethylated E2BSs (Fernandez *et al.*, 2009). Nonetheless, the data set was small and was determined using the imprecise Methylation Specific PCR (MS-PCR) method. Further study may confirm a role for E2BS methylation as an alternative route of *E6* and *E7* deregulation in episomal infections, or in integrated HPV infections where disruption of the *E2* ORF has not resulted in the loss of E2 mediated P97 repression (Fernandez *et al.*, 2009). Sp1 binding to the LCR is essential for the activation of P97 (Tan *et al.*, 1994), whilst E2 cannot bind to a methylated E2BS, Sp1 binding to the P97 enhancer region is not affected by methylation (Thain *et al.*, 1996, Fernandez *et al.*, 2009, Harrington *et al.*, 1988). As such, selective methylation of specific CpGs of the LCR might result in a variety of potential outcomes, i.e. P97 activity by specific CpG methylation or P97 repression due to conformational changes of the chromatin structure.

Even within the LCR, CpGs appear to be differentially methylated. For instance, the various E2BSs are differentially methylated, as are CpGs within the LCR but outside of the E2BSs (Vinokurova and von Knebel Doeberitz, 2011). The various CpGs of the early enhancer and promoter are frequently methylated to varying degrees (Vinokurova and von Knebel Doeberitz, 2011, Fernandez *et al.*, 2009, Kalantari *et al.*, 2004). The establishment and maintenance of the variable methylation patterns observed could plausibly be directed by the host cell or could be manipulated by HPV (Burgers *et al.*, 2007, Vinokurova and von Knebel Doeberitz, 2011).

Little research has been conducted to relate HPV gene expression with methylation status of the LCR. That which has been performed involved *in vitro* chemical demethylation of the CaSki cell line using nucleoside analogue methylation inhibitors, but the results have been varied. One study reported an increase in gene expression for some, but not all cells as a result of demethylation (Van Tine *et al.*, 2004b) another a decrease in *E6/E7* gene expression as a result of demethylation, thought to be the result of E2BS hypomethylation and subsequent E2 trans-acting repression (Fernandez *et al.*, 2009) whilst others reported no effect (Kalantari *et al.*, 2008a). In summary, the results of such experiments were inconclusive.

HPV methylation could speculatively be either a cause of HPV gene silencing through interaction with the LCR, or a consequence of HPV oncoprotein activity. HPV oncogenes are known to influence the DNMTs and there are also suggestions that HPV methylation might be associated with persistent HPV infection in the absence of clinically relevant disease (Burgers *et al.*, 2007, Wiley *et al.*, 2005, Vinokurova and von Knebel Doeberitz, 2011). In both clinical and cell line material, where multiple integration events were present, transcription was

observed primarily from a single integrated HPV genome whilst the rest were repressed via DNA methylation (Fernandez *et al.*, 2009, Rajeevan *et al.*, 2006, Van Tine *et al.*, 2004b).

In summary, HPV methylation is varied across the genome and within the LCR and the potential effects of HPV methylation could be similarly varied. A range of sites are identified as differentially methylated in association with disease grade and integration state and whilst our understanding of HPV methylation is improving, it is far from complete.

2.7. DNA Methylation Study Techniques

There are a range of methods by which to study the quantity and location of methylated cytosines within a region of DNA. These methods include techniques based on bisulfite conversion, methylation sensitive restriction enzyme digestion and affinity purification of methylated DNA (Zilberman and Henikoff, 2007). Methylation sensitive restriction mapping of various sized DNAs has provided a means by which to assess methylation status, however the utilisation of sodium bisulfite as a cytosine deaminating agent has allowed the application of existing and novel technologies to the study of DNA methylation.

2.7.1.1. Sodium Bisulfite Conversion

The most commonly used methods for assessing DNA methylation depend on treatment of the DNA with sodium bisulfite (BS) (Figure 2.4). Cytosine nucleotides but not methylcytosine nucleotides will deaminate upon treatment with sodium bisulfite to form uracil. During PCR amplification uracil is recognised by DNA polymerases as thymine and as such, adenine residues will be incorporated onto any newly synthesised antisense DNA strand where uracil is found in the template strand. Upon successive rounds of PCR, cytosine nucleotides will be replaced by thymine nucleotides (Herman *et al.*, 1996) and the result is that conserved cytosine residues indicate methylcytosines whilst those amplified as thymines indicate unmethylated cytosines (Frommer *et al.*, 1992). Methylation status can then be assessed by a variety of analyses.
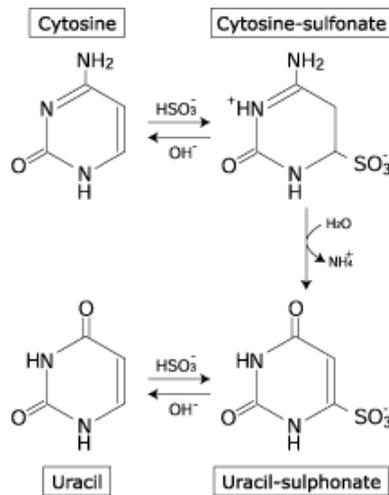
Figure 2.4: Bisulfite conversion of an unmethylated cytosine to uracil.

Methylated cytosines are resistant to this conversion.

2.7.1.2. *Methylation Specific PCR*

MS-PCR utilises the annealing specificity of two different primer sets to selectively amplify a small region of either methylated or unmethylated bisulfite converted DNA (Herman *et al.*, 1996). If the primer set specific to the methylated sequence can anneal and a PCR product is produced, the area is considered methylated. Conversely, if no product is produced by the methylated primers but a product is produced by a PCR with the unmethylated primers the area is unmethylated. One of the limitations of this technique is that DNAs with a heterogeneous methylation pattern will provide positive results for both the methylated and unmethylated primer sets. The limitations of MS-PCR on heterogeneous DNA samples can be overcome by the use of a cloning step after bisulfite treatment and prior to MS-PCR on numerous cloned DNAs. This labour intensive step allows MS-PCR to produce semi-quantitative methylation data and is known as quantitative MS-PCR or qMS-PCR.

qMS-PCR is also used to describe the use of methylation specific TaqMan probes on non-methylation-specific PCR products. TaqMan probes utilise the 5'-3' exonuclease activity of the Taq DNA polymerase to degrade bound probes and release a detectable marker (Holland *et al.*, 1991). Methylated and unmethylated TaqMan probes labelled with different fluorophores are used to probe PCRs of BS treated DNAs, if binding is achieved, the fluorophore is released. Comparing the relative intensities of the methylated and unmethylated specific fluorophore allows quantitative assessment of the methylation state (Eads *et al.*, 2000).

2.7.1.3.   *Bisulfite Sequencing*

Bisulfite sequencing is a commonly used technique that utilises conventional Sanger sequencing to determine the sequence, and therefore the methylation state of BS converted DNA. When bisulfite sequencing is used on a heterogeneous DNA mix (i.e. DNAs with a mix of methylation states), the sequencing results will contain both T and C peaks for a single position and only the highest chromatogram peak is reported. To combat this issue, cloning of the PCR products into a bacterial vector can be performed, individual clones grown and the vectors containing the cloned PCR products sequenced separately. This allows for semi-quantitative data to be produced and removes the weak banding peaks produced by partially methylated CpGs that are found in heterogeneous DNAs. The advantages of bisulfite sequencing are that the techniques are firmly established, widely used, the equipment required is widely available and the sequencing reads can also be several hundred bases long. The disadvantages are that the cloning steps are time consuming and laborious, the data is semi-quantitative at best (e.g. Six out of ten clones were methylated equals 60% methylation) and in order to get an accurate estimate of methylation a large number of clones are required, thus increasing the cost of reagents and labour.

2.7.1.4.   *Pyrosequencing*

Pyrosequencing is a real-time sequencing technique whereby short sequences can be sequenced rapidly and quantitatively (Ronaghi *et al.*, 1998). Figure 2.5 is a schematic representation of the pyrosequencing reaction. It is a sequencing by synthesis method that requires the iterative addition of each dNTP, if the dNTPs added are incorporated then light is produced by an enzymic reaction utilising luciferase (Ronaghi *et al.*, 1998). The amount of light produced is proportional to the amount of dNTP incorporated. Automated devices are now used to control the addition of the dNTPs, the detection of light signals and the analysis of the pyrograms. Whilst pyrosequencing can be used to sequence completely unknown sequences, it is more suited for detecting and quantifying small changes in an otherwise known sequence as is required during SNP and DNA methylation analysis. Pyrosequencing is also functionally limited to smaller sequencing reads and is of limited use in whole genome studies where the lower throughput and short read length are undesirable (Ronaghi, 2001).

Pyrosequencing is well suited for use in methylation studies as it provides quantitative methylation data on heterogeneous DNAs without the need for a cloning step. Recently, the cost of equipment and reagents has decreased considerably, and smaller sections of DNA can be sequenced very efficiently and with greater sensitivity than conventional Sanger sequencing. Compared to bisulfite sequencing with cloning, the cost of pyrosequencing is now very competitive.
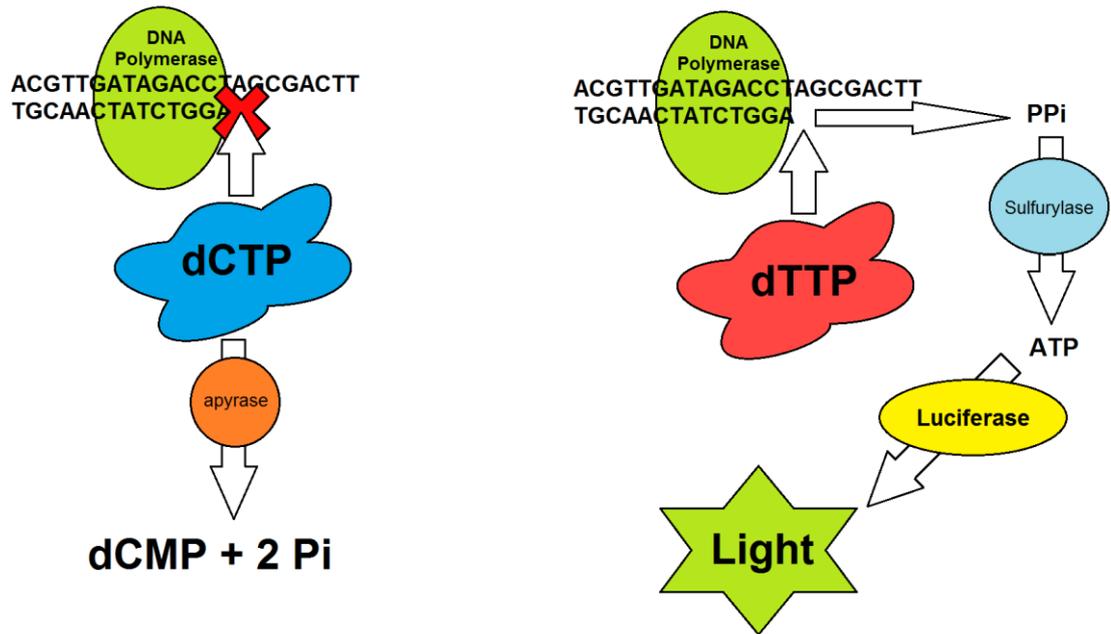
Figure 2.5: Schematic diagram of the biochemistry of pyrosequencing.

If the added dNTP does not correspond to the next base in the extending sequencing polymer then it is hydrolysed by apyrase to form a nucleoside monophosphate plus inorganic phosphate (Smith *et al.*, 2002, Ronaghi *et al.*, 1998). If the dNTP is incorporated by DNA polymerase then the diphosphate (PPi) released during incorporation is used by sulfurylase to generate ATP, this ATP is subsequently used by luciferase to generate light. The amount of light produced is proportional to the number of dNTPs incorporated (Ronaghi *et al.*, 1998).

2.7.1.5.   *Miscellaneous Methylation Assays*

Less frequently used assays include some based on mass spectrometry (Ehrich *et al.*, 2005) and methylation sensitive high resolution melting (Wojdacz and Dobrovic, 2007). Mass spectrometry is a novel approach that involves first producing an RNA transcript of the bisulfite treated/non-treated DNAs to be examined, followed by base specific RNase A cleavage at cytosine and uracil ribonucleotides (Ehrich *et al.*, 2005). Varying cleavage patterns translate into a differing mass spectrum that are representative of the CpG methylation pattern of the source DNA (Ehrich *et al.*, 2005).

Methylation Sensitive High Resolution Melting (MS-HRM) utilises bisulfite induced sequence variation in order to determine the degree of DNA methylation (Wojdacz and Dobrovic, 2007). The temperature at which dsDNA melts to ssDNA is largely dependent on the length and GC content of the DNA sequence. The GC content of BS treated DNA is altered by the level of DNA

methylation such that methylated and unmethylated sequences can be differentiated by comparing melting temperatures (Wojdacz and Dobrovic, 2007).

2.8. Biomarkers

A biological marker (biomarker) is DNA, RNA, protein, protein modification or a metabolite that reflects a biological state. It is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacologic responses to a therapeutic intervention (Nature Biotechology, 2010).

Several types of biomarker have been defined (Cancer Research UK, 2012b). Diagnostic biomarkers indicate the type or stage of cancer. Predictive biomarkers are used to identify subpopulations of patients who are most likely to respond to a given therapy. Prognostic biomarkers indicate the likely course of disease and are used to estimate patient outcome. Screening/early detection biomarkers aid in identifying disease at an early stage in a screening environment. In the context of cervical disease, cytological screening for cervical cancer can be considered a screening biomarker. While a prognostic biomarker might for example, be used to identify which patients with low grade disease are likely to progress to high grade disease, regress to a disease free state or persist as low grade disease (Cancer Research UK, 2012b).

2.8.1.1. How Biomarkers are Developed

Guidelines in the form of a "roadmap" pertaining to the development of screening, diagnostic and predictive/prognostic biomarkers have been published (Cancer Research UK, 2012d, Cancer Research UK, 2012e, Cancer Research UK, 2012c). These roadmaps were produced in order to improve the quality of biomarker studies and are presented in Appendix I. The early stages of development for each type of biomarker are very similar and several criteria in the "rationale" section should be met before proceeding. Next, another series of criteria should be met prior to proceeding further with marker development. A prospective biomarker study should be mindful of good development practice when designing the study. For this study, two classes of biomarker have relevance; these are prognostic biomarkers and screening biomarkers (Appendix I).

2.8.2. DNA Methylation as a Biomarker

The 2008 UK introduction of an effective prophylactic HPV vaccine is likely to reduce the number of cytologically positive cervical samples and reduce the positive predictive value of cytological screening. HPV testing is more sensitive for detecting CIN than cytology and will be used in the UK, initially for triage of women with low grade cytological abnormalities and for test of cure following treatment for CIN (Cuzick *et al.*, 2006). HPV testing has a greater negative predictive value, might allow use of self sampling (Katki *et al.*, 2011, Brink *et al.*, 2006) and eventually, HPV testing is likely to replace primary cytological screening in the UK. Nonetheless, HPV testing is considerably less specific than cytology and HPV testing alone would identify many transient infections and HPV infections with a low risk of progression (Wentzensen *et al.*, 2009). As such, molecular markers of CIN and cervical cancer for use in conjunction with HPV testing would greatly improve the positive predictive value of HPV screening, may allow the identification of lesions likely to progress, and thus may improve the specificity of HPV testing (Nephew, 2009, Wentzensen *et al.*, 2009). The ideal potential screening biomarkers are cellular and molecular characteristics associated with high grade cellular dysplasia and carcinoma that are not associated with low grade lesions and non-infected tissues.

Aberrant genomic DNA methylation has recently been the focus of much research as a potential molecular biomarker; it is a heritable marker that is present in 1-5% of all CpG islands in human tumours, is detectable and quantifiable and may be directly implicated in the onset of carcinogenesis (Nephew, 2009). DNA methylation shows considerable potential as a biomarker for several cancers. Lung cancer has been particularly well studied. A range of gene promoters are specifically hypermethylated in lung cancer, several of which have been further studied as early detection, predictive and prognostic biomarkers (e.g. *RASSF1*, *CDKN2A*, *CYGB*, *EDNRB*, *RARbeta*, *MGMT*, *APC*, *DAPK*, *FHIT* etc) (Liloglou *et al.*, 2012). Many of these genes have also been identified for other tumour sites, for example melanoma (*RASSF1*, *RARbeta*) oesophageal (*APC*) and breast cancer (*RASSF1*, *APC*) biomarker studies (Muller *et al.*, 2003, Mori *et al.*, 2005, Kawakami *et al.*, 2000).

Numerous studies have focussed on both human and HPV DNA methylation as a marker for cervical cancer. The most promising human methylation markers appear to be *CADM1* (cell adhesion molecule 1) and *Mal* (T-lymphocyte maturation associated protein). *CADM1* and *Mal* are associated with later and earlier events in HPV-associated carcinogenesis respectively and both show promoter hypermethylation in cervical cancer (Steenbergen *et al.*, 2004, Overmeer *et al.*, 2009).

2.8.2.1.   HPV Methylation as a Biomarker

Whilst many studies have aimed to identify methylation differences between different grades of cervical disease (occasionally with conflicting results), fewer studies have attempted to evaluate the potential of HPV methylation as a biomarker of cervical disease. Brandsma *et al* identified patterns of DNA methylation that were associated with various disease grades. HPV genomic hypomethylation was associated with lower grade cervical disease (normal cytology and negative histology in most instances) and hypermethylation at several regions (*E5*, *L1* and *L2*) was associated with higher grade disease in most instances however the sample size was very small (n=13) and there was considerable variation (Brandsma *et al.*, 2009).

Mirabello *et al* used a HPV genome wide approach on a large sample set (n=154) and identified numerous CpGs where methylation was significantly different between the disease grades studied. These CpGs were located primarily in the *L1*, *L2* and *E2*/*E4* regions although CpGs throughout other areas of the genome were also significantly associated with disease grade. *L1* and *L2* methylation was deemed especially predictive of disease grade (Mirabello *et al.*, 2012).

Kalantari *et al* considered the potential of *L1* methylation as a biomarker of cervical disease. There was an association between *L1* methylation and integration, and the authors considered that integration could be a marker of disease progression (Kalantari *et al.*, 2010). Nonetheless, the study had a small sample size (n=21) and there was a degree of variation.

As previously discussed, DNA methylation of other HPV and HPV in other anogenital sites was largely concordant with HPV16 methylation in cervical specimens. This increases the likelihood of a cervical disease biomarker developed for HPV16 and CIN having wider clinical utility.

2.8.2.2.   Biomarker Investigation

The earliest stages of biomarker development (Cancer Research UK, 2012d, Cancer Research UK, 2012e) concern assay design and the evaluation of variation within the prospective sample cohort (Appendix I). This investigation was mindful of these requirements during study design and method development. Several DNA methylation assays were trialled, methylation at various regions of the HPV genome was assessed and both the basic biology of HPV and the potential clinical utility of knowledge of HPV methylation were considered throughout.

2.9. Chapter Introductions

In order to study the biology and potential clinical utility of HPV DNA methylation, four interrelated studies were devised, presented here as separate chapters. Method development is also presented as a short chapter. Aspects of each chapter are relevant to other chapters and the results of earlier studies influenced both the hypotheses and design of later studies.

2.9.1.1. Method Development

The first steps in this investigation were the development of assays to determine HPV methylation status and gene expression status of HPV. The major steps involved in method development of qPCR and pyrosequencing are detailed and discussed in this chapter.

2.9.1.2. HPV Methylation and Disease Grade (the NSC Study)

The relationship between HPV methylation and disease grade was determined. Three sample series were used, these were: Liquid Based Cytology (LBC) material displaying normal cytology, LBC material showing severe dyskaryosis, and formalin fixed cervical cancers. This study was primarily devised to determine the differences in DNA methylation between various stages of cervical disease and consider the potential clinical utility that a knowledge of HPV DNA methylation status may provide. In addition, this study constituted a means by which to facilitate further method development and compare data produced by pyrosequencing to published methylation data.

2.9.1.3. HPV Methylation and Clinical Outcome (the CRISP Study)

There is a distinct lack of longitudinal HPV methylation data in the literature. The CRISP study was proposed in order to relate HPV methylation of low grade cytological abnormalities with a variety of clinical outcomes 6 months later. Three groups of samples were extracted from a large cohort. The first group represented those patients that had cleared a HPV16 infection during the course of 6 months. The other two groups represented patients that had low grade histology at 6 months, or high grade histology at 6 months. This study allowed further elaboration of the potential clinical utility of HPV DNA methylation, and also allowed changes in HPV methylation over time to be considered.

2.9.1.4. DNA Methylation, Integration and Gene Expression in vitro (the Cell Lines Study)

The cell lines study included methylation assessment of a variety of HPV-associated human cell lines and was primarily designed to make use of a well characterised series of VIN/VAIN short term culture cell lines. The use of cell lines allowed the assessment of changes in methylation over time (i.e. with passages) and an *in vitro* assessment of the relationship between DNA methylation, integration and HPV gene expression. In addition to the short term culture cell

lines, a W12 cell line which underwent integration whilst in passage allowed the relationship between methylation and integration to be further investigated. This study examined the relationship between methylation, integration and gene expression *in vivo*.

2.9.1.5. DNA Methylation, Integration and Gene Expression in vivo (the VR Study)

The VR study utilised a series of predominantly VIN2/3 biopsies for which both DNA and RNA were available. HPV methylation, integration and gene expression assessment were performed and the relationship between these three characteristics was explored. Hypotheses generated in previous studies were tested here and differences/similarities between the studies were explored further. The VR study also permitted characterisation of the methylation status of the VIN2/3 series from the perspective of biomarker development and HPV type distribution.

2.9.1.6. General Discussion

Several aspects of the previous 5 chapters had broader relevance or commonalities among the studies. In order to avoid repetition, topics common to several chapters are discussed in the general discussion.

Chapter 3. Methods and Materials

3.1. Methods

3.1.1. Nucleic Acid Extraction

Several methods of nucleic acid extraction from VIN biopsy material were used during this investigation.

3.1.1.1. TRIzol DNA and RNA Extraction

TRIzol Reagent (Invitrogen, Paisley, UK) extraction of DNA and RNA was performed for several VIN biopsy samples before an alternative VIN biopsy protocol was used. Biopsies were cut into pieces of 1 mm^3 or smaller using a disposable sterile scalpel and placed into 1 ml of TRIzol. Biopsy material was then disrupted using a TissueRuptor (Qiagen, Hilden, Germany) at full speed for 20 s using disposable probe tips. Samples were incubated at room temperature for 10 mins then 200 µl of chloroform was then added, mixed and the samples incubated for a further 5 mins at room temperature.

Sample mixes were centrifuged for 15 mins (10,000 rpm, 4 °C). The aqueous phase was used for RNA extraction whilst the organic phase and interphase were used for DNA extraction.

3.1.1.1.1. RNA Extraction

To the aqueous phase, 500 µl of isopropanol (100%) and 1 µl of glycogen (20 mg/ml) were added the tube was then vortexed and incubated at -20 °C overnight before being centrifuged for 10 mins (8,000 rpm, 4 °C). The supernatant was discarded and 1 ml of 70% ethanol was added, vortexed and centrifuged for 10 mins (8,000 rpm, 4 °C). The supernatant was removed again before the pellet was air dried and resuspended in 28 µl of water. RNAs were later purified further using a NucleoSpin RNA Cleanup XS Kit (Machery-Nagel, Duren, Germany) performed according to manufacturer's instructions and eluted into 30 µl of RNase free water.

3.1.1.1.2. DNA Extraction

After all of the aqueous phase was removed, 0.3 ml of 100% ethanol was added to the organic phase and interphase; this was mixed by inversion and incubated at room temperature for 5 mins. Samples were then centrifuged for 10 mins (4,000 rpm, 4 °C). The supernatant was discarded and 1 ml of 0.1 M Sodium citrate in a 10% ethanol solution was added and incubated at room temperature for 30 mins with periodic mixing. The samples were centrifuged and the Sodium citrate wash and centrifugation steps were repeated. The supernatant was removed and the DNA pellet was allowed to air dry for 15 mins before being resuspended in 250 µl of

NaOH (8 mM) and 21.5 µl of HEPES (0.1 M). The mix was centrifuged at for 10 mins (10,000 rpm) to collect cellular debris and the supernatant was removed for further purification.

DNAs were further purified using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) protocol. To the 271.5 µl sample, 180 µl of ATL buffer was added and the DNA purification from animal tissues, spin column protocol followed precisely from step 2 onwards (Handbook July/2006, p.29). The protocol involved binding of DNA to a spin column membrane, several washing steps and final elution into 2 x 100 µl of AE buffer.

3.1.1.2. VIN RNA/DNA Extraction Protocol

This protocol was adopted part way through the investigation as it was shown to produce higher quantity/quality DNA than TRIzol extraction during a small scale nucleic acid yield test. Two stages of nucleic acid extraction were performed, one stage using the RNeasy Fibrous Tissue Kit (Qiagen, Hilden, Germany) for RNA extraction and the other using the QIAamp DNA extraction kit (Qiagen, Hilden, Germany) for DNA extraction. Initial processing was performed using the RLT buffer from the RNeasy Kit.

The biopsies were placed into locking microcentrifuge tubes and rinsed with 1 ml of sterile water. Tissue pieces larger than 3 mm in diameter were cut into pieces using a sterile scalpel to ensure efficient disruption. The water was aspirated and 360 µl buffer RLT (with 10 µl/ml β-mercaptoethanol) was added to each tube. Biopsy samples were then disrupted using a TissueRuptor (Qiagen, Hilden, Germany) at full speed for 20 s using disposable probe tips.

708 µl of RNase free water and 20 µl of proteinase K were added to the tissue suspension. This was followed by incubation at 55 °C for 10 mins before centrifugation for 3 mins (13,000 rpm) to form a pellet. A volume of 650 µl of the supernatant was transferred to a labelled tube and the remaining supernatant/cell pellet were resuspended and returned for further incubation at 55 °C. The protocol then split into two stages, one each for the elution of RNA and DNA.

3.1.1.2.1. RNA Extraction

The RNA extraction stage used the 650 µl of supernatant processed previously. To the cleared lysate, 325 µl of ethanol (100%) was added and the solution was repeat pipetted to mix. The remainder of the extraction protocol was performed as per the instructions of the RNeasy Fibrous Tissue Mini Kit (Qiagen, Hilden, Germany), from step 10 onwards (Handbook Oct/2010, p.17). This consisted of washing the bound RNA with a series of buffer solutions and eluting in 2 x 50 µl of water.

3.1.1.2.2. DNA Extraction

Using the remaining cell pellet and lysate, DNA was extracted using the QIAamp DNA kit (Qiagen, Hilden, Germany). To each sample, 350 µl Buffer AL was added, mixed and incubated at 70 °C for 10 mins; 350 µl ethanol (100%) was then added to each sample. Up to 600 µl of supernatant was added to each spin column, which as then centrifuged at 13,000 rpm for one minute. The flow through was discarded and the remaining sample was added to the spin column and centrifuged. A new collection tube was used and the instructions of the QIAamp DNA kit, tissue extraction spin protocol were followed from step 8 onwards (Handbook April/2010, p.36). This consisted of washing the bound DNA with a series of buffer solutions and eluting in 2 x 100 µl of AE buffer.

3.1.1.3. *Qiagen AllPrep DNA/RNA Extraction Kit*

DNA and RNA were extracted from cell line material using the AllPrep DNA/RNA Mini kit (Qiagen, Hilden, German) according the manufacturer's instructions (animal cells protocol). Briefly, the protocol consisted of lysing the cells within the culture vessel. The lysate was then homogenised using a 20 gauge needle and a disposable syringe. DNA was then bound to a spin column, washed and eluted into 2 x 50 µl of Buffer EB. RNA was bound to a second column, washed and eluted into 2 x 50 µl of RNase free water.

3.1.1.4. *QIAamp MinElute Media Kit*

DNA was extracted from LBC material using the QIAamp MinElute Media kit (Qiagen, Hilden, Germany). LBC cell pellets were resuspended in 500 µl of TRIS (10 mM, pH 7.4) and a 100 µl aliquot was taken for DNA extraction. The remaining material was stored at -80 °C. To the sample, 150 µl of TRIS (10 mM, pH 7.4), 80 µl of ATL buffer and 20 µl of proteinase K were added and the mix was then incubated at 56 °C for 30 mins and periodically vortexed. Carrier RNA was prepared in AVE/AL buffer according to the instructions, 250 µl of AL (with carrier RNA) was added to the sample and after vortexing, incubated at 70 °C for 15 mins.

To each sample, 300 µl of ethanol (100%) was added. The samples were then vortexed and incubated at room temperature for 5 mins. To each QIAamp MinElute column, 500 µl of lysate was added and centrifuged (8,000 rpm, 3 mins). The eluate was discarded and the remaining lysate was added and centrifuged. Bound DNAs were washed with 650 µl of AW2 and centrifuged (8,000 rpm, 3 mins) and then washed with 650 µl of ethanol and centrifuged (8,000 rpm, 3 mins). The collection tubes were emptied and the membranes were dried by centrifugation (13,000 rpm, 3 mins) followed by placing the column in a clean tube DNA/RNA and heating at 56 °C for 3 mins to evaporate residual ethanol. DNA was then eluted into 100 µl of AVE elution buffer (13,000 rpm, 3 mins).

3.1.2. General Techniques

3.1.2.1. Nucleic Acid Concentration

DNA and RNA concentration were quantified using a ND-1000 spectrophotometer (NanoDrop products, Wilmington, DE, USA). DNA/RNA (1.5 µl) was tested against a blank of DNA/RNA free elution buffer/water.

3.1.2.2. RNA Handling

RNA is highly susceptible to enzymatic degradation. Several aspects of this study included RNA work (qPCR, nucleic acid extraction, APOT). All RNA work was performed within a HEPA filtered PCR cabinet treated with 30 mins UV light and all surfaces and equipment were thoroughly treated with RNaseZAP (Sigma-Aldrich Company Ltd, Dorset, UK). Separate sets of reagents/consumables were used for RNA work and where applicable, equipment/materials were also treated with UV light. RNase/DNase free microcentrifuge tubes and pipette tips and DEPC treated or RNase free water were used throughout.

3.1.2.3. DNA Handling

DNA is susceptible to enzymatic degradation. PCR and DNA work were conducted in a HEPA filtered PCR cabinet sterilised between uses with 15 mins UV light. All water used for DNA/PCR was DNA grade water that had been further sterilised with UV light. DNase free tubes and pipette tips were used throughout.

3.1.2.4. PCR

PCR and gradient PCR were performed using Techne TC-412 and Techne TC-512 thermal cyclers (Bibby Scientific Ltd, Staffs, UK).

3.1.3. HPV Typing: Greiner Bio-One PapilloCheck

The PapilloCheck assay (Greiner Bio-One GmbH, Frickenhausen, Germany) allows the simultaneous identification of 24 HPV types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73, 82, 6, 11, 40, 42, 43, and 44). The assay uses multiplex PCR with fluorescent primers to amplify a 350 bp fragment of the HPV *E1* gene. HPV type specific sequences are then detected using a microarray consisting of 5 replicate spots of 28 probes. Hybridisation of PCR products to probes was then assessed using the CheckScanner array reader. Valid results were those which passed each of the numerous PapilloCheck internal controls (including DNA adequacy, sample inhibition, and hybridization). The assay includes amplification of the human *ADAT1* gene in order to identify false negatives. The manufacturer's instructions were followed

throughout. For each multiplex PCR reaction, 20 μl of PCR master mix (Table 3.1) was added to a 5 μl of sample DNA and run according to the PCR timing/temperatures in Table 3.2.

	1 reaction Volume (μl)	13 reactions (1 chip) Volume (μl)
PapilloCheck Master Mix	19.8	257.4
HotStarTaq Polymerase (5 U/μl)	0.2	2.6
Uracil-N-Glycosylase*	1	13
Total volume	21	273

Table 3.1: Greiner PapilloCheck PCR master mix recipe.

Included are the volumes required for one reaction, or a chip of (n=13) reactions. * Uracil-N-Glycosylase was first diluted 1/200 with sterile water.

Temperature (°C)	Time (s)	Cycles
95	900	1
95	30	
55	25	40
72	45	
95	30	15
72	45	

Table 3.2: Greiner PapilloCheck PCR timing and temperature settings.

Following PCR, PCR products were hybridised to DNA arrays and washed. Wash solution was prepared using 140 ml of distilled water, 14 ml of PapilloCheck buffer A and 1.75 ml of buffer B, mixed and then divided equally among three 50 ml falcon tubes labelled 1-3. Wash buffer tube 2 was heated to 50 °C before use. Five μl of PCR product was added to 30 μl of PapilloCheck hybridisation buffer. After mixing, 25 μl of the hybridisation mix was transferred to a compartment of the chip. The chip was incubated at room temperature in the PapilloCheck humidifier box for 15 mins. Three wash steps were then performed which involved vigorous agitation for 10 s in wash buffer tube 1, 60 s in tube 2, 10 s in tube 3 before drying by centrifugation.

Chips were then scanned and data analysed using the GBP CheckScanner and the CheckReport Software as per the manufacturer's instructions.

3.1.4. Quantitative Real Time Reverse Transcriptase PCR

3.1.4.1. Reverse Transcription
Prior to Reverse Transcription (RT), RNAs were DNase treated to remove residual DNA.

3.1.4.1.1. DNase Treatment
DNase treatment was performed using amplification grade DNase I (1 U/µl) and reaction buffer (Invitrogen, Paisley, UK). One µg RNA was treated using 1 µl of DNase I, 1 µl of 10x Reaction Buffer and DEPC treated water in a final volume of 10 µl. The reagents were mixed gently (DNase I is susceptible to physical denaturation) and then incubated at room temperature for 15 mins. Inactivation of DNase I was then performed by adding 1 µl of EDTA (25 mM) and incubating at 65 °C for 10 mins.

3.1.4.1.2. Reverse Transcription (cDNA synthesis)
RT was performed using random primers (hexamers) and the SuperScript III Reverse Transcriptase (RT) kit (Invitrogen, Paisley, UK). RT was performed using 1 µl of random primers (200 ng/µl), 0.5 µg of RNA (DNase treated), 1 µl of dNTP mix (10 mM) and RNase free water to 13 µl. This mix was incubated at 65 °C for 5 mins, and then kept on ice for 1 min.

The following components of the SuperScript III RT kit (Invitrogen, Paisley, UK) were then added: 4 µl of 5x First Strand Buffer, 1 µl of 0.1 M DTT (Dithiothreitol), 1 µl of RNaseOUT and 1 µl of SuperScript III RT. This was mixed gently, incubated at 25 °C for 5 mins, 50 °C for 60 mins and then inactivated at 70 °C for 15 mins. To correct for undigested genomic DNA in later qPCR steps, all reactions were performed in duplicate using H_2O instead of RT.

3.1.4.2. qPCR Reaction Protocol
Quantitative Real Time Reverse Transcriptase PCR (qRTRTPCR or qPCR) was performed using the LightCycler carousel-based qPCR system (Roche Applied Science, Mannheim, Germany). LightCycler DNA Master SYBR Green I reagent kits and LightCycler reaction capillary tubes (Roche Applied Science, Mannheim, Germany) were used. FS mix was produced by adding 10 µl of LightCycler reagent 1a to a full vial of defrosted reagent 1b. After mixing, this reagent was kept at 4 °C and not refrozen. The qPCR master mix (Table 3.3) was prepared in a cooling block.

	Primer pair		
	E2/E6/HPRT	*E7*	*TBP2*
Reagent	μl per reaction		
Forward primer (5 μM)	2	2	2
Reverse primer (5 μM)	2	2	2
FS Mix	2	2	2
MgCl$_2$ (25 mM)	1.6	2	2.4
Water	10.4	10	9.6
Template cDNA	2	2	2

Table 3.3: qPCR master mix preparation protocol.

The table represents the required reagents per reaction for each of the 5 primer pairs.

18 μl of the qPCR master mix was dispensed into the top of each glass capillary tube, and then 2 μl of the appropriate cDNA was added to each tube before a short centrifugation step (3,000 rpm, 5 s) was performed. cDNAs from the RT reaction were typically diluted 1/10 for clinical samples and 1/100 for cell lines. Water was used as a negative control and all runs included 1/100 CaSki cDNA in triplicate. To correct for undigested genomic DNA, RT negative controls were also run.

The carousel was loaded with reaction capillaries, placed into the LightCycler and the PCR reaction performed according to the timing/temperature in Table 3.4.

	Temperature (°C)	Time (s)	Cycles
Initial denaturation	95	600	1
Denaturation	95	10	
Primer annealing	58/60/62*	5	60
Extension	72	5	
	Temperature (°C)	Rate of change (°C/s)	Cycles
Meting Curve	65-95	0.1	1

Table 3.4: qPCR temperature and timing protocol.

Details are listed for both the qPCR reaction and the subsequent melting curve. *Annealing temperatures were specific to each primer pair. *E2* (58 °C), *E6/HPRT/TBP* (60 °C) and *E7* (62 °C).

3.1.4.2.1. Data Analysis

The LightCycler software's capacity for data analysis was limited. It does not allow for the analysis of data using calibration curves from more than one standard curve or from sample data run in duplicate. There were also RT negative controls which could not be accounted for in the analysis. Instead, calibration curves were produced (Chapter 4) and imported into the LightCycler software. The LightCycler software was then used to generate Ct values and express those values as CaSki cDNA concentrations (e.g. if a sample generated the same Ct as CaSki 1/10 cDNA, the calculated concentration would be 0.1).

The difference in calculated concentration between the RT positive and the RT negative samples was equal to the amount of signal attributable to cDNA, corrected for undigested genomic DNA. These values were converted back into Ct values and used for relative quantification. Relative quantification was performed with two reference genes using the PCR efficiency corrected equation described by Vandesompele *et al* (Vandesompele *et al.*, 2002). This is discussed in greater detail in the methods development chapter (Chapter 4)

3.1.4.3. Primer Sequences
The qPCR primer sequences are detailed in Table 3.5.

HPRT	
Name	Hypoxanthine guanine phosphoribosyl transferase (*HPRT*)
Sense primer	TGACACTGGCAAAACAATGCA
Antisense primer	GGTCCTTTTCACCAGCAAGCT
Amplified DNA	4982 bp (nt 133627546-133632438 of Chr X)
Amplified RNA	94 bp (nt 496-589 of M31642.1)
Reference	(Allen, Winters *et al.*, 2008)
TBP2	
Name	TATA binding protein, 2nd pair of primers (*TBP*2)
Sense primer	TCAAACCCAGAATTGTTCTCCTTAT
Antisense primer	CCTGAATCCCTTTAGAATAGGGTAGA
Amplified DNA	803 bp (nt 170880539-170881341 of Chr 6)
Amplified RNA	122 bp (nt 1128-1224 of M55654.1)
Reference	(Minner and Poumay 2008)
E6	
Sense primer	CTGCAATGTTTCAGGACCCA
Antisense primer	TCATGTATAGTTGTTTGCAGCTCTGT
Amplified DNA/RNA	80 bp (nt 99-178 of NC001526.1)
Reference	(Wang-Johanning, Lu *et al.*, 2002)
E7	
Sense primer	AAGTGTGACTCTACGCTTCGGTT
Antisense primer	GCCCATTAACAGGTCTTCCAAA
Amplified DNA/RNA	78 bp (nt 739-816 of NC001526.1)
Reference	(Wang-Johanning, Lu *et al.*, 2002)
E2	
Sense primer	AACGAAGTATCCTCTCCTGAAATTATTAG
Antisense primer	CCAAGGCGACGGCTTTG
Amplified DNA/RNA	82 bp (nt 3362-3426 of NC001526.1)
Reference	(Roberts, Ng *et al.*, 2008)

Table 3.5: qPCR primer sequences and original source for each primer pair.

The primer sequences and source of the sequences are listed for each primer pair. For the human reference genes two amplicon lengths are given because the primer pairs are junction spanning, the mRNA amplicons are shorter than the DNA amplicons. Regions amplified are accompanied by GenBank ID.

3.1.5. Pyrosequencing

3.1.5.1. Bisulfite Treatment

Treatment of DNA with Sodium bisulfite (BS) was conducted using the EZ DNA Methylation Kit D5001 (Zymo Research Corporation, CA, USA). The protocol included modifications recommended by the manufacturer to improve conversion efficiency. Kit reagents were prepared by adding 24 ml of ethanol (100%) to the 6 ml of M-Wash Buffer concentrate, and by adding 750 µl of water and 210 µl of M-Dilution Buffer to a tube of CT Conversion Reagent. CT Conversion Reagent was then vortex mixed until required.

Each reaction required 500 ng of DNA made up with H_20 to 32.5 µl. For several low concentration DNA samples less than 500 ng was used. To each sample 7.5 µl of M-Dilution Buffer was added and incubated at 42 °C. After 30 mins of incubation, 97.5 µl of prepared CT Conversion Reagent was added to each reaction and the samples were incubated at 50 °C overnight (12-16 hrs). For negative controls, water was used in place of CT Conversion Reagent.

After incubation, the samples were cooled on ice for 10 mins, loaded into Zymo-Spin columns containing 400 µl of M-Binding Buffer, mixed by inversion and centrifuged. All centrifugation steps were performed at 13,000 rpm for 30 s and collection tubes were emptied when necessary. Samples were washed by adding 100 µl of M-Wash Buffer and centrifuging the tubes. Desulphonation buffer (200 µl) was added and samples were incubated at room temperature for 15-20 mins before centrifugation. Final wash steps were performed by adding 200 µl of M-Wash Buffer to each column and centrifuging this was performed twice.

Finally, DNAs were eluted into 10 µl of M-Elution Buffer by centrifugation.

3.1.5.2. Pyrosequencing PCR

Pyrosequencing PCRs were performed in 50 µl reactions. Two µl of 1/10 diluted BS converted sample DNA was used for each PCR reaction using clinical DNAs. CaSki BS treated DNA was typically used at 1/100 dilution. ZymoTaq hot start DNA polymerase premix was used (Zymo Research Corporation, CA, USA).

For each primer pair, the PCR reaction mixes are listed in Table 3.6.

	Primer pair			
	E2	LCR1	LCR2	*L1/L2*
Reagent	Volume per 50 µl reaction (µl)			
ZymoTaq premix	25	25	25	25
F primer (5 µM)	2	2	2	2
R primer (5 µM)	2	2	2	2
Water	17	16	16	16
MgCl$_2$ (25 mM)	2	3	3	3
Template DNA	2	2	2	2
Reaction [Mg^{2+}] (mM)	2.75	3.25	3.25	3.25

Table 3.6: Pyrosequencing PCR master mix recipes.

Values in the table represent the volume required per reaction. Template DNA refers to appropriately diluted BS treated DNA. Final [Mg^{2+}] refers to the final reaction concentration of Mg^{2+} ions in mM for each master mix.

Into each well of a PCR plate, 48 µl of master mix and 2 µl of diluted BS treated DNA were pipetted. A negative control (water or BS negative controls) and a CaSki positive control were included for each primer pair. PCR was performed according to the timing and temperatures listed in Table 3.7.

	Temperature (°C)	Time (s)	Cycles
Initial denaturation	95	600	1
Denaturation	95	30	
Primer annealing	48/51/54*	45	40
Extension	72	30	
Final extension	72	600	1

Table 3.7: Pyrosequencing PCR temperature and timing protocol.

*Annealing temperatures were specific to each primer pair. *E2*/LCR1 (48 °C), LCR2 (51 °C) and *L1/L2* (54 °C).

3.1.5.3. Pyrosequencing Reaction

Most equipment and reagents used for the pyrosequencing reaction were supplied by Qiagen (Qiagen, Hilden, Germany). This included the PyroMark Q96 ID Instrument, PyroMark CpG Software v1.0.11, PSQ96 Reagent Cartridges, PSQ Plate Low (PSQ reaction plates), PyroMark Q96 Vacuum Prep Workstation, PyroMark Gold Q96 Reagents, PyroMark buffer solutions. In addition, Streptavidin Sepharose beads (30% w/v) (GE Healthcare Life Sciences, Buckinghamshire, UK) were also required.

3.1.5.3.1. Process

PCR products were diluted by adding 27 µl of PCR product to 13 µl DNA grade water. The vacuum workstation was assembled and the trays were filled with the appropriate solutions

(Figure 3.1) (1/10 diluted wash buffer, 70% ethanol, water (deionised and autoclaved), denaturation buffer).
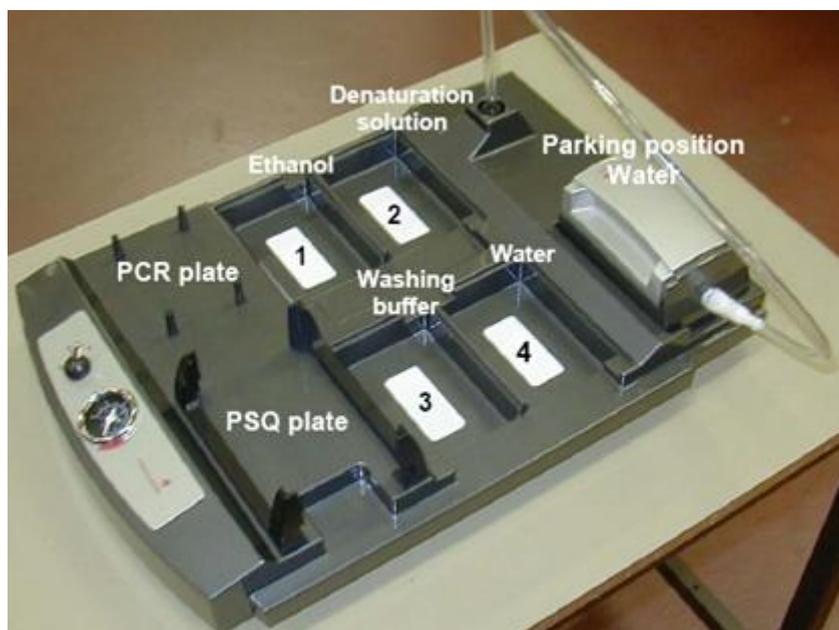
Figure 3.1: The correct tray and plate positions on the PyroMark Q96 Vacuum Prep Workstation.

The vacuum prep tool head was rinsed through with clean water for 20 s. Details of the reaction and samples were inputted into the PyroMark CpG Software.

Pyrosequencing PCR was performed using a biotin labelled primer. PCR products were immobilised by adding 1.75 μl of streptavidin sepharose bead suspension and 38.25 μl of PyroMark Binding Buffer per reaction. The mixes were then shaken on a shaking hot plate for at least 5 mins (1,400 rpm, 22 °C). Sequencing primers were made up by diluting (per reaction) 1.5 μl of 10 μM sequencing primer with 43.5 μl of PyroMark Annealing Buffer. Into each well of a PSQ reaction plate, 45 μl of sequencing primer mix was dispensed and the plate was placed into the correct compartment of the vacuum workstation (Figure 3.1).

The PCR product/sepharose bead mix was removed from the hotplate and placed into the correct position on the vacuum workstation. Within 30 s of cessation of shaking, vacuum was applied to the workstation and the sepharose beads with immobilised PCR products were captured by slowly lowering the vacuum prep tool into the PCR plate. The vacuum prep tool was then placed into each the following solution trays (in order) for 5 s each; 70% v/v ethanol solution, PyroMark Denaturation Solution, 10% v/v PyroMark Washing Buffer. The prep tool was held vertically and any residual fluid was aspirated. The vacuum switch was closed and the

prep tool was lowered into the PSQ reaction plate containing sequencing primer and agitated to release the captured PCR products. By this point, PCR products had been denatured into single stranded products and the complementary strand had been removed.

The PSQ plate containing beads and sequencing primer was heated at 80 °C for 2 mins then cooled to room temperature. The PyroMark Gold Q96 Reagent kit contained lyophilised enzyme and substrate pellets as well as dNTP mixes. Enzyme and substrate were reconstituted with the volume of water specified on the container 10 mins prior to use. The PSQ96 Reagent Cartridge was filled using the volumes specified by the pyrosequencing software and the wells specified in Figure 3.2.
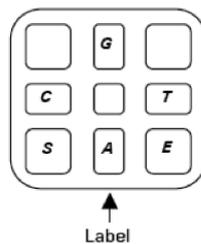
Figure 3.2: A diagram of the PSQ96 Reagent Cartridge detailing compartments of the cartridge where each reagent was loaded. The cartridge label was positioned at the front, facing the operator. Nomenclature: S = substrate mix; E = enzyme mix; G = dGTP solution; T = dTTP solution; A = dATP solution; C = dCTP solution.

The cartridge and reaction plate were secured within the PyroMark Q96 ID Instrument and the run was started. At the end of the run, the data was analysed automatically and a pyrogram was produced for each pyrosequencing reaction (e.g. Figure 3.3).
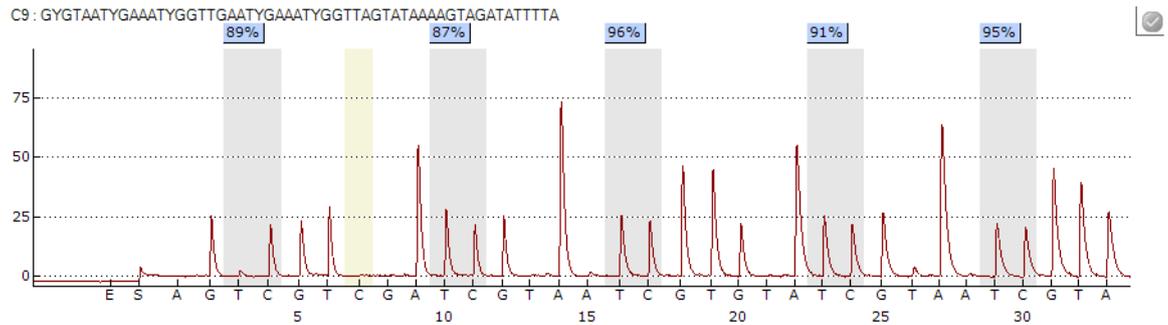
Figure 3.3: The LCR2 pyrogram for CaSki.

The string of sequence across the top is the sequence to analyse. The string of sequence across the X-axis is the dNTP dispensation order (E=enzyme, S=substrate). The Y-axis scale is Relative Light Units (RLU). The height of peaks represents the strength of the signal, here 25, 50 and 75 RLU are equal to approximately 1, 2 and 3 incorporated nucleotides respectively. Areas shaded grey represent sites of CpGs; both C and T are dispensed and the height of the peaks compared to determine DNA methylation % of that CpG. Methylation % is indicated in the coloured box above each CpG and the colour indicates the data quality (blue = good, yellow = check for acceptability and red = significant inadequacies).

Values considered by the pyrosequencing software to have failed the internal controls (indicated by a red box (Figure 3.3)) were not included in the data analysis. Due to the nature of the pyrosequencing reaction, it was more difficult to accurately determine the amount of methylation at some CpGs than it was at others. As such, for some sample cohorts there were missing values for some CpGs, some assays or even entire samples. All samples were repeated in duplicate in order to improve the likelihood of producing a complete data set and to improve the reliability of the data analysis. Failed reactions were repeated (in duplicate) at least once, data of the highest quality was selected and where two runs of equal quality (as determined by passing/failing the internal controls) existed, to avoid selection bias runs were chosen for inclusion in chronological order.

3.1.5.3.2. Clean Up

Careful clean up was especially important as blockage of the reagent cartridge and/or vacuum prep tool would result in failed reactions. Immediately after the sequencing reaction had finished, the PSQ96 Reagent Cartridge was thoroughly cleaned with sterile water.

Clean up of the vacuum workstation required all solution trays to be emptied and rinsed. The vacuum prep tool was rinsed with water, then ethanol before the head was fully aspirated of fluid and allowed to dry. All equipment was protected from the ingress of particulates by being stored in clean plastic containers in a lint free environment.

3.1.5.4. Assay Design

Pyrosequencing assays were designed within PyroMark Assay Design Software 2.0 (Qiagen, Hilden, Germany). The manufacturer's instructions were followed and primer sets were chosen based on primer length, melting temperature, GC content, GC clamp, presence of secondary structure (primer and template), and the presence of single and dinucleotide repeats. For *L1/L2*, a very limited range of alternative primers were available due to the amplimer size limits and the presence of secondary structures either side of the region of interest.

3.1.5.5. Primer Sequences

For each pyrosequencing assay, the primer sequences, regions amplified and dispensation order for sequencing are listed in Table 3.8.

E2	
Sense primer	GTGAAATTATTAGGTAGTATTTGG
Antisense primer	BTN-CAACAACTTAATAATATAACAAAAA
Sequencing primer	GTGAAATTATTAGGTAGTA
Dispensation order	ATTGCTATGATTCTGTCAGTCGATTATATAATGTCAGTCGTTTGGTGATCGAGAA TATAGTATCGATATTATGTCGA
Amplified DNA	148 bp (nt 3378-3525 of NC001526.1)
Sequenced CpGs	nt 3411, 3414, 3416, 3432, 3435, 3447, 3461, 3472
Reference	Dr T Liloglou, University of Liverpool, personal communication (Dec 7th 2010)
LCR1	
Sense primer	BTN-ATTGTATTATGTGTAATTATTGAA
Antisense primer	CCAAAAATATATACCTAACAAC
Sequencing primer	CCAAAAATATATACCTAACAAC
Dispensation order	CAGATAGTATAAATCGATATCGACAT
Amplified DNA	104 bp (nt 7611-7714 of NC001526.1)
Sequenced CpGs	nt 7691, 7679, 7673
Reference	Dr T Liloglou, University of Liverpool, personal communication (Dec 7th 2010)
LCR2	
Sense primer	GTAAAATTGTATATGGGTGT
Antisense primer	BTN-TAAAATATCTACTTTTATACTAACC
Sequencing primer	TAATTTATGTATAAAATTAAGG
Dispensation order	AGTCGTCGATCGTAATCGTGTATCGTAATCGTA
Amplified DNA	156 bp (nt 7832-83 of NC001526.1)
Sequenced CpGs	nt 31, 37, 43, 52, 58
Reference	Dr T Liloglou, University of Liverpool, personal communication (Dec 7th 2010)
L1/L2	
Sense primer	BTN-TTATTGTTGATGTAGGTGATTT
Antisense primer	CCCAATAACCTCACTAAACAACC
Sequencing primer	TAACCTCACTAAACAACCAA
Dispensation order	GCAAAAACAATCTAAAAAAATGATAATAAATCGATTATCGATACGATTTACGATAC
Amplified DNA	118 bp (nt 5551-5668 of NC001526.1)
Sequenced CpGs	nt 5615, 5609, 5606, 5600

Table 3.8: The primer sequences, regions amplified and CpGs sequenced for the pyrosequencing assays.

BTN refers to the position of a biotin label. NC001526.1 is the HPV16 reference sequence.
Dispensation order is the order used by the pyrosequencing instrument to sequence, by synthesis, the pyrosequencing PCR products.

3.1.6. DIPS

Detection of Integrated Papillomavirus Sequences (DIPS) (Luft *et al.*, 2001) was used to determine the integration status of HPV in the VR series clinical cohort (Figure 1.8).

3.1.6.1. DNA Digestion and Ligation of Adapter Sequence

Two aliquots of 1.2 µg DNA were digested using 10 U of Sau3AI or Taq1 (New England Biolabs, Beverly, MA.) in a 20 µl reaction according to Table 3.9. Reactions were prepared in PCR plates, sealed, incubated at 37 °C for 15 hrs and then incubated at 80 °C for 20 mins on a thermocycler to inactivate the restriction enzymes. HPV negative (Mono Mac 6) and HPV positive (CaSki) cell line DNAs were included in each run as controls.

Sau3AI digest		Taq1 digest	
Reagent	Volume (µl)	Reagent	Volume (µl)
DNA (1.2 µg)	-	DNA (1.2 µg)	-
Sau3AI (10 U)	2.5	Taq1 (10 U)	0.5
Buffer 1 (10x)	2	Buffer 4 (10x)	2
BSA (100x)	0.2	BSA (100x)	0.2
H_2O	to 20 µl	H_2O	to 20 µl

Table 3.9: Sau3AI and Taq1 DNA digestion reaction volumes.
Units listed are the volumes required per reaction.

Ligation was performed using double stranded DNA adapters specific to either the Sau3AI or Taq1 digested DNAs. Taq1 and Sau3AI adapters were produced using the 25 µM adapter recipe in Table 3.10. The volumes listed produced enough adapter for 200 reactions. Reaction mixes were heated to 90 °C for 2 mins then allowed to cool in an overnight protocol to 4 °C in a thermocycler. Adapter mixes were then stored at -20 °C.

Reagent	Volume (µl)
AL1 primer (100 µM)	25
TaqI or Sau3AI AS primer (100 µM)	25
Sterile TrisHCl (66 mM)	50

Table 3.10: The reagent volumes required to produce a 25 µM adapter stock solution.

The Sau3AI and Taq1 adapters were ligated to the DNA digests using T4 Ligase (New England Biolabs, Beverly, MA). The volumes and reagents used for each reaction are listed in Table 3.11. This mix was incubated at 22 °C for 2 hrs, then inactivated by heating to 65 °C for 10 mins before 6 µl of sterile water was added to bring the final volume to 40 µl.

Reagent	Volume (µl)
TaqI or Sau3AI digested DNA	20
Ligase Buffer (10x)	2.4
T4 Ligase (400 U/µl)	1
TaqI or Sau3AI Adapter (25 µM)	0.48
Water	0.12
Total volume	24

Table 3.11: The reagent volumes required per DIPS reaction to ligate adapters to digested genomic DNA.

The 20 µl of digested DNA refers to the complete mix listed in Table 3.9.

3.1.6.2. Linear PCR

Linear PCR was performed using 9 separate reactions; 8 pairs of HPV specific primers and one pair of control primers. Separate master mixes were prepared for each primer (n=9) for both DNA digests/ligations (Sau3AI and Taq1) according to Table 3.12. For each reaction 23 µl was aliquoted into PCR plates and 2 µl of digested/ligated DNA was pipetted into each reaction. Each sample had 18 PCR reactions, i.e. 9 for each DNA digest/ligation. Plates were sealed and the PCR reaction performed according to the time/temperatures listed in Table 3.13.

Reagent	Volume (µl)
10x buffer with 15 mM $MgCl_2$	2.5
dNTP (10 mM)	0.5
HPV PCR 1 primer (1-9) (10 µM)	0.5
Hot star Taq (1 U)	0.125
H_2O	19.375
Total volume	23

Table 3.12: Linear PCR master mix recipe.

Volumes are given for each reaction.

	Temperature (°C)	Time (s)	Cycles
Initial denaturation	95	900	1
Denaturation	94	30	
Primer annealing	66	30	40
Extension	72	180	
Final extension	72	420	1

Table 3.13: PCR timing and temperature for the DIPS linear PCRs.

3.1.6.3. Nested PCR

For the Nested PCRs, the linear PCR products were amplified exponentially using an adapter specific primer and a HPV specific primer. As such, 18 reactions are required for each sample;

one reaction mix for each HPV primer for both DNA digests. For each sample, the reaction mix in Table 3.14 was prepared and 20 µl aliquoted into a PCR plate. The plates were loosely sealed and taken to the post-PCR laboratory where 5 µl of linear PCR product was dispensed into the appropriate well of the second PCR plate. The plates were sealed and PCR reactions run according to the temperature/timing in Table 3.15. Remaining linear PCR products and completed nested PCR products were stored at -20 °C until required.

Reagent	Volume (µl)
10x buffer with 15 mM MgCl$_2$	2.5
dNTP (10 mM)	0.5
Hot Star Taq (1 U)	0.125
HPV PCR 2 Primer (1-9) (10 µM)	1
AP1 Primer (10 µM)	1
H$_2$0	14.875
Total volume	20

Table 3.14: Nested PCR master mix recipe.

Volumes are given for each reaction.

	Temperature (°C)	Time (s)	Cycles
Initial denaturation	95	900	1
Denaturation	94	30	
Primer annealing	66	30	30
Extension	72	180	
Final extension	72	420	1

Table 3.15: PCR timing and temperature for the DIPS nested PCRs.

3.1.6.4. Gel Electrophoresis

25 µl of nested PCR products were run on a 1% w/v agarose TBE gel with 2 µl/100 ml of ethidium bromide (10 mg/ml) and alongside 100 bp DNA ladder in order to estimate product size. Gels were run at 120 V for 1-2 hrs or until adequate separation of the PCR products was achieved.

Once run, the gels were photographed and bands of interest were cut out using a clean disposable scalpel and a UV transilluminator. Suitable PPE was worn to protect the operator from UV light. As much excess gel was removed as possible and excised band gel fragments were placed in clean microcentrifuge tubes. This was performed for all bands of interest for all samples, for all primer sets.

Following excision, gel fragments containing PCR products were purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare Life Sciences, Buckinghamshire, UK).

Extraction was performed according to the manufacturer's instructions. Briefly, this involved dissolving the agarose gel pieces using heated buffers, binding the PCR product to a membrane, washing the PCR product with a series of wash buffers and final elution into 15 μl of sterile nuclease free water.

Upon purification, 5 μl of eluate was run on a 2% w/v agarose gel containing 20 mg/100 ml ethidium bromide. This was performed to check for successful PCR product purification, check for the presence of a single band and to roughly estimate DNA concentration by comparing against known concentration/intensity 100 bp ladder.

3.1.6.5. Sequencing

Sanger sequencing of PCR products was performed by Source BioScience UK Ltd (Nottingham, UK). HPV primer 2 (i.e. the HPV specific nested primer) was used as the sequencing primer. The requirements for sequencing are presented in Table 3.16.

	Sequencing requirements	
	Volume (μl)	Concentration
PCR products	5	1 ng/μl/100 bp*
Primers	5	3.2 μM

Table 3.16: The sequencing requirements for purified PCR products.

Values listed are per reaction. * For PCR products, for every 100 bp in length 1 ng/μl of DNA was required.

3.1.6.5.1. Data Analysis

Sequence data was returned as a chromatogram with base calling. Sequences were checked manually to improve the base calling and to check for the parallel sequencing of multiple fragments (indicated by overlapping sequence). Sequence of poor quality or with background noise was removed from the sequence reads and not considered for data analysis.

Several stages of sequence analysis were performed. First, alignment of sequences against human genomic DNA sequences, and against the HPV16 reference sequence (GenBank ID NC001526.1) was performed using the NCBI BLAST (Basic Local Alignment Search Tool). Two BLAST algorithms were used, first megablast, then blastn. Megablast worked well for larger closely related sequences, blastn worked better for shorter and/or less closely related sequences. For the VR series, megablast identified the majority of potential integration events and blastn identified two events where a shorter section of human DNA was present.

To identify the region of the human genome where integration events were detected, identified human sequences were aligned using the UCSC BLAT (Blast Like Alignment Tool).

3.1.6.6. Primer Sequences

Table 3.17 contains the primer sequences used for DIPS. Essentially, the DIPS protocol described by Luft *et al* was used in conjunction with the restriction site PCR HPV specific primers described by Thorland *et al* as these were found to produce better results. Other primer/adapter sequences are as described by Luft *et al* (Thorland *et al.*, 2000, Luft *et al.*, 2001)

Adapter and adapter specific primers	
AL1 primer	GGGCCATCAGTCAGCAGTCGTAGCCCGGATCCAGACTTACACGTTG
AP1 primer	GGCCATCAGTCAGCAGTCGTAG
TaqI AS primer	PO_4-CGCAACGTGTAAGTCTG-NH_2
Sau3AI AS primer	PO_4-GATCCAACGTGTAAGTCTG-NH_2
Primary PCR primers	
HPV PCR 1 primer 1	ACAAAGCACACACGTAGACATTCG
HPV PCR 1 primer 2	AGTAATAAATCAACGTGTTGCGATTG
HPV PCR 1 primer 3	TTTGGTTACAACCATTAGCAGATGC
HPV PCR 1 primer 4	GTGCCAACACTGGCTGTATCAAAG
HPV PCR 1 primer 5	TACCAATTCTACTGTACCTAATGCCAG
HPV PCR 1 primer 6	ACTTATTGGGGTCAGGTAAATGTATTC
HPV PCR 1 primer 7	AGTAGATATGGCAGCACATAATGAC
HPV PCR 1 primer 8	GTTGGCAAGCAGTGCAGGTCAG
Control PCR 1 primer 9	TTCTCTATGTGCGTTCTCTCCCTG
Nested PCR primers (PCR2)	
HPV PCR 2 primer 1	CGTACTTTGGAAGACCTGTTAATGG
HPV PCR 2 primer 2	GGACTTACACCCAGTATAGCTGACAG
HPV PCR 2 primer 3	AATAGGTATGTTAGATGATGCTACAG
HPV PCR 2 primer 4	ACAAGCAATTGAACTGCAACTAACG
HPV PCR 2 primer 5	GAGGTTAATGCTGGCCTATGTAAAG
HPV PCR 2 primer 6	CCCTGTATTGTAATCCTGATACTTTAGG
HPV PCR 2 primer 7	TGCGTGTAGTATCAACAACAGTAAC
HPV PCR 2 primer 8	TTAAACCATAGTTGCTGACATAGAAC
Control PCR 2 primer 9	CAAACTCCAGGTCTCCAACCAG

Table 3.17: Primer and adapter sequences used for DIPS.

3.1.7. APOT

Amplification of Papillomavirus Oncogene Transcripts (APOT) was used briefly during this project and was trialled on cell line material and some VR series samples (Figure 1.7). APOT was performed as detailed by Vinokurova *et al* (Vinokurova *et al.*, 2008). The VR samples failed the APOT assay repeatedly. VR RNAs were tested for quality/integrity using a 2100 Bioanalyser (Agilent Technologies Inc, Santa Clara, CA) and the mean fragment size appeared inadequate for APOT. APOT was abandoned for DIPS.

3.2. Materials

3.2.1. Cell Lines

3.2.1.1. CaSki

The Ca Ski (CaSki) cell line was sourced from the European Collection of Cell Cultures (ECACC) (catalogue number 87020501). The CaSki cell line was originally derived from an epidermoid carcinoma of the cervix, metastatic to the small bowel mesentery of a 40 year old Caucasian. CaSki DNA and RNA were extracted using the Qiagen AllPrep (Qiagen, Hilden, Germany) protocol. The identity of the CaSki cell line was confirmed using STR typing. CaSki contains multiple HPV16 genomes inserted as a concatemer. There are 30 reported point mutations and a single 1 nt deletion in the vast majority of HPV16 genomes (Meissner, 1999).

3.2.1.2. SiHa

The SiHa cell line was sourced from the American Type Culture Collection (ATCC) (ATCC number HTB-35). The SiHa cell line was originally derived from a cervical squamous cell carcinoma of a 55 year old Asian. SiHa is reported to contain integrated HPV16, with between 1 and 2 copies per cell. Integration is associated with disruption of the HPV16 genome and a deletion at nt 3133-3385, there are also deletions at nt 3460-3512 and nt 7757-7794 (Meissner, 1999).

3.2.1.3. W12

The W12 cervical keratinocyte model is a nonclonal cell culture derived from a cervical lesion with CIN1/LSIL histology (Stanley *et al.*, 1989). Early passage W12 in organotypic raft culture exhibits an LSIL like phenotype and contains approximately 100 HPV16 genomes/cell in episomal form (Pett *et al.*, 2004), however E2 silenced integrated forms can also be detected (Pett *et al.*, 2006). Long term cultures of W12 undergo clearance of episomal HPVs and cells containing integrated HPV16 genomes dominate (Gray *et al.*, 2010). Genomic instability occurs as a result and the organotypic raft cultures adopt a HSIL/SCC phenotype. In long term culture of discrete clones of W12, different integration events are selected for, resulting in a series of diverse cell lines (Pett *et al.*, 2004). The DNAs used in this study were from W12 Series 7 (W12Ser7), sourced from Professor Nicholas Coleman (University of Cambridge, personal communication) (Gray *et al.*, 2010).

3.2.1.4. PC08 and PC09

Primary Culture 08 and 09 (PC08 and PC09) were established as primary culture of anogenital biopsies. The primary culture lines were grown for several passages, after the first passage the

lines are no longer considered primary culture and the term "short term culture" is more accurate. PC08 was derived from a vulval biopsy of a 46 yr old woman with histologically confirmed VIN3. PC09 was derived from a vaginal biopsy of a 31 year old woman with histologically confirmed VAIN3.

The PC08 and PC09 cell lines were initially established as heterogeneous lines, single cell cloning was then performed to establish homogeneous lines. Six clones were provided for this study, the PC08 clones M, P and Y and the PC09 clones A, D and H. Each homogeneous clone was cultured further and at each passage, DNA and RNA were extracted. For the PC08 clones, single cell cloning was performed at two different passages. Cloning of early passage heterogeneous lines produced the M and P clones whilst cloning of later passage heterogeneous lines produced the Y clone. For this reason, clone Y8 starts at passage 9 whilst the others start at passage 5. All characterisation and cell culture was performed by Tiffany Onions, (School of medicine, Cardiff University). For this project, Tiffany Onions supplied DNAs for DNA methylation analysis and provided APOT results and Ct values from qPCR. DNAs were extracted using the Qiagen AllPrep protocol.

3.2.2. Clinical Material

3.2.2.1. The CRISP Study

The Cervical Randomised Intervention Study Protocol (CRISP) clinical trial was conducted to determine the effect of diindolylmethane (DIM) on low grade cervical abnormalities (Castanon *et al.*, 2012). It was a double blind, randomised controlled trial to study the effect of 6 months of DIM supplementation women with newly diagnosed, low grade cervical abnormalities. Of the 600 women eligible for inclusion in the study, 439 women had paired LBC samples at 0 and 6 months for HPV typing. DIM supplementation was found to have no effect as a treatment for cervical abnormalities and was not considered for this study of DNA methylation. HPV typing was performed for the CRISP clinical trial using the PCR ELISA HPV typing system described by Jacobs *et al* (Jacobs *et al.*, 1997, Castanon *et al.*, 2012).

For this investigation, DNAs were extracted using the QIAamp MinElute Media kit protocol on frozen cell pellets. All samples used were HPV16 positive at 0 months, samples were divided into three groups. First, those infections that were HPV16 positive at 0 months, but HPV16 negative at 6 months were classified as the Cleared (C) group. Twenty-four samples met this

criteria, one of these was excluded on the basis that the histology at 6 months revealed a CIN2 lesion.

The second group of samples were HPV16 positive and were associated with high grade histology (CIN2 and CIN3) at 6 months. This group comprised of 11 patients with both 0 and 6 month DNA samples and were termed the Persistent High grade (PH) group.

Finally, those samples that were HPV16 positive and were associated with low grade histology (i.e. CIN1) at 6 months were selected. A total of 48 patients met this criteria however only 12 patients' DNAs were required for the study. The first 9 patients were selected because they had histologically confirmed CIN1 at 6 months, the final 3 were chosen at random. This group was termed the Persistent Low grade (PL) group and represents samples that had maintained a HPV16 infection for more than 6 months in the absence of significant disease.

3.2.2.2. The Prevalence Study

The prevalence study described the prevalence of HPV in South Wales in relation to age, cytology and social depravation. A total of 1,911 patients were included in the study. DNA was extracted by Hibbitts *et al* and involved disruption of LBC material using proteinase K and purification by centrifugation as described (Hibbitts *et al.*, 2006). These DNAs were then HPV typed using the PCR ELISA HPV typing system described by Jacobs *et al* (Jacobs *et al.*, 1997, Hibbitts *et al.*, 2006). For this study, 20 samples were identified with normal cytology and 20 with severe dyskaryosis and DNAs were provided for the purpose of performing DNA methylation analysis (Chapter 5).

3.2.2.3. Archival Cancers

A series of DNAs previously extracted from Formalin Fixed Paraffin Embedded (FFPE) cervical cancers were used in this study (Chapter 5). The samples were initially collected for prevalence studies of HPV in cervical cancers (Powell *et al.*, 2011, Powell *et al.*, 2009). The FFPE cervical cancers were collected from various hospitals before sectioning and DNA extraction using the Qiagen DNeasy Blood and Tissue kit without xylene washes (Powell *et al.*, 2011).

The samples for inclusion in the study were selected using the following criteria: HPV16 positivity was required, samples collected from University Hospital of Wales (n=57) were excluded as they were fixed using low pH unbuffered (rather than neutral buffered) formalin which typically provides DNAs of poor quality. DNA concentrations of the remaining samples were determined and arranged in descending order (n=79), the top 20 sample DNAs were selected. Seven additional samples were later sought for inclusion to account for samples that

had failed to amplify by PCR, or had persistently failed to produce a usable result in the pyrosequencing assay. Failure was likely the result of DNA degradation of DNA in storage and/or prior to extraction. A total of 24 samples provided DNA methylation data of sufficient quality to be included in the analysis. Of these there were SCC (n=19), adenocarcinoma (n=3), adenosquamous carcinoma (n=1) and carcinosarcoma (n=1) DNAs.

3.2.2.4. VIN Biopsy Series

A large series of anogenital biopsies (n=171) was available for this project and were used for an investigation into HPV distribution. Of these, 118 were DNAs and 53 were biopsies fixed in ThinPrep LBC media. The 53 biopsies were coined the VR series as they were of vulval origin and both DNA and RNA was extracted. The VR series is a series of VIN and VAIN biopsies (primarily VIN3). Various nucleic acid extraction methods were used for the series in order to try and improve the quality and quantity of DNA and RNA produced. The first 37 samples (VR1-37) were extracted using the TRIzol extraction protocol, the remaining 16 samples (VR38-53) were extracted using the VIN RNA/DNA extraction protocol.

VIN DNAs were HPV typed using the PapilloCheck assay described previously. The HPV typing data was used as the basis of a HPV type distribution in VIN study. The sample selection criteria involved exclusion of non-VIN biopsies, cancers and samples with incomplete clinical data. Finally, where duplicates existed (i.e. more than one sample from the same patient) the earliest samples were excluded. Sixty-two samples remained for analysis.

3.3. Statistical Analysis

Statistical analysis was performed in Minitab 16.1.0 using procedures described by Bowker (Bowker, 2007). This section contains a brief explanation of the tests used whilst Table 3.18 contains a glossary of terms.

Term	Meaning
H_0 – Null hypothesis	The default position i.e. No relationship/difference between factors/means
H_1 – Alternative hypothesis	The alternative position i.e. A significant relationship/difference between factors/means
P-value	Probability value, typically reject the H_0 and accept H_1 at P<0.05
DF	Degrees of freedom
Type 1 error	When H_0 is true but rejected, i.e. a false positive
Type 2 error	When H_0 is false but accepted, i.e. a false negative

Table 3.18: A glossary of statistical terms used and their meanings.

3.3.1.1. Correlation

Correlation analysis objectively tests for a linear relationship between two variables derived from independent, random samples. The H_0 is that there is no significant correlation between the two variables (Bowker, 2007).

Spearman's rank non-parametric correlation coefficient tests for an association between two quantitative or categorical variables that can be ranked from low to high. Normal distribution of the variables is not required (Bowker, 2007). Correlation coefficients range from -1 to +1. The further from zero, the stronger the correlation. A positive or negative number represents an upwards or downwards sloping scatter plot respectively (Bowker, 2007).

3.3.1.2. Linear Regression

Linear regression analysis is used to predict and draw a straight line of best fit between two quantitative variables according to the equation:

$$Y = m.X + c \pm \varepsilon$$

Here: X=the cause/independent variable, Y=the effect/dependent variable, m=gradient of the slope, c=intercept of the line through Y and ε = error in Y around the line of best fit. The residuals should be normally distributed (Bowker, 2007).

3.3.1.3. Anderson Darling

The Anderson Darling normality test tests the H_0 that the data is not normally distributed. Many statistical analyses rely on normal distribution of either the data or the residuals. Where normal distribution is not achieved (P>0.05), data transformation may improve the distribution (Bowker, 2007).

3.3.1.4. Homogeneity of Variances

Many statistical analyses require that groups of data have equal variances. There are several tests depending on the application and the number of groups being compared, each test tests the H_0 that the groups have equal variances. Bartlett's test for equal variances was used prior to ANOVA and the F-test variance ratio statistic was used prior to T-tests (Bowker, 2007).

3.3.1.5. Data Transformation

Where data has not met the assumptions of a statistical test, the data is transformed and tested for suitability again. Data transformation for this study included the following: untransformed (X) data, $Log_{10}(X)$, $Log_{10}(X+1)$, X^2, (X-mean)/standard deviation). Unless otherwise noted in the text, $Log_{10}(X+1)$ transformation provided the best DNA methylation data and $Log_{10}(X)$ provided the best qPCR data and were used prior to data analysis.

3.3.1.6. T-test and Associated Tests

The Student's T-tests are parametric tests to compare two groups of independent data, paired data before and after a known treatment/event and one group of data to a standard value. T-tests assume that the data is normally distributed, and the two groups have equal variances. For this study, the two-sample T-test was used and tests the H_0 that no difference between the means of the groups exist (Bowker, 2007).

Where the assumptions of the two-sample T-test were not met, the non-parametric equivalent Mann-Whitney U-test was used to test the H_0 that the two groups had the same median. Where comparison of multiple groups of data is required alternative tests should be used (i.e. ANOVA) instead of multiple two-sample tests as multiple tests increase the chance of a Type 1 error.

3.3.1.7. Bonferroni Correction

The Bonferroni correction is used to reduce the likelihood of a Type 1 error occurring when multiple statistical analyses are used (i.e. P-value of 0.05 = likely to have one type 1 error every 20 analyses). The Bonferroni correction reduces the P-value to compensate for multiple analyses by dividing the original P-value by the total number of comparisons. The Bonferroni correction is often criticised for being too conservative, in this study interpretation of results significant at both P=0.05 and a Bonferroni corrected P-value was considered.

3.3.1.8. One-way ANOVA and Kruskall-Wallis

One-way Analysis Of Variance (ANOVA) is applied to quantitative variables separated into two or more groups. ANOVA is a parametric test that requires normally distributed residuals and homogenous variances and assumes independent random samples. ANOVA tests the H_0 that there is no difference between the means of the groups (Bowker, 2007).

The ANOVA does not perform a comparison of means between pairs of groups. Pairwise comparison is performed using Tukey-Kramer a posteriori tests for unplanned differences or Fisher's a priori tests for planned differences (Bowker, 2007).

The Kruskall-Wallis test is a non-parametric equivalent of the ANOVA and tests the H_0 that there is no difference between the medians of the groups. Kruskall-Wallis is far less powerful than the ANOVA and has no function to perform pairwise comparison of means (Bowker, 2007).

It is common practice in exploratory, hypothesis generating studies to lower the requirements of the ANOVA and proceed even with violations of the assumptions on the basis that the results are interpreted with caution (Bowker, 2007). For this study, several tests violated one or more assumptions of the ANOVA. Because these statistics were exploratory and hypothesis generating in nature, it was felt that the advantages of the one-way ANOVA and Tukey-Kramer over the Kruskall-Wallis test (and multiple Mann-Whitney U tests for pair wise comparisons) far outweighed the risks involved with using the one-way ANOVA with violations. Where the analysis was continued, the results were interpreted with caution, frequently reported alongside an additional Kruskall-Wallis test for additional confirmation and the violations were listed in the text.

3.3.1.9. *Multifactor ANOVA*

Multifactor ANOVA allows the analysis of data grouped with respect to multiple factors. A two-way ANOVA tests the H_0 that neither factor has a significant effect and that the means of each group are the same. Several models of two-way ANOVA were used in this study. The two-way ANOVA General Linear Model (GLM) allows for the analysis of data with an unbalanced design (the majority of analyses). Several instances occurred where a balanced design without replication existed, for these analyses a standard two-way ANOVA was used.

Two-way ANOVA assumes equal variances in each factor combination and normally distributed standardised residuals. No suitable non-parametric equivalent of the GLM exists; where the assumptions were violated, a decision was made in each instance whether to continue with the analysis or use multiple one-way ANOVA/Kruskall-Wallis tests with a Bonferroni correction. Multiple one-way ANOVA are not advised both because of the increased rate of type 1 error, and because they remove a level of detail from the analysis. Where multifactor ANOVA was continued, the results were interpreted with caution, often presented alongside additional analyses and the violations were listed in the text.

When analysing DNA methylation data both on a per-assay and a per-CpG basis using multifactor ANOVA, the individual CpGs (i.e. E2-1...E2-8) were nested within the assay (i.e. *E2*, LCR1, etc) to allow for a more thorough analysis acceptable to the GLM.

Chapter 4. Method Development

4.1. HPV Gene Expression

This investigation required assessment of the expression of three HPV genes, the *E6* and *E7* oncogenes and the *E2* repressor. A literature search identified a study in which *E6* and *E7* mRNA level were quantified in cell line and clinical material (Wang-Johanning *et al.*, 2002). *E6* and *E7* primers were complementary to regions included in various splice variants for each gene (Zheng and Baker, 2006). *E2* primers were identified from a separate study in the same literature search (Roberts *et al.*, 2008a). The primers were trialled briefly and satisfactory results were produced which warranted further optimisation.

Selection of suitable reference genes required more thorough consideration. Quantifying the stability of a panel of reference genes in diseased and disease free epithelial tissues was outside of the scope of this project. In order to improve the reliability of the results two reference gene assays that covered a range of expression levels were sought (i.e. one highly expressed and one lowly expressed). Potential reference genes were identified from published studies that had quantified the stability and ranked the suitability of prospective reference genes. Two genes (and primer pairs) were chosen on the basis of high stability in a range of keratinocyte lines, these were *TBP* (TATA box binding protein) and *HPRT* (Hypoxanthine guanine phosphoribosyl transferase) (Allen *et al.*, 2008). Further literature searches confirmed the apparent suitability of these reference genes for epithelial RNA studies. The *TBP* assay was repeatedly problematic and inconsistent during optimisation and was eventually replaced with a second primer pair amplifying a different region of *TBP* (Minner and Poumay 2008). *HPRT* and both *TBP* assays were intron spanning such that the DNA amplimer was considerably larger and more difficult to amplify than the RNA. This was not possible for HPV genes where a complex array of potential splice variants exist.

4.1.1. qPCR Optimisation and Method Development

Several significant issues were encountered during the course of qPCR optimisation. The main issues and solutions are summarised below.

4.1.1.1. Initial Failure

During initial trial of the qPCR assays, failure of the PCR reaction was very common. Failure could have been caused by several factors. First, the RT reaction was tested to determine whether the RT was functioning properly and in each instance, a demonstrable difference between positive and negative RT controls was noted. Similarly, cDNAs that were amplified in some reactions were not amplified in subsequent reactions, i.e. there was considerable

inconsistency. A large variety of conditions were trialled using gradient PCR and a series of qPCRs in order to optimise the PCR conditions. However, the success or failure of reactions appeared random and further optimisation could not continue until the issues had been solved.

To rule out operator error, a senior member of staff demonstrated, then supervised and finally performed the same experiments independently. Subtle differences in technique were initially associated with a considerable difference in success rate, however this was later deemed coincidental after the inconsistent pass/fail rate resumed. The ambient temperature might have affected the heating and cooling rate of the LightCycler instrument so a thermometer was placed into the qPCR laboratory and the ambient temperature was recorded. No relationship between ambient temperature and success rate was observed. All reagents required for qPCR (cDNAs, primers, $MgCl_2$, water, SYBR green master mix, plastics) were replaced but inconsistent qPCR reactions continued to be observed. The LightCycler instrument was beginning to be suspected so was replaced as a precaution. The results improved considerably, but were still inconsistent. Several small checks and measures suggested by Roche technical support were implemented with little effect.

Eventually, it transpired that the LightCycler capillary tubes were out of date. The capillary tubes contain a coating to prevent interaction of the glass with the reagent mix, this coating was found to degrade with time and the tubes in use were out of date by several years. Replacement tubes were trialled and the inconsistent failure/success rate was solved.

Throughout the qPCR method development, DNase treatment of sample RNAs was only partially successful. Several RNA purification kits were tested, DNases from several manufacturers were tested and a variety of treatment protocols were trialled with varying degrees of success. In order to correct for an incomplete DNA digestion during DNase treatment (residual genomic DNA was shown to contribute as much as 10% to the signal strength), RT negative (RT-ve) controls were included in duplicate for each sample alongside the RT positive (RT+ve) reactions. The linear regression equations produced from calibration curves were used to calculate the relative DNA concentration of the RT+ve and the RT-ve samples (concentration expressed relative to 1x CaSki cDNA). The mean calculated cDNA concentration of the RT-ve repeats was subtracted from that of the RT+ve repeats to leave the amount of PCR product in the RT+ve samples that was attributable to just the cDNA. This DNA concentration was then used in the linear regression equation to produce a new Ct value that reflected the amount of cDNA in the RT+ve reaction, corrected to remove the amplified

genomic DNA in the sample. Reactions where a non-specific PCR product had been amplified were excluded from the analysis.

4.1.1.2. Optimisation

Once the cause of inconsistency in the qPCR reaction had been removed, optimisation of the qPCR reactions was resumed. The cycle number, input cDNA amount, annealing temperature and Mg^{2+} ion concentration were optimised and for most assays, reliable results were produced. For *TBP*, the results were still inconsistent. Eventually, a number of suitable reference gene primer sets were trialled and a second *TBP* set were adopted.

4.1.1.3. Data Collection

The 6 VIN biopsy samples with the largest RNA yield were trialled, further optimisation was performed and the trial was extended to the remaining VR samples. RNAs from the VR series were analysed by *E2*, *E6*, *E7*, *HPRT* and *TBP* qPCR assays. Several methods of relative quantification exist, two common methods being the comparative Ct method and the efficiency corrected comparative Ct method proposed by Pfaffl (Pfaffl, 2001). The choice of relative quantification method can have a huge effect on the outcome and reliability of the qPCR data.

4.1.2. Relative Quantification

4.1.2.1. Comparative Ct Method

The comparative Cycle threshold (Ct) method of relative quantification (commonly called the delta delta Ct or ddCt method) is widely used, and is the method of analysis used by the Applied Biosystems DataAssist v3.0 software; a software package designed to analyse the output data from multiple Applied Biosystems quantitative PCR systems. Livak and Schmittgen discussed the derivation of the (RQ) = $2^{-\Delta\Delta Ct}$ formula in detail (Livak and Schmittgen, 2001). The main disadvantage of the comparative Ct method is that the PCR efficiencies are assumed to be close to 2 (i.e. a perfectly exponential PCR reaction) and that target and reference gene PCR efficiencies are equal to within 0.05 (Schmittgen and Livak, 2008, Livak and Schmittgen, 2001). Inevitably, different PCR primer pairs tend to amplify at different efficiencies. The effect of varying PCR efficiency is illustrated in Table 4.1 where five theoretical PCRs reactions of varying efficiency amplify a single template cDNA. These reactions would all reach the qPCR threshold point at very different cycle numbers, leading to diverse estimates of the number of template copies in the input cDNA.

The comparative Ct equation does not allow for the inclusion of delta Ct values of multiple reference genes. Despite using the comparative Ct equation for analysis, the Applied Biosystems DataAssist v3.0 software allows for the selection of multiple reference genes using the arithmetic mean of the reference gene Ct values as the basis of the ΔCt calculation.

Cycle number	PCR efficiency				
	100%	95%	90%	80%	70%
0	1	1	1	1	1
1	2	1.95	1.9	1.8	1.7
2	4	3.8025	3.61	3.24	2.89
3	8	7.414875	6.859	5.832	4.913
25	33554432	17818260	9307650	2408866	577062.7
30	1.07E+09	5.02E+08	2.3E+08	45517160	8193466
35	3.44E+10	1.42E+10	5.71E+09	8.6E+08	1.16E+08
40	1.1E+12	3.99E+11	1.41E+11	1.63E+10	1.65E+09

Table 4.1: The number of template copies at each cycle in 5 theoretical PCR reactions with a starting amount of 1 PCR template copy.

Each PCR reaction has a different PCR efficiency, this table illustrates the potential for error caused by differences in PCR efficiency. Each number is the number of template cDNA copies.

4.1.2.2. The Pfaffl Equation – PCR Efficiency Correction

In order to overcome the previously discussed limitations, Pfaffl (Pfaffl, 2001) proposed a modification to the comparative Ct method that accounted for the PCR efficiency of the target and reference gene PCRs. The so called "Pfaffl equation" is detailed below (Equation 4.1).

$$RQ = \frac{\left(E_{target}\right)^{\Delta Ct_{target}\,(control-sample)}}{\left(E_{reference}\right)^{\Delta Ct_{reference}\,(control-sample)}}$$

Equation 4.1: The Pfaffl equation.

E refers to the PCR efficiency, target refers to target genes (i.e. _E2_, _E6_ and _E7_), control refers to CaSki, sample refers to the sample of interest and Ct refers to the crossing point. ΔCt refers to change in the Ct.

The Pfaffl equation allows the different PCR efficiencies of target and reference genes to be considered in the generation of relative quantification ratios (Pfaffl, 2001), if both target and reference genes had a PCR efficiency of 100% (i.e. 2) the results of using the Pfaffl equation and the comparative Ct/Livak equation would be identical. The Pfaffl equation is very commonly used for relative quantification studies and forms the basis of the Qiagen, REST 2009 relative quantification software and its predecessors (Pfaffl _et al._, 2002, Qiagen, 2009). A

limitation of the Pfaffl equation, is that it allows for only one reference gene (addressed in some situations in the REST 2009 software), a slightly modified alternative equation is in use that allows for the quantification of target genes against multiple reference genes. The Pfaffl equation is only slightly updated for the Vandesompele equation, except that in the Pfaffl equation, the denominator is calculated for a single gene (Vandesompele *et al.*, 2002). In the updated Vandesompele equation, the Pfaffl equation denominator is calculated for each reference gene, and the geometric mean of these results is used in place of the denominator used in the Pfaffl equation (Vandesompele *et al.*, 2002). This approach has been used for multiple published studies (Hoebeeck *et al.*, 2005, Hellemans *et al.*, 2004, Vandesompele *et al.*, 2002, Loeys *et al.*, 2005, Poppe *et al.*, 2004) and is the basis of the qBasePlus qPCR analysis software algorithm (Hellemans *et al.*, 2007, qBasePlus, 2007). Equation 4.2 is the updated Pfaffl equation (Vandesompele equation) for multiple reference genes (Hellemans *et al.*, 2007).

$$RQ = \frac{\left(E_{target}\right)^{\Delta Ct_{target}\,(control-sample)}}{\sqrt[n]{\prod_{\emptyset}^{n}\left(E_{reference_{\emptyset}}\right)^{\Delta Ct_{reference_{\emptyset}}\,(control-sample)}}}$$

Equation 4.2: The Vandesompele equation.

E refers to the PCR efficiency, target refers to target genes (i.e. *E2*, *E6* and *E7*), control refers to CaSki, sample refers to the sample of interest and Ct refers to the crossing point, n is the number of reference genes.

Errors in the calculation of PCR efficiency could have a profound effect on the calculated expression ratios; hence it is important that efficiency is determined in an appropriate manner.

4.1.2.3. PCR Efficiency

There are several ways of calculating the PCR efficiency. These broadly fall into two categories, those based on the Ct values of a calibration curve and those based on the actual PCR dynamics of single or multiple PCR reactions.

4.1.2.3.1. Estimating PCR Efficiency Using the Calibration Curve Method

This common method is the method best supported by the Roche LightCycler used in this study. It is widely used in the literature, including the earlier work of Pfaffl (Pfaffl, 2001). This simple method relies on observing the relationship between input cDNA concentration and Ct value across a dilution series and using linear regression to produce an equation for the relationship. Figure 4.1 is an example of a calibration curve produced for a CaSki DNA dilution series using the HPV16 *E6* qPCR primer pair. Here, a total of three calibration curves were run

and plotted alongside each other. Linear regression was performed and an equation representative of the three calibration curves produced according to Equation 4.3.

$$Ct = m.(Log_{10} \, DNA \, concentration) + c$$

Equation 4.3: The Linear regression equation.
For a linear relationship, *m* represents the gradient and *c* represents the intercept.

Which for the *E6* primer pair, linear regression analysis produces the formula:

$$Ct = -3.629.(Log_{10} \, DNA \, concentration) + 15.914$$

The calibration curve enables both the estimation of DNA concentration (in relation to the cDNA used in the calibration curve (i.e. CaSki cDNA)) and the estimation of PCR efficiency. Using the slope (*m*) from the linear regression line, the efficiency (*E*) of the PCR can be calculated according to Equation 4.4 (Pfaffl, 2001, Rassmussen, 2001).

$$E = 10^{(-1/m)}$$

Equation 4.4: The PCR efficiency calculation formula.
Here, *E* represents efficiency, and m represents the gradient calculated in Equation 4.3.

For the *E6* primer pair, this becomes:

$$E = 10^{(-1/-3.629)}$$

$$E = 1.8861$$

Or

$$100(E - 1) = 88.61\%$$

A perfect doubling (i.e. every template is amplified into two more templates) is achieved at a PCR efficiency of *E* = 2 or 100%.

A limitation of the calibration curve method of PCR efficiency calculation is that it is presumed that PCR efficiency is the same for all dilutions of template cDNA. However, when a template is diluted, any inhibitors to PCR in the template will also be diluted; this can lead to more efficient amplification in more dilute samples (Ramakers *et al.*, 2003). As such, other methods that rely on the PCR dynamics of single/multiple PCR runs can be used to negate this effect.
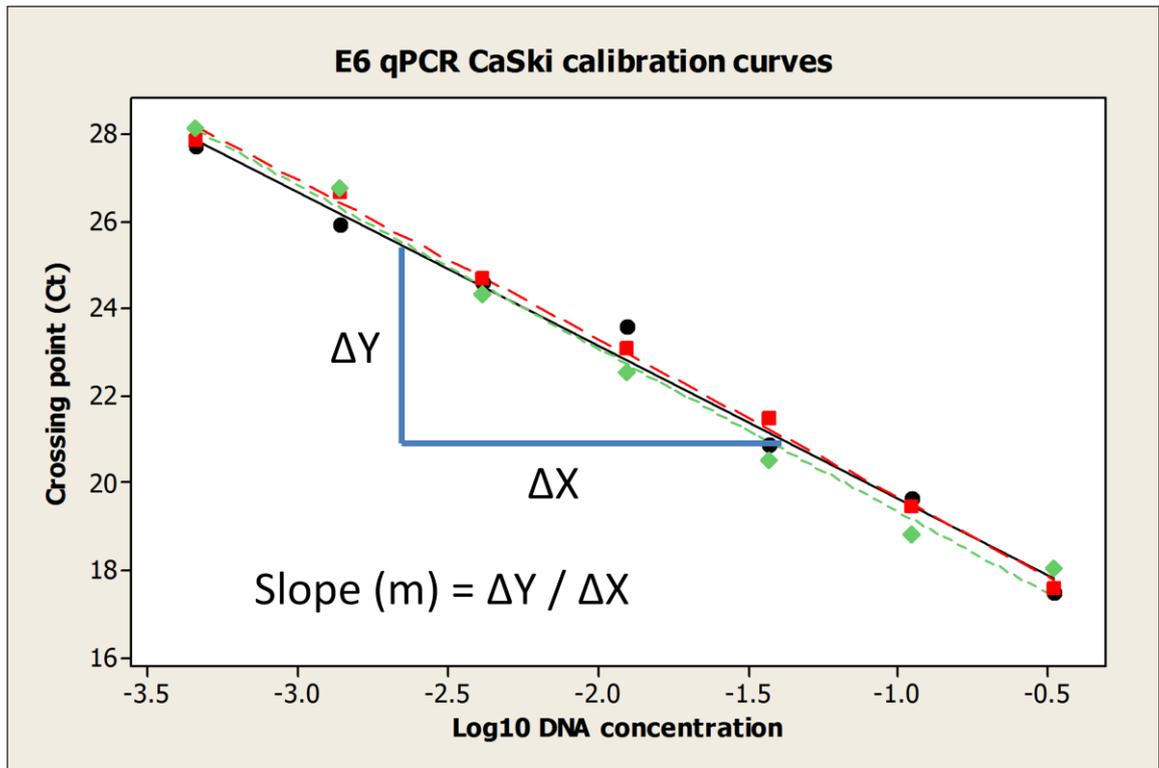
Figure 4.1: An example of a threefold dilution series of CaSki cDNA.

The dilution series was repeated in triplicate experimental repeats (represented by the three different colours/point shapes) and is used to calculate PCR efficiency and to calculate the Log_{10} CaSki cDNA DNA concentration equivalent to a sample Ct.

4.1.2.3.2. Estimating PCR Efficiency from PCR Dynamics

These approaches use the fluorescence measurement recorded at each cycle of the PCR to analyse PCR kinetics and determine the reaction efficiency of single PCR reactions. Single PCR reaction PCR efficiencies can be calculated using several different mathematical models (Liu and Saint, 2002, Ruijter *et al.*, 2009) and the choice of model has a large influence on the PCR efficiency calculated (Regier and Frey, 2010). Others suggest that the PCR efficiency of serial dilutions of samples must be used, as single sample reaction efficiency calculations are responsible for increasing the random error of qRTRTPCR quantifications and must be avoided (Nordgard *et al.*, 2006). The most reliable approach has since been shown to be to calculate the amplification efficiency from individual amplification plots and then use the mean efficiency for all samples as the amplification efficiency for that primer pair (Regier and Frey, 2010). These typically automated methods for correcting for PCR efficiency offer the best chance of accurately correcting for PCR efficiency for each primer pair.

4.1.2.3.3. PCR Efficiency Calculation for This Study

For this study, the calibration curve method was used. Calibration curves were produced with 7 points of a 3 fold dilution of CaSki cDNA, from 1/3 diluted down to 1/2187. The calibration curves were produced in triplicate and linear regression was performed on the calibration curve Ct values to produce an equation for the curves in the format of Equation 4.3.

Linear regression was performed on the results of the three calibration curves separately, and the mean of the resulting m and c values was used in calculations of PCR efficiency and the Ct value correction applied to account for reverse transcriptase negative samples. Performing one linear regression analysis for the triplicate repeats produces an identical line equation. Scatter plots with linear regression lines are included for each gene in Appendix II.

Using the slope (m) from the linear regression line, the efficiency (E) is calculated using Equation 4.4. A summary of the linear regression line equations data and the PCR efficiencies calculated from the slope value of the line equations are shown in Table 4.2.

Primer pair	Slope (m)	Intercept (c)	PCR efficiency	PCR efficiency %
E2	-3.297	15.960	2.0105	101.05
E6	-3.629	15.914	1.8861	88.61
E7	-3.533	15.843	1.9189	91.89
HPRT	-3.538	18.110	1.9171	91.71
TBP2	-3.742	23.568	1.8503	85.03

Table 4.2: The mean slope and intercept values produced by linear regression of the dilution series calibration curves for each qPCR primer pair.

Also included are the PCR efficiencies calculated according to the formula E = $10^{-1/slope}$, expressed both as a decimal and as a percentage.

4.1.3. Issues with qPCR: the Biopsy Dosage Effect

One of the potential issues with using qPCR on VIN biopsies is that several characteristics of the biopsy, both biological and experimental can affect the outcome. A large determinant of RNA yield is the size of the biopsy, i.e. larger biopsies tend to produce more RNA. A problem when looking at HPV16 gene expression is that not every cell in the biopsy will contain HPV16, the total number of cells containing HPV will be determined by the biology of the infection and the biopsy sampling surgery. In human gene expression studies, every cell contains the potential to express both the target gene and the reference gene. In HPV containing biopsy qPCR studies, some of the cells in each biopsy will not contain HPV16. This can lead to several scenarios that can affect the RQ ratio of HPV genes.

In Figure 4.2, scenario A, two biopsies with an equal number of HPV infected cells, each cell expressing the same amount of HPV gene are sampled, but the two biopsies vary in size. This would mean that for the larger biopsy, more human mRNA is produced than the smaller biopsy. Despite expressing the same expression of HPV genes in the HPV infected cells, the RQ ratio of HPV-human gene expression will be lower for the larger biopsy because of larger number of human mRNAs in the sample.

In Figure 4.2, scenario B, two identically sized biopsies have a different number of HPV infected cells; each cell in both biopsies is expressing the same amount of HPV gene. The biopsy with more HPV infected cells will have a higher RQ, purely as a result of a larger number of HPV infected cells, not because of higher HPV gene expression.

In Figure 4.2, scenario C, two biopsies of equal size are sampled. One has a small number of HPV infected cells with a high level of HPV gene expression per cell, the other sample contains a large number of HPV gene expressing cells that are expressing low amounts of HPV genes per cell. The RQ results of both biopsies could be very similar, the larger number of cells in the second biopsy making up similar numbers of mRNA as the fewer higher HPV gene expressing cells. The number of HPV containing cells masks the true differences in HPV gene expression between the two biopsies.

 In Figure 4.2, scenario D, both samples are identical in size, gene expression and the number of HPV infected cells. One sample has a large amount of adipose tissue, yielding less mRNA than epithelial tissue. This alters the amount of human gene mRNA in comparison to the HPV gene mRNA, potentially affecting the RQ ratios.

Several other factors are also masked by biopsy sampling. It is widely accepted that within a HPV lesion, different regions exhibit different HPV biology. Laser Capture Microdissection (LCM) studies for instance have revealed that within one lesion, there can be distinct areas containing different HPV types (Quint *et al.*, 2012), and variation in HPV DNA methylation and viral copy number in different parts of the same tumour (Kalantari *et al.*, 2009, Vinokurova and von Knebel Doeberitz, 2011). It is also well established that there is variation in gene expression in different epithelial layers (Doorbar, 2006). How much of the biopsy contained different cell types, or even identical types of cells with distinct patterns of HPV mRNA expression is difficult to establish. qPCR on mRNA from biopsies does not show differences in gene expression between different parts of the biopsy, rather, an average expression for the whole biopsy is returned.

LCM would have solved several of these issues however the technique is expensive, time consuming and is highly unlikely to be applicable to routine cytology/histology. It was also outside of the scope of this project both financially and in terms of time commitments.
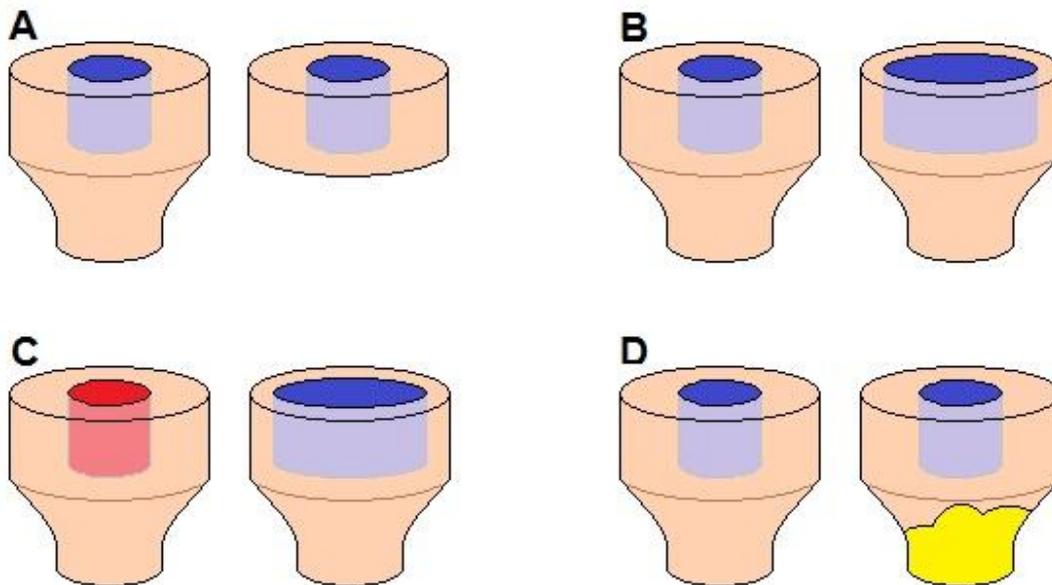
Figure 4.2: The biopsy dosage effect illustrated by punch biopsies for 8 hypothetical patients.

The shape of the biopsy images represents a typical punch biopsy sample, the elastic epidermal top layer of tissue retains its shape, whilst the softer and less elastic underlying tissue tends to constrict into a cone shape. Pink is human biopsy tissue without HPV infected cells, blue is HPV infected tissue expressing HPV genes and red is HPV infected tissue highly expressing HPV genes. In scenario A, two biopsies with an equal number of HPV infected cells are sampled, but the two biopsies vary in size. In scenario B, two identically sized biopsies have different numbers of HPV expressing cells. In scenario C, two biopsies of equal size are sampled, one has a small number of high HPV gene expressing cells, the other sample contains a large number of low HPV gene expressing cells. In scenario D, two identically sized biopsies have an identical number of HPV gene expressing cells but one sample has a large amount of adipose tissue (yellow).

4.1.4. Method Validation

4.1.4.1. Comparing RQ with Deep Sequencing

One of the VR series samples (VR49) was used for deep sequencing of the human and HPV transcriptome. The data and the analysis did not constitute part of this investigation and the data/results are not reported here. However, for the vulval sample, the RQ ratio and RPKM (Reads Per Kilobase of exon model per Million mapped reads) of *E2*, *E6* and *E7* can be compared. RPKM is the current de facto standard for quantifying expression levels for RNA deep sequencing data sets (Mortazavi *et al.*, 2008). A simple visual comparison of the difference in RQ between *E2*, *E6* and *E7* against the difference in RPKM suggests that RQ ratio difference between *E6* and *E7* are fairly accurate whilst *E2* RQ was slightly overestimated. RQ ratios are expressed relative to CaSki. That *E2* expression is slightly overestimated suggests

that CaSki *E2* mRNA transcription is slightly lower level than *E6* and *E7*. Nonetheless, the overestimation is not huge and the relative ratios of *E6* and *E7* appear to be very similar considering the methodological differences.

HPV gene	Deep sequencing		qPCR	
	Exon (nt)	RPKM	Region (nt)	RQ
E6	83-559	1397	99-178	1.08
E7	562-858	2347	739-816	2.61
E2	2755-3852	5124	3362-3426	10.30

Table 4.3: RPKM and RQ for *E6*, *E7* and *E2* of VR49.

4.1.4.2. *Reproducibility and Sensitivity*

The reproducibility of the qPCR assays is demonstrated by the CaSki calibration curves produced to determine PCR efficiency (Appendix II). There was little variation between runs. To further support this observation, the Ct values of 0.01x diluted CaSki cDNA from several independent RT treatments and independent qPCR runs are displayed in Figure 4.3. This figure shows that little variation was present between independent runs. The sensitivity of the assays was also determined, the lowest dilution of CaSki cDNA at which consistently repeatable results were produced for all assays was 1/78125 diluted (i.e. 0.000128x).
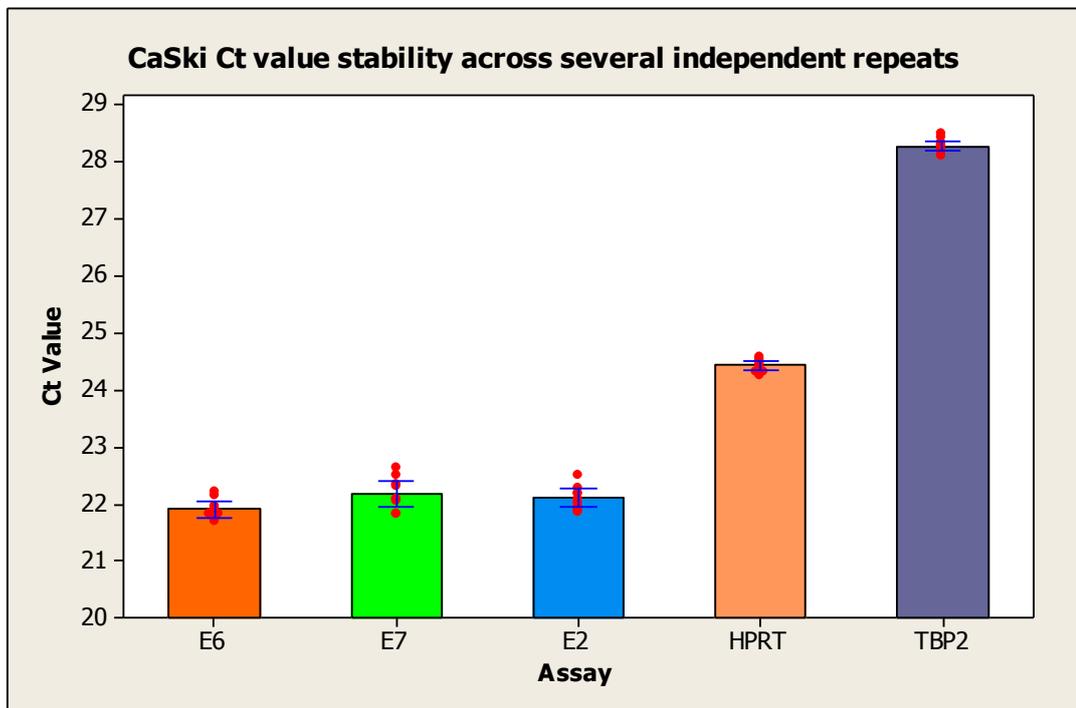
Figure 4.3: The stability of Ct value for multiple 0.01x CaSki repeats.

Included are 3 full independent repeats, repeated in triplicate (i.e. n=9). 95% confidence interval bars (blue bars) and individual values (red circles) are also included.

4.2. DNA Methylation

4.2.1. Method Development

Initial attempts at determining HPV methylation status utilised Methylation Specific PCR (MS-PCR). The EZ DNA methylation kit (Zymo Research Corporation, CA, USA) was the most commonly used BS conversion kit identified during a literature search. The conversion efficiency of the EZ DNA methylation kit was doubted at various stages during the study and the manufacturer's suggestions to improve conversion efficiency were adopted. BS conversion was tested using a fully methylated human control DNA (Zymo Research Corporation, CA, USA) and BS positive and negative controls and in each instance was found to be efficiently converted.

4.2.1.1. Methylation Specific-PCR

MS-PCR primers specific to part of the *L1*, *L2* and LCR regions of HPV as described by Fernandez *et al* (Fernandez *et al.*, 2009) were tested. PCRs specific to methylated or unmethylated DNA regions were performed (M and U primer pairs respectively). CaSki and SiHa had various methylated and unmethylated primer sets amplify, in some instances both primer sets would amplify whilst in others, neither would amplify. Rather than a range of methylation statuses, the results were more likely down to incomplete BS conversion, heterogeneous DNA methylation, a lack of primer specificity and/or incorrect PCR optimisation. Further optimisation was performed with little improvement. A fully methylated and a fully unmethylated HPV16 positive control DNA were required.

To produce the unmethylated control DNAs, two approaches were taken. First, CaSki and SiHa were treated with the demethylating agent 5-aza-2'-deoxycytidine using two treatment protocols (Fernandez *et al.*, 2009, Kalantari *et al.*, 2008a) with limited degrees of success. In addition, an *Escherichia coli* pSP64 plasmid vector containing a cloned HPV16 genome was acquired. To produce the methylated control, treatment of pSP64 with M.SssI methyltransferase was performed according to the protocol detailed by Brandsma *et al* (Brandsma *et al.*, 2009). Whilst this appeared to offer a suitable series of control DNAs, the limitations of MS-PCR had become apparent. MS-PCR was insensitive, required cloning to offer data of a quantitative nature and analysed only a few CpGs per assay.

4.2.1.2. Bisulfite Sequencing

Bisulfite sequencing was adopted and performed as described by Brandsma *et al* (Brandsma *et al.*, 2009). A series of bisulfite sequencing primers were trialled. CaSki and SiHa DNAs were bisulfite sequenced without cloning and aligned against published data. PCRs were optimised

successfully and included various annealing temperatures, MgCl$_2$ concentrations, PCR cycles and various attempts at nested PCR. It was apparent that quantitative methylation data was essential for the successful assessment of methylation status. Cloning of BS treated DNAs prior to bisulfite sequencing would provide semi-quantitative methylation data (Fernandez *et al.*, 2009) however access to a pyrosequencer was secured and the bisulfite sequencing assays were eventually abandoned.

4.2.1.3. Pyrosequencing

Pyrosequencing had the advantage of producing quantitative data, being rapid and high throughput, accurate, assessed multiple CpGs within a short region of DNA and did not require a cloning step. *E2*, LCR1 and LCR2 assay details and primer sequences were provided by Dr T Liloglou (University of Liverpool, personal communication, Dec 2010) and a basic PCR protocol was adapted from the ZymoTaq (Zymo Research Corporation, CA, USA) instruction booklet with some modifications.

The *E2*, LCR1 and LCR2 assays were chosen for several reasons. LCR2 represented the HPV promoter including E2BS$_3$ and E2BS$_4$ which have a suspected role in P97 regulation and were observed to be hypermethylated in most cervical cancers (Fernandez *et al.*, 2009). LCR1 represented the viral enhancer, another interesting region of the LCR, but importantly a region without E2BSs such that LCR methylation could be assessed independent of the potentially strong influence of E2BS methylation. *E2* was chosen as there was reason to suspect *E2* intragenic methylation might affect *E2* transcription/translation and it was a region outside of the LCR with which to assess HPV genomic methylation.

Initial attempts at pyrosequencing PCRs produced multiple secondary products. Further optimisation using gradient PCR and the addition of Qiagen HotStarTaq Q-mix (Qiagen, Hilden, Germany) appeared to minimise secondary product formation in some instances. Pyrosequencing was successfully applied to a range of DNAs. We had attempted to test the quantitative nature of the pyrosequencing reaction by mixing increasing amounts of hypomethylated control DNAs with hypermethylated control DNAs but the attempts were unsuccessful. Pyrosequencing had revealed that the 5-aza-2'-deoxycytidine treatment regime was not particularly effective at demethylating CaSki DNAs and whilst *in vitro* methyltransferase treatment worked well, pyrosequencing of the cloned pSP64 HPV16 genome consistently failed. Sequencing of the pSP64 HPV16 *E2* region revealed considerable sequence diversity that was likely responsible for this failure. Several DNA extracts and clones were trialled until this avenue of investigation was abandoned. Pyrosequencing works in a

"sequencing by synthesis" manner, each dNTP is added sequentially, incorporated and a light signal is produced. If inefficient extension of the sequencing product occurs (e.g. as a result of sequence diversity), part of the sequencing product will remain unextended until the next addition of the relevant dNTP. This causes amplification of sequencing products in several phases/frames, creating background noise and causing the runs to fail.

Subtle modifications to the dispensation protocol were performed in order to improve sequencing efficiency and diluting the PCR product 2/3 with water prior to sequencing improved data quality considerably. DNA input amount was optimised alongside PCR conditions and the volume of sepharose beads required for PCR product immobilisation.

Two additional assays were developed to target the *L1/L2* overlap and E2BS$_1$. The *L1/L2* overlap was chosen for several reasons. Whilst most studies considered methylation of only a few CpGs of the HPV genome, genome wide studies have assessed the entire methylome of HPV16 in a range of cervical disease states. Several regions of the HPV genome (especially *L1* and *L2*) were progressively methylated with increasing disease severity whilst the LCR remained largely unmethylated in all samples (Fernandez *et al.*, 2009). *L1/L2* was away from areas of transcriptional regulation (i.e. LCR) and potential transcriptional regulation (i.e. *E2* intragenic methylation). *L1* and *L2* were also regularly cited as areas with potential clinical utility in terms of a methylation biomarker. E2BS$_1$ was sought because it was the activator E2BS, we hypothesised that this E2BS might be differentially methylated from the repressor E2BSs in a manner which might be associated with high transcriptional activity. The E2BS$_1$ and *L1/L2* assays were difficult to design as both regions were flanked by long TA repeats and had considerable secondary structure. Whilst *L1/L2* was eventually optimised and applied to various cohorts, the E2BS$_1$ assay was deemed unsuitable due to repeated failure of the pyrosequencing internal controls.

4.2.2. Method Validation

The early stages of biomarker design require the development of an accurate and reproducible assay that is suitable for the range of material that it will be applied to. As such, the sensitivity and reproducibility of pyrosequencing were tested.

First, the limits of sensitivity were tested. Two independent dilution series of CaSki BS treated DNA were performed in duplicate. CaSki was diluted from 1x (BS treated DNA) down to 0.0001x. The success of these PCRs is depicted in Figure 4.4. CaSki 0.001x was the lowest concentration that produced a band that might be expected to be successfully pyrosequenced;

this represented an input of BS treated DNA of approximately 0.01 ng/PCR. For data collection, CaSki DNA was typically used at 0.01x concentration in PCR, whilst clinical samples were typically used at 0.1x concentration.
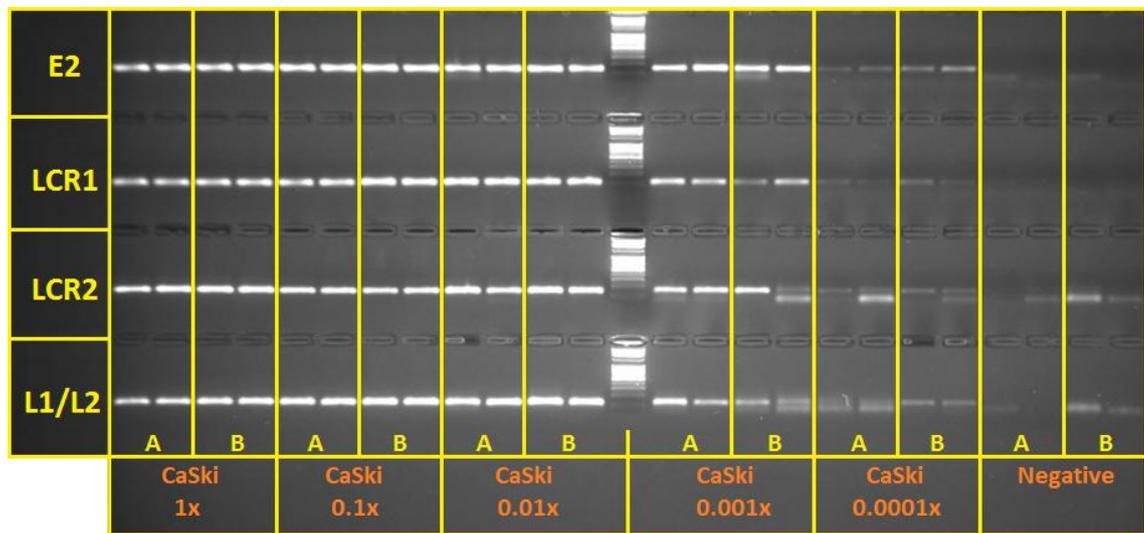
Figure 4.4: PCR success for a 1/10 dilution series of CaSki BS treated DNAs.

Five levels of a CaSki BS DNA 1/10 dilution were tested in the pyrosequencing PCRs. A and B represent two fully independent repeats (i.e. independent BS treatment and PCR). Negatives were full process negatives, i.e. water instead of DNA for BS treatment. The bands produced for negative reactions were small and most likely represented primer dimers.

These PCR products were then pyrosequenced. No difference in methylation level was observed between the diluted DNAs. The lowest concentration to produce a successful pyrosequencing result was 0.001x for all assays, and 0.0001x for *L1/L2* and LCR1. In summary, pyrosequencing can function with a very small amount of input DNA.

To test variability between pyrosequencing reactions, 6x CaSki repeats were performed. The results are summarised in Appendix III. There was little variation between runs.

4.3. Conclusions

qPCR and various assays for measuring DNA methylation were evaluated. Protocols were produced and optimised. Relative quantification was performed using a modified variant of the Pfaffl equation. For methylation analysis, pyrosequencing was chosen because it measures multiple CpGs and is quantitative, high throughput, sensitive and reproducible.

Studies in the various cohorts demonstrate that pyrosequencing can be applied to DNA extracted from several types of material including LBC samples.

Chapter 5. The NSC Study - DNA Methylation and Disease Grade

5.1. Results

This study was performed in order to investigate the relationship between DNA methylation and cervical disease. A series of HPV16 positive LBC samples with normal cytology (normal group, n=20) and severe dyskaryosis (severe group, n=20) and a series of FFPE cervical cancer DNAs (cancer group, n=27) were analysed using the pyrosequencing assays targeting the *E2*, LCR1, LCR2 and *L1/L2* regions. Several of these samples failed various pyrosequencing assays (most likely as a result of DNA quality). Poor quality data was excluded from the analysis which meant that the results of 17 normal samples, 20 severe samples and 24 cancer samples were used in the final statistical analysis.

The DNA methylation data was analysed by addressing a hierarchical series of hypotheses about HPV16 DNA methylation and the relationship between methylation and cervical disease grade. A description of the statistical analyses used, including a description of the requirements is presented in the materials and methods section.

5.1.1. Hypothesis 1: DNA Methylation Varied among Regions of the Genome

There appeared to be considerable variation in the level of DNA methylation among the four HPV16 regions (Figure 5.1). To test for a significant difference among the regions, a two-way ANOVA GLM was ideal as it considered DNA methylation for all three regions simultaneously. The two-way ANOVA found significant differences among the mean DNA methylation of each region (P=0.000). In addition, the two-way ANOVA GLM partly tested the next hypothesis and found significant differences among the three disease groups (P=0.000). One violation of the two-way ANOVA existed, the data did not have equal variances (P=0.000) however no suitable non-parametric alternative exists.
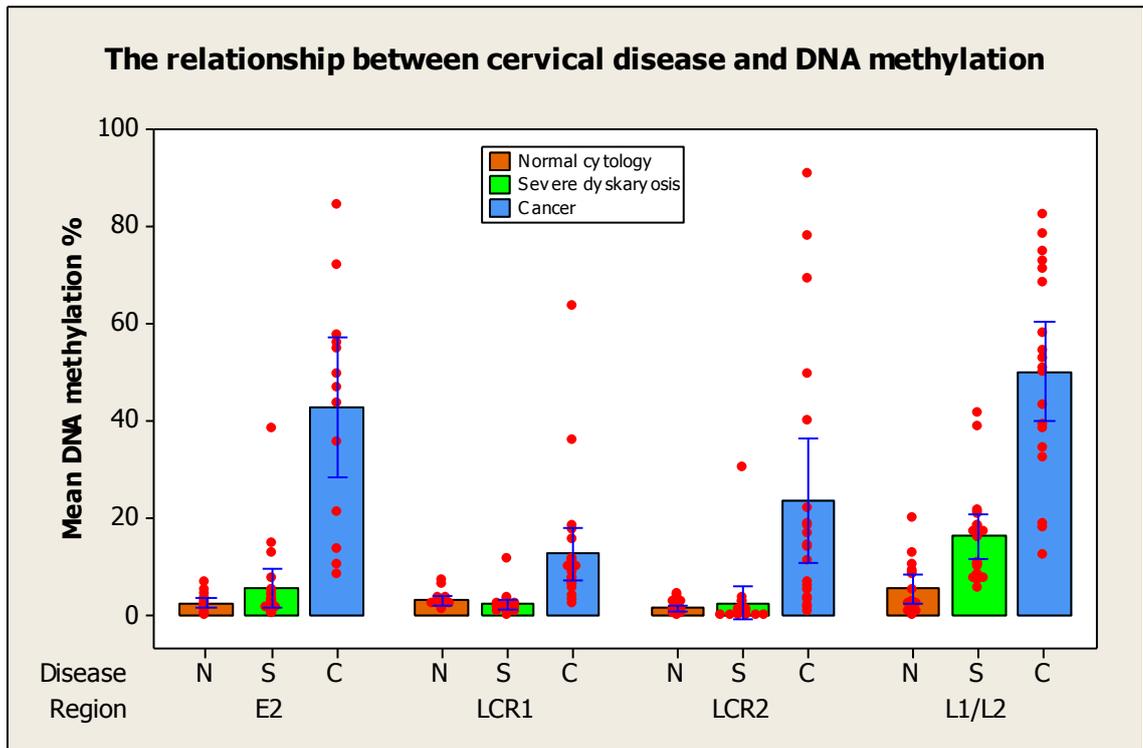
Figure 5.1: The relationship between disease grade and DNA methylation for four regions of the HPV16 genome.

Coloured bars represent mean methylation and individual values are plotted as red circles. 95% +/- confidence intervals for mean are included.

5.1.2. Hypothesis 2: There is a Difference in DNA Methylation among the Different Disease Grades

In general, there appeared to be a pronounced increase in DNA methylation within each region with increasing grades of cervical disease, with the highest seen in the cancers and the lowest in normal cytology samples (Figure 5.1). There was also considerable variation in the amount of DNA methylation at each HPV16 region within the disease groups. For example, the HPV16 *E2* region of individual cancer samples ranged from 8.5-85% methylated. This massive variation suggests that the cause or consequences of HPV DNA methylation were not common to each of the cancers or that a higher level of complexity existed.

The simplest way of comparing DNA methylation was to compare the means of the three disease groups statistically. The previously performed two-way ANOVA GLM test revealed a significant difference among the disease groups (P=0.000) but the GLM does not permit pairwise comparisons. As such, DNA methylation at each region for each of the three disease groups was compared using one-way ANOVA tests (one test for each region) and then Tukey-Kramer tests were used to identify pairwise differences between the disease groups. The methylation data for several regions violated one or more of the assumptions of the one-way

ANOVA test; however because these statistics were exploratory and hypothesis generating in nature, the advantages of the one-way ANOVA and Tukey-Kramer over the non-parametric Kruskall-Wallis test and multiple Mann-Whitney U tests outweighed the consequences of the violations. Significant results must be interpreted with caution and the one-way ANOVAs performed ideally require a Bonferroni correction of P=0.013 to be considered. To increase the reliability of the interpretation of the results, Kruskall-Wallis tests were also performed. The results of the one-way ANOVA, Tukey-Kramer and Kruskall-Wallis tests are presented in Table 5.1. The details of any violations of the assumptions of the ANOVA are included in the table legend.

There were highly significant differences among the means of the three disease groups for all tested regions. There were also significant differences between the cancer group and the normal/severe groups for every region and a significant difference between the severe and normal groups for the *L1/L2* region.

	HPV16 Region			
	E2	LCR1	LCR2	*L1/L2*
One-way ANOVA (P-value)	0.000	0.000[*1]	0.000[*2]	0.000[*3]
Significant Tukey-Kramer results	C vs. S, C vs. N	C vs. S, C vs. N	C vs. S, C vs. N	C vs. S, C vs. N, S vs. N
Kruskall-Wallis (P-value)	0.000	0.000	0.000	0.000

Table 5.1: The results of one-way ANOVA tests, Tukey-Kramer pairwise comparison tests and Kruskall-Wallis tests to compare the differences among the three disease groups.

One-way ANOVA tests were performed to compare the means of the three disease groups in each HPV region. P is significant at P<0.05 or P<0.013 (Bonferroni corrected). The Tukey-Kramer pairwise comparisons significant at P=0.05 are presented, here N is the normal cytology group, S is the severe dyskaryosis group and C is the cancer group, e.g. "C vs. S" represents a significant difference between the mean methylation of C and S for that region. Kruskall-Wallis P-values are included and represent a difference among the medians of the three groups. Violations of the one-way ANOVA assumptions *1; unequal variances (P=0.042) and non-normally distributed residuals (P=<0.005), *2; unequal variances (P=0.004), *3; unequal variances (P=0.026).

5.1.3. Hypothesis 3: DNA Methylation Varies within Regions

The pyrosequencing assays each cover several CpGs so as well as differences among regions, differences among individual CpGs within each region can be determined. There appeared to be variation in the amount of methylation of different CpGs within each site, and a rather well conserved pattern of methylation at each CpG, i.e. the same CpGs seem to be the most methylated in each disease group despite large differences between the disease groups overall. The E2-4 (the 4th CpG of the *E2* region) for instance was highly methylated in all

disease groups relative to the other CpGs, conversely, E2-3 was low in each instance despite the CpGs being separated by only 14 bp. CpGs E2-1, E2-2 and E2-3 are very close together (E2-2/E2-3 are directly adjacent and E2-1/E2-2 are separated by 1 bp) and yet DNA methylation among these CpGs varied considerably, possibly suggesting that a very specific mechanism regulates HPV DNA methylation. The CpG specific methylation patterns also appeared to be well conserved for most of the regions, the "N" shaped pattern (i.e. CpGs L1L2-1 and L1L2-3 lowly methylated and L1L2-2 and L1L2-4 more methylated) seen in the *L1/L2* region was particularly evident in each disease group (Figure 5.5).

5.1.3.1. E2

Kruskall-Wallis one-way analysis of variance was used to determine whether a significant difference existed among the median methylation values at each CpG of the *E2* region (Figure 5.2). Bonferroni correction for multiple analyses means that a P=0.006 should be considered. There was a significant difference among the median values of methylation of the 8 CpGs of the *E2* region in the normal cytology sample group (P=0.000) and the severe dyskaryosis group (P=0.001) but not the cervical cancer group (P=0.101).

5.1.3.2. LCR1

Kruskall-Wallis one-way analysis of variance was used to determine whether a significant difference existed among the median methylation values at each CpG of the LCR1 region (Figure 5.3). Bonferroni correction for multiple analyses means that P=0.017 should be considered. There was no significant difference in the median methylation at each CpG for the normal group (P=0.517), however there was a significant difference among the median methylation values of each CpG for the severe (P=0.000) and cancer (P=0.005) groups.

5.1.3.3. LCR2

Kruskall-Wallis one-way analysis of variance was used to determine whether a significant difference existed among the median methylation values at the CpGs of the LCR2 region (Figure 5.4). Bonferroni correction for multiple analyses means that a P=0.010 should be considered. There was a significant difference among the median methylation for the normal (P=0.000) and an almost significant difference for the severe group (P=0.028), but not for the cancer group (P=0.821).

5.1.3.4. L1/L2

Kruskall-Wallis one-way analysis of variance was used to determine whether a significant difference existed among the median methylation values at the CpGs of the *L1/L2* region (Figure 5.5). Bonferroni correction for multiple analyses means that a P=0.013 should be

considered. There was a significant difference among the median methylation for the normal (P=0.011), severe group (P=0.000) and cancer group (P=0.001).
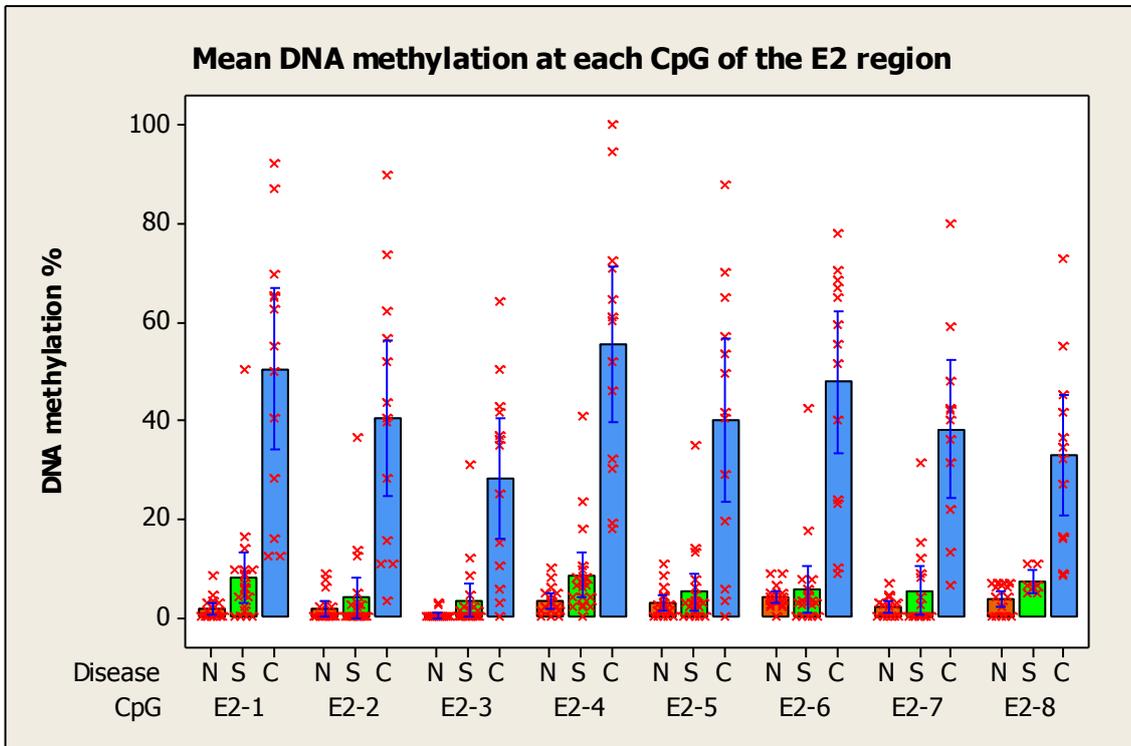
Figure 5.2: Variation in CPG methylation with disease grade for the *E2* region of HPV16.

The mean amount of methylation at each of the 8 CpGs of the *E2* region are compared for each of the three sample groups (normal=N, severe=S and cancer=C). Error bars indicate 95% confidence intervals for mean. Red crosses indicate individual sample values.
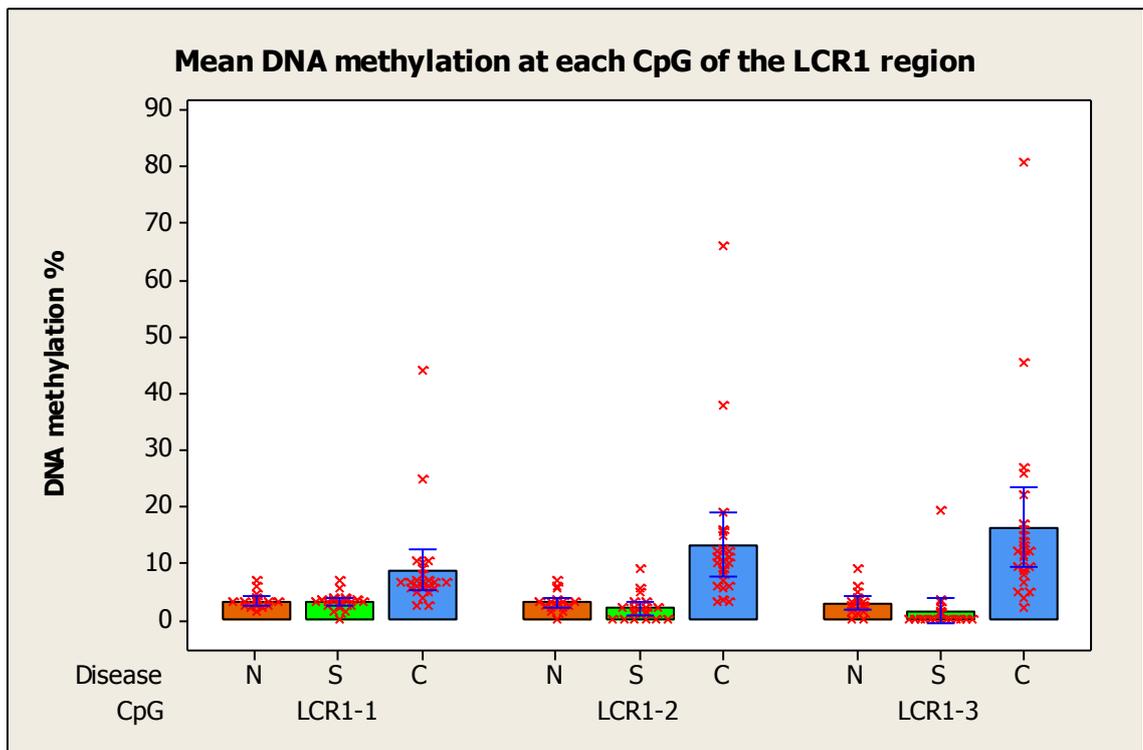
Figure 5.3: Variation in CPG methylation with disease grade for the LCR1 region of HPV16.

The mean amount of methylation at each of the 3 CpGs of the *E2* region are compared for each of the three sample groups (normal=N, severe=S and cancer=C). Error bars indicate 95% confidence intervals for mean. Red crosses indicate individual sample values.
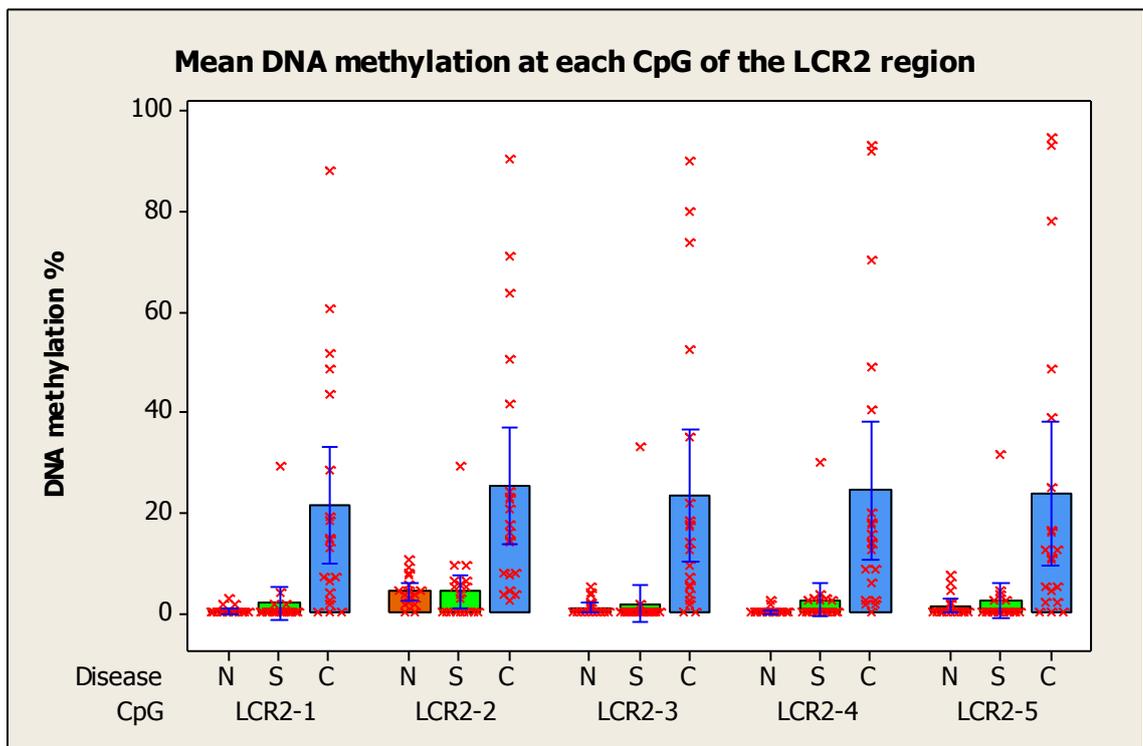
Figure 5.4: Variation in CPG methylation with disease grade for the LCR2 region of HPV16.

The mean amount of methylation at each of the 5 CpGs of the *E2* region are compared for each of the three sample groups (normal=N, severe=S and cancer=C). Error bars indicate 95% confidence intervals for mean. Red crosses indicate individual sample values.
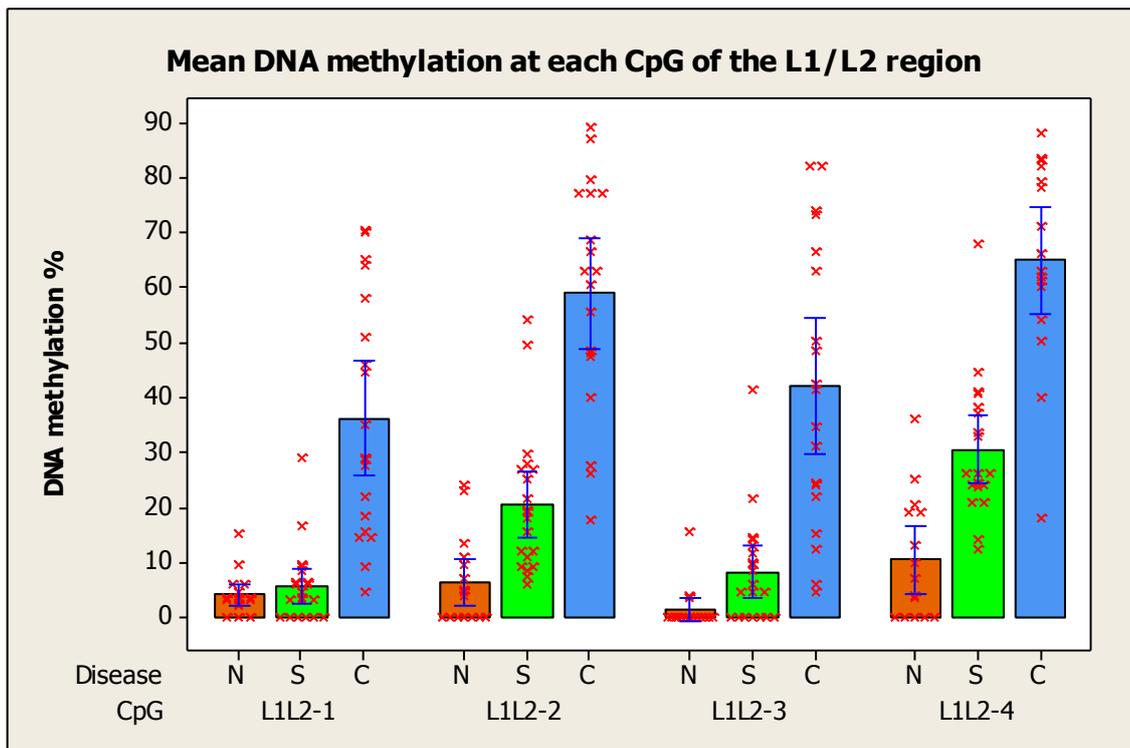
Figure 5.5: Variation in CPG methylation with disease grade for the *L1/L2* region of HPV16.

The mean amount of methylation at each of the 4 CpGs of the *E2* region are compared for each of the three sample groups (normal=N, severe=S and cancer=C). Error bars indicate 95% confidence intervals for mean. Red crosses indicate individual sample values.

5.1.4. Hypothesis 4: E2BS Methylation is Associated with Disease

Previous studies have observed a relationship between E2BS methylation and disease severity. LCR2 covers two E2BSs (E2BS$_3$ and E2BS$_4$) however LCR2 methylation was low in almost all precancerous neoplasias. Of the cervical cancers, only 5 samples had LCR2 methylation above 40% whilst the rest were considerably less methylated, there was not a strong association between LCR2 methylation and cancer.

5.1.5. Hypothesis 5: HPV DNA Methylation is Clinically Relevant

One of the primary reasons for performing pyrosequencing on the three disease groups was to determine whether HPV methylation represents a suitable biomarker of cervical disease. An aspect of this was to determine what features of the methylation data were the most informative clinically. In order to assess which of the four regions and which of the twenty CpGs analysed were the most different among the three disease grades, multiple statistical analyses were applied.

The regions or single CpGs with the greatest significant differences in methylation among the three disease grades are more likely to be of use in future biomarker studies. One-way ANOVA

and Tukey-Kramer pairwise comparison tests were applied to the methylation data for each of the regions and CpGs, and the significance of the results compared. The statistics performed in other parts of this investigation were exploratory and hypothesis generating in nature and so a degree of deviation from the assumptions of ANOVA, or P-values below the Bonferroni corrected P-value were acceptable providing that the results were interpreted with caution. Here, a direct comparison of the P-values of the tests was being performed in order to determine which comparisons were the most significant; as such the violation of assumptions may have had a considerable effect on the conclusions drawn. In order to aid interpretation of the results, Kruskall-Wallis non-parametric ANOVAs were also applied to compare differences among the median of each group. In addition to these analyses, the overall spread of the methylation data for each CpG and the size of the confidence intervals for the graphically plotted methylation data were considered when deciding which region or CpG appears to be the most different among the three disease groups.

5.1.5.1. *Which of the Four HPV Regions is the Most Informative?*

One-way ANOVA, Tukey-Kramer and Kruskall-Wallis tests were performed on the mean methylation data for each pyrosequencing assay, the results are summarised in Table 5.2. A Bonferroni corrected P-value of 0.013 was used. In order to determine which of the four regions was the most informative; the H-value, F-value and Tukey-Kramer results were considered. *E2*, LCR1 and *L1/L2* had the highest F values however LCR1 violated the assumptions of the ANOVA so the results must be interpreted with caution. Only *L1/L2* was differently methylated between the normal and severe groups and the *E2* region had the lowest H-value. Of all the regions, mean *L1/L2* methylation appeared to be the most informative.

Next, similar statistics were applied to the individual CpG methylation data of each region in order to identify the most informative CpGs.

	HPV16 Region			
	E2	LCR1	LCR2	*L1/L2*
Kruskall-Wallis P-value	0.000	0.000	0.000	0.000
H value (DF)	25.26 (2)	37.40 (2)	26.41 (2)	36.75 (2)
ANOVA P-value	0.000	0.000	0.000	0.000
F value (DF)	48.04 (2)	41.32 (2)	26.70 (2)	54.45 (2)
Equal variances	Yes	Close	No	Close
Normal residuals	Yes	No	Yes	Yes
	Significant Tukey-Kramer Pairwise Comparison?			
	E2	LCR1	LCR2	*L1/L2*
Cancer vs. Normal	Yes	Yes	Yes	Yes
Cancer vs. Severe	Yes	Yes	Yes	Yes
Severe vs. Normal	No	No	No	Yes

Table 5.2: The results of multiple statistical analyses on methylation data all regions.

The non-parametric Kruskall-Wallis P-values and H-values represent the significant differences among the medians of each disease group, higher significance is represented by a lower P-value and a larger H-value. ANOVA P-values and F-values are accompanied by degrees of freedom (DF) and two statements for violations of the ANOVA assumptions (equal variances and normally distributed residuals), here three degrees of success are present, No; a violation (P=<0.01), Close; P=0.01-0.05 and Yes; P>0.05. Finally, the significance of Tukey-Kramer tests are listed for each comparison of disease group, where Yes = significance at P=0.05 and No = no significance.

5.1.5.2. E2 CpG Methylation: Which E2 CpG is the Most Informative?

One-way ANOVA, Tukey-Kramer and Kruskall-Wallis tests were performed for each *E2* CpG, the results are summarised in Table 5.3. Each CpG was significantly differently methylated among the disease groups and met the Bonferroni corrected P-value of 0.006 in every instance. In order to decide which of the 8 *E2* CpGs was the most differently methylated among disease grades, the H-value, F-value and Tukey-Kramer results were considered. E2-1, E2-2 and E2-4 had the highest (i.e. most significant) Kruskall-Wallis H-values and ANOVA F-values. E2-2 data violated the assumptions of the ANOVA and so pairwise comparison was less reliable, there was also no difference in mean E2-2 methylation between the severe and the normal groups.

There were however significant differences in E2-1 and E2-4 methylation among each of the three disease groups. When considering the individual value plot for *E2* (Figure 5.2), it was difficult to decide which of E2-1 and E2-4 was more informative as both had overlaps among the three groups, and for the cancer group especially there was a wide range of individual methylation values. Values overlap for the severe and normal groups, but in both instances

narrower confidence intervals exist than for the *E2* region as a whole suggesting that individual CpG methylation might prove more useful.

	E2-1	E2-2	E2-3	E2-4	E2-5	E2-6	E2-7	E2-8
Kruskall-Wallis P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
H-value	30.94	26.91	25.97	30.22	17.47	26.19	21.93	22.87
(DF)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)
ANOVA P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F-value	42.49	36.61	33.48	49.61	18.41	37.30	27.05	26.20
(DF)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)
Equal variances	Yes	Yes	No	Yes	Yes	Close	Yes	Close
Normal residuals	Close	No	No	Yes	Close	No	No	No
	Significant Tukey-Kramer Pairwise Comparison?							
	E2-1	E2-2	E2-3	E2-4	E2-5	E2-6	E2-7	E2-8
Cancer vs. Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cancer vs. Severe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Severe vs. Normal	Yes	No	No	Yes	No	No	No	No

Table 5.3: The results of multiple statistical analyses on methylation data for the *E2* CpGs.

The non-parametric Kruskall-Wallis P-values and H-values represent the significant differences among the medians of each disease group, higher significance is represented by a lower P-value and a larger H-value. ANOVA P-values and F-values are accompanied by degrees of freedom (DF) and two statements for violations of the ANOVA assumptions (equal variances and normally distributed residuals), here three degrees of success are present, No; a violation (P=<0.01), Close; P=0.01-0.05 and Yes; P>0.05. Finally, the significance of Tukey-Kramer tests are listed for each comparison of disease group, where Yes = significance at P=0.05 and No = no significance.

5.1.5.3. *LCR1 CpG Methylation: Which LCR1 CpG is the Most Informative?*

One-way ANOVA, Tukey-Kramer and Kruskall-Wallis tests were performed for each LCR1 CpG and the results are summarised in Table 5.4. Each CpG was significantly differently methylated among the disease groups and met the Bonferroni corrected P-value 0.017 in every instance. LCR1-2 and LCR1-3 had the most significant (i.e. highest) H-values and F-values however there was no significant difference in LCR1-2 methylation between the severe and normal groups, and LCR1-3 data failed one of the ANOVA assumptions making interpretation unreliable. In addition, there was considerable variation among individual data points in each disease group and there was substantial overlap between the severe and normal groups (Figure 5.3). For differentiating between cancer and normal or severe samples, methylation level of the LCR1-2 and LCR1-3 CpGs may prove clinically useful.

	LCR1-1	LCR1-2	LCR1-3
Kruskall-Wallis P-value	0.000	0.000	0.000
H value (DF)	26.26 (2)	36.49 (2)	37.71 (2)
ANOVA P-value	0.000	0.000	0.000
F value (DF)	18.79 (2)	35.20 (2)	48.29 (2)
Equal variances	Yes	Yes	Yes
Normal residuals	No	Yes	No
	Significant Tukey-Kramer Pairwise Comparison?		
	LCR1-1	LCR1-2	LCR1-3
Cancer vs. Normal	Yes	Yes	Yes
Cancer vs. Severe	Yes	Yes	Yes
Severe vs. Normal	No	No	Yes

Table 5.4: The results of multiple statistical analyses on methylation data for the LCR1 CpGs.

The non-parametric Kruskall-Wallis P-values and H-values represent the significant differences among the medians of each disease group, higher significance is represented by a lower P-value and a larger H-value. ANOVA P-values and F-values are accompanied by degrees of freedom (DF) and two statements for violations of the ANOVA assumptions (equal variances and normally distributed residuals), here three degrees of success are present, No; a violation (P=<0.01), Close; P=0.01-0.05 and Yes; P>0.05. Finally, the significance of Tukey-Kramer tests are listed for each comparison of disease group, where Yes = significance at P=0.05 and No = no significance.

5.1.5.4. LCR2 CpG Methylation: Which LCR2 CpG is the Most Informative?

One-way ANOVA, Tukey-Kramer and Kruskall-Wallis tests were performed on the methylation data for each LCR2 CpG and the results are summarised in Table 5.5. All comparisons were significant even when the Bonferroni corrected P-value 0.010 was considered. None of the H-values or F-values were exceptionally high but CpGs LCR2-3 and LCR2-4 were the most significantly different among the groups. However the data for all of the comparisons did not satisfy the assumptions of the ANOVA and none of the subsequently unreliable Tukey-Kramer comparisons were able to differentiate between methylation of the severe and normal samples. Methylation data for the LCR2 region may be useful for differentiating cancer samples from normal or severe disease however the data analyses had violations and were not particularly suitable for the analyses performed. Also, the range of DNA methylation in each disease group overlapped for every CpG (Figure 5.4).

	LCR2-1	LCR2-2	LCR2-3	LCR2-4	LCR2-5
Kruskall-Wallis P-value	0.000	0.000	0.000	0.000	0.000
H value (DF)	26.97 (2)	20.27 (2)	29.63 (2)	30.72 (2)	20.86 (2)
ANOVA P-value	0.000	0.000	0.000	0.000	0.000
F value (DF)	27.08 (2)	18.00 (2)	28.88 (2)	33.71 (2)	17.26 (2)
Equal variances	No	Yes	Close	No	Close
Normal residuals	No	Close	No	No	No
	Significant Tukey-Kramer Pairwise Comparison?				
	LCR2-1	LCR2-2	LCR2-3	LCR2-4	LCR2-5
Cancer vs. Normal	Yes	Yes	Yes	Yes	Yes
Cancer vs. Severe	Yes	Yes	Yes	Yes	Yes
Severe vs. Normal	No	No	No	No	No

Table 5.5: The results of multiple statistical analyses on methylation data for the LCR2 CpGs.

The non-parametric Kruskall-Wallis P-values and H-values represent the significant differences among the medians of each disease group, higher significance is represented by a lower P-value and a larger H-value. ANOVA P-values and F-values are accompanied by degrees of freedom (DF) and two statements for violations of the ANOVA assumptions (equal variances and normally distributed residuals), here three degrees of success are present, No; a violation (P=<0.01), Close; P=0.01-0.05 and Yes; P>0.05. Finally, the significance of Tukey-Kramer tests are listed for each comparison of disease group, where Yes = significance at P=0.05 and No = no significance.

5.1.5.5. L1/L2 CpG Methylation: Which L1/L2 CpG is the Most Informative?

One-way ANOVA, Tukey-Kramer and Kruskall-Wallis tests were performed on the methylation data for each *L1/L2* CpG and the results are summarised in Table 5.6. P-values were below the Bonferroni corrected threshold of 0.013 in every instance. Both L1L2-2 and L1L2-3 had high (i.e. highly significant) F-values and H-values. L1L2-3 data was very close to satisfying the assumptions of the ANOVA and showed a significant difference among all three groups. L1L2-2 data did not satisfy the assumptions of the ANOVA so the interpretation was less reliable, but there was still a significant difference in methylation among the three disease groups.

L1L2-1 methylation was different between normal or severe and cancer, but not between normal and severe. In terms of clinical utility, the most significant differences among the three disease groups were observed for L1L2-2 methylation. Whilst methylation of L1L2-2, L1L2-3 and L1L2-4 was significantly different among the three disease groups, the smallest confidence intervals and the smallest degrees of overlap among the groups were observed for L1L2-2 and L1L2-4 (Figure 5.5). In summary, the biggest differences were observed for L1L2-2 methylation however methylation of all *L1/L2* CpGs seemed likely to be useful, and could potentially be more useful than mean *L1/L2* methylation alone.

	L1L2-1	L1L2-2	L1L2-3	L1L2-4
Kruskall-Wallis P-value	0.000	0.000	0.000	0.000
H value (DF)	30.56 (2)	37.38 (2)	34.7 (2)	33.39 (2)
ANOVA P-value	0.000	0.000	0.000	0.000
F value (DF)	32.14 (2)	48.10 (2)	43.58 (2)	33.66 (2)
Equal variances	Yes	No	Yes	No
Normal residuals	No	Close	Close	No
	Significant Tukey-Kramer Pairwise Comparison?			
	L1L2-1	L1L2-2	L1L2-3	L1L2-4
Cancer vs. Normal	Yes	Yes	Yes	Yes
Cancer vs. Severe	Yes	Yes	Yes	Yes
Severe vs. Normal	No	Yes	Yes	Yes

Table 5.6: The results of multiple statistical analyses on methylation data for the *L1/L2* CpGs.

The non-parametric Kruskall-Wallis P-values and H-values represent the significant differences among the medians of each disease group, higher significance is represented by a lower P-value and a larger H-value. ANOVA P-values and F-values are accompanied by degrees of freedom (DF) and two statements for violations of the ANOVA assumptions (equal variances and normally distributed residuals), here three degrees of success are present, No; a violation (P=<0.01), Close; P=0.01-0.05 and Yes; P>0.05. Finally, the significance of Tukey-Kramer tests are listed for each comparison of disease group, where Yes = significance at P=0.05 and No = no significance.

5.1.5.5.1. Summary

Comparison of individual CpGs determined that a number of CpGs were differentially methylated with respect to disease grade. Methylation of these CpGs might represent a potential biomarker to differentiate between either cancer and severe or normal samples, or even differentiate all three groups. A larger data set would improve the power of the analyses and increase the chances of identifying suitable CpGs. It would also permit the application of statistics more typically associated with biomarker studies. Currently, the CpGs E2-1, E2-4, LCR1-2 and LCR1-3 show potential to differentiate between cancer samples and those with a normal cytology or severe dyskaryosis whilst the L1L2-2 and L1L2-4 CpGs shows good potential as a way of differentiating among all three groups; normal cytology, severe dyskaryosis and cancers.

Differences among the disease grades were more significant when the *E2* region was considered as a whole rather than for individual CpGs. However, significant differences specifically between the normal cytology and severe dyskaryosis groups were present for individual CpGs, but not for mean *E2* regional methylation. Methylation differences among all three groups were observed for individual CpGs within the LCR and were different between the normal and severe disease groups whilst mean LCR1 methylation was not. Mean LCR2

methylation was also not different between the normal and severe groups regardless of whether single CpGs or the whole regions were considered and no LCR2 comparisons had exceptionally high F/H values.

Finally, mean *L1/L2* methylation had differences of greater significance among the three disease groups than individual CpGs did. The overall spread and degree of overlap for the mean *L1/L2* methylation data was also greater than was seen for individual CpGs of the *L1/L2* region. The *L1/L2* region appeared to have potential as a biomarker for cervical disease. Further work on a larger data set is required to better determine whether methylation at individual CpGs or the mean of all four CpGs is more informative. Ideally, HPV integration data would also be available as the high and variable methylation seen in the cancer set could be a result of HPV DNA integration.

5.2. Discussion

5.2.1.1. Aims

This study was performed as a preliminary study prior to testing for DNA methylation in other clinical cohorts. It addressed the need to determine the relationship between disease grade and DNA methylation such that data collected for other clinical cohorts could be considered in context. We were also mindful of the process of biomarker development recommended by Cancer Research UK (CRUK) (Appendix I) and this study had the capability of fulfilling several aspects of the early development requirements, i.e. that there was variation within our target population and that there was a difference among subjects with low grade disease, high grade disease and cancer. It was also used to determine whether the data from this study correlated with published studies performed using a variety of methods and targeting a variety of HPV regions. There was considerable variation in the results of previous studies especially when concerning the methylation status of the LCR, this was an issue that we had aimed to clarify.

5.2.1.2. Strengths

This study had aimed to identify potential biomarkers to discriminate between HPV positive women with and without signs of cervical neoplasia. The potential application would be on HPV positive LBC material, this was reflected by the study which was primarily conducted on LBC samples. A series of DNAs from histologically confirmed cervical cancers were also included in this analysis to determine whether the trends observed in the cytology samples were also present in cancers and to extend the severity of disease grades studied. The sample

size (n=67) was above average when compared to similar studies and various significant results were achieved when the data were tested statistically.

Until recently, pyrosequencing had not been widely applied for HPV methylation studies, likely due to the expense and limited availability of equipment. However, pyrosequencing has the advantage of producing fully quantitative results from a heterogeneous mix of DNAs. Consistent with other studies, we found bisulfite sequencing without cloning to be a poor method for determining DNA methylation (Kalantari *et al.*, 2004). Bisulfite sequencing with cloning is semi-quantitative and the results are highly affected by low clone number (commonly observed in published studies) such that methylation frequencies are not representative. Pyrosequencing does not have these issues. In summary, the study design and the assay used were a good way of testing the hypotheses.

5.2.1.3. Weaknesses

One of the weaknesses of the study was that while the normal cytology and severe dyskaryosis DNAs were extracted from LBC material collected during routine screening, the cancer group DNAs were extracted from FFPE cancer biopsies. As the LBC samples were collected as part of an anonymous prevalence study, we did not have corresponding histology for the severe dyskaryosis samples and therefore did not know the proportion that actually had CIN. However, of 3,019 women referred for moderate dyskaryosis or worse, 2,556 (84%) showed CIN2 or worse (Cervical Screening Wales, 2011). It is therefore highly likely that the majority of samples with severe dyskaryosis would be associated with clinically significant disease (i.e. CIN2+).

The degraded state of DNA from the FFPE cancer samples also affected the success of the pyrosequencing assays. Multiple repeats of the PCRs and sequencing reactions were required to achieve an acceptable quality of data for some of the cancer samples, which was still typically of a lower quality than data collected from LBC material. Samples with particularly poor quality data (as determined objectively by the pyrosequencing analysis software) were excluded from the analysis.

To fully assess the potential of DNA methylation as a biomarker would require a much larger study designed specifically to test and identify disease biomarkers in a suitable population of LBC samples, with associated histology data. This was not achieved by this study.

Finally, there were several violations of the assumptions of the ANOVA. The tests were used in spite of these violations on the basis that the multifactorial ANOVA produced data confirmed by other tests. It is common practice to use the tests with violations in an exploratory nature

to generate hypotheses under the premise that the results are interpreted with an appropriate degree of caution. The sample set size would ideally be larger as this would both improve the reliability of the data and increase the likelihood of the data being suitable for testing without violations being incurred.

5.2.1.4. Comparison to Other Studies

HPV methylation has quite recently become a popular avenue of investigation, several published studies have related HPV methylation and HPV-associated disease grade and in several instances, there was considerable variation in the results.

The association between LCR methylation and cervical disease varies among studies. Most studies reported increased LCR methylation to be associated with higher grade disease (Kalantari *et al.*, 2004, Bhattacharjee and Sengupta, 2006a, Ding *et al.*, 2009, Brandsma *et al.*, 2009, Hong *et al.*, 2008) whilst some report the opposite (Badal *et al.*, 2003, Hublarova *et al.*, 2009, Patel *et al.*, 2012, Piyathilake *et al.*, 2011). One of the major points of contention appears to be LCR methylation of very low grade disease samples; several studies, including ours did not see LCR methylation in LSIL/normal cytology/no CIN samples (Ding *et al.*, 2009, Fernandez *et al.*, 2009, Bhattacharjee and Sengupta, 2006a) whilst others saw higher methylation in these samples than in higher grade precancerous disease (Hong *et al.*, 2008, Patel *et al.*, 2012, Kalantari *et al.*, 2004, Badal *et al.*, 2003). Nonetheless, whilst there was variation in the published methylation data, there was no valid reason to doubt the data produced by this study as the results were consistent across a diverse range of material. For the cell lines study for instance (Chapter 7: 7.1.2 & 7.1.3.3), cell lines derived from precancerous lesions (i.e. W12Ser7, PC08 and PC09) were rarely methylated in the LCR at early passage and LCR methylation was not commonly observed in VIN2/3 (Chapter 8: 8.2.3).

The published studies encompass a range of material, a range of methodologies and determine the significance of DNA methylation from a variety of perspectives (study methodologies summarised in Appendix VI). Differences between LBC and biopsy material is liable to contribute to some of the differences as LCR methylation is known to be both intrinsically associated with cellular differentiation and to be heterogeneous between different parts of the same lesion (Vinokurova and von Knebel Doeberitz, 2011). LBC samples contain only desquamated cervical cells whilst biopsy material contains cells from various layers of the epithelium. Methodological differences, especially where DNA methylation was determined using non-quantitative assays (i.e. direct bisulfite sequencing/MS-PCR, or methylation sensitive

restriction enzymes) are also likely to have a large influence on the results and is observed in the literature (Badal *et al.*, 2003, Fernandez *et al.*, 2009). Further, the method of reporting results, even where the same assays have been used varied considerably; e.g. Hong *et al* used pyrosequencing to evaluate LCR methylation, including 5 CpGs covered by our LCR2 assay (Hong *et al.*, 2008). Methylated samples were those with "some degree of methylation", if some means >0% then every sample in all disease groups for our study had "some degree of methylation" at LCR1 and/or LCR2.

Other studies have observed an association between increased *L1* and/or *L2* methylation and higher disease grades (Fernandez *et al.*, 2009, Kalantari *et al.*, 2009, Sun *et al.*, 2011, Mirabello *et al.*, 2012). Genome wide studies have also revealed that *E2*/E4 methylation appears to be related to disease grade (Fernandez *et al.*, 2009, Mirabello *et al.*, 2012). Fernandez *et al* used bisulfite sequencing, with cloning to assess the entire methylome of HPV16 in cytologically normal, CIN and SCC samples. Several regions of the HPV genome (especially *L1* and *L2*) were progressively methylated with increasing disease severity whilst the LCR remained largely unmethylated in all samples (Fernandez *et al.*, 2009). Sections of missing data were observed between the *E2* and *L2* regions (i.e. parts of *E2*, *E4*, *E5*, *L1* and *L2*) for most of the SCC and CIN samples, these were explained as deletions (Fernandez *et al.*, 2009) which seems somewhat unlikely. Nonetheless, there was a good agreement between the results of this study and our study. E2BS methylation has been observed for cancer samples in several studies; whilst 5 cancer samples had what might be described as LCR2 hypermethylation, the rates of hypermethylation of the LCR2 E2BSs were lower than expected and lower than reported elsewhere (up to 90% E2BS methylation in an SCC cohort) (Fernandez *et al.*, 2009, Snellenberg *et al.*, 2012).

Several published studies have approached DNA methylation from a biomarker perspective considering multiple CpGs. Brandsma *et al* identified patterns of DNA methylation that were associated with disease grades. HPV genomic hypomethylation was associated with lower grade cervical disease (normal cytology and negative histology in most instances) and hypermethylation at several regions (*E5*, *L1* and *L2*) was associated with higher grade disease (Brandsma *et al.*, 2009). There was however considerable variation among the samples (likely a result of small sample number) and no *E2* hypermethylation was observed for any of the cases; one LSIL sample had methylation in the LCR1 and LCR2 regions however the rest were hypomethylated throughout the LCR (Brandsma *et al.*, 2009). Mirabello *et al* used a HPV genome wide approach and identified numerous CpGs where methylation was significantly different among the disease grades studied. These CpGs were located primarily in the *L1, L2*

and *E2*/E4 regions although CpGs throughout other areas of the genome were also significantly associated with disease grade. *L1* and *L2* methylation could predict disease grade (Mirabello *et al.*, 2012).

5.2.2. Meaning

5.2.2.1. Variation in Methylation

The variation in methylation level among different HPV regions and among CpGs within regions was also seen for cell lines and other clinical cohorts. In most instances there was *E2* and *L1*/*L2* methylation and LCR hypomethylation, this is consistent with several of the studies described above. Similarly, the pattern of CpG methylation in each region (i.e. when methylation at each CpG is plotted graphically) was also common to the other clinical/cell line cohorts. The pattern of *L1*/*L2* CpG methylation for instance was common for all three disease groups but was also seen elsewhere in our study. The CpG specific and region specific methylation pattern, and possible causes/consequences are discussed in further detail in the general discussion (Chapter 9: 9.1.4 & 9.1.5).

Because there was considerable variation among individual CpGs and because it would be much simpler to assess large numbers of samples using single CpG assessing methods like MS-PCR, any methylation biomarker study would be well advised to consider individual CpG methylation as well as regional methylation.

5.2.2.2. There was a Progressive Increase in DNA Methylation with Increasing Severity of Disease

There was a progressive increase in methylation of the *E2* and *L1*/*L2* regions with increasing disease severity, the LCR was hypomethylated in most samples except for the cancer group where mean LCR methylation was higher and several samples had very high levels of LCR methylation. Other studies have identified a more prominent association between E2BS hypermethylation (4/5 CpGs of the LCR2 region are in E2BSs) and cancer which may be down to methodological differences (Fernandez *et al.*, 2009, Snellenberg *et al.*, 2012). Whether HPV DNA methylation changes are a cause of or a consequence of disease progression is yet to be established, but the variation within groups does suggest that any relationship is likely to be complex.

Integration of HPV is common in higher grade disease and there is an association between DNA hypermethylation and integration (Fernandez *et al.*, 2009, Kalantari *et al.*, 2008b, Doerfler, 2006, Van Tine *et al.*, 2004b, Melsheimer *et al.*, 2004). The higher frequency of

integration events typically observed in the more severe disease groups might account for some of the DNA methylation increase observed. In our other studies (Chapter 7: 7.1.3.5 & Chapter 8: 8.2.3.3), most integrated HPV were very heavily methylated in the *E2* and/or *L1*/*L2* regions. The differences between normal cytology and severe dyskaryosis could therefore be due to the increased integration frequency. However, whilst there were a few examples of *E2* or *L1*/*L2* hypermethylation of the severe dyskaryosis group, most were not heavily methylated. Instead, small increases in methylation of the severe dyskaryosis group compared to the normal group were seen for most samples. These small increases in methylation are not consistent with an increased integration rate in the higher disease grades being the only explanation. As such, integration is not the only factor involved in increased methylation in higher disease grades.

There is an association between HPV DNA methylation and other aspects of HPV molecular biology related to cervical disease; i.e. gene expression (Chapter 8: 8.2.3.4) and cellular differentiation (Vinokurova and von Knebel Doeberitz, 2011, Ding *et al.*, 2009). LCM studies have determined that cellular differentiation in a productive infection is associated with changes in LCR methylation. In undifferentiated basal cells, the 5' LCR and the P97 promoter (i.e. LCR2) region were hypomethylated whilst the enhancer region (i.e. LCR1) was hypermethylated (Vinokurova and von Knebel Doeberitz, 2011). In more differentiated cell layers, the 5' LCR remained hypomethylated but there was a switch to a hypomethylated enhancer and a hypermethylated promoter (Vinokurova and von Knebel Doeberitz, 2011). Vinokurova *et al* also suggested that a "latent" form of infection might be associated with a completely hypermethylated LCR whilst cells with high grade disease appeared to be associated with a hypermethylated 5'LCR and enhancer, but a hypomethylated promoter (Vinokurova and von Knebel Doeberitz, 2011).

The DNMTs are responsible for methylating DNA. Several oncogenic DNA viruses influence DNA methylation and HPV are also known to affect DNMTs both directly and indirectly (Au Yeung *et al.*, 2010, Leonard *et al.*, 2012, Burgers *et al.*, 2007, McCabe *et al.*, 2005, Hsu *et al.*, 2012). It has also been proposed that HPV can actively promote self methylation in order to regulate gene expression (De-Castro Arce *et al.*, 2011) so it is not unreasonable to suspect that HPV DNA methylation may be at least partly self determined.

Human cells actively target foreign DNA elements for silencing by DNA methylation and the association between cancer and aberrant human DNA methylation has been known for many years (Doerfler, 2006, Doerfler, 2008, Doerfler *et al.*, 2001). It is difficult to determine whether the profile of HPV DNA methylation observed with increasing disease severity is a consequence

of the action of HPV, or the result of general DNA methylation changes as there are numerous reports of changes in human DNA methylation also being associated with HPV related cervical disease (Henken *et al.*, 2007, Sun *et al.*, 2012, Eijsink *et al.*, 2012, Wentzensen *et al.*, 2009, Leonard *et al.*, 2012, Apostolidou *et al.*, 2009). In summary, the observed methylation pattern represents a heterogeneous mix of DNAs, likely under the influence of a variety of factors of both viral and human origin.

5.2.2.3. HPV Methylation as a Biomarker of Disease

Several chapters of this investigation considered the potential clinical utility of HPV DNA methylation, in order to avoid repetition and provide a concise discussion; a more thorough consideration is presented in the general discussion (Chapter 9: 9.3).

There is a pressing need for suitable biomarkers to be used in conjunction with HPV testing (Wentzensen and von Knebel Doeberitz, 2007). For this study, there were significant differences in DNA methylation among the three disease groups. The most significant differences were observed for the *L1*/*L2* region, for which pairwise differences in methylation existed between all three groups. There were however significant overlaps in the ranges of each group. Several cancer samples for instance had methylation levels more commonly observed in the normal cytology group which is consistent with the variation observed in similar studies (Brandsma *et al.*, 2009, Fernandez *et al.*, 2009, Mirabello *et al.*, 2012).

Nonetheless, our results are consistent with the results of several studies, in that the *E2*/E4, *L1* and *L2* regions were associated with disease grade (Fernandez *et al.*, 2009, Kalantari *et al.*, 2009, Sun *et al.*, 2011, Mirabello *et al.*, 2012, Kalantari *et al.*, 2010, Brandsma *et al.*, 2009). Similarly, our observation of differences in LCR methylation among the disease grades was consistent with most studies (Kalantari *et al.*, 2004, Bhattacharjee and Sengupta, 2006a, Ding *et al.*, 2009, Brandsma *et al.*, 2009) but not with others (Badal *et al.*, 2003, Hublarova *et al.*, 2009). The inconsistent association between LCR methylation and disease grade is considered the result of methodological differences, heterogeneous and small sample sizes, and/or disease misclassification (Mirabello *et al.*, 2012).

L1 and LCR methylation have been previously tested as a diagnostic biomarker of disease grade (Turan *et al.*, 2007, Kalantari *et al.*, 2010). The results appeared to indicate that *L1* methylation was a surrogate marker of integration, and that integration was associated with disease grade. We however observed incremental increases in DNA methylation with disease

grade which was not consistent with *L1/L2* methylation only being a surrogate marker of integration.

A screening biomarker could have relevance as a triage test within HPV based cervical screening. Pyrosequencing is not currently a very high throughput means of evaluating methylation status and is also somewhat expensive. Therefore, the magnitude of differences between the disease grades for individual CpGs was determined. Of all of the CpGs tested, the L1L2-2 CpG appeared to be the most informative. Using just one CpG or a small number of CpGs as a biomarker would allow less expensive, simpler and higher throughput assays such as qMS-PCR to be used to quantity DNA methylation.

5.3. Conclusions

The aims of this study were to determine the methylation profile of HPV in a range of disease states, compare data produced using our assays to published data and apply the pyrosequencing assays to the early development sections of the CRUK biomarker development roadmaps (Cancer Research UK, 2012e, Cancer Research UK, 2012d). In these respects it achieved what was planned and clearly demonstrated that distinct methylation differences between disease of different cytological grades and between precancerous and cancerous disease exist. There was also an attempt to take the data analysis a stage further and determine which regions/CpGs are likely to represent the most informative biomarkers of disease.

There were some promising results produced by this study, DNA methylation did vary among disease grades for several regions/CpGs however the differences were not simple and there was wide variation within each group which represented an undesirable characteristic for a biomarker. If HPV methylation is used as a biomarker, it will likely be in the form of an algorithm considering methylation at various HPV regions and CpGs (Brandsma *et al.*, 2009, Mirabello *et al.*, 2012).

5.3.1. Further Work

The results of this study suggest several avenues for future work. First, it is unknown whether transformation and disease progression are a result of or cause of DNA methylation changes. Similarly, it is unknown whether integration is the cause of initial aberrant gene expression or a consequence of genomic instability (Vinokurova and von Knebel Doeberitz, 2011). A longitudinal study, following HPV positive patients over several years might better put the relevance of integration and methylation changes to disease progression into context.

LCM studies have revealed considerable methylation differences among different parts of a lesion, and changes with cellular differentiation (Vinokurova and von Knebel Doeberitz, 2011). With this in mind, a more uniform set of samples (i.e. a set consisting entirely of LBC or biopsy material in conjunction with microdissection techniques) would allow better elucidation of the association between DNA methylation and HPV-associated disease. Whilst microdissection allows investigation of the relationship between DNA methylation and different aspects of disease and cellular biology, further investigation using LBC material is the most clinically relevant with respect to primary screening triage tests.

HPV DNA methylation has great potential as a diagnostic and prognostic biomarker. There have been some promising studies but there are also several contradictions in the literature. Further investigation of HPV DNA methylation, possibly used in conjunction with another biomarker (i.e. p16, integration, human methylation) might prove useful and would present HPV methylation in context with other biological characteristics.

Chapter 6. The CRISP Study – DNA Methylation and Clinical Outcome

6.1. Results

Material for this study was acquired from the CRISP clinical trial (Castanon *et al.*, 2012). Two LBC samples taken 6 months apart existed for each patient (the 0 and 6 month samples respectively). Every 0 month LBC sample had low grade cervical abnormalities as it was a requirement for inclusion into the clinical trial. At 6 months, cytology was repeated and the majority of patients then underwent colposcopy, with histology performed on any biopsied material. All cytology and histology results were available for this study. DNA was extracted from LBC cell pellets, BS treated and pyrosequencing was performed on the treated DNAs.

 Samples were identified from the CRISP sample set that corresponded to each of three different outcome groups. The first group comprised of samples that were HPV16 positive at 0 months only and represent those samples that had lost the HPV16 infection, these were the Cleared group (C group). The second group was the Persistent Low grade group (PL group), these were samples that had a persistent HPV16 infection (i.e. HPV16 positive at 0 and 6 months) and low grade histology (CIN1 or lower) at 6 months. Finally, samples that were HPV16 positive at both 0 and 6 months and had high grade histology (CIN2/3) at 6 months were termed the Persistent High grade group (PH group). The PH group represented samples with high grade disease, potentially as a result of persistent HPV16 infection. In total, 23 C, 11x2 PL (i.e. 0 and 6 months) and 12x2 PH samples were assessed by pyrosequencing. Several samples in each instance failed the pyrosequencing assays and the data from 14 cleared, 9 PL and 10 PH patients were used in final analysis.

This chapter aimed to generate hypotheses and answer hypothesis driven questions relevant to the sample cohort. The study was designed to determine differences in DNA methylation over time, determine differences in DNA methylation between samples with different clinical outcomes and to assess the potential of HPV DNA methylation as a biomarker.

6.1.1. Hypothesis 1: DNA Methylation Varies among Regions of the HPV Genome and among the CpGs within Each Region

There was considerable variation in the amount of DNA methylation among regions, and among CpGs within each region (Figure 6.1). A two-way ANOVA GLM was used to determine whether the regions and the CpGs within the regions had a significant effect on DNA methylation for the C, PL and PH groups, the results are presented in Table 6.1. Any violations

of the assumptions of the ANOVA are detailed in the figure legend. Highly significant differences in mean DNA methylation were observed both among the HPV16 regions and the CpGs. Further statistics performed later in this chapter support these results. The Bonferroni corrected P-value for significance is P=0.017.

	Outcome group		
	C	PL	PH
CpG (P-value)	0.000*[1]	0.003*[2]	0.000*[3]
Assay (P-value)	0.000*[1]	0.000*[2]	0.000*[3]

Table 6.1: The results of three two-way ANOVA GLM tests with CpG and pyrosequencing assay as factors.

Each value in the table is a P-value, in the CpG row this represents the difference in DNA methylation among different CpGs within each assay; in the Assay row this represents the difference in DNA methylation among the four HPV regions. Violations of the assumptions of the ANOVA: *1, unequal variances (P=0.003), non-normally distributed residuals (P=0.022); *2, non-normally distributed residuals (P=<0.005); *3, unequal variances (P=0.000).
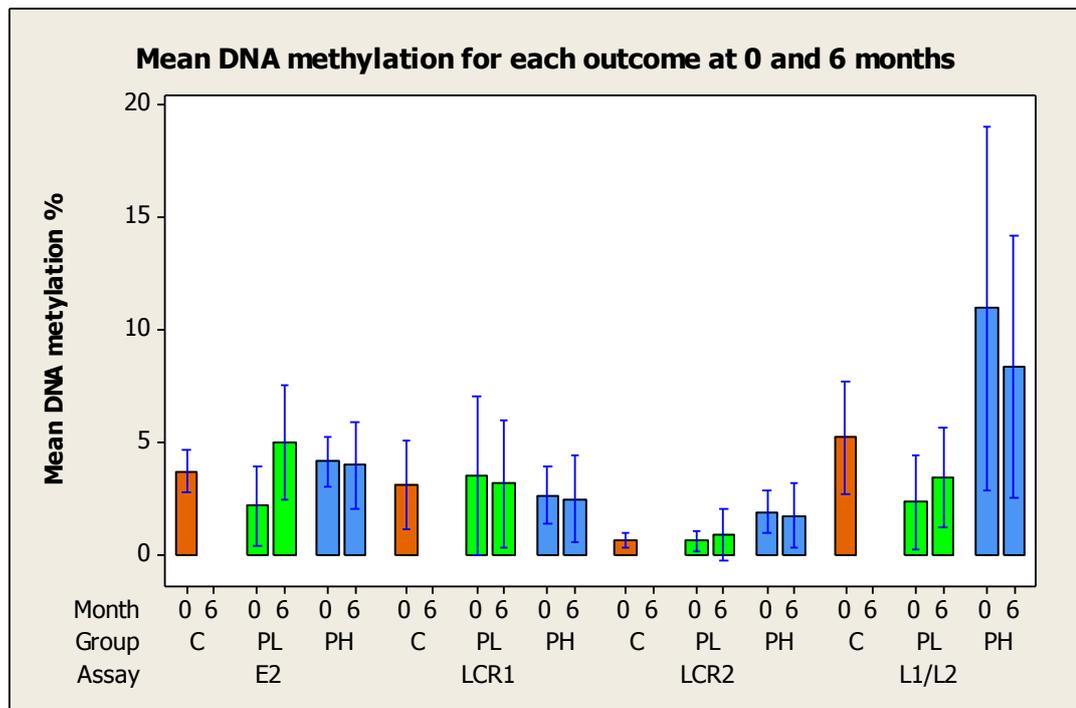
Figure 6.1: Mean DNA methylation of each HPV region, for each outcome group for both the 0 and 6 months time points.

Coloured bars represent mean DNA methylation for each region. Error bars represent +/- 95% confidence intervals for mean.

6.1.2. Hypothesis 2: DNA Methylation of the HPV16 Genome Varied over Time

For the PL and PH groups, DNA methylation data was produced for the 0 and 6 month samples. It was of interest to determine whether changes in DNA methylation occurred over the 6

month time period and whether or not any changes observed were different for the PL and PH groups. The 0 and 6 month data of the PL group (Figure 6.2) and the PH (Figure 6.3) are presented below. There appeared to be changes between the 0 and 6 month samples in several instances, especially for the *E2* and *L1/L2* regions (Figure 6.1). The PL group had an increase in methylation between the 0 and 6 month samples for the *E2* and *L1/L2* regions whilst the PH group had a slight decrease in the *L1/L2* region.
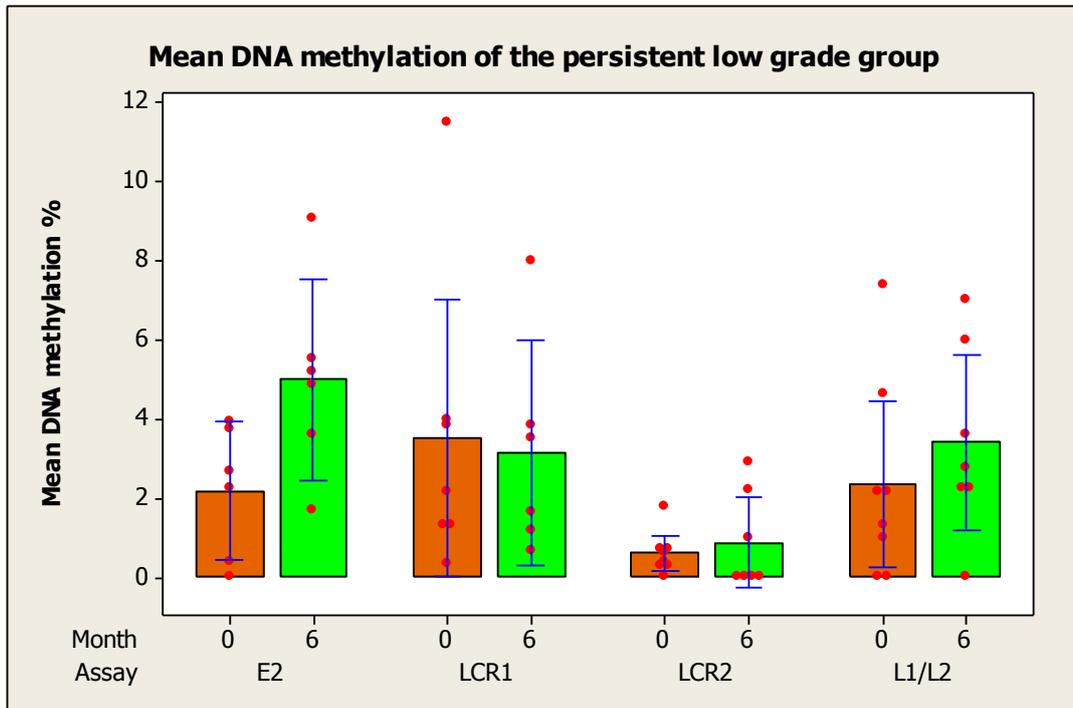
Figure 6.2: The mean amount of methylation at each assay for the persistent low grade group.

The bars represent the overall mean methylation at each assay for each sample time. The error bars represent 95% confidence intervals for mean and the red circles represent mean methylation at each assay for each of the individual samples.
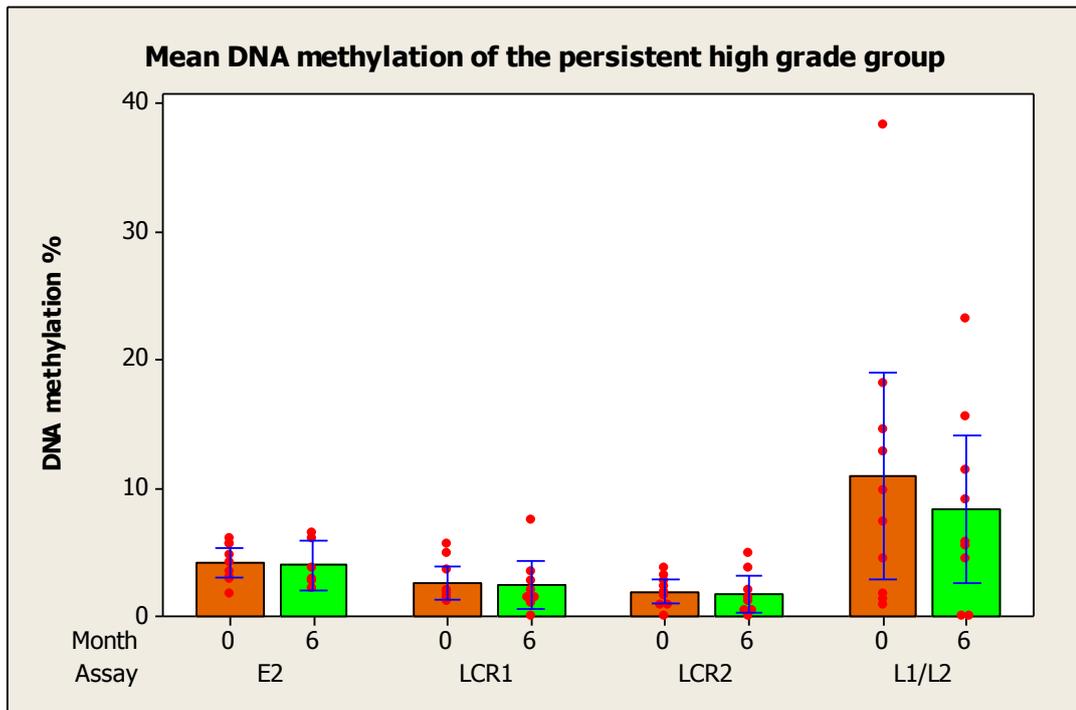
Figure 6.3: The mean amount of methylation at each assay for the persistent high grade group.
The bars represent the overall mean methylation at each assay for each sample time. The error bars represent 95% confidence intervals for mean and the red circles represent mean methylation at each assay for each of the individual samples.

A two-way ANOVA GLM was applied to the DNA methylation data of the PL and PH groups to determine whether a significant difference in DNA methylation existed between the 0 and 6 month samples. The two-way ANOVA was ideal as additional analytical power was afforded by the consideration of individual CpGs, and the differences in DNA methylation among the individual CpGs of each assay can also be determined. Because no suitable non-parametric alternative exists and as the investigation was exploratory and hypothesis generating in nature, where the data had not satisfied the assumptions of the two-way ANOVA GLM, the analysis was continued and the contraventions are listed in the table legend. The results of these analyses are in Table 6.2 and graphical representation of the data used for the interpretations are presented below for the PH group (Figure 6.4, Figure 6.5, Figure 6.6 and Figure 6.7) and the PL group (Figure 6.8, Figure 6.9, Figure 6.10 and Figure 6.11). As a total of 8 analyses were performed, a Bonferroni corrected P-value (P=0.006) was used however comparisons significant at P=0.05 were still considered. For every HPV region of the PH group, there were no significant differences between the amount of methylation at 0 months and 6 months. There was however a significant difference among the methylation level of individual CpGs within each region. For the PL group, there was a significant difference between the 0 and 6 month samples within the *E2* region only. There was also a significant difference in DNA

methylation between the CpGs of the *E2*, LCR2 and *L1/L2* region. The observed differences among CpGs within each HPV region support the results of hypothesis 1.

		HPV16 region			
		E2	LCR1	LCR2	*L1/L2*
PL group (P-value)	CpG	0.000*[1]	0.873	0.006*[2]	0.003*[3]
	Month	0.000*[1]	0.288	0.771*[2]	0.824*[3]
PH group (P-value)	CpG	0.000*[4]	0.002*[5]	0.000*[6]	0.001*[7]
	Month	0.490*[4]	0.603*[5]	0.505*[6]	0.266*[7]

Table 6.2: The results of two-way multifactorial ANOVA GLM analyses.

The persistent low grade (PL group) and persistent high grade (PH group) group methylation data was tested to determine the effect of month (0 vs. 6) and each CpG of each assay on mean DNA methylation. Significance is assumed at P<0.05 or P<0.006 (Bonferroni corrected). A significant P-value for CpG indicates a significant difference in mean methylation among CpGs within a region. A significant P-value for month indicates that a significant difference existed between the 0 and 6 month samples. Violations of the ANOVA assumptions: *1, non-normally distributed residuals (P=0.037); *2, non-normally distributed residuals (P=<0.005), unequal variances (P=0.029); *3, non-normally distributed residuals (P=<0.005); *4, non-normally distributed residuals (P=0.006); *5, unequal variances (P=0.04); *6, non-normally distributed residuals (P=0.036); *7, non-normally distributed residuals (P<0.005).

The results of these analyses must be interpreted with caution because of the contraventions of the assumptions of the analyses. In order to improve the strength of any conclusions drawn, multiple two-sample T-tests were used to compare the differences between mean DNA methylation of each assay for the 0 and 6 month samples. The Bonferroni corrected P-value is P=0.006 however comparisons significant at P=0.05 would also be considered. None of these comparisons were significant, the P values for the 0 vs. Six month comparisons are summarised in Table 6.3. There appear to be no differences between the 0 month and 6 month samples however it is important to consider that multiple T-tests exclude a level of detail (i.e. the methylation data for each CpG).

0 vs. 6 months	*E2* mean	LCR1 mean	LCR2 mean	*L1/L2* mean
Persistent low grade (P-value)	0.058	0.986	0.903*[1]	0.354
Persistent high grade (P-value)	0.800	0.612	0.664	0.617

Table 6.3: The P-values of multiple two-sample T-tests comparing the mean methylation values of each assay for the 0 and 6 month samples.

The values given are P-values. Violation of the assumptions of a T-test: *1, this comparison included a data set that was not quite normally distributed (P=0.020).

Overall, there was no difference between the 0 and 6 month samples in every instance except for the persistent low grade samples, where a difference was observed in the DNA methylation of the *E2* region. For this comparison, the two-way ANOVA GLM had only a minor violation of the assumptions and a highly significant P-value (Table 6.2) and the T-test had a low P-value that was close to significant (without a Bonferroni corrected P-value) (Table 6.3). This discrepancy in significance between the two types of tests (one considering every CpG, the other type considering only mean methylation for that region) suggest that CpG specific changes occurred between the 0 and 6 month samples for the PL group. Indeed, for the PL group, every *E2* CpG was more methylated in the 6 month samples than the 0 month samples however the difference varied among CpGs (Figure 6.8).
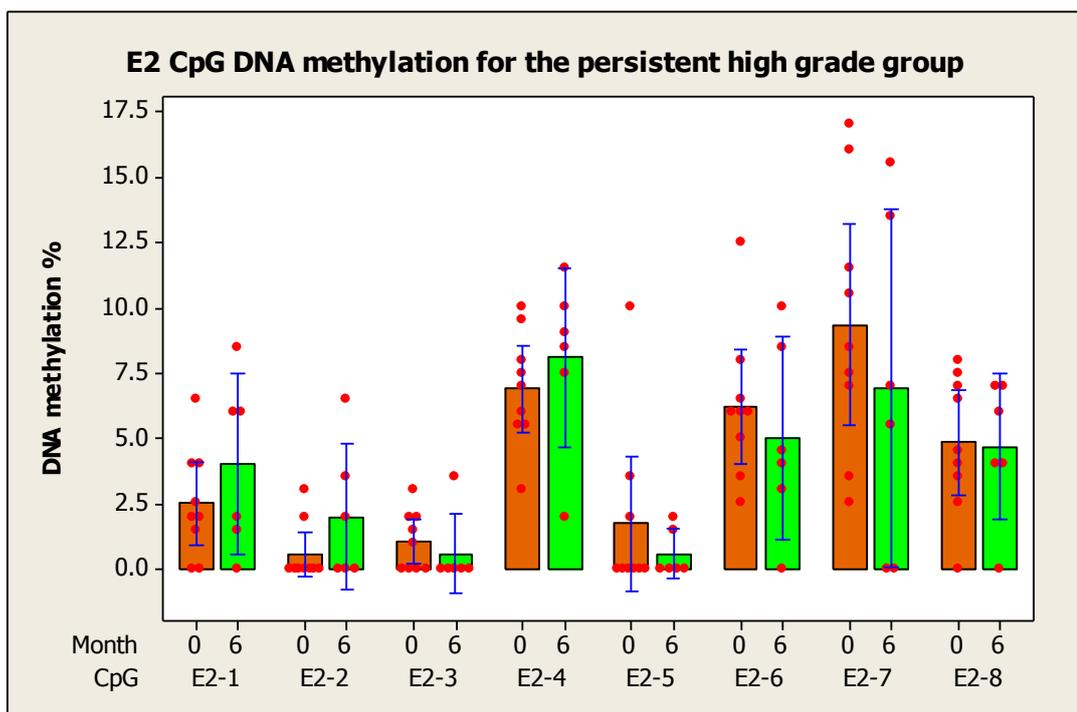
Figure 6.4: *E2* CpG DNA methylation for the PH group.

The DNA methylation data of both the 0 and the 6 month samples are presented. Each bar represents the mean DNA methylation of the month/CpG of interest, blue error bars represent 95% confidence intervals for mean and the red circles represent individual sample data.
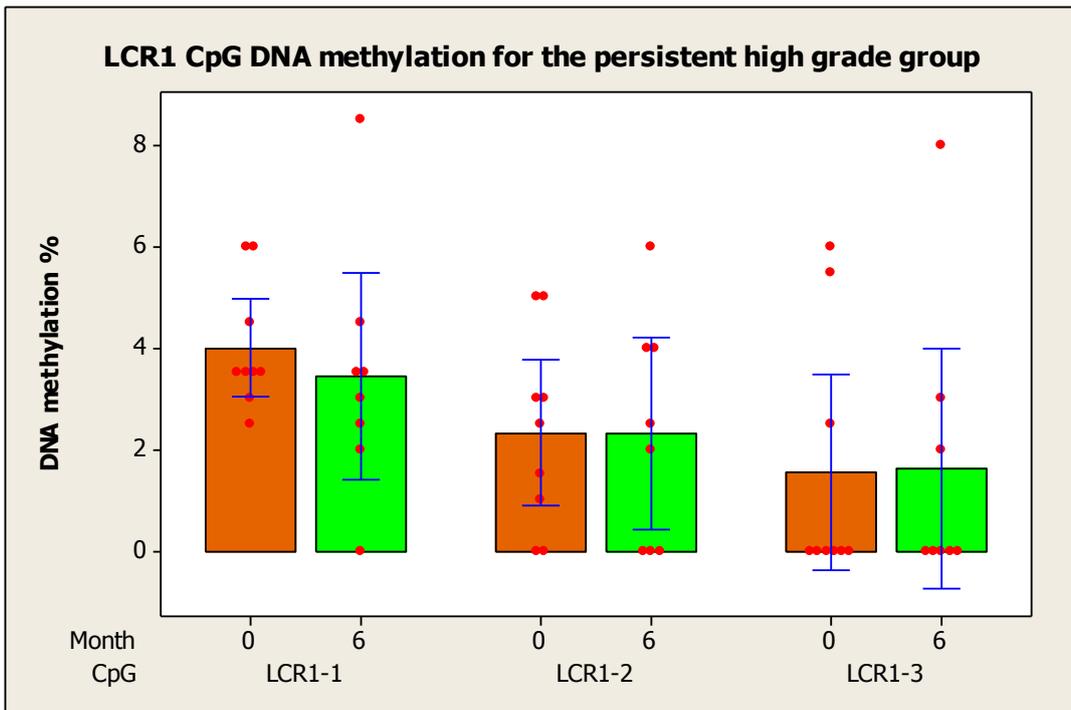
Figure 6.5: LCR1 CpG DNA methylation for the PH group.

The DNA methylation data of both the 0 and the 6 month samples are presented. Each bar represents the mean DNA methylation of the month/CpG of interest, blue error bars represent 95% confidence intervals for mean and the red circles represent individual sample data.
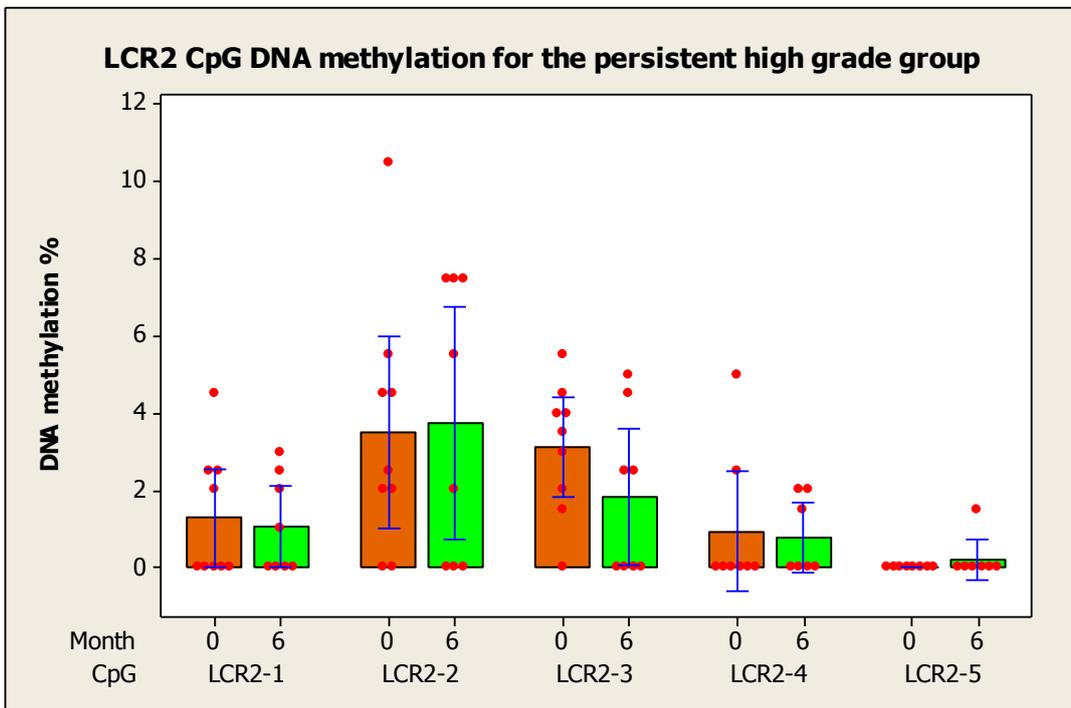
Figure 6.6: LCR2 CpG DNA methylation for the PH group.

The DNA methylation data of both the 0 and the 6 month samples are presented. Each bar represents the mean DNA methylation of the month/CpG of interest, blue error bars represent 95% confidence intervals for mean and the red circles represent individual sample data.
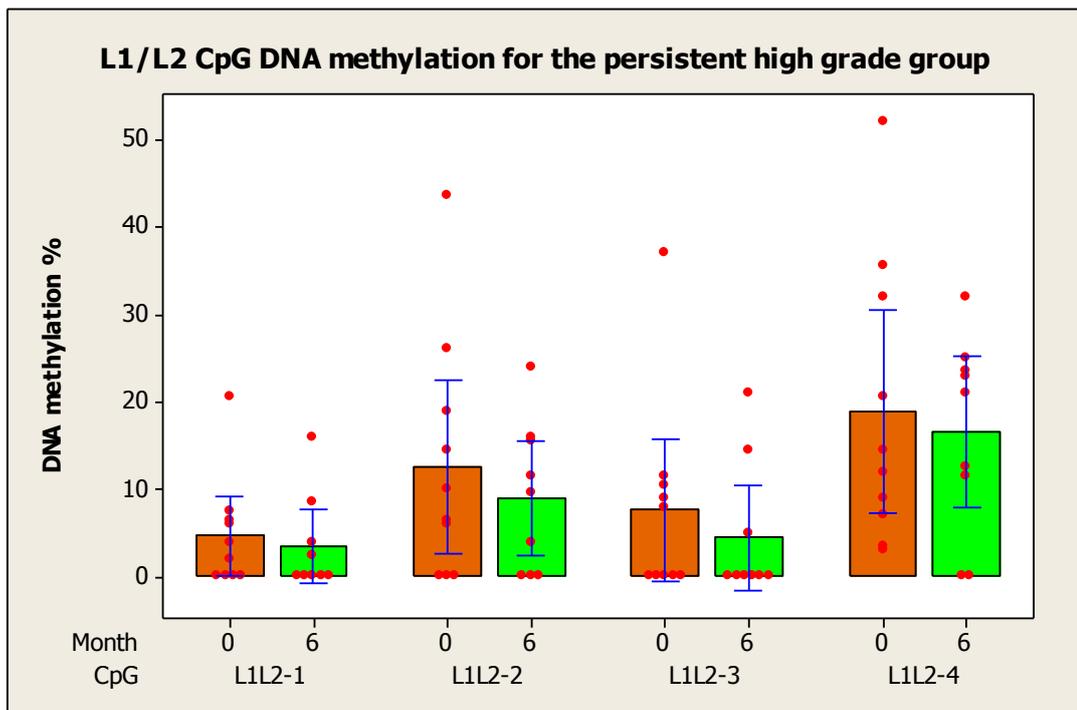
Figure 6.7: *L1/L2* **CpG DNA methylation for the PH group.**

The DNA methylation data of both the 0 and the 6 month samples are presented. Each bar represents the mean DNA methylation of the month/CpG of interest, blue error bars represent 95% confidence intervals for mean and the red circles represent individual sample data.
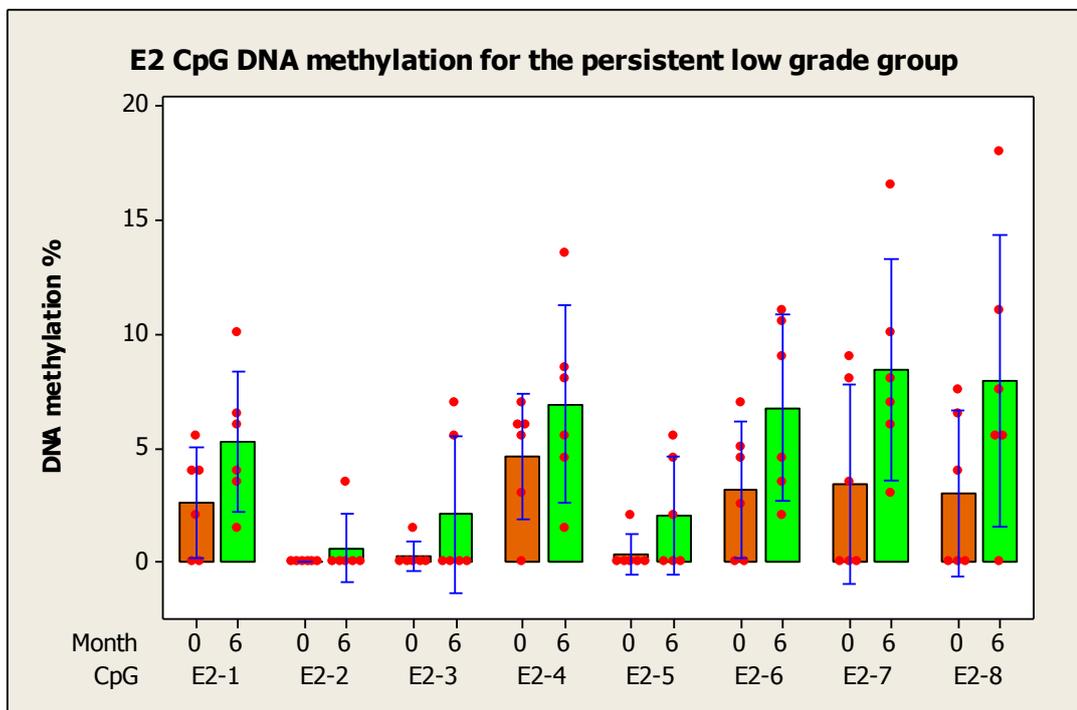
Figure 6.8: *E2* **CpG DNA methylation for the PL group.**

The DNA methylation data of both the 0 and the 6 month samples are presented. Each bar represents the mean DNA methylation of the month/CpG of interest, blue error bars represent 95% confidence intervals for mean and the red circles represent individual sample data.
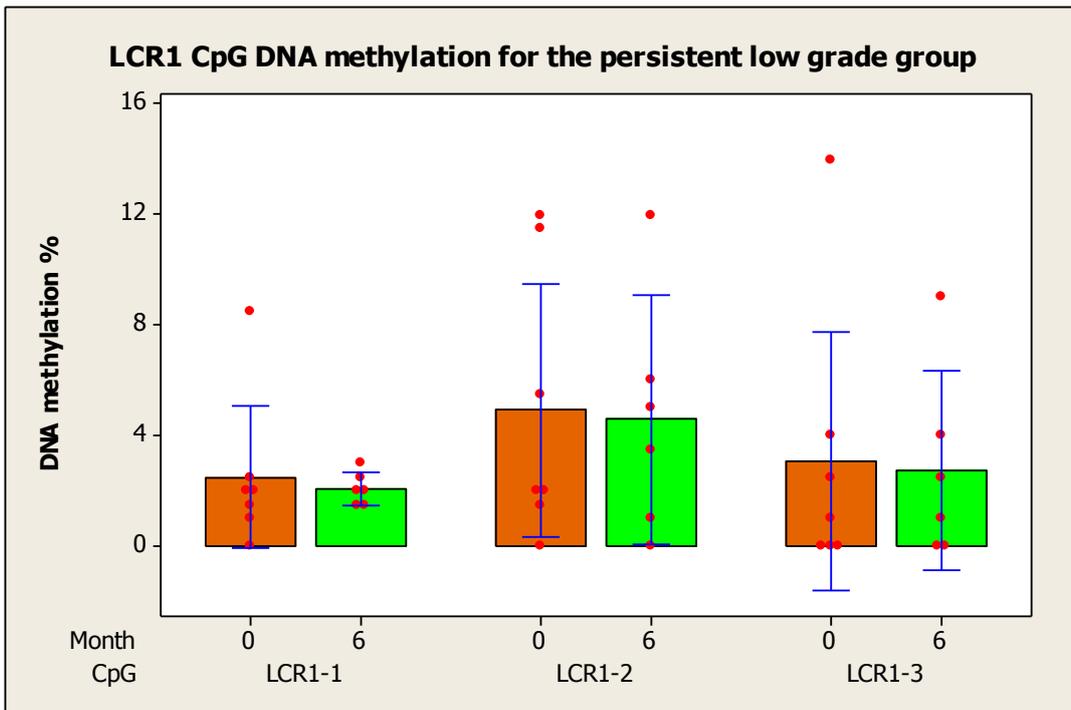
Figure 6.9: LCR1 CpG DNA methylation for the PL group.

The DNA methylation data of both the 0 and the 6 month samples are presented. Each bar represents the mean DNA methylation of the month/CpG of interest, blue error bars represent 95% confidence intervals for mean and the red circles represent individual sample data.
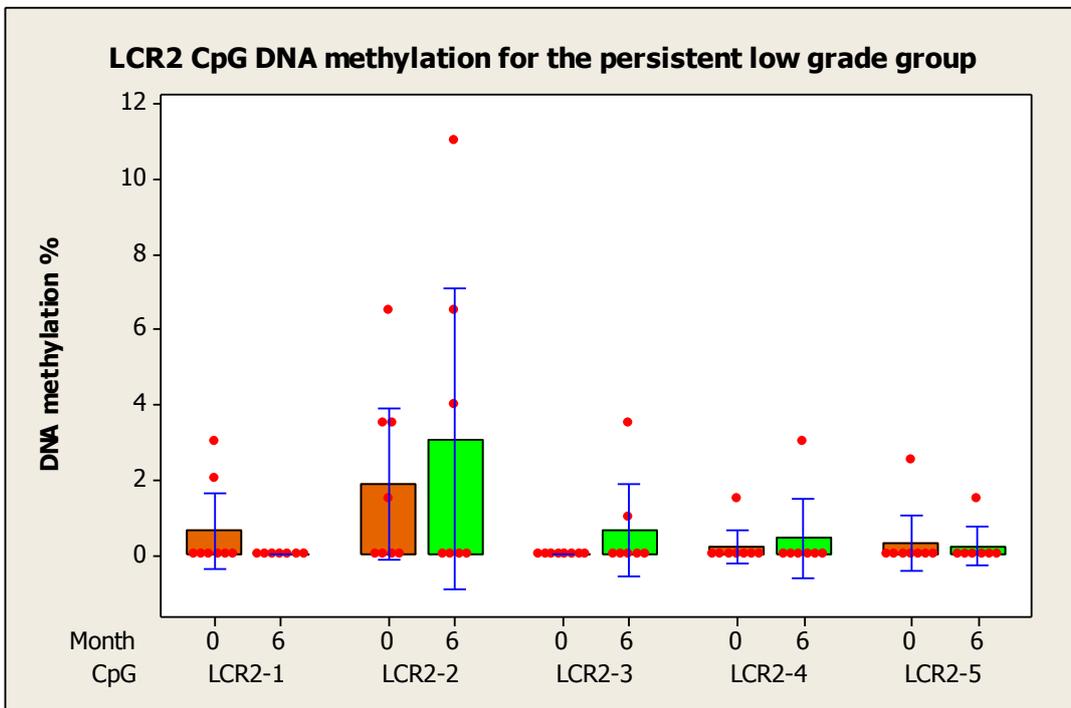
Figure 6.10: LCR2 CpG DNA methylation for the PL group.

The DNA methylation data of both the 0 and the 6 month samples are presented. Each bar represents the mean DNA methylation of the month/CpG of interest, blue error bars represent 95% confidence intervals for mean and the red circles represent individual sample data.
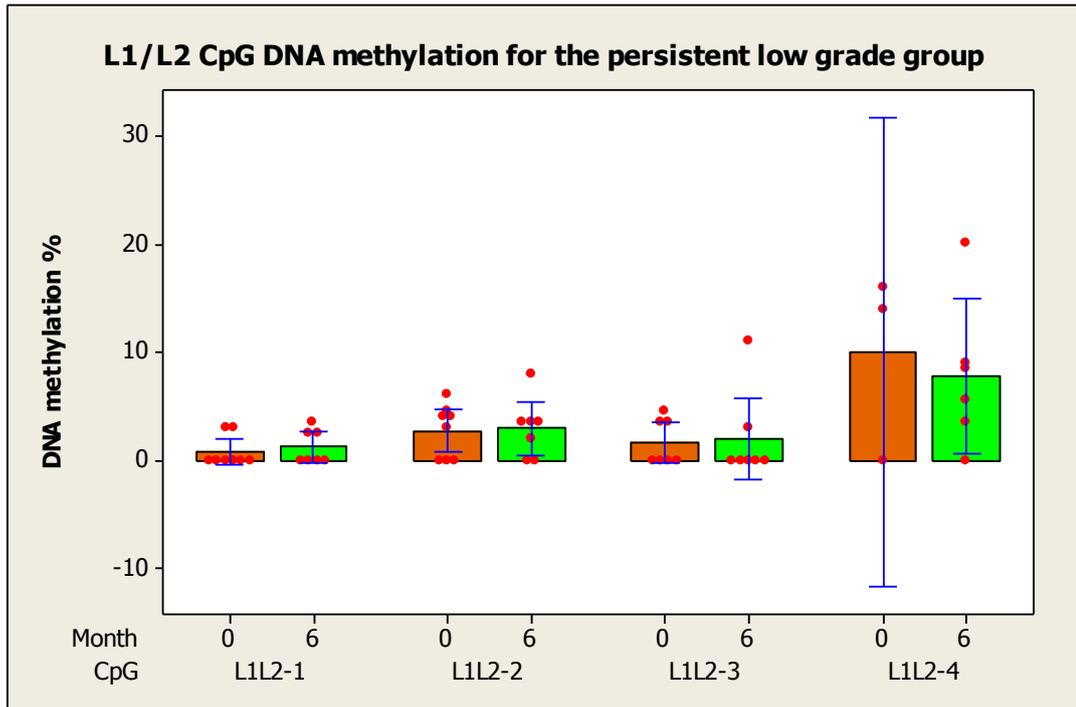
Figure 6.11: *L1/L2* **CpG DNA methylation for the PL group.**

The DNA methylation data of both the 0 and the 6 month samples are presented. Each bar represents the mean DNA methylation of the month/CpG of interest, blue error bars represent 95% confidence intervals for mean and the red circles represent individual sample data.

6.1.3. Hypothesis 3: There are DNA Methylation Differences among the Outcome Groups

Another subject of interest was to determine whether DNA methylation differences existed between HPV infections that were cleared, and those that were maintained and had low grade or high grade disease. In order to address this question, methylation of the 0 month samples for each outcome group was compared (Figure 6.12). The effect of disease grade on DNA methylation has already been established (Chapter 5: 5.1.2). As such, PH group DNA methylation may already be expected to be higher than that of the C or PL group because of the confirmed high grade (CIN2/3) disease at 6 months that may have also been present in the 0 month samples. The comparison of the PL and C groups was of particular interest as it could have provided insight into differences between maintained and cleared infections without the influence of high grade disease on DNA methylation.
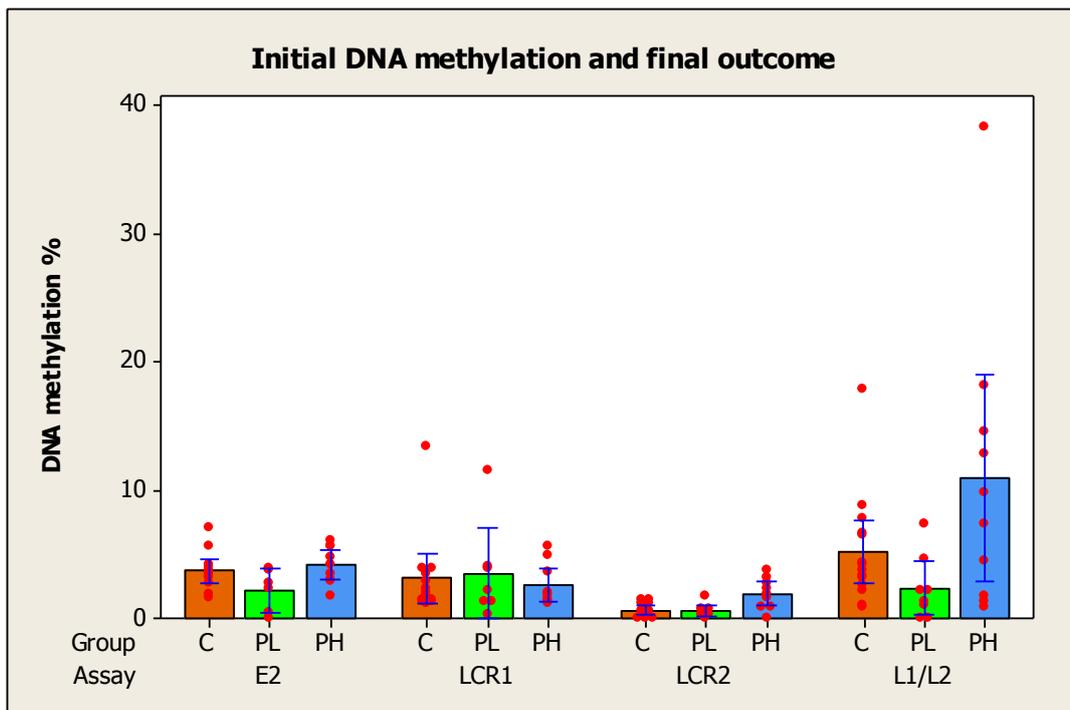
Figure 6.12: The mean amount of methylation at 0 months for each assay for each of the three outcome groups.

C=cleared, PL=persistent low grade, PH=persistent high grade disease. The bars represent the mean for the factor combination presented whilst the error bars represent 95% confidence intervals for mean.

The 0 month samples were not highly methylated regardless of final outcome. As may be expected, there did appear to be methylation differences between the PL and PH groups at the *E2*, LCR2 and *L1/L2* regions. Methylation was typically higher for the PH group. What was particularly intriguing was that the C group seemed to have more DNA methylation for the *E2* and *L1/L2* assays than the PL group. One of the aims of this study was to determine if a knowledge of HPV DNA methylation could be used predict the outcome of an infection, i.e. were there DNA methylation differences between individuals that would go on to lose the infection or develop disease. From this perspective, the most informative region might be the *L1/L2* region for which stark changes between the C, PL and PH groups existed. The best indicator of high grade histology at 6 months appeared to be methylation of the LCR2 region. Both the C and PL groups were methylated at a very low level and the PH group had slightly more methylation, albeit with a considerable variation among individual samples. These observations were tested statistically using a two-way ANOVA GLM and associated pairwise tests.

Two-way ANOVA GLM tests were used to compare the mean amount of methylation in each group with respect to both the final outcome (C, PL and PH) and individual CpG methylation

values. In total, four comparisons were performed (one test for each region), so a Bonferroni corrected P-value of P=0.0125 should be considered. A summary of the results are presented in Table 6.4, where the comparisons violated one or more of the ANOVA assumptions, the details are listed in the table legend.

	E2	LCR1	LCR2	*L1/L2*
CpG	0.000[*1]	0.000[*2]	0.000[*3]	0.000[*4]
Outcome	0.001[*1]	0.951[*2]	0.000[*3]	0.008[*4]

Table 6.4: The results of two multifactorial ANOVA GLM tests.

The 0 month sample data was compared with respect to individual CpGs in each assay (CpG) and final outcome (Outcome). A P-value of P<0.05 for CpG indicates a significant difference among the mean methylation at each CpG within the assay exists, similarly, a P-value of P<0.05 for Outcome indicates that a significant difference exists among the mean methylation for each outcome group (C, PL and PH). Violations of the assumptions of the ANOVA: *1, non-normally distributed residuals (P<0.005); *2, non-normally distributed residuals (P=0.023), unequal variances (P=0.002); *3, non-normally distributed residuals (P<0.005); *4, non-normally distributed residuals (P=0.449).

Each comparison violated one or more assumptions of the ANOVA. As such the results of comparisons should be observed with caution. Nonetheless, a significant difference was observed among the CpGs in each assay in every comparison and among the outcomes for all but the LCR1 assay. Both as a confirmation of the previously observed results and as a way of performing pairwise comparisons (i.e. C vs. PL); multiple one-way ANOVAs and Kruskall-Wallis tests were performed on mean regional DNA methylation (rather than individual CpG methylation). The results of these tests are presented in Table 6.5. A Bonferroni corrected P-value of P=0.013 was used however comparisons significant at P=0.05 would still be considered; one violation of the ANOVA assumptions was encountered and is detailed in the table legend.

	HPV16 Region			
	E2	LCR1	LCR2	*L1/L2*
One-way ANOVA (P-value, F-value, DF)	0.025, 4.30, 2	0.972[*1], 0.03, 2	0.007, 5.98, 2	0.022, 4.35, 2
Significant Tukey-Kramer results (at P=0.05)	PL vs. PH	None	C vs. PH, PL vs. PH	PL vs. PH
Kruskall-Wallis (P-value, H-value, DF)	0.116, 4.31, 2	0.997, 0.01, 2	0.018, 8.07, 2	0.060, 5.64, 2

Table 6.5: The results of tests to determine whether a difference in mean/median DNA methylation existed among 0 month samples of the three outcome groups.

The tests used were one-way ANOVA with Tukey-Kramer tests and multiple Kruskall-Wallis tests. Significant Tukey-Kramer results indicate which pairwise comparisons were significant at P=0.05 where C=cleared group, PL=persistent low grade group and PH=persistent high grade group. Violations of the assumptions of the ANOVA: *1, non-normally distributed residuals (P=0.007).

The one-way ANOVAs confirmed the observations seen for the two-way ANOVAs, however without the assumption violations and at a lower significance (significant at P=0.05, but in some instances not at Bonferroni corrected P=0.013). The lower significance suggests that CpG specific methylation differences were more pronounced than mean methylation differences. For the *E2* and *L1/L2* region, differences were observed between the PL and PH groups only. The observation that the C group was more methylated in *L1/L2* than the PL group was not significant, possibly as a result of relatively small sample size and therefore large confidence intervals. In terms of differentiating between the patients that cleared the infection (C) and those that maintained the infection (PL and PH), none of the HPV regions were particularly informative and no significant differences were observed between either C vs. PL or C vs. PH. That said the LCR2 region produced the most significant difference between the C and PL groups (low grade disease) vs. the PH group (high grade disease).

6.1.3.1. For Which CpG Were the Most Significant Differences among the Three Outcome Groups Observed?

The previous one-way ANOVA tests revealed that greater significance may be produced when the DNA methylation of the three outcome groups is compared on a per CpG basis. Multiple one-way ANOVA tests with associated Tukey-Kramer pairwise comparison tests, and multiple Kruskall-Wallis tests may be able to further expand upon this observation and determine which of the CpGs might become the most useful for differentiating eventual outcome group of 0 month samples. In order to provide a concise summary of the results, the results tables are presented in Appendix VI. For the *E2* region, there did not appear to be a discernible pattern of CpG methylation for the three outcome groups and there was a wide variation in DNA methylation among individual samples (Figure 6.13). The only CpGs where a significant difference in mean DNA methylation of the three outcome groups were observed were E2-4 (one-way ANOVA P=0.034; Kruskall-Wallis P=0.012) and E2-7 (one-way ANOVA P=0.049; Kruskall-Wallis P=0.123 (not significant)). None of the P-values exceeded the Bonferroni corrected P-value. For E2-4 there was a significant difference for C vs. PH and for E2-7 there was a significant difference for PL vs. PH (Tukey-Kramer, P=0.05). There were several violations of the ANOVA assumptions. The *E2* region does not appear to be a good region for predicting outcome group at 6 months in this sample series.
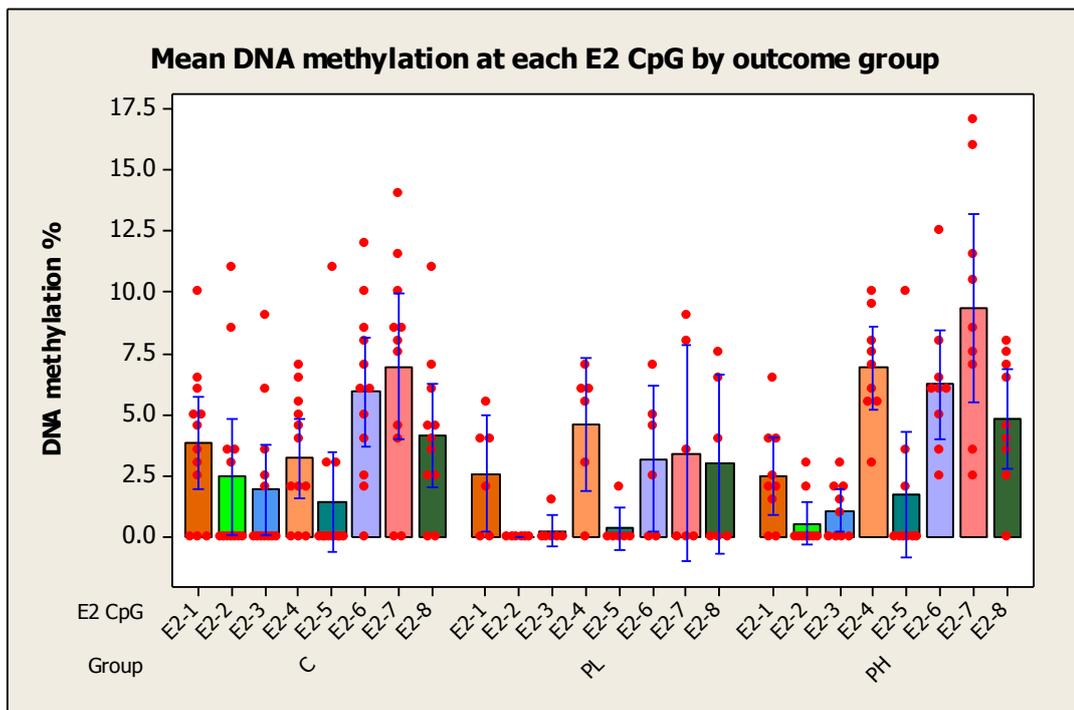
Figure 6.13: The mean amount of DNA methylation at each *E2* CpG for the 0 month samples of each of the three outcome groups.

C=cleared group, PL=persistent low grade and PH=persistent high grade. Each bar represents mean DNA methylation for the indicated CpG/outcome group combination, blue error bars represent 95% confidence intervals for mean and red circles represent individual sample data.

For the LCR1 region (Figure 6.14), there did not appear to be any discernible pattern of CpG methylation. The mean values were low, there was considerable variation among individual samples and there were large confidence intervals, especially for the PL group. The only CpG where significant differences were observed among the groups was LCR1-1 (one-way ANOVA, P=0.014; Kruskall-Wallis, P=0.017), a significant pairwise difference was observed between C vs. PL, and PL vs. PH (Tukey-Kramer P=0.05) and both P-values exceeded the Bonferroni corrected P-value. There were several violations of the ANOVA assumptions. The LCR1-1 CpG methylation data may have potential for predicting outcome at 6 months in this sample series. In this instance however there were no significant differences between the C and PH groups.
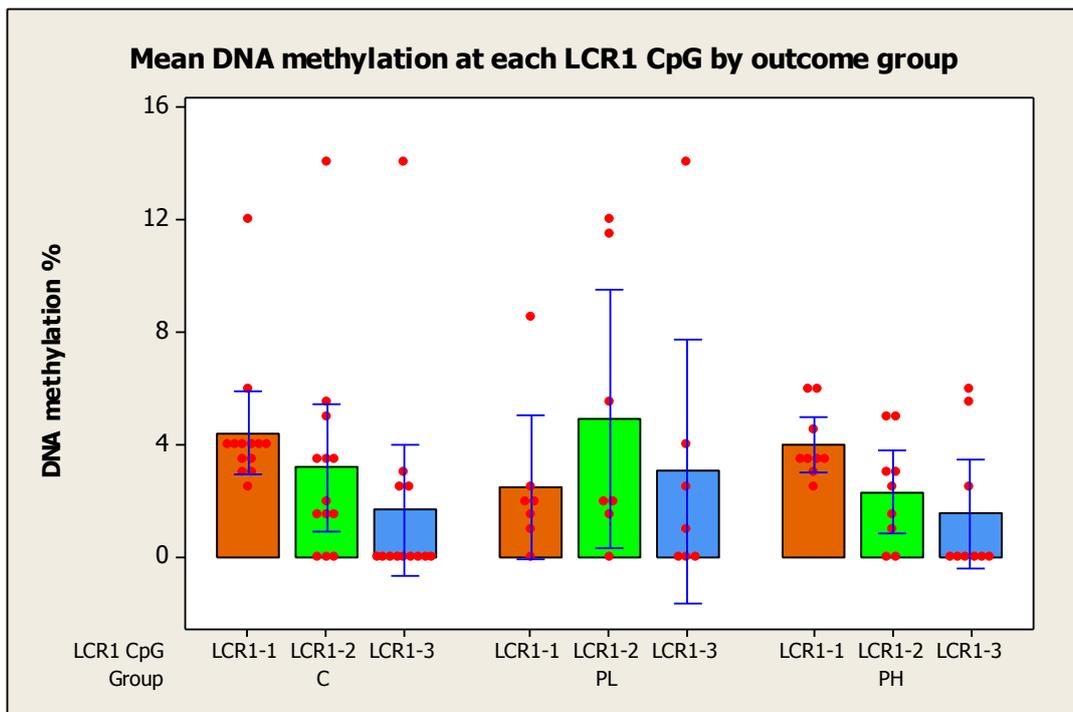
Figure 6.14: The mean amount of DNA methylation at each LCR1 CpG for the 0 month samples of each of the three outcome groups.

C=cleared group, PL=persistent low grade and PH=persistent high grade. Each bar represents mean DNA methylation for the indicated CpG/outcome group combination, blue error bars represent 95% confidence intervals for mean and red circles represent individual sample data.

For the LCR2 region (Figure 6.15), LCR2-5 was the lowest methylated in every sample group. LCR2-3 was methylated at a very low level in groups C and PL, but higher in group PH. The mean values were rather small and there was considerable variation among individual samples. Highly significant differences among the outcome groups were observed for LCR2-3 only (one-way ANOVA, P=0.000; Kruskall-Wallis, P=0.000). Significant pairwise differences were also observed between C vs. PH, and PL vs. PH (Tukey-Kramer P=0.05) and both P-values exceeded the Bonferroni corrected P-value. There was one violation of the ANOVA assumptions. Methylation of LCR2-3 was different between the C/PL and PH outcomes and may be a marker of high grade histology at 6 months in low grade cytology samples.
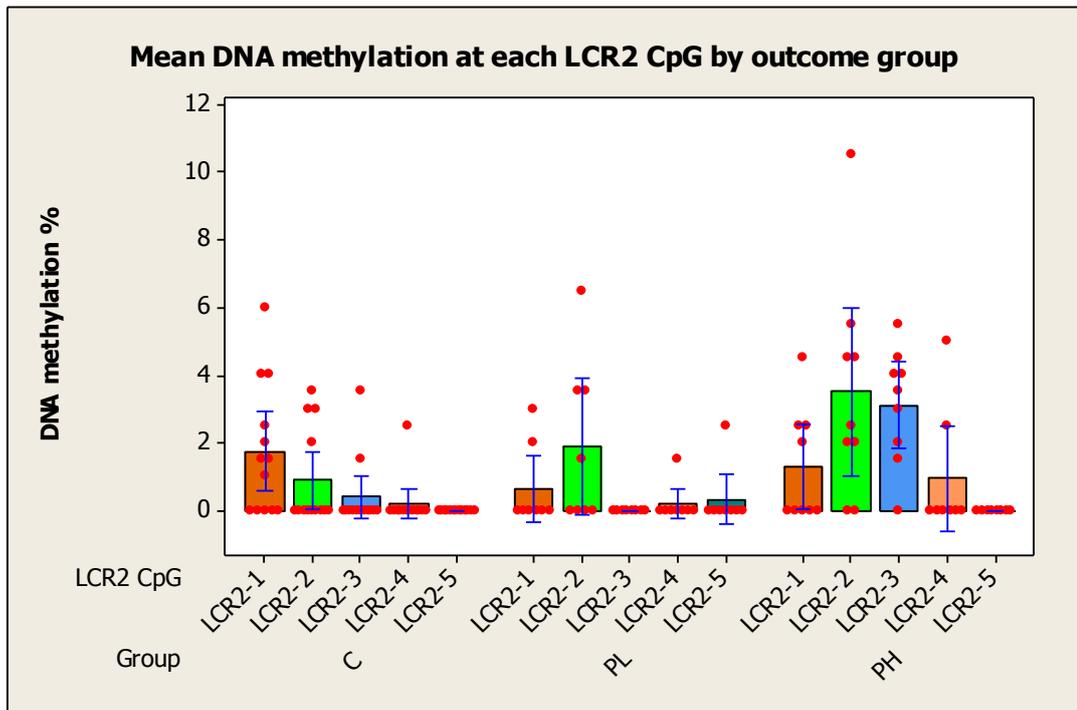
Figure 6.15: The mean amount of DNA methylation at each LCR2 CpG for the 0 month samples of each of the three outcome groups.

C=cleared group, PL=persistent low grade and PH=persistent high grade. Each bar represents mean DNA methylation for the indicated CpG/outcome group combination, blue error bars represent 95% confidence intervals for mean and red circles represent individual sample data.

For the *L1/L2* region (Figure 6.16) the characteristic "N" shaped pattern (L1L2-1 low, L1L2-2 high, L1L2-3 low and L1L2-4 high) of CpG methylation was present that was also seen elsewhere in this study. No significant differences were observed among the three outcome groups for any of the *L1/L2* CpGs. The *L1/L2* region did not appear to be very informative in this instance.
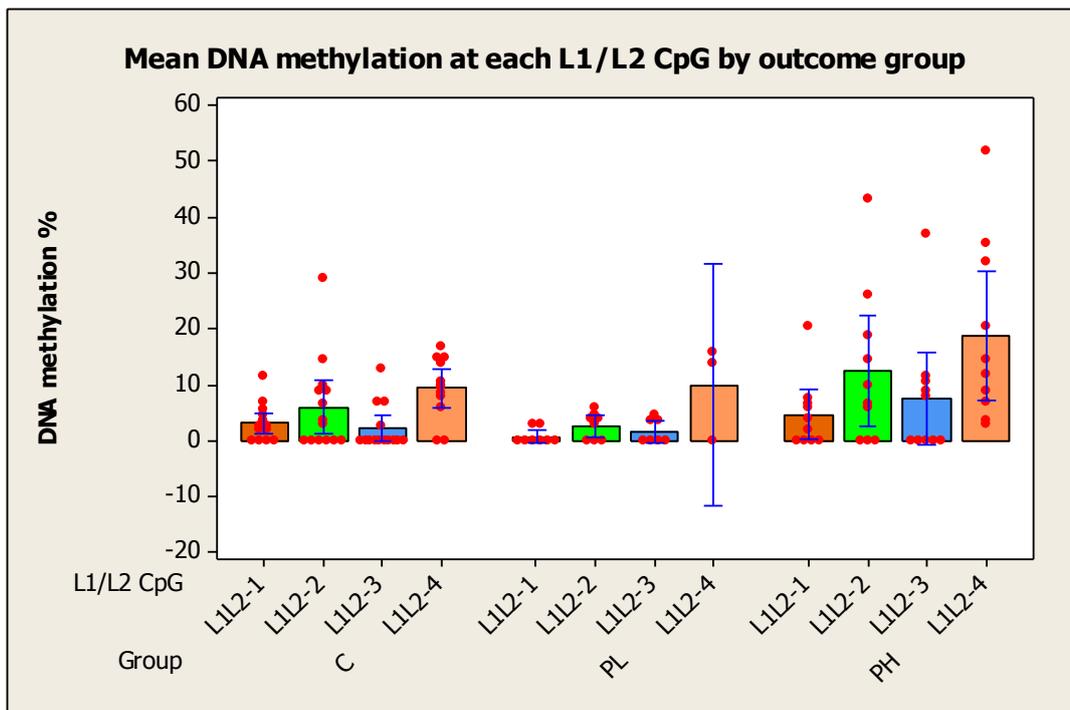
Figure 6.16: The mean amount of DNA methylation at each *L1/L2* CpG for the 0 month samples of each of the three outcome groups.

C=cleared group, PL=persistent low grade and PH=persistent high grade. Each bar represents mean DNA methylation for the indicated CpG/outcome group combination, blue error bars represent 95% confidence intervals for mean and red circles represent individual sample data.

6.1.4. Hypothesis 4: The Level of DNA Methylation of the CRISP Samples Should be Comparable to the Disease Grade Study

This question was posed because despite all three outcome groups originally being cytologically classified as low grade cervical abnormalities, the histology of the infections at 6 months differed considerably. Some patients lost the infection completely whilst others maintained the infection and had either low or high grade histology. The normal cytology vs. severe dyskaryosis DNA methylation comparison found significant differences in DNA methylation among the cytology grades (Chapter 5: 5.1.2). It was of interest to consider whether or not the 0 month samples of the three CRISP outcome groups were more similar to the normal cytology group or the severe dyskaryosis group both in terms of mean DNA methylation level and of CpG specific methylation profile. If the DNA methylation level/profile of one outcome group was more similar to the DNA methylation level/profile of a more severe cytology result then DNA methylation data might indicate higher grade disease at 0 months than the 0 month cytology had suggested.

For the *E2* region, the CpG DNA methylation profile of the three CRISP outcome groups did not appear to relate to the methylation profile of either the normal cytology or severe dysplasia

groups. Similarly, there did not appear to be a discernible similarity for the LCR1 or LCR2 regions either. For the *L1/L2* region however, the distinctive "N" shaped profile was observed in all outcome groups.

Overall DNA methylation of the C and PL groups was most similar to the normal cytology group. This was to be expected because at 0 months, all of these samples had low grade cytological abnormalities and indeed for the PL group, 6 months later had low grade histology (Table 6.6). The PH group however were methylated at a level between the normal cytology and the severe dyskaryosis samples. Six months later, these samples were found to have high grade (CIN2/3) histology.

	Mean DNA methylation (%)			
	E2	LCR1	LCR2	*L1/L2*
C group (0 month)	3.7	3.1	0.7	5.2
PL group (0 month)	2.2	3.5	0.6	2.3
PH group (0 month)	4.1	2.6	1.9	10.9
Normal cytology	2.5	3.1	1.5	5.7
Severe dyskaryosis	6.6	2.3	2.6	16.3
Cancer	42.7	12.8	23.7	50.1

Table 6.6: Mean DNA methylation of each HPV region for the CRISP groups and the previously tested disease groups.

Each number represents a percentage methylation which is calculated as an average of all CpGs in each region.

6.1.5. Hypothesis 5a: There is an Association between DNA Methylation and Outcome Cytology

For the CRISP samples, cytology was performed at both 0 and 6 months. All of the cytology results at 0 months were either second borderline or mild abnormalities. At 6 months, various cytology results were produced which included negative, borderline, mild, moderate and severe dyskaryosis. It was of interest to determine if DNA methylation data for the 0 month samples was different among the various grades of 6 month outcome cytology. For this comparison, the CRISP samples were not separated by outcome group but were instead grouped together (Figure 6.17). Outcome cytology results were grouped into negative cytology, borderline/mild abnormalities (i.e. no change from 0 months), moderate dyskaryosis and severe dyskaryosis. Very few patients had moderate dyskaryosis or severe dyskaryosis at 6 months, which resulted in large confidence intervals. Nonetheless, multiple one-way ANOVAs were performed and determined that a significant difference in mean DNA methylation existed among the various 6 month cytology types for the LCR2 region (P=0.010) and almost for the *L1/L2* region (P=0.030) (Table 6.7). ANOVA assumption violations are detailed in the table

legend, a Bonferroni corrected P-value of P=0.017 should be considered. The only significant Tukey-Kramer pairwise comparisons were for the LCR2 region, where DNA methylation of the moderate cytology samples was significantly lower than the samples where no change was observed, or those where a negative cytology result occurred (Table 6.7).

The samples where a moderate or severe cytology result was detected at 6 months could be more methylated in the *L1/L2* region however it is impossible to determine from this data set because of the small sample number, large confidence intervals and for the moderate cytology samples, a large range of DNA methylation values. A larger sample series would be required in order to further investigate this possibility.
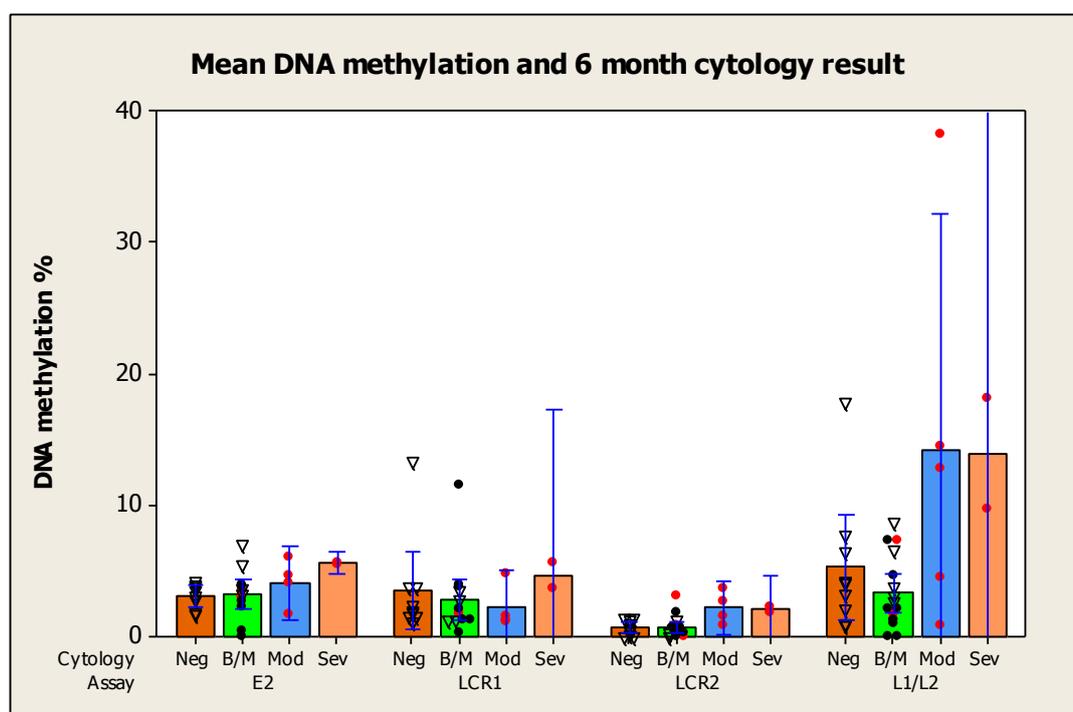
Figure 6.17: The relationship between mean assay 0 month DNA methylation and 6 month cytology result.

Four abbreviations are used to represent the cytology at 6 months; Neg = negative cytology, B/M = borderline/mild abnormalities (i.e. no change from 0 months), Mod = moderate dyskaryosis and Sev = severe dyskaryosis. Coloured bars represent mean DNA methylation and blue error bars represent 95% confidence intervals for mean. Individual values are plotted as three symbols, representing the samples of the cleared (C) group (white triangles), the persistent low grade (PL) group (black circles) and the persistent high grade (PH) group (red circles). Because only two samples had severe cytology at 6 months the error bars were very large and if plotted, would disrupt the scale of the graph hindering interpretation. The upper 95% confidence interval of the *L1/L2* severe dyskaryosis group was 67.1% methylation.

	HPV16 region			
	E2	LCR1	LCR2	*L1/L2*
One-way ANOVA (P-value)	0.373*[1]	0.513*[2]	0.010	0.030
Significant Tukey-Kramer results	B/M vs. Mod			
	Mod vs. Neg			

Table 6.7: The results of four one-way ANOVAs to test for a difference in mean 0 month DNA methylation among the four different 6 month cytology results.

B/M = borderline/moderate, i.e. no change between 0 and 6 month cytology, Mod = moderate dyskaryosis, Neg = negative cytology. Violations of the assumptions of the ANOVA: *1, non-normally distributed residuals (P<0.005), unequal variances (P=0.025); *2, non-normally distributed residuals (P<0.005).

6.1.6. Hypothesis 5b: There is a Relationship between Outcome Cytology and Outcome Histology

At 6 months, both cytology and histology were performed. Samples were assigned to the three groups (C, PL and PH) based on the histology and/or HPV16 positivity result. The individual value points in Figure 6.17 show the relationship between histology, HPV16 positivity and cytology in the 6 month samples. It is quite telling that all of the samples with moderate or severe dyskaryosis at 6 months also had high grade histology. Similarly, all of the samples with a negative cytology at 6 months also belonged to the cleared group, representing the loss of HPV16 positivity. The samples with 6 month borderline/mild abnormalities were quite variable and some samples belonged to each outcome group (i.e. C, PL and PH). All PL group samples had no change in cytology at 6 months.

6.2. Discussion

6.2.1.1. Aims

This study was planned in order to determine whether there were characteristic differences in DNA methylation associated with different outcomes in a series of patients with low grade cytological abnormalities. It was a longitudinal investigation that aimed to correlate methylation patterns with disease in LBC material from women with histologically confirmed disease, and allow initial assessment of the prognostic potential of a knowledge of HPV DNA methylation. We were mindful of the biomarker development road maps published by CRUK, and this study partly fulfilled a component of the early prognostic biomarker development protocol (Appendix I) (Cancer Research UK, 2012d).

6.2.1.2. Strengths

Longitudinal studies looking at HPV DNA methylation are uncommon. The only other longitudinal study is very recent and the design and results were remarkably similar to this study, albeit with small differences (Mirabello *et al.*, 2012). The CRISP study results are where relevant, consistent with the NSC study (Chapter 5).

The implications of this study are that HPV methylation assessment has potential as a method for differentiating low grade cytology samples which are at higher risk (i.e. those that will have high grade cytology/histology in 6 months time) and there are several aspects of the data that might suggest an association between HPV DNA methylation and episomal clearance and/or HPV persistence.

Pyrosequencing is a fully quantitative method for assessing DNA methylation and we have targeted four regions of the genome, each identified by previous studies as regions where DNA methylation is potentially associated with cervical disease.

6.2.1.3. Weaknesses

The biggest weaknesses of the study are the relatively short duration of follow up and the lack of histology at 0 months. These limitations are present because the initial samples were collected during routine screening and as part of a clinical trial, not as a longitudinal biomarker study for DNA methylation. Histology is not routinely performed for the first low grade/second borderline result and as such was not available.

The CRISP samples were from patients entered into a clinical trial to study the effects of diindolylmethane (DIM), a dimer of Indole-3-carbinol. Both compounds occur in cruciferous vegetables and are commonly marketed as dietary supplements. It was hypothesised that DIM might inhibit cervical dysplasia (Castanon *et al.*, 2012). It could be questioned whether samples from a treatment trial can reliably provide information on prognosis, however the trial results were negative; DIM supplementation did not significantly affect disease progression/regression, or cause clearance of HPV infection (Castanon *et al.*, 2012). Nonetheless, there was a possibility that DIM might affect HPV DNA methylation however it was not possible to control for this, as the methylation study was performed blind to the DIM treatment regime.

Several samples failed to produce high quality pyrosequencing data. This was especially common for the cleared group where 9 samples failed all assays. The primary reasons for high failure rate of the cleared samples were likely to be a false HPV16 positivity result during the initial HPV typing (by PCR ELISA) and/or sensitivity limitations of the pyrosequencing assays.

6.2.1.4. Comparison to Other Studies

Longitudinal studies of HPV DNA methylation are uncommon. Only one such study was identified by a literature search. Mirabello *et al* used a sample cohort grouped into three outcomes, similar to the outcome groups used by this study. These outcomes were; a control group where HPV16 clearance over the course of 2 years was observed (i.e. similar to the cleared group), a persistence group where persistent HPV16 infection was associated with low grade disease (i.e. similar to PL group) and a CIN3 group, where persistence was associated with CIN3 (i.e. similar to PH group) (Mirabello *et al.*, 2012).

The study also used pyrosequencing to assess multiple sites of the HPV genome however differed considerably in the statistical analysis and aims/approach of the study. The main findings were that methylation of several CpGs of the *L1*, *L2* and *E2*/E4 regions corresponded with an increased risk of persistence and CIN3. This is largely consistent with the results of our study, albeit our C group was more (but not significantly) methylated in the *E2* and *L1/L2* regions than our PL group, which was not observed by Mirabello *et al.* It is difficult to explain the differences between the studies; both used pyrosequencing, both used desquamated cervical cells (i.e. LBC) and for both studies only HPV16 positives were considered. However they differ in the observed methylation status of infections destined to be cleared of HPV16. It is worth noting that the analysis was conducted differently and the duration of follow up was also different (6 months vs. 2 years). A longitudinal study of viral clearance of HPV16 positive women observed a 32% clearance rate after 6 months, and 50% clearance after 12 months (Hildesheim *et al.*, 2007). This suggests that most HPV16 infections take longer than 6 months to resolve and that our shorter duration of follow up may have for instance, identified DNA methylation profiles associated with an imminent loss of infection.

The three groups derived from the CRISP study were based on outcome histology at 6 months. The PL (low grade histology) group was less methylated at 0 and 6 months than the PH group (high grade histology) for the LCR2 and *L1/L2* regions, and at 0 months for the *E2* region. This is consistent with our NSC study (where *E2* and *L1/L2* methylation were higher in more severe disease) and also with many published studies (Mirabello *et al.*, 2012, Kalantari *et al.*, 2004, Bhattacharjee and Sengupta, 2006a, Ding *et al.*, 2009, Brandsma *et al.*, 2009, Hong *et al.*, 2008, Fernandez *et al.*, 2009, Sun *et al.*, 2011). It is not consistent with several studies, which have reported LCR hypermethylation in low grade disease (Badal *et al.*, 2003, Hublarova *et al.*, 2009, Patel *et al.*, 2012, Piyathilake *et al.*, 2011). It is notable that there are a range of methodologies, sample materials, target regions, sample numbers, as well as a range of study aims in these conflicting reports (Appendix VI).

One of the points of conflict for our NSC study was the presence/absence of LCR methylation in HPV positive cytologically normal samples. Substantial LCR methylation was not observed in either our CRISP or NSC samples which is consistent with several studies (Ding *et al.*, 2009, Fernandez *et al.*, 2009, Bhattacharjee and Sengupta, 2006a) but inconsistent with others (Hong *et al.*, 2008, Patel *et al.*, 2012, Kalantari *et al.*, 2004, Badal *et al.*, 2003). Again, a range of analytical methods and sample materials were used in these studies. Our *in vitro* studies (Chapter 7: 7.1.2 & 7.1.3.3) observed changes in HPV methylation with increasing passage number (i.e. number of cell culture cycles/time). This study observed rather stable DNA methylation over the course of 6 months *in vivo* suggesting that in some instances at least, the cell lines may not be representative of *in vivo* methylation processes or dynamics.

6.2.2. Meaning

6.2.2.1. DNA Methylation was Stable over 6 Months However Persistence may be Associated with Increased E2 DNA Methylation.

DNA methylation did not change for the PH group and methylation did not change for the PL group at any region other than *E2*. PL group *E2* methylation was seen to differ significantly when analysed using a multifactorial ANOVA; there was only a very minor violation of the analysis assumptions and the analysis appeared robust. When multiple single factor analyses were performed, the difference in PL group *E2* methylation was not significant. The variation between analyses was likely the result of variable methylation differences among individual CpGs and the additional analytical power afforded by the multifactorial analysis. For the PL group, methylation at every *E2* CpG increased over the course of 6 months. This was not observed for any other region where DNA methylation at both time points was very similar.

The PL group had low grade histology and cytology at 6 months, suggesting that disease for most samples was low grade and had not progressed over the course of the study. That these persistent infections exhibited an increase in *E2* DNA methylation in later samples might hint at a mechanism of persistence in the absence of disease progression, i.e. that increasing *E2* DNA methylation is associated with or even involved with persistence. Other studies have suggested a mechanism of HPV latency by DNA hypermethylation, albeit of the LCR and in the absence of visible disease (Vinokurova and von Knebel Doeberitz, 2011). We however observed low and consistent LCR methylation in the PL group. The differences between the studies might be related to differences between a persistent infection with and without signs of disease. DNA methylation is also implicated in the latency lifecycle of EBV (Paulson and Speck, 1999), although such a well defined latency period is not present for HPV. Alternatively,

as the 6 month study period is relatively short (i.e. the development of cervical cancer is often said to take 10-20 years (Cancer Research UK, 2012a, McIndoe *et al.*, 1984), *E2* methylation might represent an early event in the progression to a higher disease state.

6.2.2.2. DNA Methylation was Associated with Clinical Outcome

6.2.2.2.1. There is a Need for Biomarkers

Low grade cytology (borderline/mild) is a common result (6% of the UK primary screening results) and has a complex management strategy (Castanon *et al.*, 2012). These lesions cause considerable distress to women in whom they occur, and in several thousand low grade cytology cases referred to colposcopy there was a mix of HPV positivity and histology (albeit predominantly low grade histology) (Table 6.8). In England, HPV testing is being adopted as a triage test for these patients. HPV testing is likely to replace cytology as a primary screening method, potentially with cytology being used to triage the high number of HPV positive patients (Naucler *et al.*, 2009, Rijkaart *et al.*, 2012). In these instances, further triage with a prognostic biomarker with which to determine a suitable management strategy will be of substantial value. DNA methylation of the 0 month samples varied significantly among the outcome groups and may represent a means by which to risk stratify mild/second borderline cytology cases.

	Referral Cytology			
	Inadequate/ borderline/ mild		Moderate or worse	
Histology	Number	%	Number	%
Total	4535	100	3271	100
Cancer	3	0.1	80	2.5
CGIN	2	0.1	31	1
CIN3	227	5.9	1256	40
CIN2	377	9.8	681	21.7
CIN1	905	23.5	366	11.6
HPV only	413	10.7	118	3.8
No HPV/CIN	1065	27.6	323	10.3
Normal colposcopy, no biopsy	820	21.3	266	8.5
Inadequate biopsy	46	1.2	22	0.7
Unknown	677	14.9	128	3.9

Table 6.8: The relationship between cytology and histology for women referred to colposcopy. Figures represent data provided by (Cervical Screening Wales, 2004).

6.2.2.2.2. Differences Were Observed among the Outcome Groups

Significant differences in methylation among outcome groups were observed for the *L1/L2* region. Methylation of the *L1* and *L2* regions is frequently associated with more severe disease (Fernandez *et al.*, 2009, Kalantari *et al.*, 2009, Sun *et al.*, 2011, Mirabello *et al.*, 2012), and CpGs of the *L1*, *L2* and *E2*/E4 region were also identified as informative by a similar longitudinal study (Mirabello *et al.*, 2012). Because cleared infections represent low level disease, we had expected to see low level methylation in the C group. However, C group mean *L1/L2* methylation was at a level between the PL and PH groups, this is considered in further detail later in this discussion chapter. When individual *L1/L2* CpGs were assessed, none were significantly differently methylated in the three outcome groups; the difference between mean methylation and individual CpGs is likely the result of considerable and significant CpG specific methylation differences in the *L1/L2* region.

Differences in 0 month LCR2 methylation were also observed; mean LCR2 methylation was higher in samples that had high grade histology at 6 months (PH) than those that had low grade histology (PL) or cleared the infection (C). When individual CpGs were considered rather than mean regional methylation, only two CpGs of the 20 studied were differentially methylated in the three outcome groups (LCR1-1 and LCR2-3). Methylation of LCR1-1 was different among the groups, but surprisingly not between C and PH. Highly significant differences between the PH group and the C/PL groups were observed for methylation of the LCR2-3 CpG, which might have clinical significance.

In a similar study, differences among cleared and persistent infections and CIN3 were identified in CpGs of *L1* and *L2*, and *E2*/E4 (Mirabello *et al.*, 2012). That they did not identify LCR methylation, and we did not identify *E2* methylation or *L1/L2* CpG specific methylation as informative may stem from differences in samples size, disease misclassification, group classification, method of statistical analysis and duration of follow up. Nonetheless, consideration of methylation at multiple CpGs would likely represent an effective means by which to differentiate the samples in the three outcome groups. Further analysis and generation of a suitable algorithm is a prospect for future work.

6.2.2.2.3. Cause/Consequence of the Differences

There are several possible explanations for the associations between DNA methylation in the 0 month samples and outcome at 6 months. All CRISP samples had low grade cytology at 0 months, however cytology is occasionally a poor indicator of histology (Cervical Screening Wales, 2004). If the PH group underwent progression from low grade disease to high grade disease, we might expect DNA methylation changes over the course of the 6 months, which

were not observed. Rather than the differences in DNA methylation between the PL and PH groups representing a progression of the PH group from an initially low grade disease to an eventual high grade disease, it seems more plausible that there were existing differences in disease severity (i.e. histology) that were not represented by the first cytology result, but were associated with differences in DNA methylation. That is to say, the low/high grade histology seen at 6 months were likely also present in the 0 month samples but were not accurately represented by the 0 month cytology. This explanation is supported by the known differences in DNA methylation for different disease grades (NSC study) (Fernandez *et al.*, 2009, Mirabello *et al.*, 2012) and by the diagnosis of CIN2+ disease in women referred for low grade cytological abnormalities (15.9% of Inadequate/ borderline/ mild cytology cases) (Table 6.8). If this is correct, then assessment of DNA methylation in conjunction with cytology might be capable of identifying low grade cytology cases where further investigation is warranted.

6.2.2.3. Outcome Cytology was Associated with Initial DNA Methylation

Whilst at 0 months, all CRISP samples had either mild cytological abnormalities or were a second borderline result, the 6 months cytology reports were more diverse. We compared 0 month methylation and 6 month cytology, to determine whether DNA methylation was associated with 6 month follow up cytology. The most significant differences were observed for the LCR2 region. Differences in *L1/L2* methylation were not quite significant when the Bonferroni correction was applied and the two-way analyses (considering individual CpG data) suffered several violations of the assumptions so should be considered with caution. Whilst not quite significant, *L1/L2* methylation did have the largest absolute differences among the groups, albeit with large confidence intervals and the data was difficult to analyse as a result of the relatively small sample size. A combination of LCR2 and *L1/L2* methylation data might prove useful as a means of differentiating low grade cytology cases that will be low grade/negative cytology or moderate/severe dyskaryosis in 6 months time.

6.2.2.4. Differences between 6 Month Cytology and Histology Were Observed

At 6 months, all of the negative cytology cases were HPV16 negative, there was a mix of histology/HPV positivity for the mild/borderline cytology cases and all of the moderate/severe dyskaryosis samples had high grade histology. Moderate/severe dyskaryosis is typically associated with high grade CIN in most (62-78%), but not all instances (Cervical Screening Wales, 2004, Cervical Screening Wales, 2011); our 100% association is likely the consequence of assessing a small number of cases with moderate/severe cytology.

6.2.2.5. *The Cleared Group had More L1/L2 and E2 Methylation than the PL Group*

More severe cervical disease was associated with higher methylation of the *E2* and *L1/L2* regions (Chapter 5: 5.1.2) (Fernandez *et al.*, 2009, Kalantari *et al.*, 2009, Sun *et al.*, 2011, Mirabello *et al.*, 2012). Nonetheless at 0 months, the cleared group had a higher degree of *E2* and *L1/L2* methylation than the PL group. Whilst these differences were not significant, the differences were quite large and were contrary to the expected trend. They also differed from recently published data, where *E2/E4*, *L1* and *L2* methylation were greater in persistent (low grade) HPV16 infections than in infections that would be cleared and greater again in CIN3 lesions (although these differences were small and frequently non-significant) (Mirabello *et al.*, 2012).

Most of the C group samples had negative cytology at 6 months so the loss of HPV16 positivity seems genuine. A pertinent consideration however is that DNA methylation of the C and PH groups could be the result of host cell attempts to target DNA methylation to viral episomes in order to silence viral gene expression and/or promote HPV clearance. In different backgrounds of disease, episomal clearance could have diverse outcomes; for the C group, HPV16 clearance was achieved and the patients had predominantly negative cytology. For the PH group, representing patients with CIN2/3, a stage of disease where HPV DNA integration is quite common, clearance of the episomal forms could represent a growth advantage to the host cell or even result in disease progression. Several prospective treatments for HPV infection rely on the ability to induce an antiviral response that results in the clearing of episomal HPV genomes, for example, exogenous type I interferon and imidazoquinolone compounds (Cirelli and Tyring, 1994, Stanley, 2002). Viral clearance could result in the loss of inhibitory episomal HPV genomes, *in vitro* this can result in the emergence of cells containing integrated HPV genomes (Pett *et al.*, 2006, Herdman *et al.*, 2006). There are also accounts of elevated disease progression in small scale clinical trials and increased HPV oncogene expression in cells lines as a result of antiviral response inducing compounds (Kim *et al.*, 2000, Koromilas *et al.*, 2001). A relationship between episomal methylation and clearance has not been proven, however is worthy of further consideration.

The differences between DNA methylation of the cleared group for our study and for Mirabello *et al* might lie in the duration of follow up. The shorter 6 month follow up may have identified samples that were soon to be cleared of HPV16 infection. The higher methylation level observed in our study might represent changes in HPV methylation that immediately precede rapid episomal clearance. There is currently no data concerning HPV methylation and episomal

clearance in the absence of integration. When integration is present, clearance of episomes is often rapid and associated with changes in DNA methylation (Chapter 7: 7.1.2) (Gray *et al.*, 2010).

6.3. Conclusions

This study attempted to assess the prognostic potential of methylation in a longitudinal study. The methylation profile of the three outcome groups was compared and differences were observed, particularly at the *L1/L2* region and parts of the LCR. However, these differences could be caused by several mechanisms so the results must be interpreted with caution. The observation of higher methylation of the cleared group than the more severe disease PL group was unexpected, however could represent a mechanism of episomal clearance so is worth further investigation. We have at least partly fulfilled a component of the prognostic biomarker development protocol. The potential of HPV methylation as a biomarker is discussed in further detail in the general discussion (Chapter 9: 9.3).

6.3.1. Future Work

An interesting avenue for further work would be to establish whether a relationship between HPV methylation and episomal clearance exists. This could potentially be investigated *in vitro* using models like the W12 cell line or PC08/09 short term cell lines. A larger clinical study might involve short follow up in a series of HPV positive normal cytology samples in order to correlate methylation with spontaneous regression in the absence of visible disease.

This study was performed blind to DIM treatment. Although the putative mechanism of action of DIM (modulation of oestrogen metabolism) does not suggest an obvious link with DNA methylation, it would be pertinent to determine whether an association existed between DIM treatment and 6 month DNA methylation. The relevant data was not available for this study however if access could be granted it would be an interesting avenue to explore.

Finally, this study used a relatively small sample number and a single follow up point with a short time span. Expanding the number of samples, the time between points or the number of consecutive time points would likely improve the results as changes in disease grade might be observed and correlated with changes in DNA methylation.

Chapter 7. The Cell Lines Study - DNA Methylation, Integration and Gene Expression *in vitro*

7.1. Results

To supplement the studies conducted using clinical material, a small series of additional *in vitro* studies were performed using cell lines. These studies facilitated assay design and protocol development and allowed parallels to be drawn between *in vivo* and *in vitro* studies. Three sets of cell line material were used; the immortal cell line CaSki, the short term cell lines PC08/PC09 and the CIN progression model W12.

7.1.1. CaSki

CaSki has been used in several HPV DNA methylation studies, it was important to consider how the DNA methylation data produced by pyrosequencing in this study corresponded to published DNA methylation data. Table 7.1 contains the DNA methylation data for CaSki produced during this study. This data is presented graphically in Figure 7.1. In summary, CaSki HPV16 genomes are hypermethylated in all regions, especially the *E2* and LCR2 regions.

E2 CpG	E2-1	E2-2	E2-3	E2-4	E2-5	E2-6	E2-7	E2-8
Nucleotide	3411	3414	3416	3432	3435	3447	3461	3472
Methylation %	95.5	98.5	80	100	96.5	93	81.5	77

Table 7.1.a

LCR1 CpG	LCR1-1	LCR1-2	LCR1-3
Nucleotide	7691	7679	7673
Methylation %	63	67.5	76

Table 7.1.b

LCR2 CpG	LCR2-1	LCR2-2	LCR2-3	LCR2-4	LCR2-5
Nucleotide	31	37	43	52	58
Methylation %	89.5	91.5	96.5	95.5	97

Table 7.1.c

L1/L2 CpG	L1L2-1	L1L2-2	L1L2-3	L1L2-4
Nucleotide	5615	5609	5606	5600
Methylation %	74.5	87.5	76	87.5

Table 7.1.d

Table 7.1.a-d: DNA methylation of CaSki HPV16.

Four tables are present, one for each tested region of the HPV genome. Included are the corresponding nucleotide positions (GenBank ID NC001526.1).
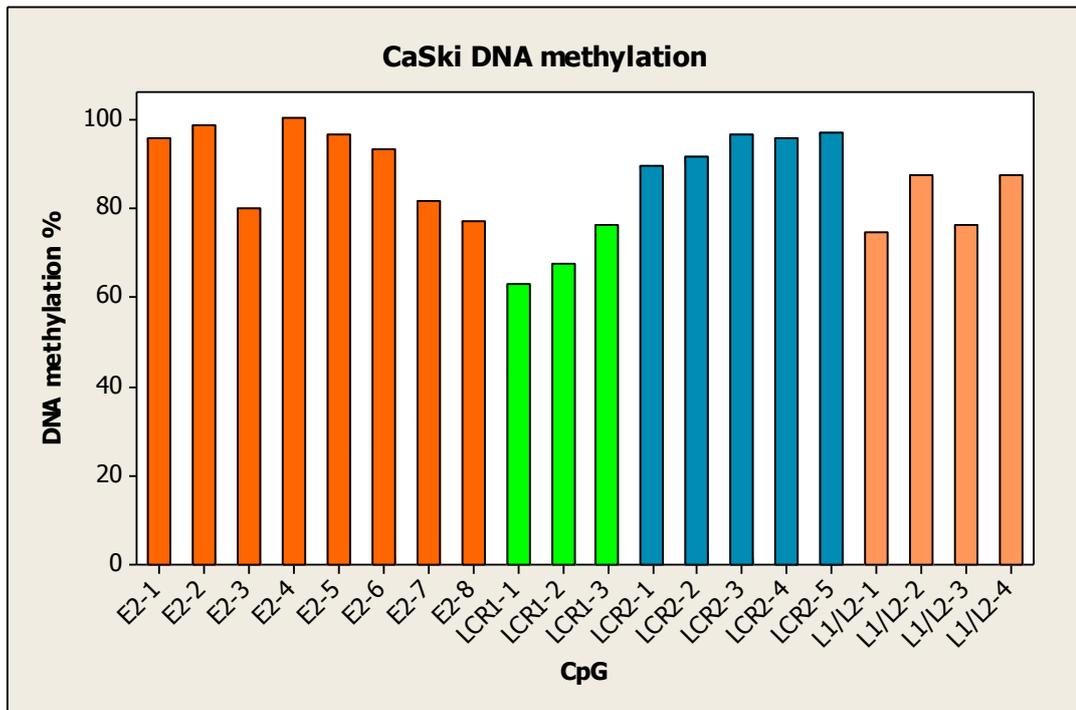
Figure 7.1: CpG methylation of HPV16 in CaSki.
Each of the 20 CpGs tested are listed, bars are colour coded to aid interpretation.

7.1.2. The W12 Series 7 Cell Line

Pyrosequencing was applied to DNAs from several passages of the W12 series 7 cell line (W12Ser7). The term passage refers to the successive subculturing of cells into new flasks, i.e. a flask is inoculated with cells, the cells are grown to near confluence and then a small number are re-inoculated into a new flask. One round of subculture represents one passage. DNAs from passage 11 (p11) to p50 were available. In series 7, integrated HPV16 was first detected at p13. Between p13-17 episomal HPV16 constituted 10-90% of viral load however after p17, episomal HPV16 constituted less than 10% of the viral load (Gray *et al.*, 2010). No further detail about the site(s) of HPV integration was available for W12Ser7. Nonetheless, there appeared to be changes in DNA methylation starting during this period (p13-17). *E2* methylation was quite closely matched to LCR1 and LCR2 DNA methylation and was much lower than seen in the *L1/L2* region (Figure 7.2).

In addition to the low overall *E2* methylation, there was a wide variation in methylation within the *E2* region (Figure 7.3). DNA methylation at p50 for instance varied from 72.5% for CpG E2-5 down to 8% for E2-2. Particularly intriguing was that E2-4 is separated from E2-5 by only 3 bp, yet DNA methylation differed by over 50%.This suggests that a very specific mechanism could be responsible for methylating the region. For E2-5 and E2-6, DNA methylation increased

substantially between p15-17 which corresponded to the integration events between p13-17. DNA methylation of other *E2* CpGs remained constant and low.

Similar to the *E2* region, CpGs LCR1-2 and LCR1-3 appeared to become increasingly methylated after the integration event at p13-17 (Figure 7.4). LCR1 CpGs are present in the viral enhancer and are not within E2BSs. DNA methylation of the LCR2 CpGs increased following the integration event at passage 13 but in later passages (p42+) appeared to decline (Figure 7.5). Four of the LCR2 CpGs (LCR2-2... LCR2-5) are within two of the inhibitory E2BSs of the LCR.

For the *L1/L2* region, p11 data was missing so it was difficult to see when the increase in DNA methylation at each CpG began; the increase in methylation between p15-17 was greater than between p13-15, suggesting that the majority of the changes occurred somewhere between p15-17 (Figure 7.6). For *L1/L2*, there was at times more than a two-fold difference between the levels of DNA methylation at different *L1/L2* CpGs (i.e. L1L2-4 vs. L1L2-1 for p17-31). Nonetheless, there was considerably less disparity among the level of DNA methylation at each of the *L1/L2* CpGs than there was for the *E2* CpGs.
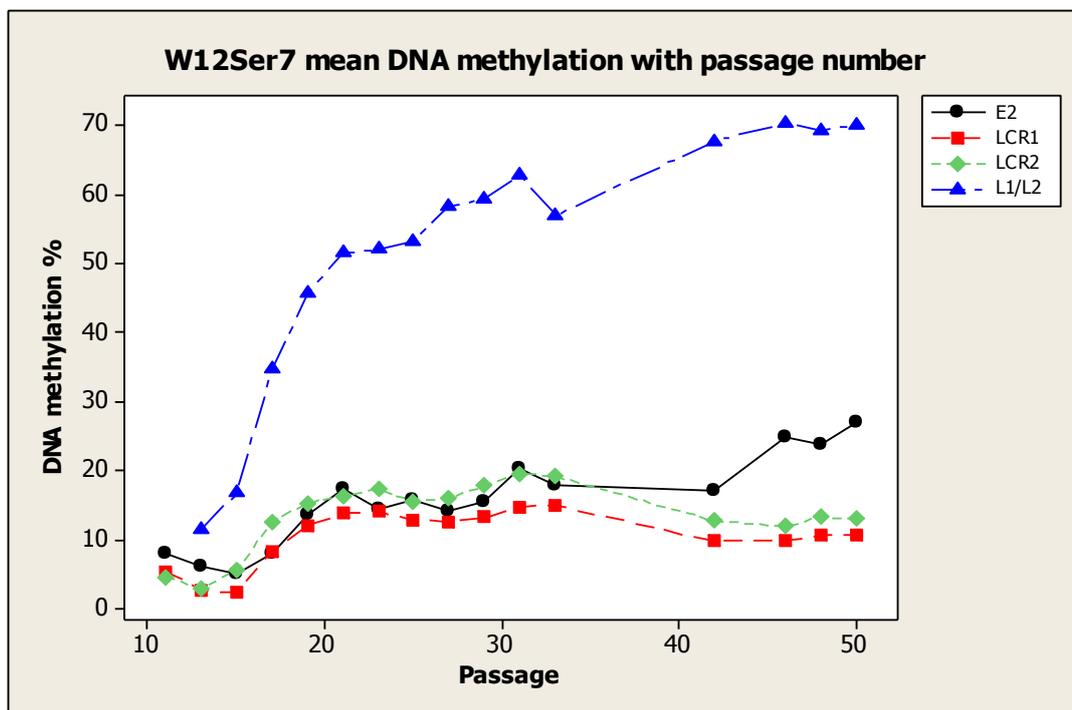
Figure 7.2: Mean DNA methylation at each site of the HPV16 genome for the W12Ser7 cell line with passage number.

Each coloured point represents a different HPV region, connecting lines are present to aid the interpretation of the data.
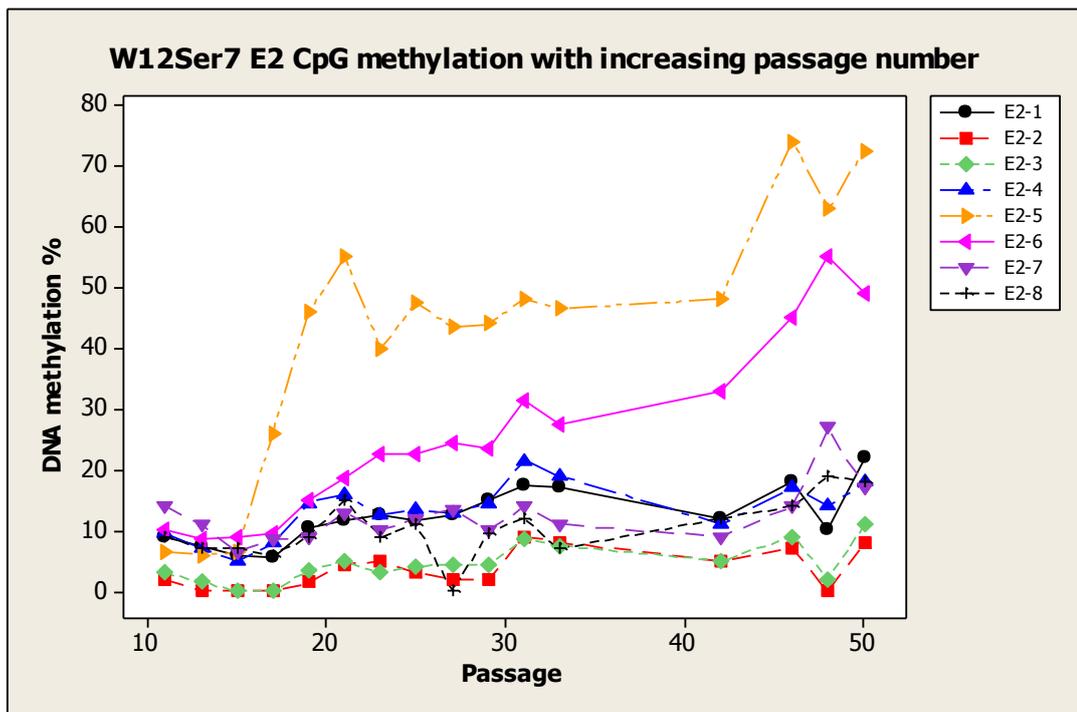
Figure 7.3: The mean amount of DNA methylation at each *E2* CpG for the W12Ser7 cell line with passage number.

Each coloured point represents a different CpG, connecting lines are present to aid the interpretation of the data.
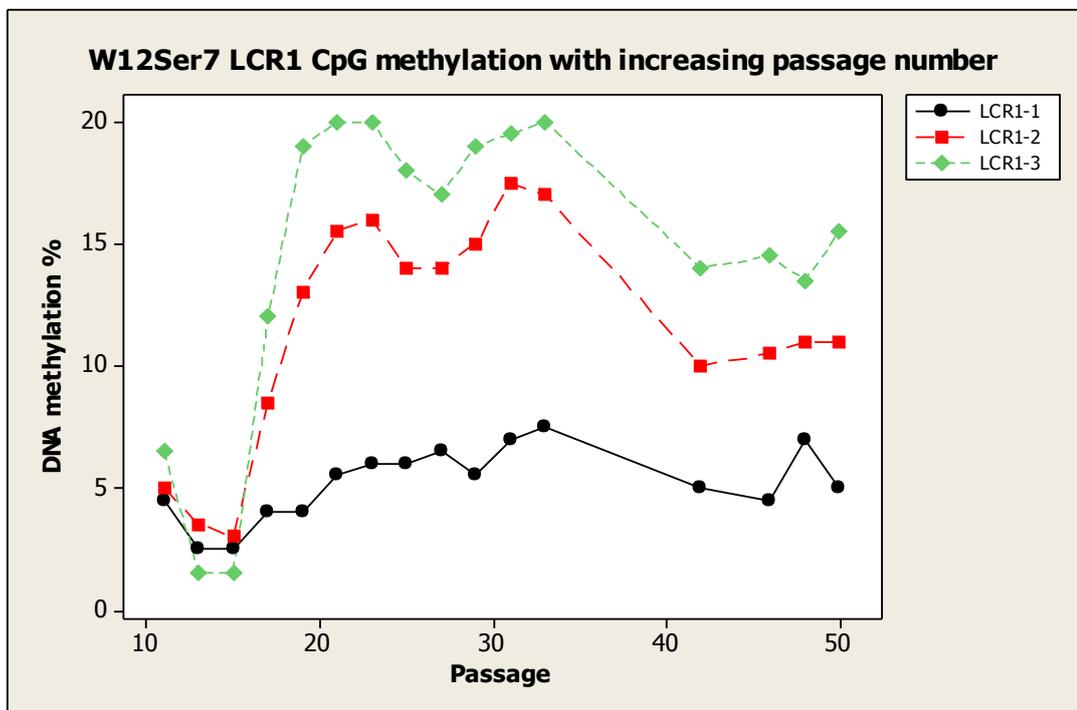
Figure 7.4: The mean amount of DNA methylation at each LCR1 CpG for the W12Ser7 cell line with passage number.

Each coloured point represents a different CpG, connecting lines are present to aid the interpretation of the data.
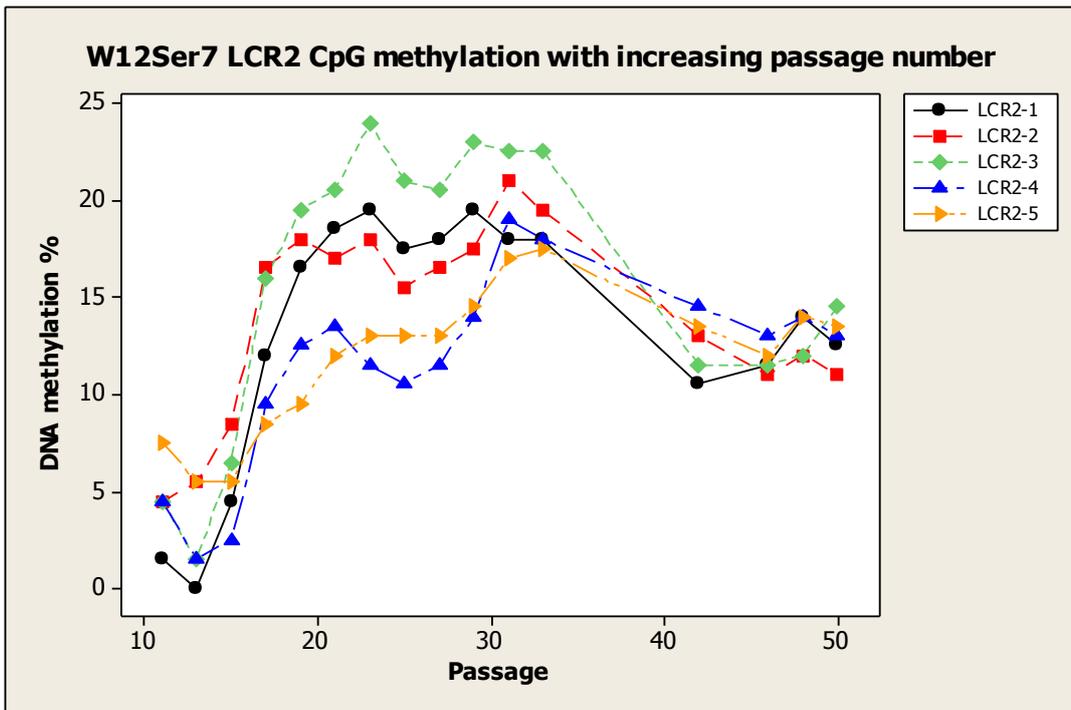
Figure 7.5: The mean amount of DNA methylation at each LCR2 CpG for the W12Ser7 cell line with passage number.

Each coloured point represents a different CpG, connecting lines are present to aid the interpretation of the data.
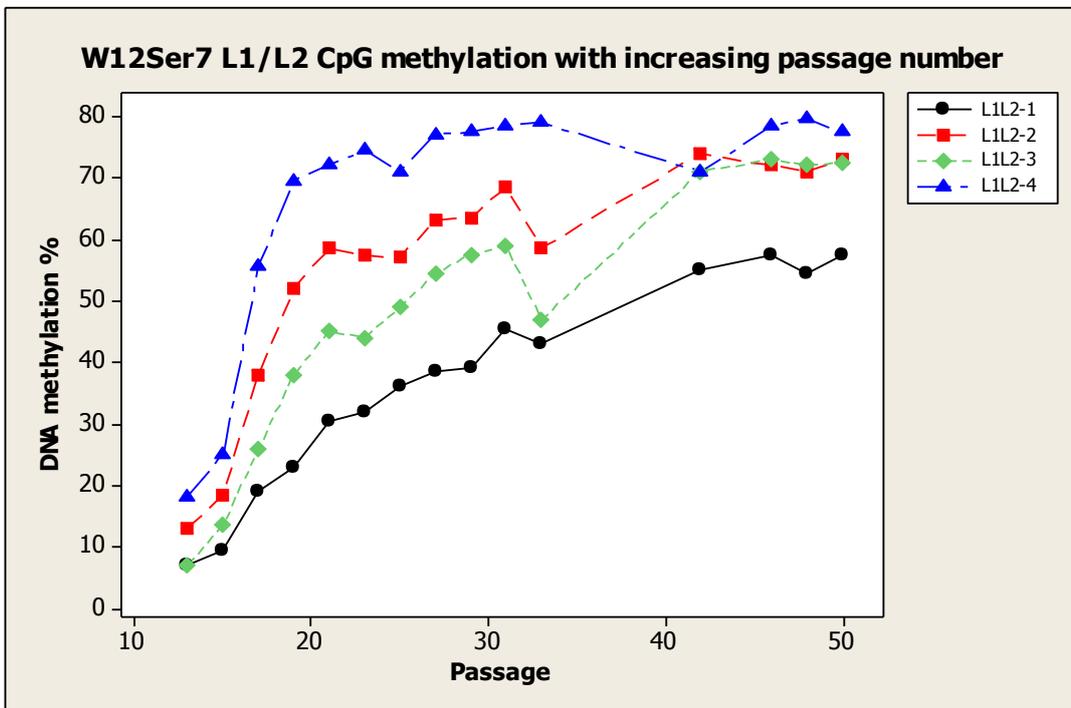
Figure 7.6: The mean amount of DNA methylation at each *L1/L2* CpG for the W12Ser7 cell line with passage number.

Each coloured point represents a different CpG, connecting lines are present to aid the interpretation of the data.

7.1.3. The PC08 and PC09 Short Term Cell Lines

DNAs for this study were provided by another student (Tiffany Onions (TO)) from a series of short term culture cell lines. The PC08 and PC09 cell lines were established as primary cultures of biopsies from two different patients. PC08 was derived from a VIN3 biopsy whilst PC09 was derived from a VAIN3 biopsy. Initially, these two cultures were established as heterogeneous lines, then single cell cloning was performed to establish homogenous lines. In total, 3 clones for each biopsy were provided for use in this study, these are the PC08 clones M, P and Y and the PC09 clones A, D and H. These clones were then cultured and at each passage, DNA and RNA was extracted and gene expression, growth rate, integration status, telomere length were assessed (by TO). DNAs for DNA methylation analysis using the pyrosequencing assays were provided by TO.

For the rest of this study, the sample names will be generated according to passage number, clone name and biopsy of origin, i.e. 13Y9 is the PC09 clone Y, at passage 13. This data set was of particular interest because it represented a unique *in vitro* model that allowed comparisons to be made between several major components of HPV biology. The PC08 and PC09 data was analysed by testing a series of hypotheses.

7.1.3.1. HPV DNA Integration

Integration analysis by APOT was conducted by TO (Vinokurova *et al.*, 2008). A limitation of APOT is that only actively transcribed HPV sequences are detected and APOT is unable to confirm the absence of integration, i.e. not detecting an integration event would not necessarily mean that an integration event was not present. Similarly, detected episomally derived transcripts could in theory be derived from transcriptionally active genomes integrated as concatemers. Another limitation of the APOT assay stems from the use of transcript sequences, HPV16 mRNAs are frequently spliced at the *E7/E1* 880 bp site (Figure 1.3) (Zheng and Baker, 2006) which would essentially mask the site of HPV genome disruption due to splicing.

Two passages (early and late) were chosen for each clone and assessed using the APOT assay. The samples included in these runs and a summary of the results are shown in Table 7.2. The sites of integration both of the HPV and human genome varied among clones whilst other clones had no detectable integration events; this suggests that the integration events were

acquired after initial primary culture or that a variety of integration events existed in various cells of the original biopsies.

Clone	Passage	Details of detected fusion transcripts		
		Junction region	HPV16 site (nt)	Human region
A9	5	-	-	-
A9	11	-	-	-
D9	5	*E2*	3716	Chr.18 (p11.31)
D9	9	*E2*	3716	Chr.18 (p11.31)
D9	11	*E2*	3716	Chr.18 (p11.31)
H9	5	*E7/E1*	882	Chr.5 (p11.2)
H9	11	*E7/E1, E2/E4*	882, 3494	Chr.5 (p11.2), Chr.1 (p36.13)
M8	5	*E7/E1*	880	Chr.3 (q28)
M8	9	-	-	-
P8	5	-	-	-
P8	9	-	-	-
Y8	9	*E7/E1*	882	Chr.3 (q21.31)
Y8	13	*E7/E1* x 2	882 x 2	Chr.3 (q21.31) x 2

Table 7.2: A summary of the detected HPV-human fusion transcripts detected using the APOT assay.

The results are listed as the junction region and HPV16 site, i.e. the HPV gene and the HPV16 nt (GenBank ID NC001526.1) position respectively at which the junction between HPV and human sequence was detected. Other transcripts were detected that did not include human transcripts, these are not listed. 13Y8 had two detected fusion transcripts comprising of two proximal sites of the human genome.

7.1.3.2. *PC08/PC09 Gene Expression*

For this series of short term culture DNAs, both the methylation data collected in this study and the gene expression data collected by TO were available. The DNA used for DNA methylation analysis and the RNA used for qPCR analysis were extracted from the same sample of cells. Expression data was provided by TO as Ct values. Ct values for *E2*, *E6* and *E7* and the two reference genes *HPRT* and *TBP* were provided.

Gene expression is presented as Relative Quantification ratios (RQ) and represents a fold difference in gene expression compared to the calibrator sample (CaSki), normalised against the reference genes *HPRT* and *TBP*. RQ ratios were calculated using the Pfaffl/Vandesompele equation (Vandesompele *et al.*, 2002). The fold difference (i.e. RQ) of *E2*, *E6* and *E7* expression for the PC08 and PC09 clones are presented in Figure 7.7. RQ is interpreted such that an RQ of 10, represents a 10 fold higher level of gene expression than CaSki and an RQ of 0.1 represents a 10 fold lower level of gene expression between the sample in question and CaSki.

The data is simpler to interpret when presented on a Log_{10} scale (Figure 7.8) as RQ ratios lower than CaSki (RQ of 1; i.e. 1 x CaSki gene expression) appear as negative. Log_{10} RQ ratio is

interpreted such that a $Log_{10}RQ$ of +1, represents a 10 fold higher level of gene expression and an RQ of -1 represents a 10 fold lower level of gene expression between the sample in question and CaSki. When interpreting the data it is important to consider what the results actually represent. Considering sample 7D9 for example, all three bars are approximately the same height. This does not necessarily mean that each transcript is being expressed at the same level, only that they are being expressed more than in CaSki (and other samples) and that the ratio of expression between the genes is roughly equal to that of CaSki. The data is interpreted by comparing changes and differences between samples and clones i.e. clone D9 expresses more of every gene than clone P8.

Figure 7.7: The fold difference (RQ ratio) in *E2*, *E6* and *E7* gene expression for the PC08 and PC09 series.

Each bar represents an RQ ratio for the PC08/PC09 sample listed on the X-axis. Each gene expression assay is colour coded according to the figure legend.

Figure 7.8: The Log$_{10}$ fold difference (Log$_{10}$ RQ ratio) in *E2*, *E6* and *E7* gene expression for the PC08 and PC09 series.

Each bar represents a Log$_{10}$ RQ ratio for the PC08/PC09 sample listed on the X-axis. Each gene expression assay is colour coded according to the figure legend.

7.1.3.2.1. Hypothesis 1: RQ Ratios Varied among Clones and among HPV Genes

Clone D9 appeared to transcribe *E2/E6/E7* mRNA at a higher level than any other clone.

Among the clones there was substantial variation both in overall level compared to other

clones, and in the relative level of expression of *E2*, *E6* and *E7*. A two-way ANOVA was applied

to the passage 9 RQ values; clone M8 data was excluded from the analysis on the basis of

missing data and as sample 9A9 had incomplete RQ data, it was substituted by sample 8A9.

The data satisfied the assumptions of the two-way ANOVA. A significant difference was

observed among the mean Log$_{10}$ RQ values for each clone (P=0.003). The other factor tested

was gene expression assay (i.e. *E2*, *E6* and *E7*), somewhat surprisingly, no significant difference

in mean Log$_{10}$ RQ as a result of the assay was observed (P=0.475).

7.1.3.2.2. Hypothesis 2: HPV Gene Expression is Associated with HPV DNA Integration

A two-way ANOVA GLM was applied to the same passage 8/9 RQ data for the five clones used

previously. No significant difference in mean RQ was found with respect to either integration

status (P=0.857) or gene expression assay (P=0.832), one violation of the ANOVA assumptions

154

occurred (unequal variances, P=0.003) however multiple two-sample T-tests confirmed the lack of significant effect of DNA integration on mean RQ. If there was a relationship between DNA integration and gene expression, it was more complex than could be determined by a simple comparison of RQ means.

7.1.3.2.3. Hypothesis 3: There is a Linear Relationship between *E2*, *E6* and *E7* RQ

The classical model of P97 regulation is that *E2* is a repressor of P97 activity and that the loss of *E2* function is responsible for *E6* and *E7* up-regulation. A simple way of testing for this scenario in this data set was to correlate *E2*, *E6* and *E7* RQ. Spearman's rank correlation analysis was applied to the RQ data of every passage of each clone. There was a significant positive correlation between all three pairs of RQ values such that overall, *E2*, *E6* and *E7* expression were correlated with each other (Table 7.3).

	E2 RQ vs. *E6* RQ	*E2* RQ vs. *E7* RQ	*E6* RQ vs. *E7* RQ
Spearman's rank correlation coefficient	0.568	0.706	0.736
P-value	0.001	0.000	0.000

Table 7.3: The P-values and correlation coefficients of Spearman's rank correlation coefficient tests between the RQ values of *E2*, *E6* and *E7* for the PC08 and PC09 clones.

7.1.3.3. PC08 and PC09 DNA Methylation

Pyrosequencing was used to determine the DNA methylation status of the PC08 and PC09 clone DNAs. Mean DNA methylation (i.e. mean of the multiple CpGs) of each region of the HPV genome for the PC08 and PC09 clones are presented in Figure 7.9 and Figure 7.10.
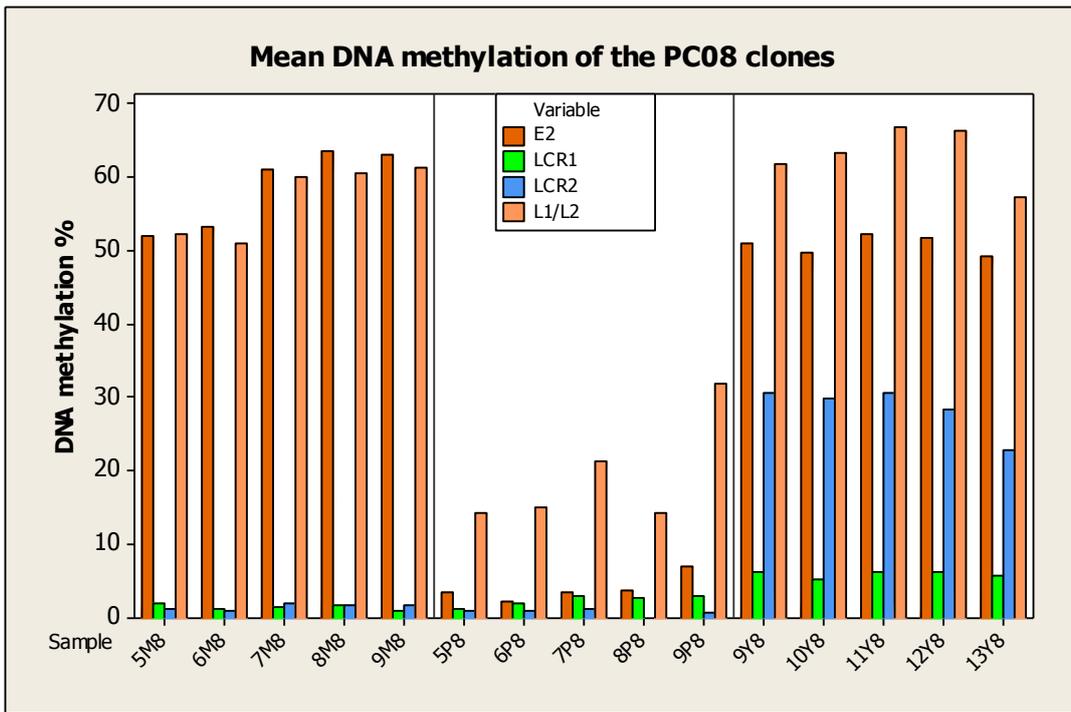
Figure 7.9: Mean DNA methylation of the PC08 clones.

Here each bar represents the mean DNA methylation for the region of the HPV genome indicated by the bar colour.
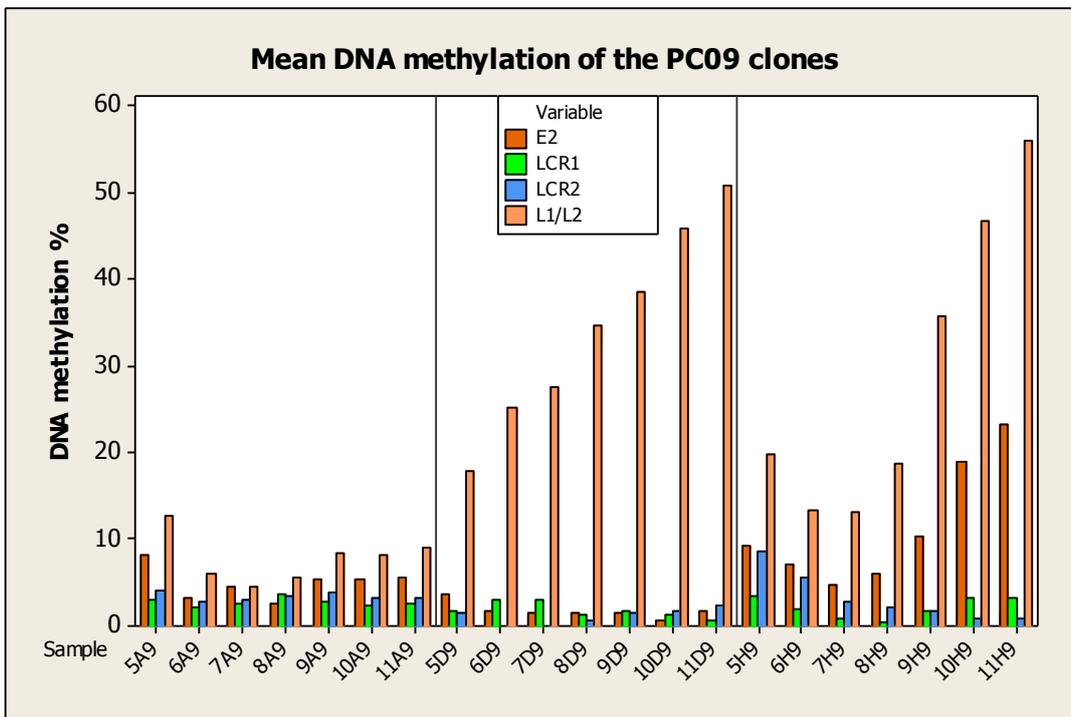
Figure 7.10: Mean DNA methylation of the PC09 clones.

Here each bar represents the mean DNA methylation for the region of the HPV genome indicated by the bar colour.

7.1.3.3.1. Hypothesis 4a: DNA Methylation Differs among HPV Regions

For other sample cohorts, large differences in DNA methylation level were observed among the different regions of the HPV genome, and also among the different CpGs within regions.

The one passage number common to every clone was p9, the DNA methylation data of these samples are presented graphically in Figure 7.11. A two-way ANOVA GLM was applied to the p9 DNA methylation data. This confirmed that statistically significant differences in DNA methylation existed among each HPV region (P=0.000) and within each clone. It also confirmed significant differences in mean DNA methylation of each clone (P=0.037). The significant differences among the clones are discussed later.
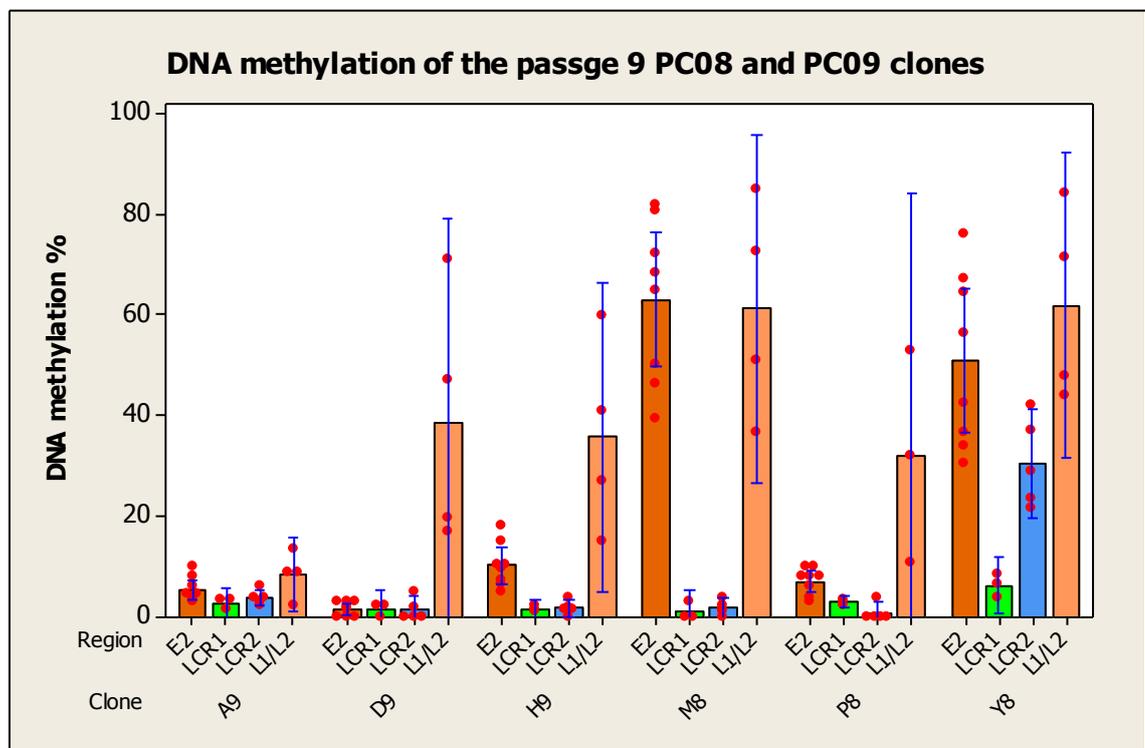
Figure 7.11: DNA methylation of the passage 9 PC08 and PC09 clones.

Bars represent the mean DNA methylation for the clone/assay combination depicted, error bars indicate 95% confidence intervals for mean methylation and red circles represent the individual DNA methylation values for each CpG of each region.

7.1.3.3.2. Hypothesis 4b: DNA Methylation Differs among CpGs within Each HPV Region

For the passage 9 samples, there were noticeable differences in DNA methylation among the CpGs within each region (Figure 7.11). To test whether statistically significant differences existed, two-way ANOVA GLMs were applied to the p9 data, one test for each of the four

regions. The tests were used to determine whether mean methylation differed among CpGs and among clones, the results are presented in Table 7.4. The statistical differences among the various clones are discussed later. A Bonferroni corrected P-value of P=0.013 was applied. There was a significant difference among the amount of DNA methylation of each CpG of the *E2*, LCR2 and *L1/L2* regions.

Factor	*E2*	LCR1	LCR2	*L1/L2*
Clone (P-value)	0.000	0.031*[1]	0.000	0.000
CpG (P-value)	0.000	0.157*[1]	0.156	0.000

Table 7.4: The results of two-way ANOVA GLM tests on the PC08/PC09 passage 9 methylation data. The two-way ANOVA tested for differences in the mean values with respect to both clone and CpG. Violations of the assumptions of the ANOVA: *1, unequal variances (P=0.005).

It was also worth considering whether or not differences among CpGs were consistent for all of the clones. CpG DNA methylation data for the p9 samples is displayed in Figure 7.12, Figure 7.13, Figure 7.14 and Figure 7.15. Large differences in *E2* methylation existed among each CpG and there was a remarkable conservation of the overall methylation profile as shown by the shape of the line for the interconnected individual values. CpG E2-1 was typically highly methylated, E2-2 and E2-3 considerably less so, E2-4 again highly methylated and there was a general decrease in the amount of methylation from there onwards (Figure 7.12). For LCR1, LCR1-1 was the most methylated and LCR1-3 the least for all but the highly methylated Y8 clone for which the converse was true (Figure 7.13). The LCR2 data was similar, showing a consistent pattern for the five hypomethylated clones, while the profile of the Y8 clone was almost the exact opposite (Figure 7.14). The DNA methylation profile of the *L1/L2* region had a similar shape for every clone. Methylation of L1L2-1 and L1L2-3 were the lowest, and L1L2-2 and L1L2-4 were the highest, this gave rise to a distinctive "N" shaped profile (Figure 7.15).
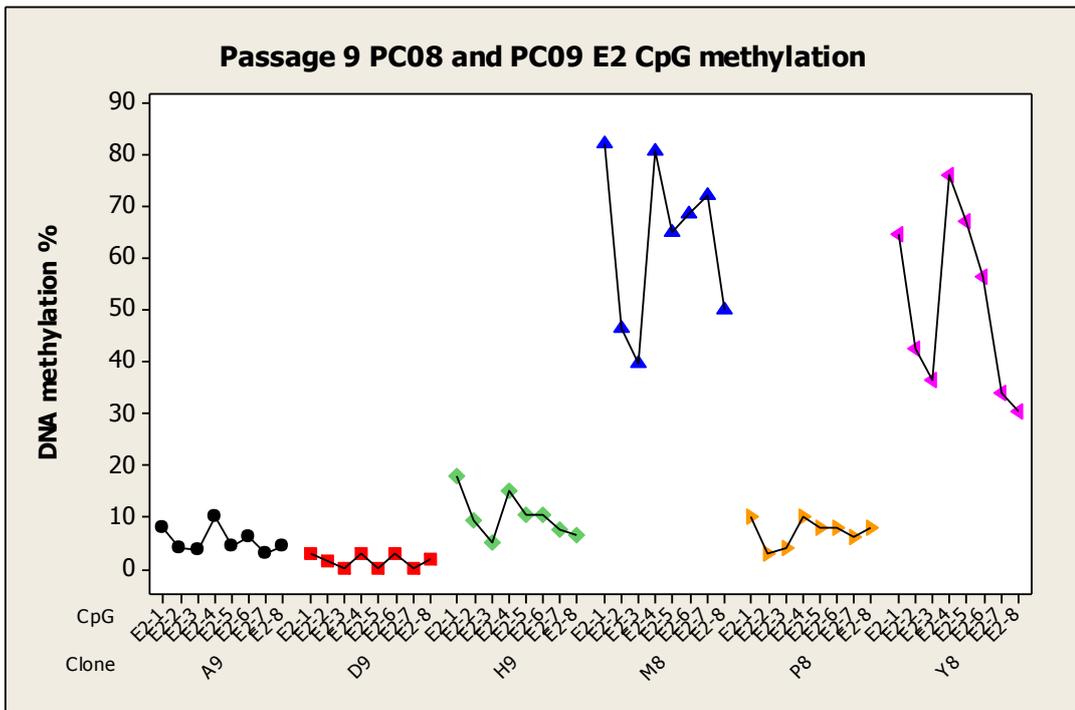
Figure 7.12: DNA methylation of the *E2* CpGs for the PC08 and PC09 clones at passage 9.

The connecting lines between CpG DNA methylation values are included to aid interpretation only, no relationship between the points is inferred.
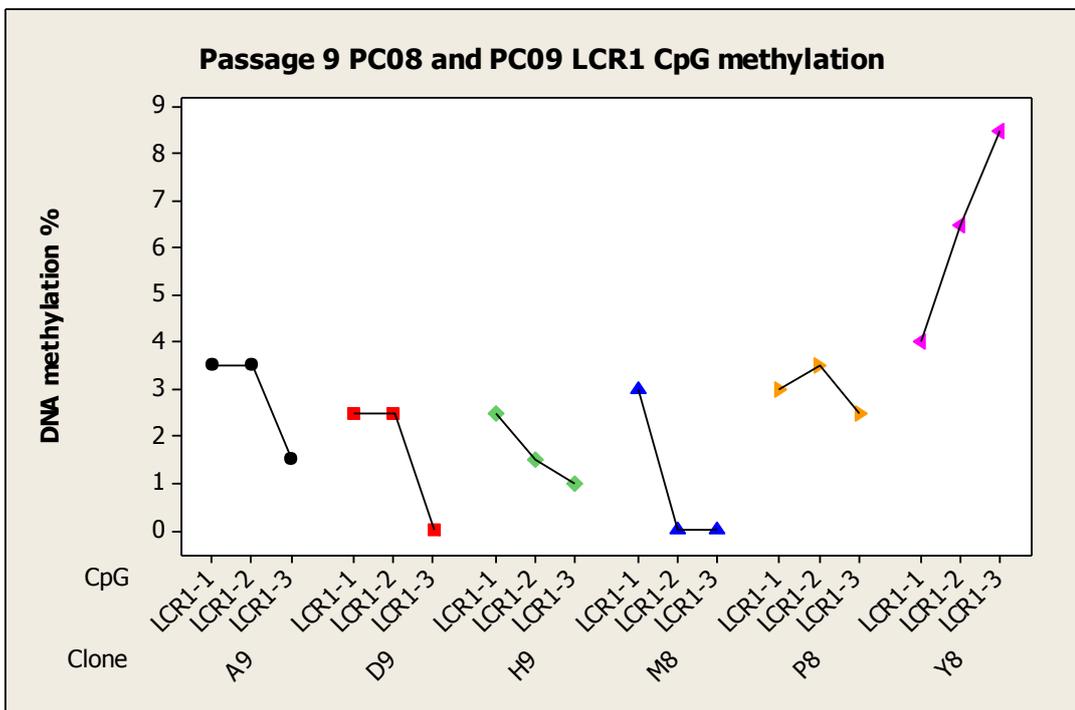
Figure 7.13: DNA methylation of the LCR1 CpGs for the PC08 and PC09 clones at passage 9.

The connecting lines between CpG DNA methylation values are included to aid interpretation only, no relationship between the points is inferred.
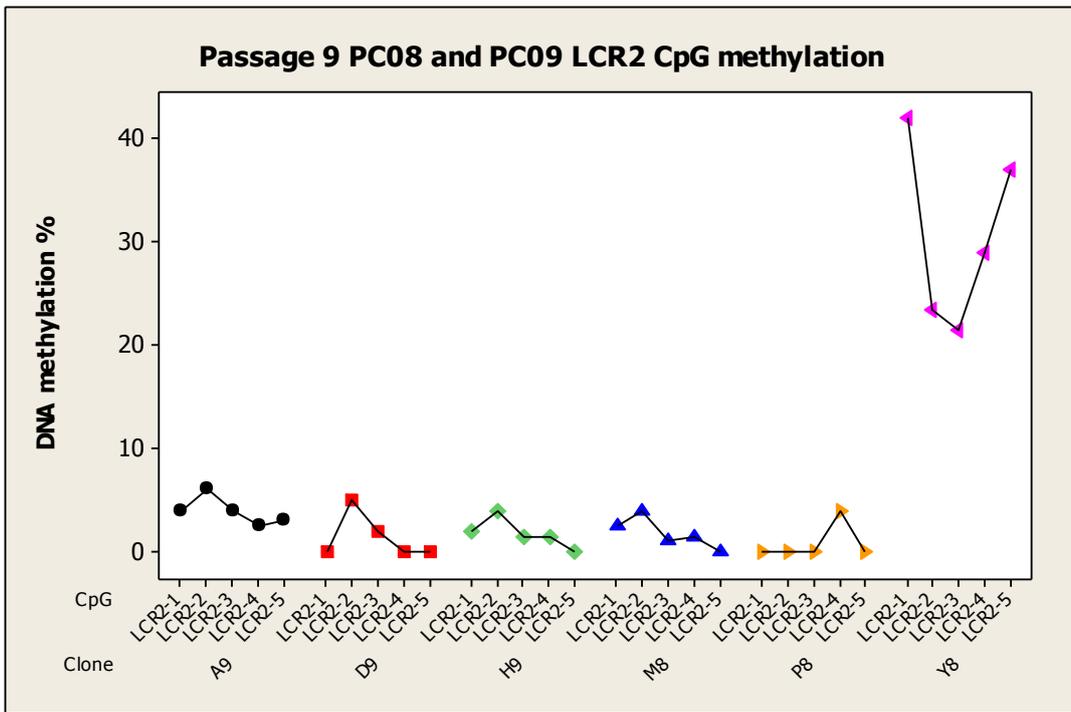
Figure 7.14: DNA methylation of the LCR2 CpGs for the PC08 and PC09 clones at passage 9.

The connecting lines between CpG DNA methylation values are included to aid interpretation only, no relationship between the points is inferred.
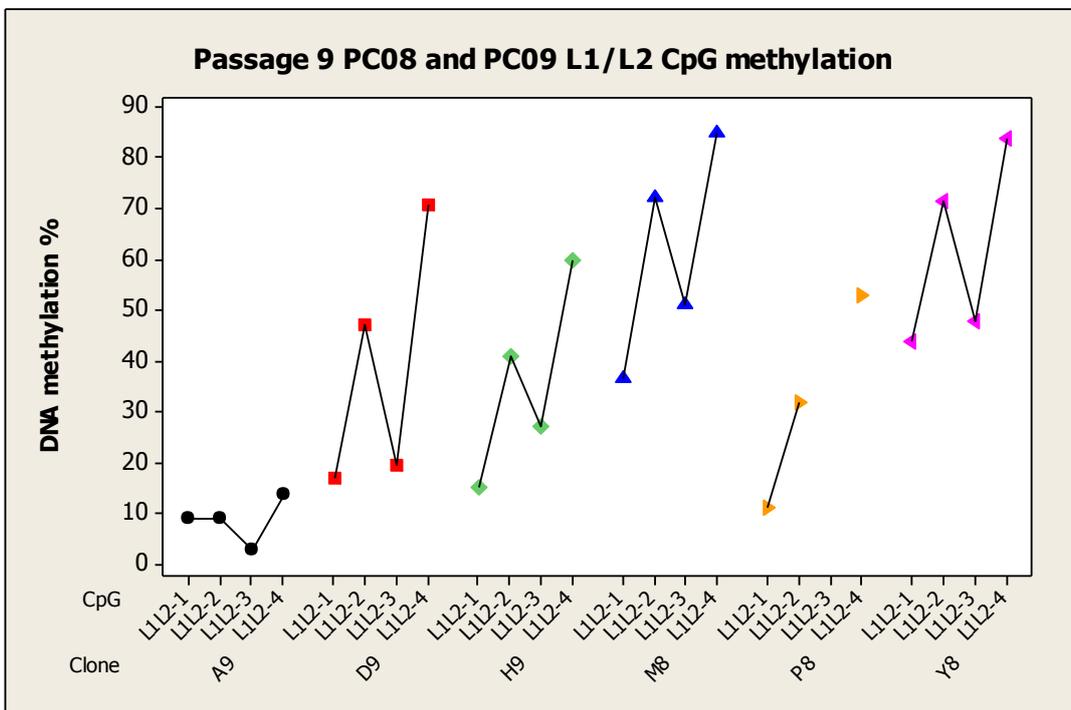
Figure 7.15: DNA methylation of the *L1/L2* CpGs for the PC08 and PC09 clones at passage 9.

The connecting lines between CpG DNA methylation values are included to aid interpretation only, no relationship between the points is inferred.

160

7.1.3.3.3. Hypothesis 5a: Differences in DNA Methylation are Detectable when Comparing Clones from PC08 with Those from PC09

The PC08 and PC09 cell lines were derived from two patients with disease at different anogenital sites. It is reasonable to suspect that DNA methylation might be different between the two original biopsies, and therefore between the two sets of short term culture lines. The passage 9 cell line data for the 6 clones was tested using a two-way ANOVA GLM to determine if a statistically significant difference in mean DNA methylation between the two original biopsies existed. The other factor included in the analysis was HPV region, such that all four HPV regions were tested simultaneously. There was a statistically significant difference in mean DNA methylation both as a result of HPV region (P=0.000) but also as a result of original biopsy (P=0.029). There was a difference between cell lines from the two biopsies.

7.1.3.3.4. Hypothesis 5b: There are DNA Methylation Differences among the Six Clones

LCM studies have demonstrated that HPV DNA methylation level varies throughout a biopsy depending on the characteristics of the HPV infection in different areas (Vinokurova and von Knebel Doeberitz, 2011). Given that these clones originated from two heterogeneous mixes of cells, each potentially representing a diverse part of the same biopsy, it was interesting to consider if the methylation profile was different among the different clones. Previously performed two-way ANOVA GLM tests (passage 9 data only, one test per region, clone and CpG as factors) found that a significant difference among the clones was observed for the *E2*, LCR and *L1/L2* regions (Table 7.4). The Bonferroni corrected P-value was P=0.013, if the Bonferroni correction was not applied then differences among the clones are also seen for the LCR1 region.

As well as differences in mean DNA methylation among the clones, further differences were observed when every passage number was considered. The changes in DNA methylation with passage number are discussed in further detail later in this chapter, however it was of particular interest to note that the relationship between DNA methylation and passage number varied considerably for different clones.

7.1.3.3.5. Hypothesis 6: HPV DNA Methylation Changes with Passage

For the *E2* region (Figure 7.16), clones M8 and Y8 were the most highly methylated, the other PC08 clone (P8) however was hypomethylated in every passage. All three PC09 clones were lowly methylated in early passages. The D9 and A9 clones remained hypomethylated however for the H9 clone, after p8 there was a gradual increase in *E2* DNA methylation.
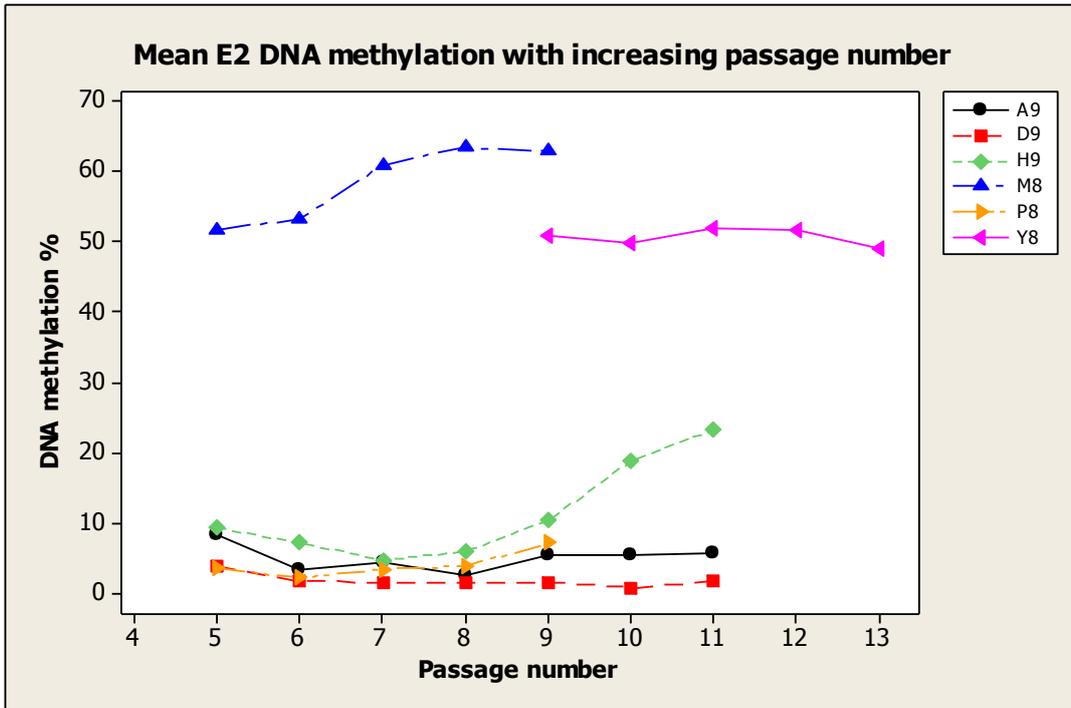
Figure 7.16: Mean *E2* DNA methylation of the PC08 and PC09 clones with increasing passage number. Each clone is represented by a different colour. Different passages of the same clone are connected with a coloured line in order to aid interpretation.

For the LCR1 region (Figure 7.17), every clone was methylated at a low level, changes between passages were difficult to determine accurately because the differences were so small that they could be due to random variation between pyrosequencing runs and experimental noise. The Y8 clone, as per some other regions was more methylated than the other clones. LCR1 methylation did not change drastically with passage number in any discernible manner for any of the clones.

For the LCR2 region (Figure 7.18), all clones except for the Y8 and H9 clones were less than 5% methylated throughout. LCR2 methylation for the Y8 clone began to decrease at p11. The H9 clone had a gradual decrease in DNA methylation from p5 until approximately p10.
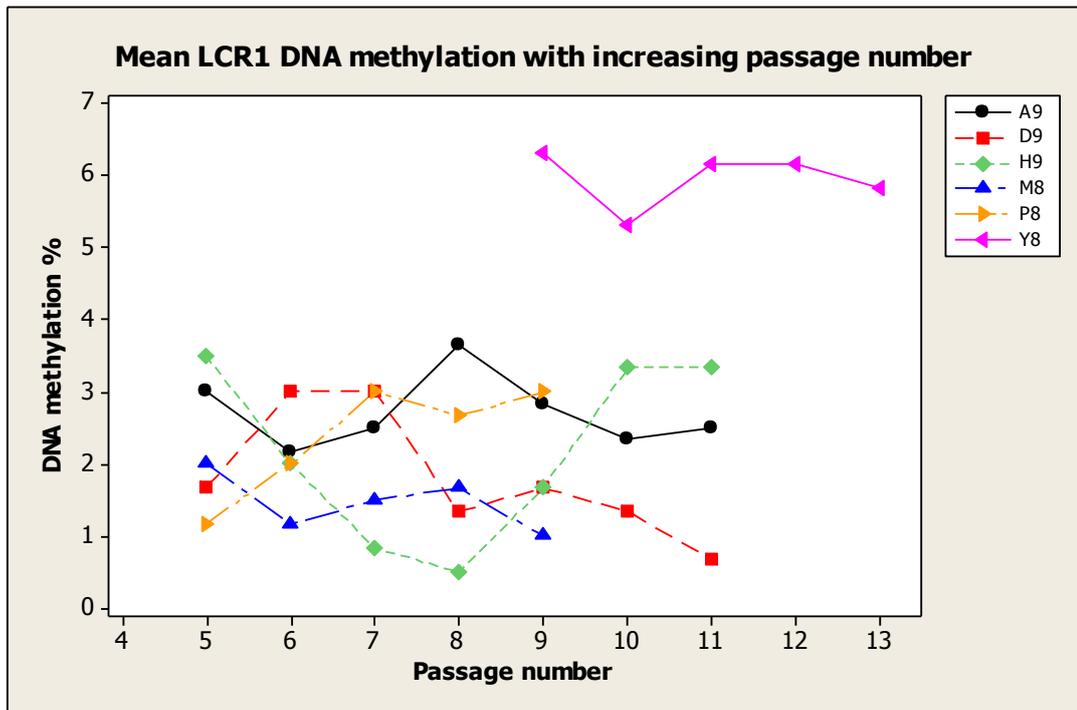
Figure 7.17: Mean LCR1 DNA methylation of the PC08 and PC09 clones with increasing passage number.

Each clone is represented by a different colour. Different passages of the same clone are connected with a coloured line in order to aid interpretation.
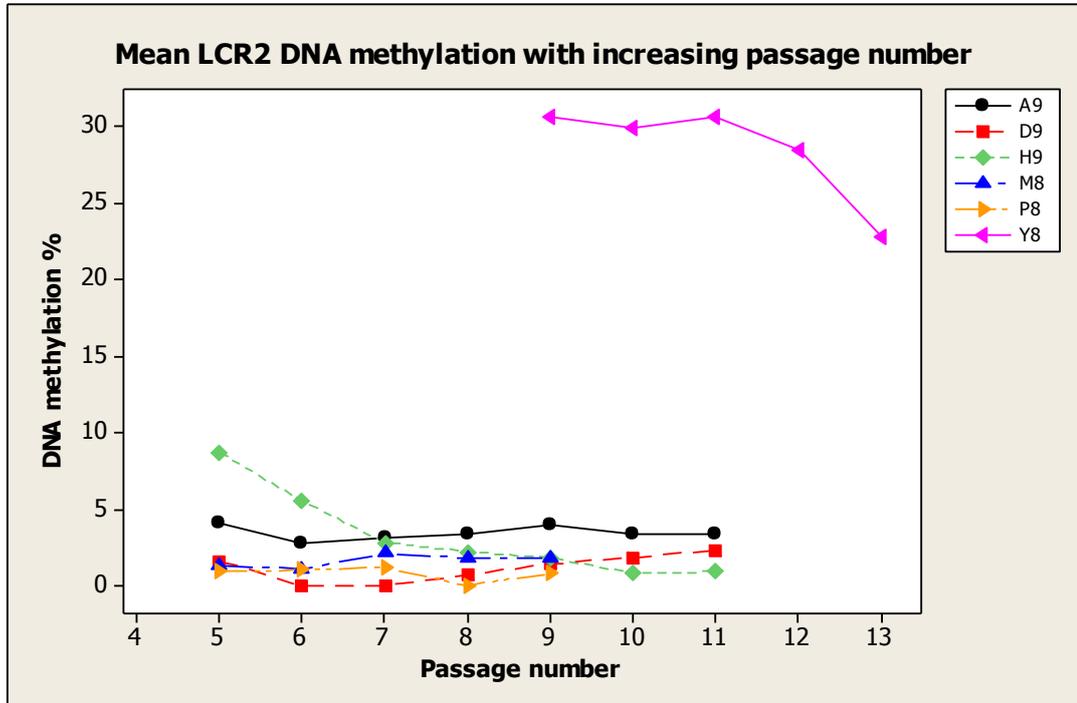
Figure 7.18: Mean LCR2 DNA methylation of the PC08 and PC09 clones with increasing passage number.

Each clone is represented by a different colour. Different passages of the same clone are connected with a coloured line in order to aid interpretation.

For the *L1/L2* region (Figure 7.19), DNA methylation was much more variable than for other HPV regions, every clone appeared to experience changes in DNA methylation with increasing passage number. As for other regions, the Y8 clone was the most methylated and as per the LCR2 region, there was a decrease in DNA methylation for the later passage. The M8 clone was also heavily methylated and became more methylated with increasing passage number (*E2* methylation also increased with passage number). The P8 clone became in general, more methylated with passage number however there was some variation at passage 8, where DNA methylation was lower than might be expected giving rise to an odd shaped curve. The A9 clone was the least methylated of the clones and after an initial decrease, DNA methylation remained low. The D9 clone was particularly intriguing in that a positive, linear relationship between DNA methylation and passage number was observed; it is worth considering that whilst *L1/L2* methylation level was high, the *E2* region of D9 was hypomethylated throughout. The H9 clone showed a similar overall increase in DNA methylation as D9, however the changes were non-linear, initially experiencing a decrease before a more rapid increase. The increased *L1/L2* DNA methylation for the H9 clone, coincided with the increase in *E2* DNA methylation discussed previously.
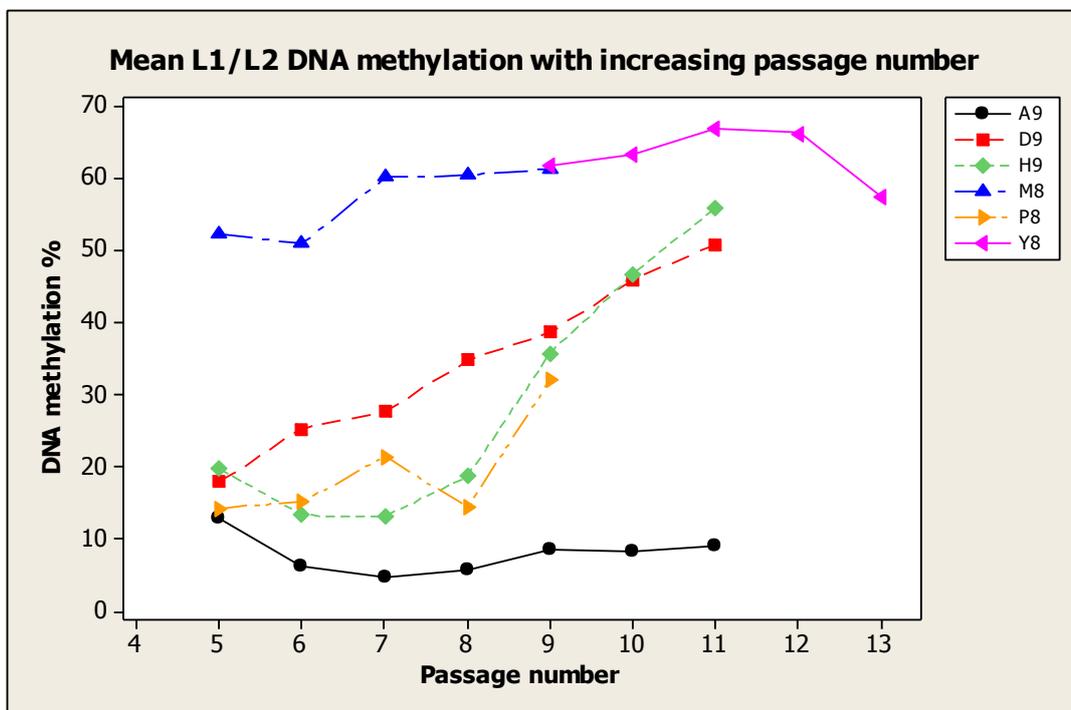
Figure 7.19; MeanL1/*L2* DNA methylation of the PC08 and PC09 clones with increasing passage number.

Each clone is represented by a different colour. Different passages of the same clone are connected with a coloured line in order to aid interpretation.

7.1.3.4. The Interaction of DNA Methylation, Gene Expression and HPV Integration

Comparison of DNA methylation, gene expression and HPV integration for the PC08 and PC09 cell lines was conducted in two main stages, the first stage considered all six clones simultaneously and the second stage considered each clone individually.

7.1.3.4.1. Hypothesis 7: There is a Relationship between DNA Methylation and Gene Expression

The simplest way of determining whether a direct relationship existed between DNA methylation and gene expression was to perform correlation analysis. Spearman's rank correlation analysis was performed on the mean DNA methylation and RQ results for every passage of every clone to determine whether a general relationship existed (Table 7.5). A significant relationship existed between *E2* DNA methylation and *E2*, *E6* and *E7* RQ, LCR1 methylation and *E6* RQ, LCR2 methylation and *E2* RQ and a very close to significant correlation existed between *L1/L2* methylation and *E6* RQ. The data was highly unsuitable for multifactorial ANOVA analysis. Further investigation into the relationship between DNA methylation and gene expression was performed on an individual clone basis.

		E2 RQ vs.	*E6* RQ vs.	*E7* RQ vs.
E2 methylation	Spearman's rank	-0.593	-0.679	-0.702
	P-value	0.001	0.000	0.000
LCR1 methylation	Spearman's rank	-0.234	-0.357	0.028
	P-value	0.212	0.035	0.873
LCR2 methylation	Spearman's rank	-0.036	-0.421	-0.150
	P-value	0.851	0.012	0.388
L1/L2 methylation	Spearman's rank	-0.216	-0.332	-0.238
	P-value	0.251	0.051	0.169

Table 7.5: The results of Spearman's rank correlation analysis to investigate the relationship between DNA methylation and gene expression.

Each column represents a gene expression assay whilst each row represents a DNA methylation assay. Data from every clone at every passage was included. Mean DNA methylation of the multiple CpGs in each region was used in the analysis. Spearman's rank refers to the Spearman's rank correlation coefficient. P-values are significant where P<0.05. Significant correlations are highlighted yellow and correlations close to significant are coloured blue.

7.1.3.4.2. Hypothesis 8: There is a Relationship between DNA Methylation and HPV Integration

Passage 9 DNA methylation data was tested to determine whether a significant difference in mean DNA methylation existed among HPV regions (previously confirmed), and HPV integration status using a two-way ANOVA GLM. For integration, two classes were defined; integration detected and integration not detected. No significant relationship between mean DNA methylation and integration status was identified (P=0.168), suggesting that if DNA integration had an effect on DNA methylation, it was more complex than a simple increase/decrease in mean DNA methylation.

7.1.3.5. Individual Clone Comparisons

The relationship between DNA methylation, gene expression and integration appeared to be more complex than a simple comparison of means or correlation could determine. Visual comparisons of the DNA methylation mean data (Figure 7.16, Figure 7.17, Figure 7.18 and Figure 7.19), the Log_{10} RQ plot (Figure 7.8) and DNA integration results (Table 7.2) were performed and used to generate hypotheses that could be tested statistically.

7.1.3.5.1. Clone A9

Clone A9 *E2/E6/E7* transcription was fairly consistent across all passages (p9 data missing). There were only small changes observed with increasing passage number, however there were slightly greater differences between p10-11 than there were for the other passages. DNA methylation level of clone A9 was low and consistent with increasing passage, it had the lowest overall methylation of *L1/L2* and one of the lowest levels of methylation for the other regions. No integration events were detected for clone A9 in either p5 or p11.

7.1.3.5.2. Clone D9

Of the six clones tested, clone D9 displayed the highest mRNA levels for each gene; there was also a progressive increase in the amount of *E2* and *E6* mRNA right up until the final sample at p11 where there was a sharp and pronounced decrease in the expression of all genes. This was a little surprising because we typically consider that increased *E2* expression should have a negative impact on the expression of the *E6* and *E7* genes, as such we may expect increases in *E2* to be followed by decreases in *E6* and *E7* and vice versa.

The DNA methylation of clone D9 was also interesting in that for *E2*, LCR1 and LCR2 there was very little DNA methylation whereas for *L1/L2* there was a steady and linear increase with

166

increasing passage number. *E2* hypomethylation coinciding with *E2* integration was of particular interest because a potential association between site of integration and methylation was identified *in vitro* (Chapter 8: 8.2.3.3).

Passage 11 excluded, the increase in *E2*/6/7 mRNA in D9 with passage appeared to coincide with an increase in *L1/L2* DNA methylation (Figure 7.20). Spearman's rank correlation analysis was applied to determine whether gene expression and *L1/L2* methylation correlate significantly, the results are summarised in Table 7.6. Significant correlation was observed for *E2* RQ vs. *L1/L2* mean methylation and *E6* RQ vs. *L1/L2* mean methylation. None of the correlations were significant when 11D9 (p11) was included; the p11 data may have been erroneous. Several explanations for the increase in DNA methylation with passage number could exist. It could be that the HPV16 DNA is becoming gradually more methylated in each cell, or alternatively, that the integrated genomes are heavily methylated and are becoming more common as they are selected for due to *E2* disruption, *E6/E7* overexpression and an enhanced growth rate.

	E2 RQ vs. *L1/L2* Methylation	*E6* RQ vs. *L1/L2* Methylation	*E7* RQ vs. *L1/L2* Methylation
Spearman's rank correlation coefficient	+0.943	+1.000	+0.543
P-value	P = 0.005	P < 0.000	P = 0.266

Table 7.6: The P-values and correlation coefficients of Spearman's rank correlation coefficient tests between the RQ values of *E2*, *E6* and *E7* and the mean methylation values at the *L1/L2* region for the D9 clone.
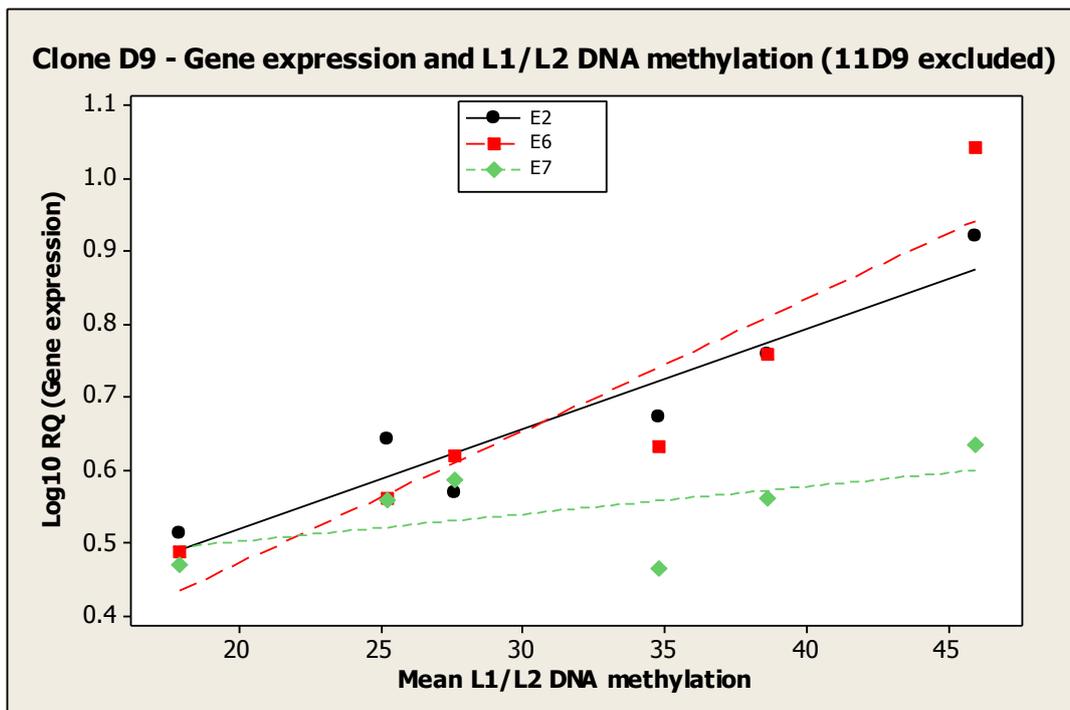
Figure 7.20: The log$_{10}$ RQ ratios for *E2*, *E6* and *E7* expression plotted against *L1/L2* mean DNA methylation.

Passage 11 data was excluded from the diagram and from subsequent correlation analysis. Included are regression lines of best fit for each of the three gene expression assays.

7.1.3.5.3. Clone H9

For clone H9 there was a slight increase in expression of *E6* and *E7* and a considerable decrease in *E2* expression with increasing passage. This was more in line with what might be expected, i.e. a decrease in *E2* expression alleviates some of the repression on the P97 promoter allowing *E6* and *E7* expression to rise. Nonetheless, despite an apparent negative correlation between *E2* RQ and *E6/E7* RQ for the H9 clone, the correlation was not significant (Table 7.7).

	E2 RQ vs. *E6* RQ	*E2* RQ vs. *E7* RQ	*E6* RQ vs. *E7* RQ
Spearman's rank correlation coefficient	-0.429	-0.357	+0.857
P-value	P = 0.337	P = 0.432	P = 0.014

Table 7.7: The P-values and correlation coefficients of Spearman's rank correlation analysis to determine a relationship between the RQ values of *E2*, *E6* and *E7* for the H9 clone.

The methylation profile of clone H9 was of particular interest because of the pronounced changes at and around p8-9 (large increases in *E2* and especially *L1/L2* methylation). Between p5-8 the expression of *E2* was considerably higher than between p9-11, this coincided with changes in methylation in the *E2* and *L1/L2* regions. In addition, LCR2 methylation decreased

relatively rapidly between p5-7 and decreased continually for the rest of the passages available. Between the early and late passages there appeared to be a large difference in gene expression and DNA methylation that may be a consequence of the integration event detected in p11, but not in p5. The integration event was detected in the *E2/E4* region.

Two-sample T-tests were used to determine whether gene expression and DNA methylation were significantly different between the early passage (p5-8) and the late passage (p9-11), the results are summarised in Table 7.8. There was a significant difference between the early and late passage for *E2* RQ and for mean *L1/L2* DNA methylation, however the difference for *L1/L2* methylation was not significant if the Bonferroni correction (P=0.007) was applied.

	Difference between passages 5-8 and 9-11 with respect to:						
	Gene expression (RQ)			Mean DNA methylation			
Two-sample T-test	*E2*	*E6*	*E7*	*E2*	LCR1	LCR2	*L1/L2*
P-value	0.007	0.593	0.365	0.112	0.290	0.097	0.039

Table 7.8: The results of two-sample T-tests to assess differences in mean gene expression and mean DNA methylation between the early and late passages of H9.

The table is separated into DNA methylation and gene expression assays, the P-value represents the significance of the difference between the early passage (passage 5-8) and the late passage (passage 9-11) samples for the assay of interest. No violations of the assumptions of the T-test were encountered.

7.1.3.5.4. Clone M8

Clone M8 was lacking *E2* expression data so parallels between *E2* and *E6/E7* expression could not be drawn (Figure 7.8). Nonetheless there did appear to be a slight increase in *E7* expression with somewhat more variable *E6* expression. As previously discussed, clone M8 had high *E2* and *L1/L2* methylation that increased slightly with passage number; it was difficult to draw further conclusions without the *E2* RQ data. Integration was detected in the *E7/E1* region at p5 but not at p9. This could be a selective loss of cells containing the integration event, silencing of the HPV sequence expressing the integration event or a failing of the APOT assay.

7.1.3.5.5. Clone P8

Clone P8 had a low level of DNA methylation in all regions, fairly constant *E6* and *E7* RQ and a slight increase in *E2* RQ with increasing passage number. Interestingly, p9 had the highest *E2* RQ and the highest *L1/L2* methylation. No integration was detected.

7.1.3.5.6. Clone Y8

Finally, clone Y8 showed increasing mRNA levels for all genes of interest up until p12 where there was a 1 Log_{10} decrease in expression. By p13, mRNA levels had returned to previous levels except for a considerable increase in *E6* RQ. Clone Y8 had the highest level of DNA methylation for most of the assays out of any of the clones analysed and also the lowest *E6* RQ out of all of the clones analysed. HPV integration was detected at *E7/E1* and an additional *E7/E1* integration event was detected in p13. The increase in gene expression for p13 (especially for *E7*) may be related to a decrease in DNA methylation at p13 that was observed at every site (*E2*, LCR1, LCR2 and *L1/L2*).

7.2. Discussion

7.2.1. CaSki

7.2.1.1. Aims

CaSki DNA was used as a HPV16 positive control DNA for method development and to compare the results produced by pyrosequencing to published data.

7.2.1.2. Comparison to Other Studies

Fernandez *et al* used bisulfite sequencing with cloning in order to assess methylation in a semi-quantitative manner (Fernandez *et al.*, 2009) and produced results similar to those produced by this study (Table 7.9, Table 7.10). Where differences are present, Fernandez *et al* typically had a low number of clones sequenced (e.g. LCR1). Even with a larger number of bisulfite sequenced clones there are wide confidence intervals. E2-5 for example had 8/12 clones sequenced, indicating methylation of 67% however the 95% confidence intervals range from 40.4-93.6% making direct comparison with a quantitative method unreliable.

	E2								LCR1		
	1	2	3	4	5	6	7	8	1	2	3
Nucleotide	3411	3414	3416	3432	3435	3447	3461	3472	7691	7679	7673
CaSki repeat 1	96	97	80	100	98	92	82	76	63	68	75
CaSki repeat 2	95	100	80	100	95	94	81	78	63	67	77
Mean	95.5	98.5	80	100	96.5	93	81.5	77	63	67.5	76
Fernandez *et al* Methylated clones	11/12	12/12	11/12	11/12	8/12	12/12	10/12	9/12	2/2	0/2	2/2
Methylation %	92	100	92	92	67	100	83	75	100	0	100
Difference %	**3.5**	**-1.5**	**-12**	**8**	**29.5**	**-7**	**-1.5**	**2**	**-37**	**67.5**	**-24**

Table 7.9: A comparison of CaSki methylation data collected in this study with published CaSki methylation data for the *E2* and LCR1 region (Fernandez *et al.*, 2009).

The published methylation data was bisulfite sequencing data and was calculated by sequencing multiple bisulfite converted cloned DNAs.

	LCR2					L1/L2			
	1	2	3	4	5	1	2	3	4
Nucleotide	31	37	43	52	58	5615	5609	5606	5600
CaSki repeat 1	90	94	98	95	95	73	86	75	85
CaSki repeat 2	89	89	95	96	99	76	89	77	90
Mean	89.5	91.5	96.5	95.5	97	74.5	87.5	76	87.5
Fernandez *et al* Methylated clones	8/8	8/8	8/8	6/8	7/8	7/10	10/10	7/10	10/10
Methylation %	100	100	100	75	88	70	100	70	100
Difference %	**-10.5**	**-8.5**	**-3.5**	**20.5**	**9**	**4.5**	**-12.5**	**6**	**-12.5**

Table 7.10: A comparison of CaSki methylation data collected in this study with published CaSki methylation data for the LCR2 and *L1/L2* (Fernandez *et al.*, 2009).

The published methylation data was bisulfite sequencing data and was calculated by sequencing multiple bisulfite converted cloned DNAs.

Another study considered DNA methylation of the CaSki HPV16 LCR and *L1* region using bisulfite sequencing with cloning and found almost complete hypermethylation of LCR2 region (93% methylation of LCR2-1 and LCR2-5, 100% methylation of the other LCR2 CpGs) and LCR1 region (87% methylation of LCR1-1, 100% methylation of the other LCR1 CpGs) (Kalantari *et al.*, 2004). They also observed heavy methylation of the *L1* region, albeit in an area not covered by our assays (Kalantari *et al.*, 2004). Direct bisulfite sequencing without a cloning step suggested there was methylation of all LCR1 and LCR2 CpGs in CaSki (i.e. the result is methylated or unmethylated) (Badal *et al.*, 2003). Both of these studies are consistent with our results.

7.2.1.3. Meaning

CaSki HPV16 was heavily methylated in all CpGs but the level of methylation varied both among region of the genome and among CpGs within each region. The *E2* and LCR2 (promoter) region were the most densely methylated, whilst the LCR1 (enhancer) and *L1/L2* regions were hypermethylated to a lesser degree. CpG specific patterns of DNA methylation were observed in *L1/L2* and *E2* which appeared to match those observed in other sample cohorts. The "N" shaped *L1/L2* CpG methylation pattern observed in CaSki was also present in published CaSki work (Fernandez *et al.*, 2009), as was the LCR1 pattern (Ding *et al.*, 2009).

For our study, CaSki was a crucial part of the method development, as a source of methylated control DNA, a means by which to compare independent experimental repeats, and a means by which to compare data produced using our assays to published data. One of the first observations was that two experimental repeats (i.e. DNAs were separately BS treated, amplified by PCR and then pyrosequenced) exhibited levels of methylation that were very closely matched, the greatest individual difference of any CpG being 5% (Table 7.9, Table 7.10). The close similarities between this data and data produced by the widely used bisulfite sequencing technique were also reassuring.

HPV16 genomes are integrated into the CaSki human genome at around 500 copies per cell as head-to-tail hypermethylated concatemers and located in several sites of the human genome with only a small number of transcriptionally active copies (Fernandez *et al.*, 2009, Van Tine *et al.*, 2004b, Van Tine *et al.*, 2001, Mincheva *et al.*, 1987). This is consistent with our observation of heavily methylated LCR, *E2* and *L1/L2* regions.

7.2.2. W12 series 7

7.2.2.1. Aims

W12 series 7 (W12Ser7) DNA was used as a means to consider DNA methylation over a large number of cell passages and to consider the effect of HPV integration on DNA methylation.

7.2.2.2. Comparison to Other Studies

Kalantari *et al* studied DNA methylation of W12 cells in both episomal and integrated forms, and after induction of cellular differentiation (Kalantari *et al.*, 2008a). They observed a low level of enhancer and promoter methylation and some methylation of the 5'LCR and *L1* region in episomal HPV16. Upon differentiation, DNA methylation was reduced considerably in all regions for episomal, but not for integrated HPV16 genomes (Kalantari *et al.*, 2008a). The reduction in W12 LCR methylation upon differentiation was also observed in another study

(Kim *et al.*, 2003). Low level methylation of the enhancer, promoter and *L1* region in these studies is consistent with our observation of low level LCR1, LCR2 and *L1/L2* methylation in early passage W12, and prior to the integration events.

Two types of integration event were detected (single insertions and concatemeric repeats) that corresponded to the methylation profiles of SiHa (single insertion) and CaSki (concatemer) respectively (Kalantari *et al.*, 2008a, Fernandez *et al.*, 2009). Our W12 methylation profile corresponds best to that of the concatemeric insertion (Kalantari *et al.*, 2008a). Further comparison revealed that where we saw LCR1-1 to be less methylated than LCR1-2/3 at every passage, this was also reflected in published data where LCR1-1 was typically unmethylated in most instances (Kalantari *et al.*, 2008a). LCR1-1 hypomethylation and LCR1-2/LCR1-3 methylation was not frequently observed in our other sample cohorts/cell lines. In summary, where parallels can be drawn, our data align well to previously published W12 data from other W12 series.

7.2.2.3. Meaning

The information and material available for the W12Ser7 line was limited. We had no access to RNA with which to perform gene expression analysis, only very basic integration data has been published (Gray *et al.*, 2010) and detail pertaining to the site of integration within the human and HPV genome was not available. There was also insufficient DNA with which to perform DIPS.

W12Ser7 integration was detected from passage 13 onwards and by p17 the vast majority of genomes were integrated (Gray *et al.*, 2010); this corresponded to large and distinct changes in DNA methylation beginning around p13. The changes observed were both region specific (i.e. assay) and CpG specific (i.e. within assays). All regions of the HPV genome were hypomethylated (i.e. approx 10%) in the first passage available (p11/p13). Changes were most pronounced for the *L1/L2* region. LCR and *E2* methylation did change with passage but not to the same extent.

For the *E2* region there was a large increase in DNA methylation from p13 onwards almost exclusively for the E2-5 and E2-6 CpGs. Adjacent CpGs, separated by only a few bp were seen to differ considerably, hinting at a CpG specific mechanism of DNA methylation. This was also observed at LCR1 and to a degree for LCR2. An interesting difference between W12Ser7 and other cohorts was that W12Ser7 LCR1-1 CpG was hypomethylated whilst methylation of LCR1-2 and LCR1-3 varied with passage number. This was reflected in studies using W12 of an

unknown series (Kalantari *et al.*, 2008a). LCR1-1 does not correspond to an E2BS, so it was unusual that it was persistently hypomethylated in W12Ser7. LCR1-1 hypomethylation possibly corresponded to a growth advantage to the cell hence its continued hypomethylation (as observed in a range of W12 lines). LCR1-1 is close to the binding sites of many cis-acting elements, the binding of which could be influenced by DNA methylation. Previous studies have noticed an association between nucleosome linker position and sites of continual hypomethylation (i.e. CpG 7862) (Kalantari *et al.*, 2004) however LCR1-1 corresponds to a site positioned within a nucleosome (Stunkel and Bernard, 1999) and was not routinely hypomethylated (compared to other LCR1 CpGs) in our other work. In summary, the conserved methylation differences between proximal CpGs suggest that specific mechanisms of methylation and/or strong selection pressure could be responsible. Some of the more pronounced CpG specific differences for W12 were not however routinely observed in other cohorts, which could be due to several reasons including integration state, integration site and characteristics of the W12 cell lines that are not shared with any of our other cohorts.

7.2.2.3.1. Integration and Methylation

W12Ser7 HPV methylation was seen to undergo pronounced increases at the onset of integration which continued for over 35 passages. Methylation increased in both a region specific and CpG specific manner. Methylation of the *L1*/*L2* region increased the most after the integration event, methylation of the other three regions occurred but not as substantially.

Published W12, SiHa and CaSki HPV16 integration, copy number and methylation data may give a degree of insight into the association between methylation and integration for W12Ser7. HPV16 is present in SiHa at around 1-2 copies per genome and in CaSki at approximately 500 copies per genome (Van Tine *et al.*, 2001, Fernandez *et al.*, 2009). SiHa and CaSki HPV16 differ significantly; whilst both cell lines contain HPV16 which is heavily methylated in the E2-*L1* regions, the LCR of SiHa HPV16 is hypomethylated (Kalantari *et al.*, 2004, Fernandez *et al.*, 2009). In CaSki, HPV16 genomes exist as integrated concatemers, the majority of which are repressed by heavy DNA methylation. The few HPV16 integrants in SiHa are transcriptionally active with an unmethylated LCR (Kalantari *et al.*, 2008a, Van Tine *et al.*, 2001). Similarly, W12 single HPV16 insertions were methylated in *L1*, but hypomethylated in the LCR whilst concatemeric insertions were typically hypermethylated in the enhancer, promoter and *L1* region as per SiHa (single insertion) and CaSki (concatemer) respectively (Kalantari *et al.*, 2008a, Fernandez *et al.*, 2009). Our observation of methylation at all regions

(with hypermethylation of *L1/L2*) better corresponds to that seen for the concatemeric insertion type integration event (Kalantari *et al.*, 2008a).

Pyrosequencing is unable to differentiate between a mix of heterogeneously methylated genomes, and a mix of hypomethylated and hypermethylated genomes. An intriguing explanation for the W12Ser7 methylation profile was that the progressive increase in DNA methylation was due to a selection for, and eventual predominance of cells containing methylated, integrated HPV, with the apparent gradual increase in methylation being attributable to the gradual loss of hypomethylated episomes. The highest rate of change in DNA methylation was observed between p13 and p17 however methylation continued to increase in much later passages too. W12 contained almost entirely integrated forms from p17 onwards (Gray *et al.*, 2010) which suggests that episomal loss did contribute to the change in DNA methylation, but also that integrants are methylated in a gradual manner, rather than immediately hypermethylated upon integration.

The selection for, and eventual dominance of W12Ser7 cells containing integrated genomes (Gray *et al.*, 2010) implies that integration confers a growth advantage for the host cell (Jeon *et al.*, 1995). Integration of HPV must occur in a background of episomal HPV which would initially be repressed by episomal *E2*. In this instance, loss of the episomes would confer a strong selective benefit which might explain why episomes were quickly lost *in vitro* once the integration event had occurred (Pett and Coleman, 2007, Pett *et al.*, 2006).

7.2.2.3.2. Unusual Observations

Prior to integrant selection, there was an initial decrease in DNA methylation from p11-p15 at all regions. This is difficult to explain; DNA methylation is seen to be reduced during differentiation (Kalantari *et al.*, 2008a) but this is not occurring, there is reason to suspect that HPV oncogene expression can alter DNA methylation through their action on DNMTs (Burgers *et al.*, 2007, Leonard *et al.*, 2012) so this initial change in DNA methylation might reflect known changes in oncogene expression prior to integration and selection for integrants (Gray *et al.*, 2010).

CpG methylation of the *E2* region was also different than seen elsewhere; only two *E2* CpGs were hypermethylated in association with integration and passage number whilst CpGs adjacent to these two remained hypomethylated throughout. This was observed in multiple passages, so there is no reason to suspect it is erroneous however it is difficult to explain. DNA methylation of integrated genomes may operate in a manner that requires an initial seeding of specific sequences (Doerfler, 2008, Turker, 2002), however this seems unlikely in this instance

as other *E2* CpGs might be expected to be methylated in passages after this initial seeding and this unusual pattern was not observed elsewhere. A more plausible suggestion might be that HPV16 integration resulted in topological changes of the *E2* DNA structure and affected its ability to be methylated and might possibly because of the chromatin structure of adjacent human DNA (Doerfler, 2011).

7.2.3. PC08 and PC09

7.2.3.1. Aims
The PC08 and PC09 cell lines are short term culture lines derived from VIN3 and VAIN3 lesions respectively. The aim of the study was to determine the relationship between DNA methylation, integration and gene expression in a tractable and well characterised *in vitro* model.

7.2.3.2. Strengths
The biopsy dosage effect described in the qPCR method development section (Chapter 4: 4.1.3) would not apply to cell line material. The biopsy dosage effect is where variation in the size of the biopsy, the number of HPV containing cells and tissue type alters the RQ ratios produced. This is not a concern for a homogeneous cell line as every cell is of the same tissue type and every cell contains HPV16.

CaSki, SiHa and HeLa are often used to study HPV biology. These lines were derived from malignant tumours, in some instances from metastases and contain an array of chromosomal rearrangements, deletions and duplications following decades of cell culture. Other studies have utilised transformation of keratinocytes with HPV genomes or plasmid constructs containing certain HPV ORFs. Whilst useful, these cell lines represent HPV in a manner that might have little relevance to a real infection or precancerous disease. W12, and now PC08 and PC09 were derived from primary cultures of precancerous disease, such that we might observe events vital to carcinogenesis *in vitro*.

Many assays (especially the integration assays) require a large amount of high quality DNA/RNA that is not readily available for clinical material, but is not an issue for cell line studies. Most clinical material studies are simply cross sectional. Longitudinal sample sets are less common and are very rarely used for HPV methylation studies. The PC08/09 cell lines allow for changes with time to be studied.

7.2.3.3. Weaknesses

APOT only detects transcribed integrants, not those transcriptionally silenced. DNA methylation is a mechanism of transcriptional regulation and heavily methylated transcriptionally silent integrants might not be detected. This might hinder the investigation of an association between integration and methylation in instances where multiple integrants exist. Also, as APOT also utilises spliced mRNA HPV-human fusion transcripts, in many instances the true site of integration is spliced out from the transcript and the region of disruption of the HPV16 genome cannot be determined.

7.2.3.4. Meaning

7.2.3.4.1. There was Variation

The three PC08 and three PC09 cell lines are homogenous clones derived from VIN3 and VAIN3 biopsies respectively. Whilst differences between the two biopsies were expected (and observed), an intriguing feature of the PC08/PC09 series was the massive variation among clones derived from the same biopsy. Within the 6 clones, there was wide variation in gene expression, DNA methylation, and presence/absence/site of integration. That integration occurs in different sites for each clone, and that integration is not present in all clones suggests that many of the integration events occurred during cell culture. LCM studies have found large gene expression and DNA methylation differences between various tissue types within a biopsy, depending on the characteristics of the HPV infection in each area (Kalantari *et al.*, 2009, Vinokurova and von Knebel Doeberitz, 2011), which may explain why such a diverse range of clones were produced. In addition, the integration data suggests that changes have occurred both in heterogeneous cell culture and after cloning. Coupled with a different environment (i.e. cell culture) the clones are likely very different than the ancestor cells from the original biopsies.

DNA methylation, integration and gene expression were seen to vary with passage number, i.e. there were changes with time and number of cell divisions. As per the W12Ser7 data, this suggests that the relationship between these three factors was dynamic and varied over time. For the CRISP study (Chapter 6: 6.1.2) however, minimal variation in DNA methylation was observed over 6 months suggesting that *in vivo* HPV infections are more stable. This view is supported by the long duration of persistent infections and the time taken for disease progression (Cancer Research UK, 2012a). The differences likely stem from positive selection for fast growing deregulated cells in culture in an environment very different from that of low grade disease.

7.2.3.4.2. General Gene Expression

We initially compared gene expression for all clones combined; a significant, positive correlation was observed between *E2*, *E6* and *E7* RQ which was unexpected. It is possible that the classic model of P97 regulation, (i.e. *E2* downregulation results in *E6*/*E7* upregulation) could be an early, transient event in carcinogenesis. Initial oncogene deregulation might be transient in both integrated and non-integrated HPV containing cells, and might actually cause integration as a result of chromosomal instability (Vinokurova and von Knebel Doeberitz, 2011). CaSki, SiHa and HeLa cell lines require continued oncogene expression in order to avoid senescence and maintain a transformed phenotype (DeFilippis *et al.*, 2003, Schwarz *et al.*, 1985, Wells *et al.*, 2000, Yee *et al.*, 1985). Conversely, excessive expression of oncogenes might adversely affect growth and cell viability rate (i.e. through continued chromosomal instability) such that oncogene downregulation confers a growth advantage after initial deregulation (De-Castro Arce *et al.*, 2011, Gray *et al.*, 2010, Van Tine *et al.*, 2004b).

A similar mechanism appears to be present in episomal infections. Early, transient oncogene deregulation occurred for episomal W12Ser4 where early passage W12E had low *E2* expression and translation whilst *E6*/*E7* were highly expressed and translated (Gray *et al.*, 2010). At later passage, *E2* expression increased to a level similar to *E6* and *E7*, and expression of all three genes decreased considerably with increasing passage number (Gray *et al.*, 2010). This pattern was suggested to be the result of a balance between deleterious effects of overexpression and expressing sufficient level of oncoprotein to maintain the infection. In W12, this compromise appeared to be post-transcriptionally regulated in later passages, where considerable divergence between mRNA and protein level was observed (Gray *et al.*, 2010) which could partly explain why there was a positive correlation between *E2*, *E6* and *E7* mRNA in many of the PC08/09 lines. Most clones had low (relative to CaSki) and closely matched *E2*/*E6*/*E7* RQ levels which could represent a state after the initial deregulation event, or a state in which post-transcriptional regulation is important. Other clones had considerably less *E2* mRNA than *E6*/*E7* mRNA, possibly representing earlier deregulatory events.

Clone D9 had an *E2* disrupted integrant, the highest transcription levels and transcription of *E2*/*E6*/*E7* increased gradually until passage 11. There are several reasons why *E2*, *E6* and *E7* expression could be positively correlated. First, the *E2* qPCR assay amplified a region upstream of the *E2* integration event and would not have been affected by the integration event. Whilst *E2* mRNA levels appeared to increase dramatically with passage and were positively correlated to *E6*/*E7*, disruption of the *E2* ORF might result in a non-functional E2 protein or perhaps even competitively exclude functional E2 if the E2BS binding motifs were present (Kim *et al.*, 2003).

The progressive increase in gene transcription could also be explained by a gradual positive selection for cells producing non-functional E2 repressor protein due to an enhanced growth rate, and the loss of cells containing functional E2 repressor protein. In this instance, *E2*, *E6* and *E7* transcripts could all increase in abundance without the repressor effects becoming apparent.

Further evidence also suggests that *E2* mRNA level does not necessarily correlate to the amount of full length, functional E2 protein. The *E2* transcripts of many papillomaviruses encode additional translational products; products arising from translation at internal ATGs and/or fusion products with other upstream ORFs (Lambert *et al.*, 1990), which may competitively exclude E2 at the E2BSs (Stubenrauch *et al.*, 2001). HPV transcripts are frequently polycistronic so detection of *E2* mRNA might not correlate to E2 translation (Zheng and Baker, 2006).

7.2.3.4.3. DNA Methylation and Gene Expression Comparisons

There was a significant, negative correlation between *E2* methylation and *E2*/*E6*/*E7* mRNA level, i.e. high *E2* methylation was associated with low gene expression. Similarly, *E6* transcription was significantly, negatively correlated to *E2*, LCR1 and LCR2, and almost *L1*/*L2* methylation. These results are somewhat unusual; integration has previously been associated with deregulated oncogene expression, *E2* hypermethylation with integration (Fernandez *et al.*, 2009) and higher grade disease with HPV methylation (Chapter 5: 5.1.2). Here however we did not observe a relationship between *E2* methylation and integration, and we observed a negative correlation between methylation and gene expression in these cell lines. These discrepancies might be the result of the small number of clones, that oncogene deregulation is an early event, or that the relationship between gene expression and DNA methylation is more complex than can be determined by simple correlation analysis. Intragenic methylation and gene expression is considered further in the general discussion (Chapter 9: 9.1.3 & 9.1.4).

7.2.3.4.4. DNA Methylation and Integration

The relationship between DNA methylation and integration of the PC08/09 clones appears complex. Whilst methylation of the *E2* and *L1*/*L2* regions varied considerably, all clones other than Y8 were hypomethylated in the LCR. Nonetheless, methylation of *E2* and *L1*/*L2* varied considerably among clones with integrated HPV16. The site of integration within the human genome could have an effect on the DNA methylation of the integrated HPV sequences as methylation may spread from areas of dense to sparse methylation (Turker, 2002, Doerfler, 2008). Furthermore newly integrated foreign DNAs are considered a target for *de novo*

methylation (Doerfler, 2008) in a manner that might reflect the number and position of integrants.

7.2.3.4.5. DNA Methylation, Gene Expression and Integration

The 6 clones used in this study represent those single cells that exhibited a considerable growth advantage over most cells in the heterogeneous mix, likely as a result of oncogene expression. No differences in oncogene expression were observed between integrated and episomal HPV in various grades of cervical disease (Hafner *et al.*, 2008). Whilst for several clones oncogene deregulation might be aided by the presence of integration events, P8 and A9 had no detectable integration so another mechanism of growth deregulation must have existed for these clones.

7.2.3.4.5.1. Methylation as a Mechanism of Oncogene Deregulation?

We hypothesised that hypermethylation of the promoter E2BSs might be present in the absence of integration. This was because of reports of increased E2BS/LCR methylation with disease progression, and the apparent inability of *E2* to bind methylated E2BSs and repress P97 which could be an alternative mechanism of transformation to integration (Snellenberg *et al.*, 2012, Kim *et al.*, 2003, Thain *et al.*, 1996, Bhattacharjee and Sengupta, 2006a, Fernandez *et al.*, 2009). For the P8 and A9 clones however, this did not appear to be the case, methylation was low in all regions and at a level roughly corresponding to low grade cervical disease (Chapter 5: 5.1.2, Chapter 6: 6.1.3).

For clone Y8, it was intriguing that LCR1 (the enhancer) was specifically hypomethylated at all passages whilst LCR2 (the promoter) was methylated at 20-30% throughout. Four of the five LCR2 CpGs are located within the two E2BSs of the promoter, the LCR1 CpGs are located near to but not within the enhancer E2BS. Hypermethylation of the promoter E2BSs blocks E2 binding (Thain *et al.*, 1996, Kim *et al.*, 2003). LCR2 hypermethylation might confer a growth advantage by blocking the repressive effects of E2 protein. For clone H9, there was also a decrease in promoter (LCR2) methylation which appeared to correspond to increased *E6* and *E7* RQ prior to passage 9. Whilst there is some evidence of an effect of E2BS methylation in these cell lines, it is far from conclusive.

7.2.3.4.5.2. DNA Methylation, Gene Expression and Integration

The D9 and H9 clones had integration events resulting in *E2* disruption. Y8 and M8, but also H9 had HPV-human mRNA fusion transcripts that mapped to *E7/E1*, and the DNAs were frequently heavily methylated. All *E7/E1* integration events mapped to 880/882 bp, regardless of clone/initial biopsy. HPV16 *E7* transcripts are spliced at 880 bp (Zheng and Baker, 2006) and APOT detects the junction between viral transcript sequence and human transcript sequence

suggesting that the fusion transcript has had the true site of integration spliced out at 880 bp. An unknown quantity of both HPV and human sequence has been spliced out such that the disrupted human/HPV site is downstream. Clone Y8 for instance had two integration events for *E7*/*E1*, both were within 1 kb (of the human genome) and at 882 bp (HPV genome) and appear to represent two splice variants. Downstream of HPV16 *E7*/*E1* is *E2* (Figure 1.1), a common site for HPV integration (Collins *et al.*, 2009). Detected *E2* integration junctions did not occur at recognised splice sites which makes it more likely that the sites represented the actual sites of integration.

There are several potential results due to this unknown site of integration for several clones. First, integration is somewhere in *E1*, in which case *E2* is intact but during integration when the genome is linearised, becomes positioned upstream of the P97 promoter. Expression of *E2* would likely require an adjacent, in frame and active human promoter, an actively transcribed concatemeric integrant or an episomal HPV genome. Alternatively, disruption could occur in *E2*, in which case production of functional E2 protein is unlikely but detection of mRNA by qPCR is possible, depending on where the disrupted site is. Finally, the integration event could occur downstream of *E2* in which case functional E2 protein might be produced. It is worth noting that two of the three clones with *E7*/*E1* integration junctions did express some *E2*, albeit at low level which could reconcile with several of the above scenarios. For the other clone, the *E2* qPCR assay failed which might correspond to an *E2* integration event, or an integration event upstream of *E2* which leads to its silencing. All integrants appear capable of translating full length, functional *E6*/*E7* oncoprotein (Zheng and Baker, 2006).

For clone H9, large changes in *E2* expression at around p8 coincided with changes in methylation at both the *E2* and *L1*/*L2* regions. A significant difference in *E2* expression and *L1*/*L2* methylation was observed between the early and late H9 passages, suggesting an event occurred around p8/p9 that triggered the changes. Whilst both early and late passages had a detected *E7*/*E1* integration junction, an additional *E2*/E4 integration was detected for the later passage and might be associated with changes in DNA methylation and *E2* transcription that are difficult to explain. Genomic instability caused by aberrant oncogene expression for instance might have promoted HPV integration or recombination of existing integrants within the genome. This is supported by *E6*/*E7* levels being the highest at p8/p9, the same passage at which a large decrease in *E2* expression and large changes in DNA methylation occurred. Alternatively, both integration events could have existed in early passage, but the second integrant might have been repressed and not detected. In this instance, episomal clearance is

a suitably rapid event (W12Ser7) and could be responsible for both the increase in *L1/L2* methylation and the decrease in *E2* expression after p8/9.

7.2.3.4.6. Unusual Observations

For clone P8, some methylation of the *L1/L2* region was observed. *L1* and *L2* methylation was more typically associated with integrated HPV genomes (VR series, other clones, W12) (Kalantari *et al.*, 2010) so in this instance APOT may have failed to detect an integration event. APOT does not detect transcriptionally silent integrants.

Passage 11 of the D9 clone was an unusual observation, all mRNA levels dropped considerably whilst DNA methylation continued to rise. The drastic changes (i.e. 10 fold) over a short period of time make experimental error (i.e. mislabelling) a possibility or the changes might represent a repressive or selective event.

The reduction in transcription for passage 12 of the Y8 clone and the subsequent increase in transcription for passage 13 are difficult to explain. They could coincide with the slight decrease in methylation, including LCR2 (promoter) methylation observed at passage 13.

E2 expression data was not present for clone M8 because the qPCR reactions consistently produced a melting peak that was atypical of the *E2* qPCR amplimer. It seemed likely that this was due to sequence diversity or deletion such that the melting temperature of amplimers was reduced. Integration in the *E7/E1* region was seen at passage 5 but not at passage 9, this could be a selective loss of cells containing the integration event, silencing of the HPV sequence expressing the integration event or a failing of the APOT assay.

7.3. Conclusions

The CaSki results were useful in that our data could be compared against published data. Our observation of the distinctive pattern of CpG methylation in each region was also seen both in our data and published data. W12Ser7 work also allowed the CpG specific pattern of methylation in LCR1 to be compared to published data. For W12Ser7, we observed dramatic methylation changes at the onset of integration that appeared to be both due to clonal selection for integrant containing cells but also progressive hypermethylation of integrated HPV genomes after clonal selection.

There were differences among the PC08/09 clones, and changes over time (passage). This was a contrast to our longitudinal CRISP study (Chapter 6: 6.1.2) where few changes with time were observed. There is some evidence to suggest that deregulation of the oncogenes might be an

early event which is then subsequently tightly regulated. This is in line with observations for W12 and our observation of positive correlation of *E2/E6/E7* mRNA transcripts in short term culture cell lines. These observations might also be explained by disparity between mRNA transcript and protein level.

There was reason to suspect that rather than causing oncogene deregulation, integration might be a consequence of chromosomal instability caused by another deregulating event. For these clones, methylation could be the cause of initial deregulation however we did not observe an immediately evident mechanism. Our initial hypothesis of E2BS hypermethylation occurring in instances where *E2* disruption had not occurred was not present for any of the non-integrated clones, but may have been observed for clone Y8.

Another hypothesis was that intragenic methylation might have an effect on gene expression. Whilst intragenic methylation of the *E2* region was observed in conjunction with reduced *E2* expression, it was also observed in conjunction with reduced *E6/E7* expression. Whether intragenic HPV methylation is associated with reduced gene expression is still unclear. In summary, the relationship between methylation, gene expression and HPV integration is complex.

Chapter 8. The VR Study - DNA Methylation, Integration and Gene Expression *in vivo*

A large series of anogenital biopsies of varying histological grade and sample site were collected for this investigation (n=171). The primary aim was to acquire a series of HPV16 positive VIN biopsies for which DNA methylation could be determined and compared to other aspects of HPV biology. HPV typing was performed initially for the purpose of selecting HPV16 positive samples for further analysis however the large cohort of HPV typing data represented a large and novel VIN HPV typing study.

As such, two main components of this study are presented. The first considers the HPV type distribution in VIN whilst the second considers the basic biology of HPV methylation and how it may relate to HPV gene expression and integration *in vivo*.

8.1. Results - HPV Type Distribution in VIN

This HPV distribution in VIN study was published (Bryant *et al.*, 2011) (Appendix VII) however since publication, more samples were collected. HPV infection was identified using the Greiner Bio-One PapilloCheck assay which detects and types 24 different HPV types in a single reaction (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73, 82, 6, 11, 40, 42, 43, and 44). The assay includes a series of internal controls to identify false negative results. Only histologically confirmed VIN samples were used in this study. A total of 62 VIN biopsies were eligible for inclusion in the type distribution study, 57 of which passed the internal controls and produced valid typing results. Of these, 46 were VIN3, 6 were VIN2 and 5 were VIN1. Four of the five samples that failed the assay were VIN3 whilst one was VIN1.

8.1.1. HPV Type Distribution

HPV infection was present in 43/57 cases (75.4%: 95% CI 62.9-84.7%), HPV16 and/or HPV18 were present in 35/57 cases (61.4%: 95% CI 48.4-72.9%) and HPV16 and/or HPV18 were the only types present in 27/57 cases (47.4%: 95% CI 35.0-60.1%). HrHPV types were present in 38 cases and LrHPV types in eight cases, three of these cases contained both HrHPV and LrHPV types. The typing distribution is presented in Table 8.1.

Nine cases had multiple HPV types present, one was a VIN1 sample with HPV18, 42 and 44/55 present whilst the others were all VIN3 samples. The VIN3 samples had the following combinations of types; two cases had HPV16 and HPV82, and there was one case each of the following combinations: (16, 18), (16, 33), (16, 6), (16, 31, 33), (16, 51), (16, 82, 6).

Of the five cases with LrHPV only, three cases contained HPV6 (one VIN2, one VIN3 and one VIN3 with well differentiated carcinoma and viral warts) and the other two cases were VIN1 lesions with HPV40 or HPV42.

			HPV type									
Histology	HPV-ve	HPV+ve	16	18	31	33	51	82	6	40	42	44/45[*1]
VIN1	1	4	0	1	0	1	0	0	0	1	2	1
VIN2	3	3	2	0	0	0	0	0	1	0	0	0
VIN3	10	36	32	1	1	4	1	3	4	0	0	0
Total	14	43	34	2	1	5	1	3	5	1	2	1
%	24.6	75.4	59.6	3.5	1.8	8.8	1.8	5.3	8.8	1.8	3.5	1.8

Table 8.1: HPV type distribution of VIN.

There were 9 cases showing multiple infection, hence the number of types identified is greater than the number of cases. HPV+ve = HPV positive, HPV-ve = HPV negative. Percentages indicate the proportion of cases that tested positive for a given type. *1: The HPV44 probe cross-hybridises to HPV45, hence these types are reported together.

8.1.1.1. Hypothesis 1: There is a Relationship between HPV Status and Age

HPV positive cases were aged 22-81 (mean=45.0) and HPV negative cases were aged 33-82 (mean=55.2). A two-sample T-test showed a significant difference between the mean ages of the HPV positive and negative groups (P=0.022).

8.1.1.2. Hypothesis 2: There is a Relationship between HPV Status and Histology

39/43 (90.7%) HPV positive cases had VIN2/3 whilst 13/14 (92.8%) HPV negative cases had VIN2 or higher. The proportion of cases with VIN2 or higher was not significantly associated with HPV positivity (Fisher's exact test, P=1.000). Interestingly, all HPV16 positive cases were associated with VIN2 or worse.

8.2. Results - HPV DNA Methylation, Integration and Gene Expression

Of the samples used for the assessment of HPV type distribution in VIN, 53 were biopsy samples from which both DNA and RNA were extracted, these samples were termed the VIN RNA (VR) series. From this series of biopsies, HPV16 containing samples (n=31) were selected for further study. Methylation and/or gene expression data were produced for 28 samples which comprised of mostly VIN3 (n=21), a small proportion of VIN2 (n=3), one VAIN3, one AIN2 and two samples without histologically confirmed disease. Patient age ranged from 22-81 years (mean = 46 years) The analysis of the HPV16 positive VR series samples was conducted by considering in a sequential manner, HPV integration status, HPV gene expression status and finally HPV methylation status in order to generate and test a series of hierarchical hypotheses. The study aimed to both determine the status of these three characteristics in the VIN series, but also to determine the relationships between the characteristics and with other variables within the data set.

8.2.1. HPV DNA Integration

DIPS was performed on all HPV16 positive VR series DNAs (n=31). In order to improve the negative predictive value of DIPS, the protocol was later changed such that a series of control PCRs were performed to check for DNA quality and samples that failed these controls were not analysed further. However, data for this study was collected prior to the protocol changes and several samples failed the DIPS controls but had detectable integration events. Four sample groups were generated based on the result of the control PCRs and the presence/absence of detected integration events (Table 8.2). For the purpose of investigating a relationship between HPV DNA integration and other characteristics of HPV biology, the X1 and 1 groups were frequently combined and groups were referred to as groups X, 0 and 1 only.

	Detected integration events	No detected integration events
Control PCR(s) passed	Group 1 (n=6)	Group 0 (n=9)
Control PCRs failed	Group X1 (n=2)	Group X (n=14)

Table 8.2: The assignment of integration group based on the passing/failing of control PCRs and the detection/non-detection of integration events.

Samples were entered into the control PCR(s) passed groups if a control PCR reaction was successful for the TaqI or Sau3AI digested DNAs. Samples that failed both PCRs were entered into the control PCRs failed groups. The number of samples that fell into each group (n) has been included.

Out of 31 samples analysed by DIPS, the failure rate of the control PCRs was high (n=16, 51.6%). Excluding those in the X1 group (Table 8.2), 6 of the remaining 15 samples that passed the control PCRs had detected integration events (40%). Details of the integration events,

including a list of the disrupted HPV genes and the region of the human genome where integration occurs have been summarised in Table 8.3. It was impossible to ascertain exactly where in the human genome the HPV16 in sample VR42 was integrated without using alternative assays. The VR42 human integration sequence identified corresponded to a satellite DNA repeat HSATII that is found in multiple locations throughout the human genome.

Sample	Gene disrupted	Disruption (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 8.3: Summary of the detected integration events for the VR series.

Gene disrupted refers to the HPV16 gene where the integration event occurs. Disruption (bp) gives the integration site into the HPV genome. Human genome region refers to the chromosome number and band number of the area of the human genome that has been detected in the integration events. *Homology was found to HSATII satellite DNA in Chr7, Chr22, Chr2, Chr16, Chr10 and ChrY.

VR37 had two detected integration sites. Whilst two HPV regions were disrupted (*E2* and *L2*), both events were within the same region of the human genome and close together. The events were facing in opposite directions so rather than representing two separate integration events, it likely represented the detection of one integration event (with a deletion between *E2* and *L2*) by two DIPS primer sets.

8.2.1.1. Hypothesis 3: There is a Relationship between HPV Integration and Patient Age

A previous study observed a relationship between patient age and integration status (Vinokurova *et al.*, 2008). Whilst the sample number in this study was relatively low, the hypothesis can still be tested for the data set. The patient age data was divided into the three integration groups (0, 1 and X) and compared to determine whether the mean age of each group differed significantly. The untransformed age data satisfied the assumptions of the one-way ANOVA and the means of the three integration groups were compared. There was not a statistically significant difference between the mean patient age of each integration group (one-way ANOVA P=0.917). In summary, a relationship between integration status and patient age was not identified.
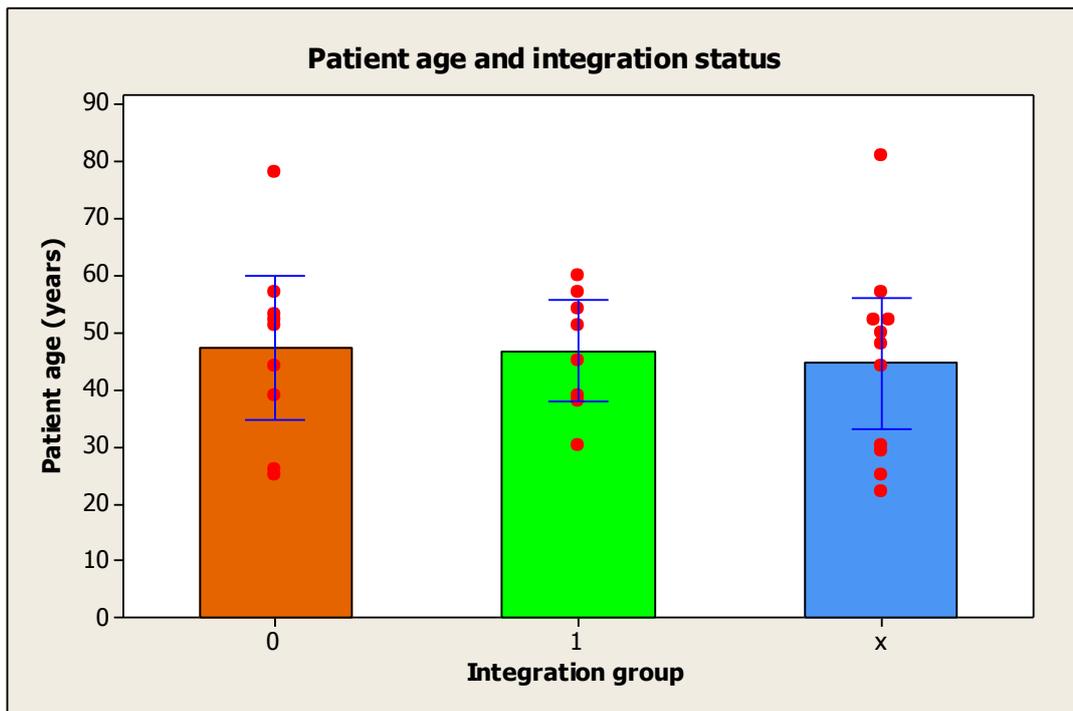
Figure 8.1: Patient age distribution of the VR series, separated by DNA integration status.

Red circles represent individual patient ages, the bars represent the mean age of each group and the blue error bars are 95% confidence intervals for mean.

8.2.2. HPV Gene Expression

RNAs from the VR series were analysed by *E2*, *E6*, *E7*, *HPRT* and *TBP* qPCR assays, relative quantification was performed using the equation described by Vandesompele *et al* (Vandesompele *et al.*, 2002).

8.2.2.1. Hypothesis 4: There is Variation in RQ among Samples

For the VR series as a whole, the three assays had a similar mean Log_{10} RQ (*E2* = -0.657, *E6* = -0.498 and *E7* = -0.608), however variation in RQ among the samples was considerable (Figure 8.2). The VR series had a large 8.5 Log_{10} range (i.e. $10^{8.5}$ = 316 million fold difference in RQ) between the highest and lowest *E2* expressing samples (Log_{10} *E2* RQ ranges from -6.70 to +1.78). Similarly, there was a 4.7 Log_{10} range (i.e. $10^{4.7}$ = 50 thousand fold difference in RQ) for *E6* RQ (Log_{10} *E6* RQ ranges from -3.16 to +1.54) and a 5.1 Log_{10} range (i.e. $10^{5.1}$ = 126 thousand fold difference in RQ) for *E7* RQ (Log_{10} *E7* RQ ranges from -3.92 to +1.18). The skewed distribution of the RQ data was highly unsuitable for multifactorial ANOVA GLM analysis and such analyses were not performed.
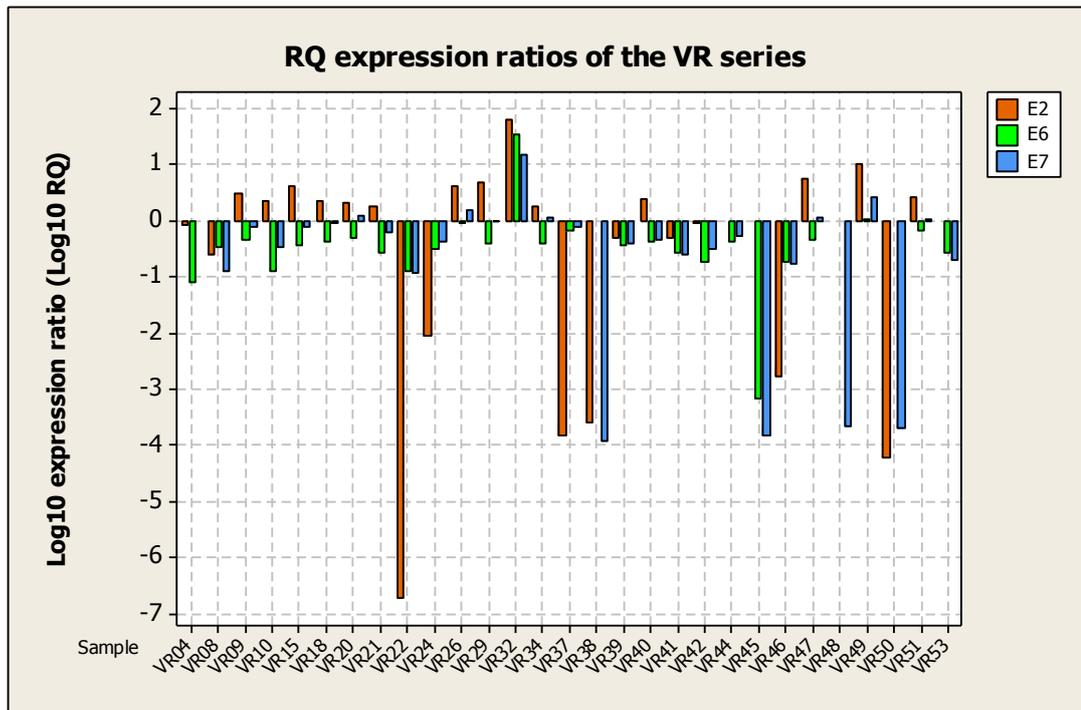
Figure 8.2: Log$_{10}$ Relative quantification expression ratios (Log$_{10}$ RQ) for the *E2*, *E6* and *E7* genes of the VR series.

The ratios are expressed relative to CaSki as a calibrator sample (CaSki has a log$_{10}$ RQ of zero for each assay).

8.2.2.2. Hypothesis 4: There is Variation in RQ among Assays

The qPCR data was unsuitable for multifactorial ANOVA so a non-parametric alternative was used. Friedman's test compares median values of data with respect to two factors (i.e. qPCR assay and sample). The Freidman's test requires a single result for each combination of factors and no missing sample data. Freidman's test was performed on the Log$_{10}$ RQ values presented in Figure 8.2 with the exception of samples where one or more RQ value was missing (i.e. VR04, VR38, VR44, VR45, VR50 and VR53).

There were significant differences in median RQ among the VR series samples (P=0.000) and among the three assays (P=0.001) meaning that RQ among the samples varied significantly, but also that the median RQ values of the three gene expression assays were different.

8.2.2.3. Hypothesis 5: There is a Relationship between gene Expression and HPV Integration

To determine whether or not a relationship existed between gene expression (RQ) and HPV DNA integration status the data was divided into testable groups. The RQ ratios of samples

falling into the three integration groups, X (sample inadequate), 0 (sample adequate, no detected integration) and 1 (detected integration) were then compared statistically.

The RQ data when separated with respect to qPCR assay and integration group were highly unsuitable for any of the multifactorial ANOVAs or non-parametric alternatives. Instead, multiple single factor Kruskall-Wallis tests were used with a Bonferroni corrected P-value (P=0.017). There were no significant differences among the median RQ values for integration group.
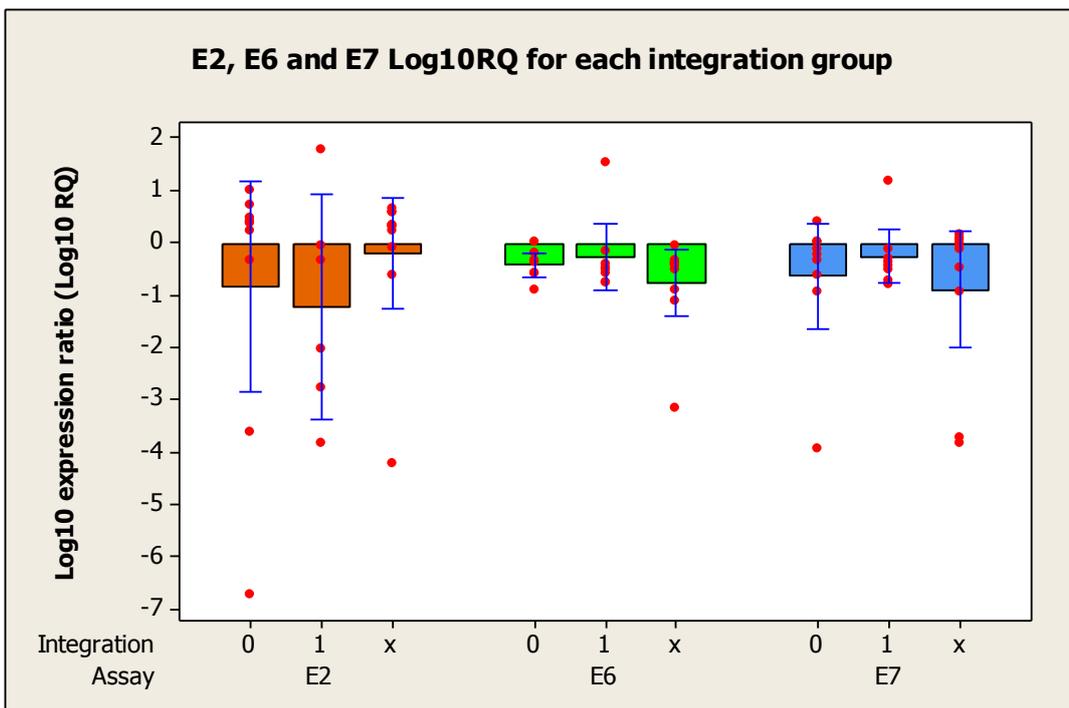
Figure 8.3: The Log$_{10}$ RQ values of the VR series for each qPCR assay, separated by integration status group.

Group 0 had no detected integration, group 1 had detected integration and group X were samples inadequate for integration analysis. Bars represent group means, red circles represent individual data points and blue bars are 95% confidence intervals.

Another approach was taken to compare just the adequate group 0 (no integration detected) and group 1 (integration detected) samples using multiple Mann-Whitney U tests. These tests compared the median RQ values for each qPCR assay. Again, no significant difference was observed between the median Log$_{10}$ RQ data for samples with and without detected integration events for any of the three qPCR assays (Mann-Whitney U test; *E2* (P=0.444), *E6* (P=0.637) and *E7* (P=0.666)).

As various HPV integration sites existed, a potential relationship between HPV genome integration site and gene expression in individual samples was investigated. There appeared to be wide variation in the RQ ratio of samples with integration events in similar areas of the HPV16 genome. Compared to the VR series mean, RQ of the integrated samples varied considerably (Figure 8.4) and there did not appear to be a clear relationship between HPV integration site and HPV gene expression (Table 8.4).
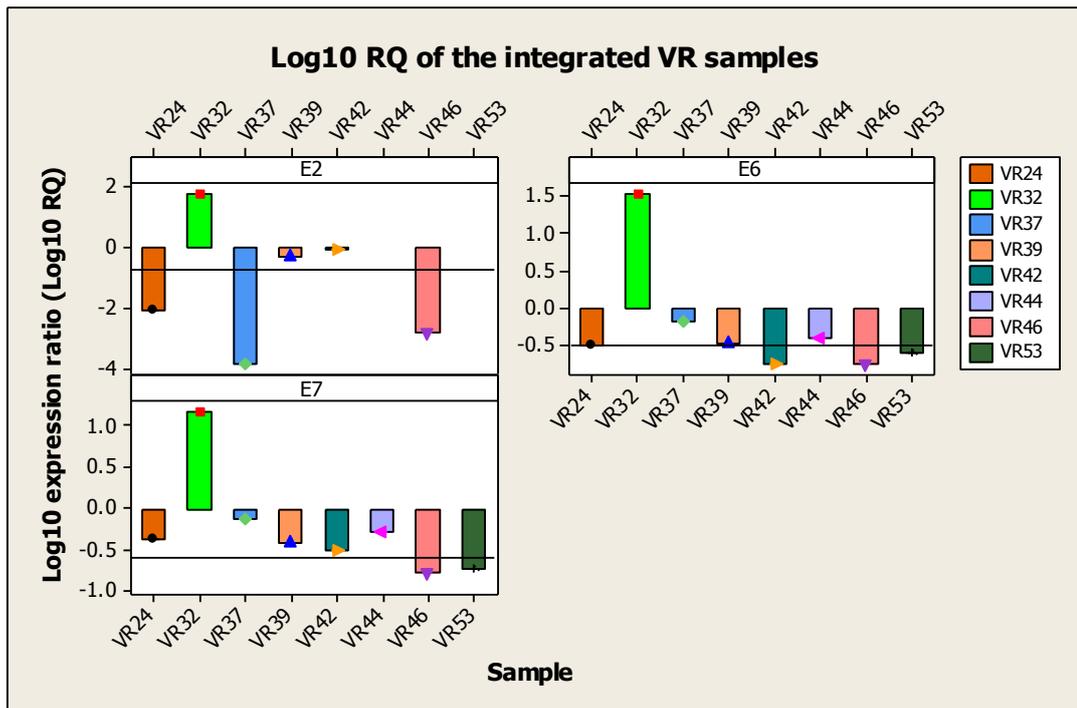
Figure 8.4: Gene expression Log$_{10}$ RQ values for the VR series integration group 1 samples.

A different graph pane is present for each gene expression assay, each bar represents Log$_{10}$ RQ for that sample. The horizontal black line represents the mean RQ for all samples (not just integrated) and is shown to aid the interpretation of Table 8.4.

Integration		Difference in Log10 from VR series mean		
Sample	Gene disrupted	ΔLog$_{10}$ E2RQ	ΔLog$_{10}$ E6RQ	ΔLog$_{10}$ E7RQ
VR24	E1	-1.38 (V low)	-0.01 (Average)	0.24 (Above average)
VR32	E2	2.43 (VV high)	2.04 (VV high)	1.78 (V high)
VR37	E2 & L2	-3.18 (VV low)	0.33 (Above average)	0.48 (High)
VR39	E1	0.34 (Above average)	0.04 (Average)	0.19 (Above average)
VR42	L2	0.6 (High)	-0.25 (Below average)	0.1 (Average)
VR44	E2	*	0.12 (Average)	0.33 (Above average)
VR46	E1	-2.11 (VV low)	-0.25 (Below average)	-0.16 (Below average)
VR53	L2	*	-0.09 (Average)	-0.11 (Average)

Table 8.4: The difference in *E2*, *E6* and *E7* RQ of the integrated VR series samples compared to the mean gene expression of the VR series.

ΔLog$_{10}$ RQ represents sample mRNA level minus VR series mean mRNA level and is used to quantify expression level statements. Colour coding is used to aid interpretation and is based on the difference from the mean (ΔLog$_{10}$ RQ): VV high (>2), V high (>1<2), High (>0.4<1), Above average (>0.15<0.4), Average (0 +/- 0.15), Below average (<-0.15>-0.4), Low (<-0.4>-1), V low (<-1>-2), VV low (<-2). * missing data.

8.2.2.4. *Hypothesis 6: The Relationship between E2, E6 and E7 Expression Might Coincide with an Accepted Model of Gene Regulation*

Relative quantification of Ct values accounts for variation in input cDNA and ratios are calibrated against CaSki gene expression Ct values. Equal RQ ratios between assays therefore represents an mRNA level ratio between the assays (i.e. *E2:E6*, *E6:E7* etc) similar to that in CaSki rather than equal *E2*, *E6* and *E7* mRNA level. Nonetheless, comparison of RQ ratios for one VR sample with mRNA counts from deep sequencing suggested that RQ ratios were fairly indicative of true mRNA level (Chapter 4: 4.1.4.1). The relationship between *E2*, *E6* and *E7* RQ and the results of Spearman's rank correlation analysis are presented in Figure 8.5 and Table 8.5 respectively. *E2*, *E6* and *E7* RQ were strongly positively correlated, as per the PC08 and PC09 clones.
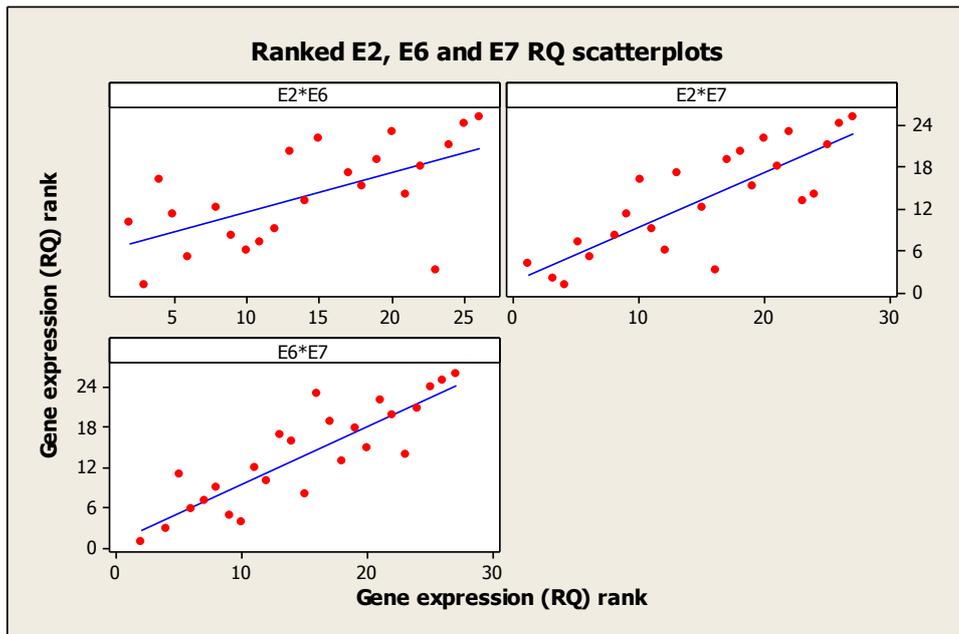
Figure 8.5: The relationship between *E2*, *E6* and *E7* gene expression is explored using scatter plots of ranked data.

The data presented are ranked RQ, the relationship being presented is indicated by the plot header, i.e. *E2*E6* represents *E2* RQ ranks on the Y-axis and *E6* RQ ranks on the X-axis. Both individual values (red circles) and linear regression lines (blue lines) are included.

	E2 RQ vs. *E6* RQ	*E2* RQ vs. *E7* RQ	*E6* RQ vs. *E7* RQ
Spearman's rank correlation coefficient	+0.641, P= 0.013	+0.824, P= 0.000	+0.909, P=0.000

Table 8.5: The results of Spearman's rank correlation analyses on the ranked gene expression data for *E2*, *E6* and *E7*.

Included are both the Spearman's rank correlation coefficients and the P-value. Correlation is considered significant at P=<0.05.

8.2.3. DNA Methylation

8.2.3.1. Hypothesis 7: There is Variation in the Amount of DNA Methylation among Samples and among Regions

DNA methylation at the *E2*, LCR1, LCR2 and *L1/L2* regions was assessed. The mean amount of DNA methylation of each region, for each VR sample is shown in Figure 8.6. There was considerable variation in the amount of methylation among the VR samples and the level of methylation varied with region of the HPV16 genome (Figure 8.7). Methylation of the *E2* region was especially variable, whilst the mean *E2* methylation was 16.5%, *E2* methylation ranged from 1.2-89.7% between the least and most methylated samples. The *L1/L2* region was similarly variable, mean *L1/L2* methylation was 24.5% however a large range of 1.6-88.0% was observed. The LCR1 and LCR2 LCR regions were considerably less variable. The mean amount

193

of methylation at the LCR1 region was 2.6% with a narrow range of 1.0-4.7%. Similarly the mean for the LCR2 region was 1.3% with a small range of 0.0-4.3%. In summary, the *E2* and *L1/L2* regions were the most variable whilst the LCR1 and LCR2 regions of the HPV16 LCR were universally hypomethylated (Figure 8.7).
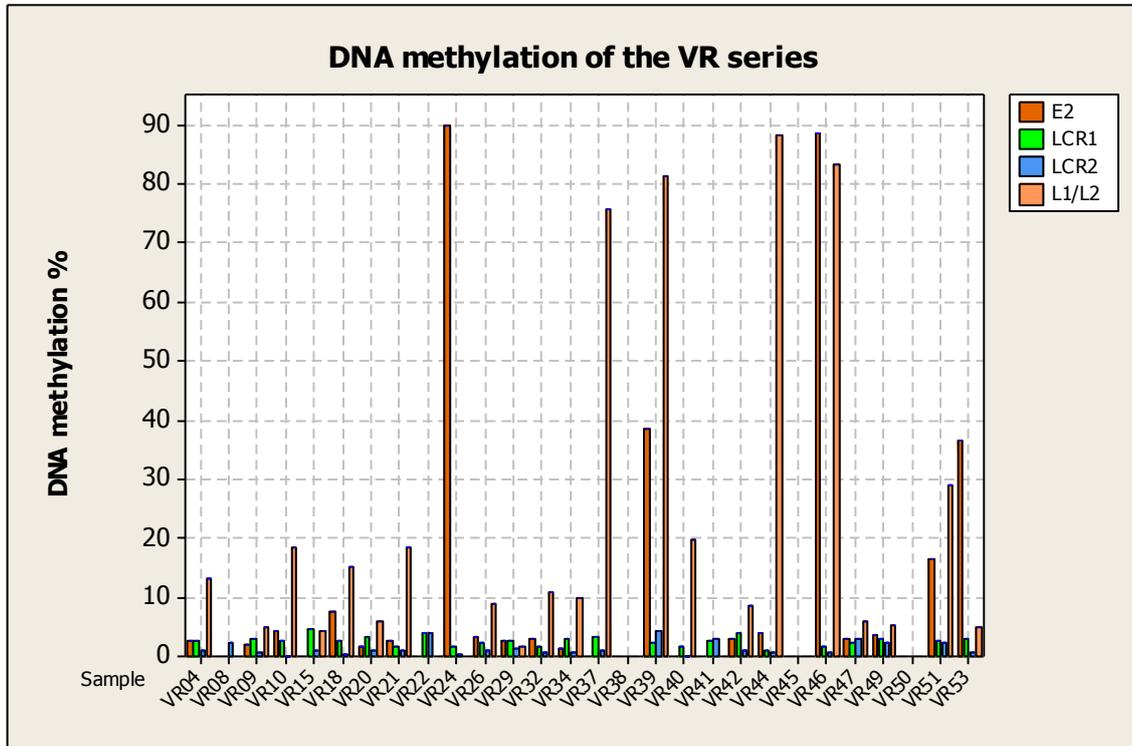
Figure 8.6: DNA methylation of HPV16 in the VR series.

Coloured bars represent mean DNA methylation for each region of the HPV16 genome for the VR series.
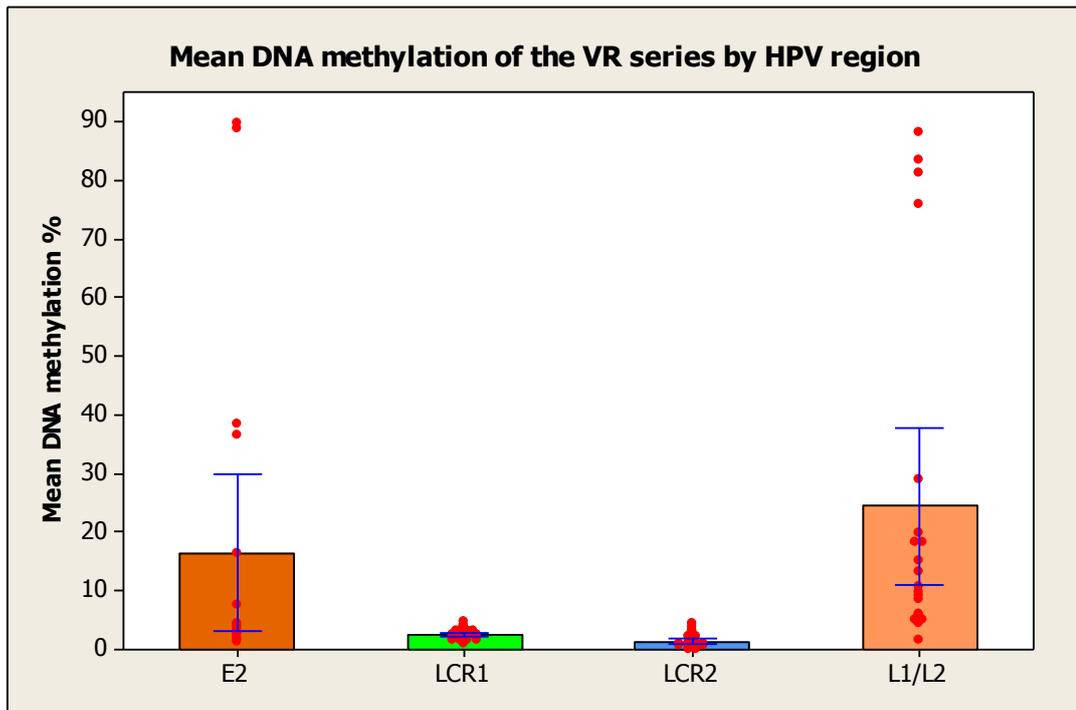
Figure 8.7: VR series mean methylation at each HPV16 region.

Bars represent the mean amount of methylation at each site for the entire VR series. The graph includes 95% confidence intervals for the mean (blue bars) and the level of methylation of each individual VR sample (red circles).

A two-way ANOVA GLM was used to test for a significant difference in mean DNA methylation among samples and each HPV region. One violation of the assumptions of the two-way ANOVA was present, the variances were unequal (Bartlett's P=0.000) however no suitable non-parametric alternative exists. There was a significant difference in mean DNA methylation among the regions (P=0.000) but not among the samples (P=0.189).

8.2.3.2. Hypothesis 8: There is CpG Specific Variation within Each Site

Previous chapters reported a variation in the methylation of each CpG within each HPV region. CpG specific differences were also conserved between sample cohorts such that patterns of methylation were discernible, especially for the *L1/L2* region where an "N" shaped profile was observed. It was of interest to consider whether or not a similar scenario existed for the VR series.

There appeared to be a somewhat conserved pattern of CpG methylation for the VR series *E2* region (Figure 8.8). CpG E2-3 tended to be the lowest methylated in most samples and had the lowest mean methylation of all of the *E2* CpGs (12.7%). E2-4 was the highest methylated in most samples and also had the highest mean methylation (20.3%), in two instances E2-4 was

almost 100% methylated. E2-1 and E2-2 were quite closely matched and in general, there was a decrease in methylation between CpGs E2-4 and E2-8. This was not paralleled perfectly by the *E2* CpG methylation pattern of the PC08 and PC09 cell line clones but there were some similarities (e.g. a decrease between E2-4 and E2-8).

For the LCR1 region, whilst the overall values were fairly low, there appeared to be a general decline in DNA methylation between LCR1-1 and LCR1-3 (Figure 8.9). This was the same as was observed for some previous studies.

For the LCR2 region, there was an "M" shaped profile to many of the methylation curves, i.e. DNA methylation of LCR2-1, LCR2-3 and LCR2-5 was low whilst DNA methylation of LCR2-2 and LCR2-4 was higher (Figure 8.10). A similar pattern was observed for some of the PC08 and PC09 passage 9 cell line samples.

The distinctive "N" shaped CpG methylation profile of *L1/L2* was highly reminiscent of the *L1/L2* methylation profile of other cohorts. Particularly fascinating about this methylation data was that the *L1/L2* CpGs are all within 17 bp of each other yet varied considerably. VR51 for instance (Figure 8.11) had large differences among CpGs; not only was there a difference of 45% between CpGs L1L2-1 and L1L2-4, but there was a 27.5% difference in DNA methylation between L1L2-2 and L1L2-3, two CpGs that are separated by only 1 bp.

In order to attempt to confirm statistically the observation of considerable variation in the methylation of different CpGs and of different HPV regions, a two-way ANOVA GLM was applied. Several violations of the two-way ANOVA GLM occurred, the standardised residuals were not normally distributed (Anderson Darling P<0.005) and the variances were not equal (Bartlett's P=0.000). As no suitable non-parametric alternative exists, the two-way ANOVA GLM was performed anyway. A significant difference among the means was observed for both HPV DNA region (P=0.000) and CpG (P=0.000), confirming the observations of variation in the methylation data.
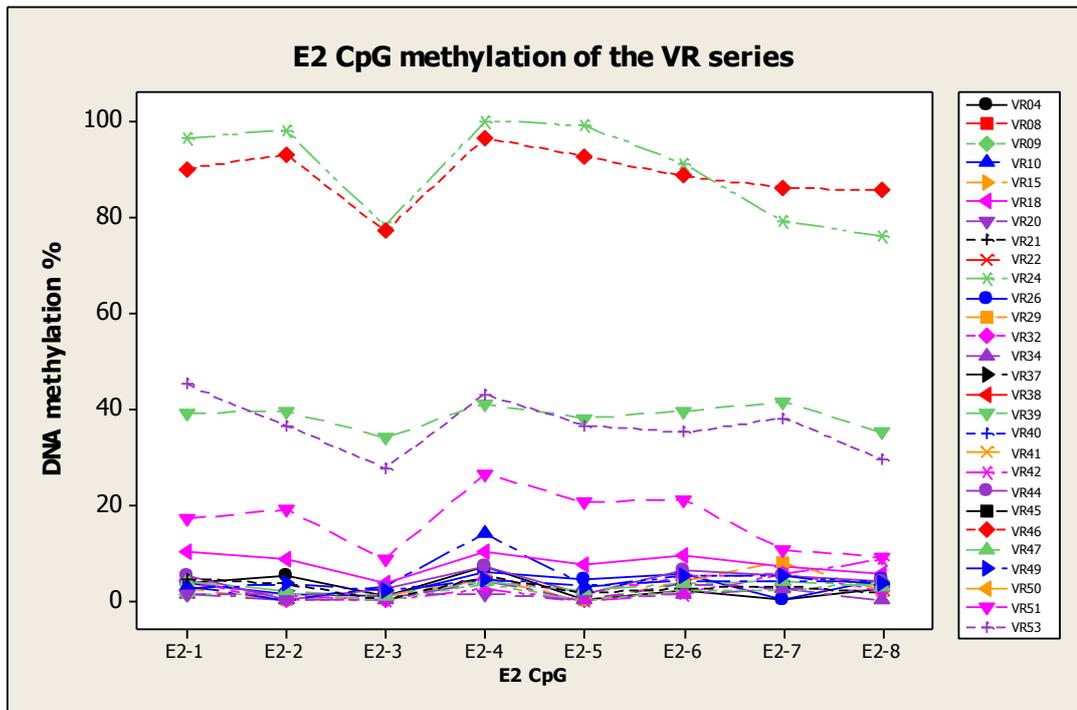
Figure 8.8: *E2* DNA methylation profile for the VR series.

Each symbol represents a different VR sample. No relationship is inferred by the connecting lines between points; they are included only to improve interpretation of the figure data.
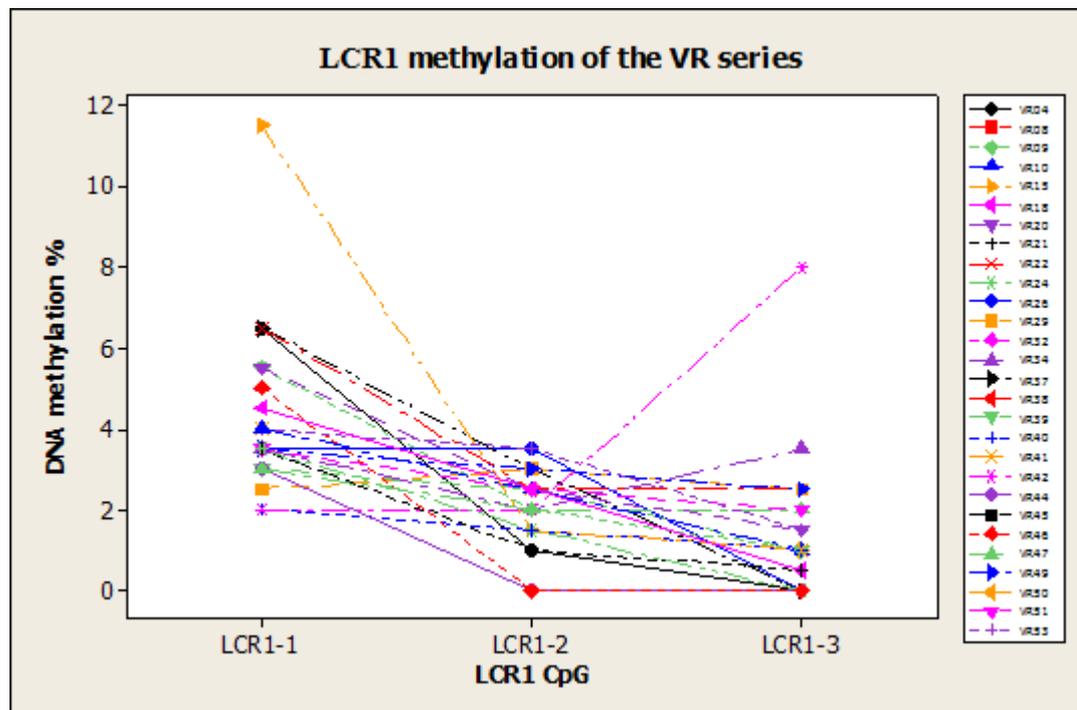
Figure 8.9: LCR1 DNA methylation profiles for the VR series.

Each symbol represents a different VR sample. No relationship is inferred by the connecting lines between points; they are included only to improve interpretation of the figure data.
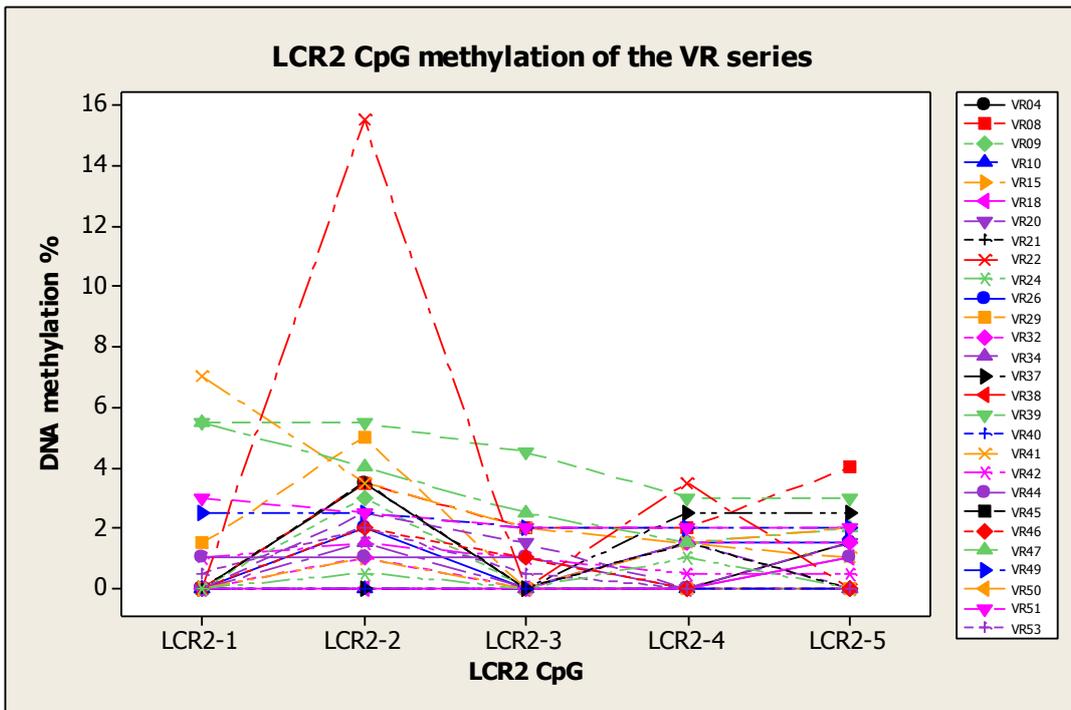
Figure 8.10: LCR2 DNA methylation profiles for the VR series.

Each symbol represents a different VR sample. No relationship is inferred by the connecting lines between points; they are included only to improve interpretation of the figure data.
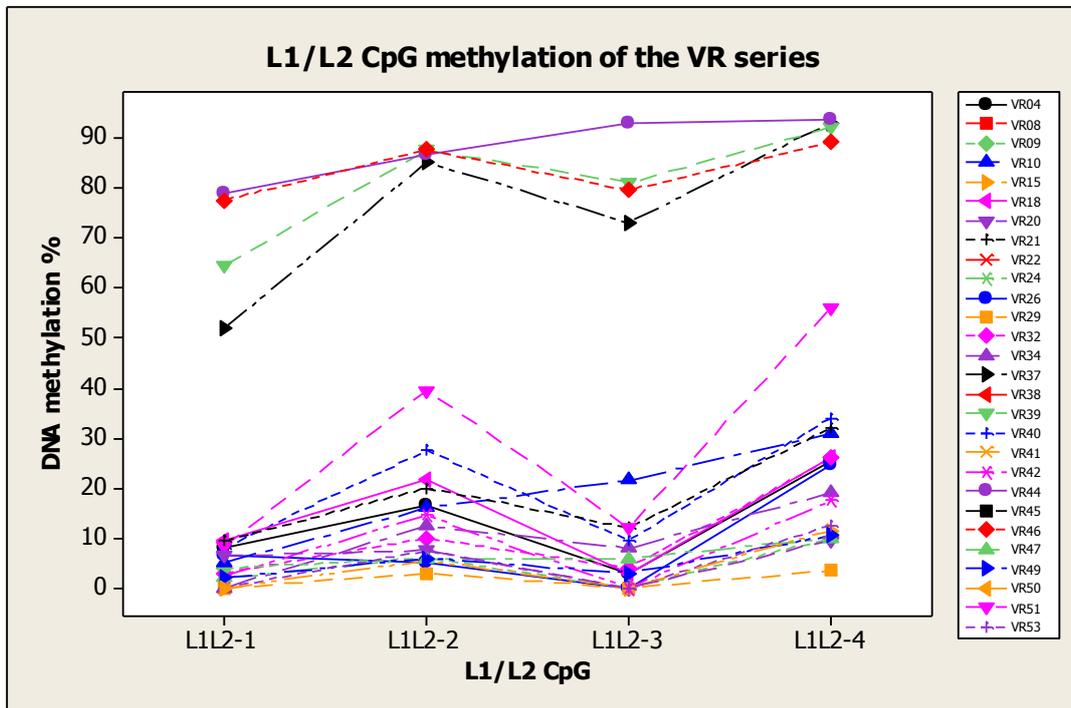
Figure 8.11: *L1/L2* DNA methylation profiles for the VR series.

Each symbol represents a different VR sample. No relationship is inferred by the connecting lines between points; they are included only to improve interpretation of the figure data.

8.2.3.3. Hypothesis 9: There is a Relationship between DNA Methylation and HPV DNA Integration

One of the most important questions that needed to be asked of the methylation data was whether or not the amount of DNA methylation could be associated with HPV16 integration into the host genome. Previous studies have found an association between DNA methylation and DNA integration whereby integrated genomes were typically hypermethylated, especially within the *L1* region (Kalantari *et al.*, 2008b).

The VR samples were separated into the three integration groups previously described; group 1 (detected integration events), group 0 (no detected integration events) and group X (sample inadequate for analysis). A notable characteristic of the VR samples with high DNA methylation (i.e. over 30%) at *E2*, *L1/L2* or both *E2* and *L1/L2* was that they all had detected integration events. The relationship between DNA methylation and HPV integration of the VR series can be seen in Figure 8.12.

The association between DNA methylation and HPV DNA integration appeared to be complex and there were considerable differences in the methylation profiles of integrated samples. Some samples had hypomethylation of both *E2* and *L1/L2*, some had hypermethylation of *E2*, some samples had hypermethylation of *L1/L2* whilst others had hypermethylation of both *E2* and *L1/L2*. It is worth noting that sample VR24 was the only sample not run through the *L1/L2* assay and that sample VR37 failed the *E2* assay repeatedly.
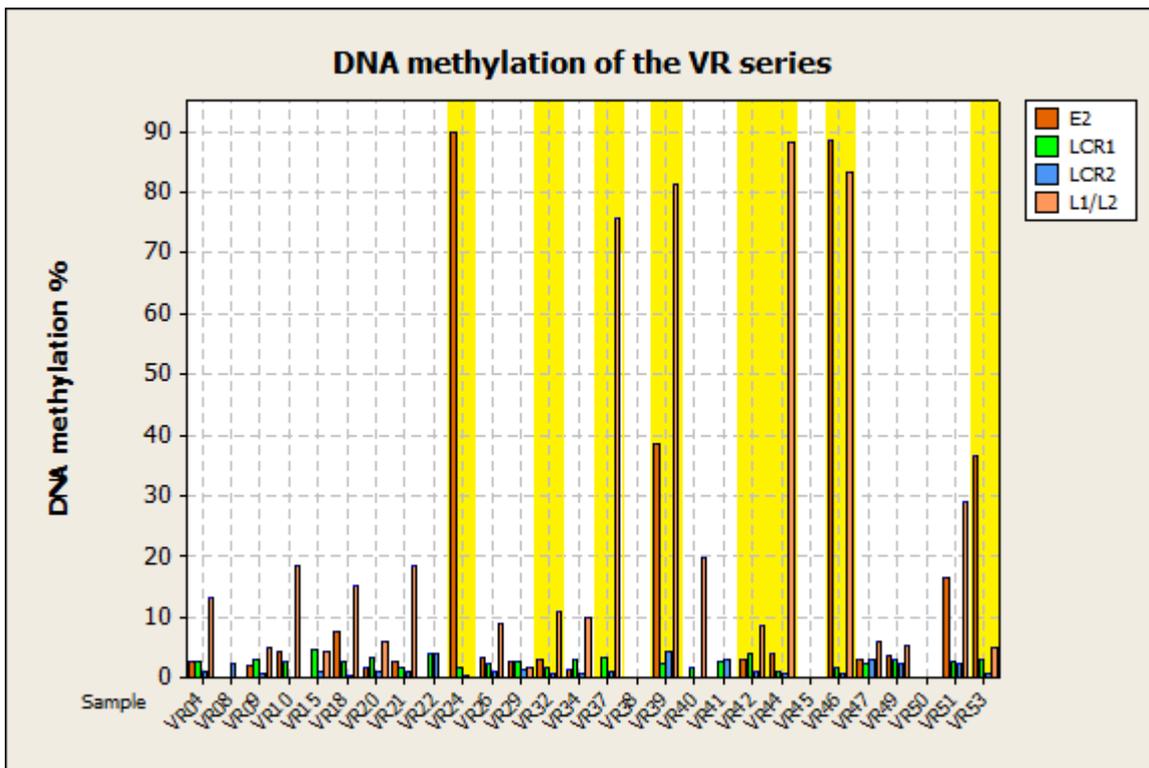
Figure 8.12: DNA methylation of the VR series with integration state.
Samples with detected integration events are highlighted in yellow.

In order to explain the different DNA methylation profiles within the integrated samples, the site of disruption of the HPV genome was compared to the site of HPV hypermethylation. Table 8.6 lists both the hypermethylated genes (hypermethylation defined in this instance by >30% methylation) and the genes disrupted by HPV integration for the group 1 samples. A possible relationship existed between disruption site and hypermethylation site. Samples with *E1* disruption (VR39 and VR46) had hypermethylation at both the *E2* and *L1/L2* regions; VR24 also had *E1* disruption and *E2* hypermethylation however the *L1/L2* result was missing. *E2* disruption for the VR44 sample was accompanied by *L1/L2* hypermethylation and *E2* hypomethylation whilst conversely, *L2* disruption for VR53 was accompanied by the exact opposite, *E2* hypermethylation and *L1/L2* hypomethylation. VR37 was a particularly interesting case because multiple integration events were detected (possibly due to deletion) and whilst it had failed the *E2* assay repeatedly, the *L1/L2* region was hypermethylated. Samples VR42 and VR32 may have had HPV present in both integrated and episomal forms; episomal HPV16s are typically maintained in very high copy number (Doorbar, 2006) and here, were lowly methylated. If VR42 and VR32 contain both integrated and episomal HPV this would allow for both the detection of integration for the DIPS assay and the observation of very low levels of DNA methylation with the quantitative pyrosequencing assay.

200

Sample	Gene disrupted	Hypermethylation E2	Hypermethylation L1/L2
VR24	E1	Yes	*
VR32	E2	No	No
VR37	E2 & L2	*	Yes
VR39	E1	Yes	Yes
VR42	L2	No	No
VR44	E2	No	Yes
VR46	E1	Yes	Yes
VR53	L2	Yes	No

Table 8.6: Details of the disrupted HPV16 gene for the integrated VR samples along with the regions of hypermethylation.

Hypermethylated regions are highlighted yellow. Samples showing both E2 and L1/L2 hypomethylation are highlighted grey. * Data missing.

8.2.3.3.1. Statistical Comparison of the Integration Groups

To determine whether or not a relationship existed between DNA methylation and HPV DNA integration status, DNA methylation of the three integration groups was compared. Mean methylation of the *E2* and *L1/L2* regions appeared higher for the integrated group than other groups however there was variation in all groups (Figure 8.13).

The mean methylation data was not quite suitable for analysis using a two-way ANOVA GLM for integration group and pyrosequencing assay as factors (Anderson Darling P=0.028, Bartlett's test for equal variances P=0.000). However, because no non-parametric equivalent of the two-way ANOVA GLM exists that is capable of analysing data of an unbalanced design, with missing values and with repeated data, the two-way ANOVA GLM was used under the premise that the results were interpreted with caution. The two-way ANOVA GLM found that a significant difference in mean methylation existed as a consequence of both the region being tested (i.e. *E2*, LCR1, LCR2 and *L1/L2*; P=0.000) and the integration status (i.e. 0, 1 and X; P=0.006). As the GLM does not permit pairwise comparison, multiple one-way ANOVAs were used. Multiple analyses should be accompanied by a Bonferroni correction, however in this instance the investigation was hypothesis generating and so the usual level of significance (P=0.05) was also considered for interpretation.
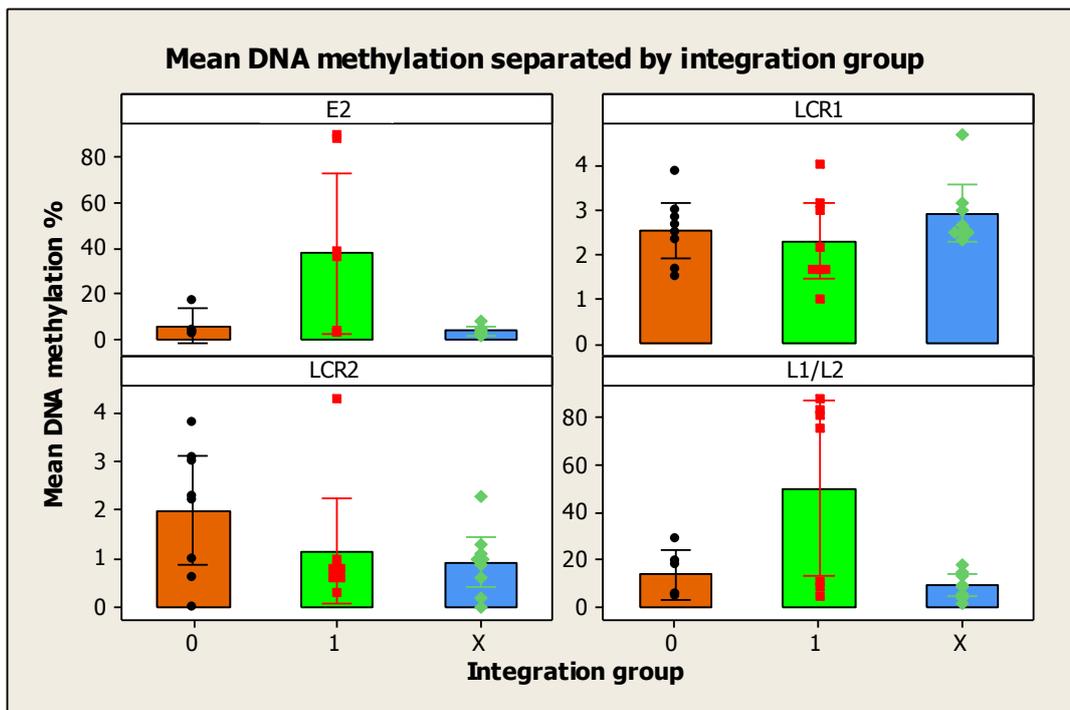
Figure 8.13: Mean DNA methylation of each region for the VR series separated by integration group. Bars represent mean methylation, symbols represent individual samples and error bars represent 95% confidence intervals.

First, the difference in mean methylation of the various integration states was compared for each of the pyrosequencing assays. For mean *E2* methylation, there was a significant difference in mean methylation between each integration group as whole (one-way ANOVA P=0.024). Pairwise comparisons however revealed that there was not quite a significant (Tukey-Kramer at P=0.05) difference in mean methylation between the integration detected (1) and no integration detected (0) groups, but there was a significant difference between groups 1 and X.

For the *L1/L2* region, a significant difference existed among the mean *L1/L2* methylation of the three integration groups (one-way ANOVA P=0.040) however pairwise comparison yielded similar results to the *E2* comparison; there was not quite a significant (Tukey-Kramer at P=0.05) difference between groups 1 and 0, there was a significant difference between groups 1 and X. This is difficult to explain, but seems likely the result of small sample number and the presence of region specific hyper/hypomethylation of the integrated samples. This further promotes the concept of a complex relationship between HPV DNA methylation and integration. Finally, there was no significant difference between the mean methylation of the three integration groups for the LCR1 region (P=0.260) or the LCR2 region (P=0.201). Pairwise

comparison by Tukey-Kramer revealed that none of the groups were significantly different from each other. The Bonferroni corrected P-value for these comparisons was P=0.0125.

Integration appeared to be highly associated with DNA methylation for the VR series; therefore, for future analyses it was worth considering the relationship of DNA methylation with other biological factors both with and without the integrated samples included in the analysis.

8.2.3.4. Hypothesis 10: There is a Relationship between DNA Methylation and Gene Expression

A simple but limited way to determine whether a relationship exists between two quantitative factors is to perform correlation analysis. In this study, the correlation between the amount of HPV DNA methylation at each genomic site and the expression of the *E2*, *E6* and *E7* genes was tested. Previously, integration status was seen to be associated with hypermethylation of *E2* and *L1/L2* but not with gene expression. As such, the analysis was performed several times, once with integrated samples included, once with the integrated samples excluded and once with the integrated samples alone. This was performed to isolate the effects of integration from the relationship between DNA methylation and gene expression.

Spearman's rank correlation analysis was performed on the ranked *E2*, LCR1, LCR2 and *L1/L2* mean methylation data and on the *E2* RQ, *E6* RQ and *E7* RQ gene expression data (Table 8.7). Several significant correlations were observed.

A significant negative correlation between *L1/L2* methylation and *E2* expression existed when all samples were considered and was conserved when the integrated samples were excluded from the analysis. There was a significant negative correlation between *E2* methylation and *E7* expression for all samples however this was lost when the integrated samples were excluded. When considering the integrated samples alone, a significant negative correlation was observed between *E2* methylation and *E2* RQ. However, for the integrated samples alone results must be interpreted with caution as the sample number was very small and an association between gene expression and integration might confound any observed correlation.

	Including integrated (group 1) samples			
	E2 methylation	LCR1 methylation	LCR2 methylation	*L1/L2* methylation
E2 RQ	P= 0.201	P= 0.720	P= 0.667	P= 0.003
	-0.327	-0.081	-0.095	-0.649
E6 RQ	P= 0.485	P= 0.814	P= 0.671	P= 0.619
	-0.171	-0.051	0.089	-0.115
E7 RQ	P= 0.018	P= 0.900	P= 0.905	P= 0.121
	-0.55	0.028	0.026	-0.359

Table 8.7.a

	Excluding integrated (group 1) samples			
	E2 methylation	LCR1 methylation	LCR2 methylation	*L1/L2* methylation
E2 RQ	P= 0.220	P= 0.659	P= 0.444	P= 0.044
	0.382	-0.12	-0.199	-0.545
E6 RQ	P= 0.484	P= 0.996	P= 0.929	P= 0.336
	0.224	0.001	0.024	-0.278
E7 RQ	P= 0.671	P= 0.765	P= 0.791	P= 0.151
	-0.145	0.084	-0.072	-0.422

Table 8.7.b

	Integrated samples (group 1) **alone**			
	E2 methylation	LCR1 methylation	LCR2 methylation	*L1/L2* methylation
E2 RQ	P= 0.029	P= 0.892	P= 0.852	P= 0.275
	-0.916	-0.072	0.099	-0.61
E6 RQ	P= 0.439	P= 0.408	P= 0.468	P= 0.613
	-0.352	-0.342	0.301	0.234
E7 RQ	P= 0.276	P= 0.601	P= 0.468	P= 0.803
	-0.145	0.084	-0.072	-0.422

Table 8.7.c

Table 8.7.a-c: The results of Spearman's rank correlation analysis of the DNA methylation data for all VR samples with and without the group 1 integrated samples included.

P-values are significant at P<0.05. Included are the Spearman's rank correlation coefficients for each comparison. Significant results are highlighted in yellow.

When these significant relationships were presented graphically, there were some interesting observations (Figure 8.14). The three comparisons where a significant relationship was observed are included (*E2* methylation vs. *E7* RQ, *E2* methylation vs. *E2* RQ and *L1/L2* methylation vs. *E2* RQ) and the figure includes the data points of all of the integration group samples for reference purposes.

8.2.3.4.1. E2 Methylation vs. E7 Gene Expression

This relationship was only deemed significant when all of the VR series were included for analysis. The integrated samples were the most variable in both gene expression and DNA methylation (Figure 8.14). *E2* DNA methylation and *E7* gene expression of the remaining samples was considerably less variable and the correlation was lost if the group 1 samples were excluded or analysed alone.

8.2.3.4.2. E2 DNA Methylation and E2 Gene Expression

A significant correlation between *E2* DNA methylation and *E2* gene expression was only observed for the integrated samples. There appeared to be a linear relationship present for these samples whereas for the group 0 and group X samples there was no discernible relationship present (Figure 8.14). *E2* DNA methylation might be associated with *E2* gene expression when HPV has integrated into the host genome.

8.2.3.4.3. L1/L2 DNA Methylation vs. E2 Gene Expression

For the comparisons with and without integrated samples, a significant correlation was observed between *L1/L2* methylation and *E2* RQ. A particularly interesting observation for this comparison was that the integrated samples occupied the extremes of both high DNA methylation and low/very high gene expression (Figure 8.14). The integrated samples appeared to be varied whilst the relationship for the group 0 and group X samples was more linear (as determined by Spearman's rank) and over a narrower range.
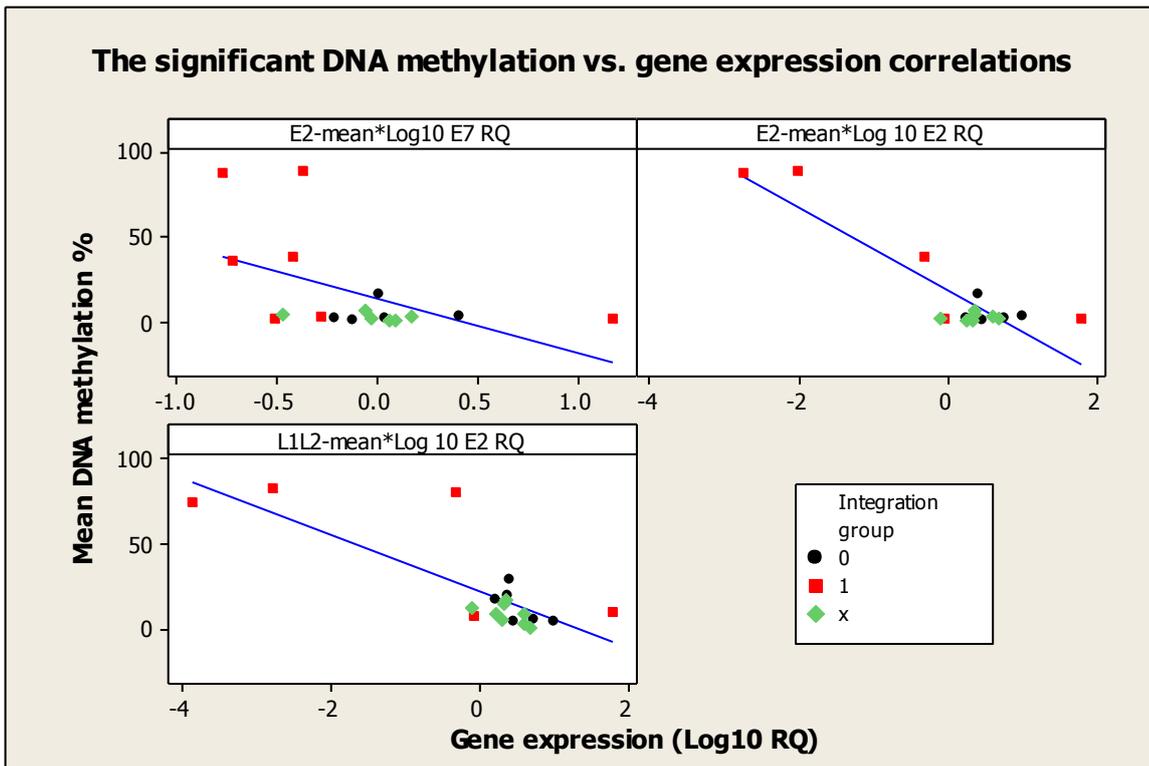
Figure 8.14: Scatter plots of the significant correlations (as determined by Spearman's rank) between HPV DNA methylation and HPV gene expression.

Three comparisons are presented, *E2* mean DNA methylation vs. Log_{10} *E7* RQ, *E2* mean DNA methylation vs. Log_{10} *E2* RQ and *L1/L2* mean DNA methylation vs. Log_{10} *E2* RQ. The data presented includes all three integration groups (0, 1 and X) in order to see the differences among the groups where a significant relationship was observed for some groups and not others. The X-axis scale of the three scatter plots differs as a result of some samples having various missing DNA methylation or gene expression RQ values.

8.2.3.5. *Further Analysis of the HPV DNA Methylation and Gene Expression*

The correlation analyses hinted at a relationship between DNA methylation and gene expression, however a more informative and robust approach to the analysis might be gained using ANOVA tests. The RQ data had a very large range and the distribution was highly skewed, non-parametric and very unsuitable for ANOVA. As such, a more suitable way of analysing a potential relationship between gene expression and DNA methylation was to divide the gene expression data (i.e. RQ) into groups and determine whether a relationship between RQ group and DNA methylation existed.

Various ways of dividing the data were considered, initial attempts used the mean of the VR series and a range of values above and below the mean to generate the groups. This however meant that very large negative values affected the grouping, the grouping was subjective and

the groups were not representative of the sample distribution. Instead, the RQ data for each assay was divided by quartile group (i.e. RQ data were divided into four equal groups to make the 1st, 2nd, 3rd and 4th quartiles (Q1, Q2, Q3 and Q4 respectively)). As per previous analyses, statistical analysis was performed multiple times, once with integrated samples included and again with integrated samples excluded. This was in order to factor into the interpretation a potential association between HPV methylation and integration.

8.2.3.5.1. Integrated Samples Included Analysis

With the VR samples separated into gene expression RQ quartiles, statistical analysis was undertaken to determine a relationship between gene expression and DNA methylation. The DNA methylation data could not be transformed to an acceptable degree for use in a multifactorial ANOVA GLM. Instead, the data was separated by region (i.e. *E2*, LCR1, LCR2 and *L1/L2*) and assessed by one-way ANOVAs with Tukey-Kramer pair wise comparisons (i.e. one ANOVA for *E2* RQ and *E2* methylation, another for *E2* RQ and *LCR1* methylation etc). The nature of this investigation was exploratory and hypothesis generating so the Bonferroni correction was used during interpretation of the statistics however significance at P=0.05 was also considered. The samples were separated by RQ quartile for each gene expression assay and DNA methylation of each quartile was compared graphically and statistically. This data is presented in Figure 8.15, Figure 8.16, Figure 8.17 and Figure 8.18.
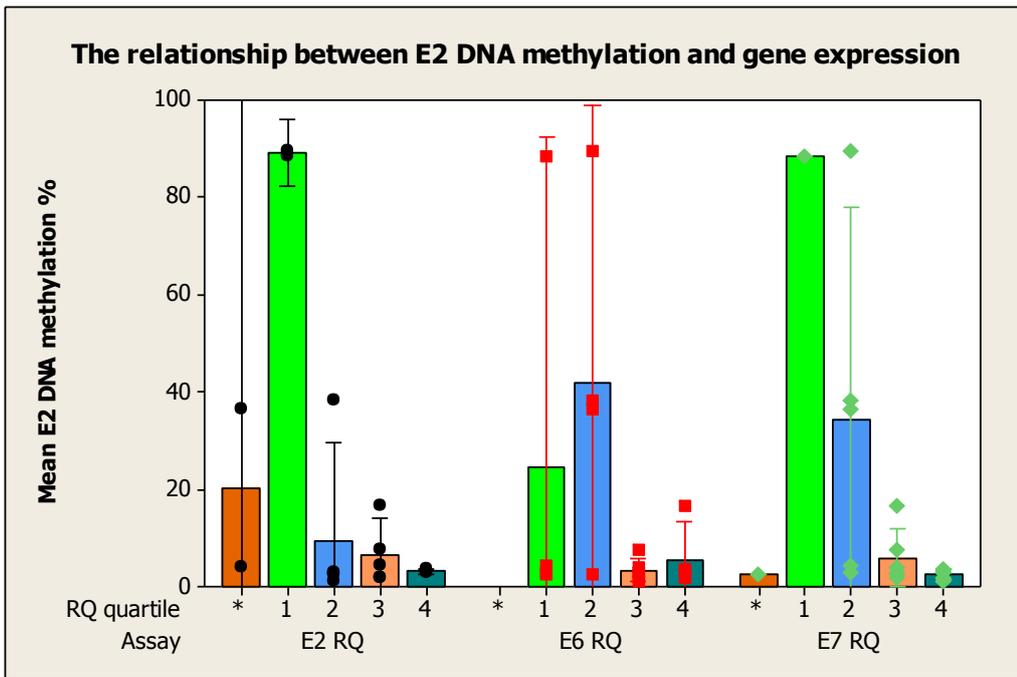
Figure 8.15: DNA methylation of the VR series *E2* region, separated by gene expression assay and by RQ quartile.

* Represents those samples where the gene expression assays have failed to produce an RQ result. Bars represent the mean methylation and 95% confidence intervals are included.
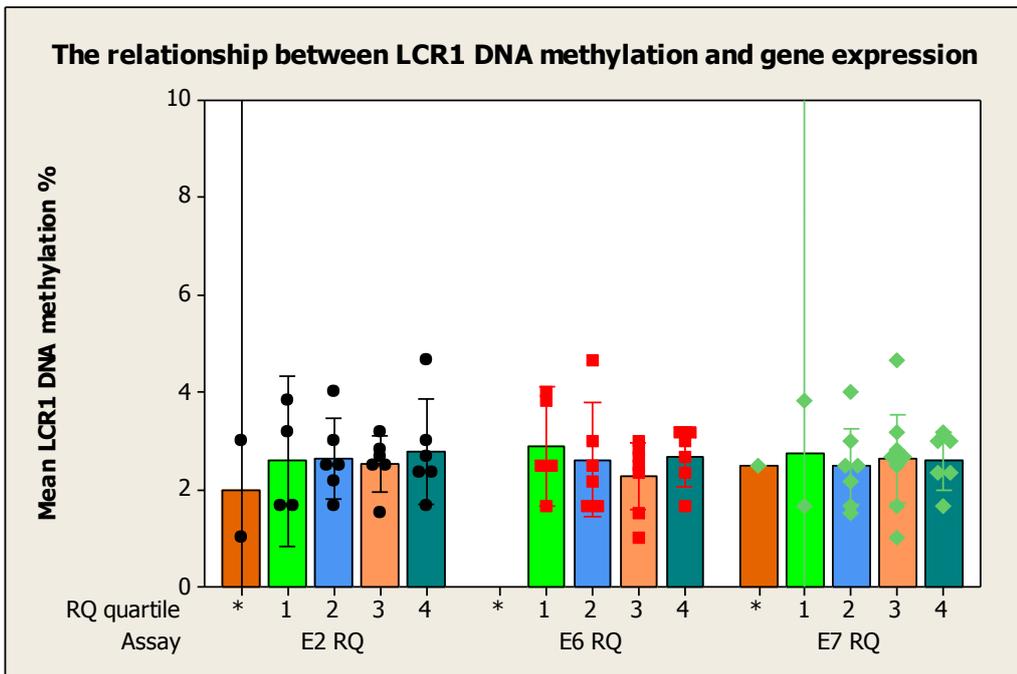
Figure 8.16: DNA methylation of the VR series LCR1 region, separated by gene expression assay and by RQ quartile.

* Represents those samples where the gene expression assays have failed to produce an RQ result. Bars represent the mean methylation and 95% confidence intervals are included.
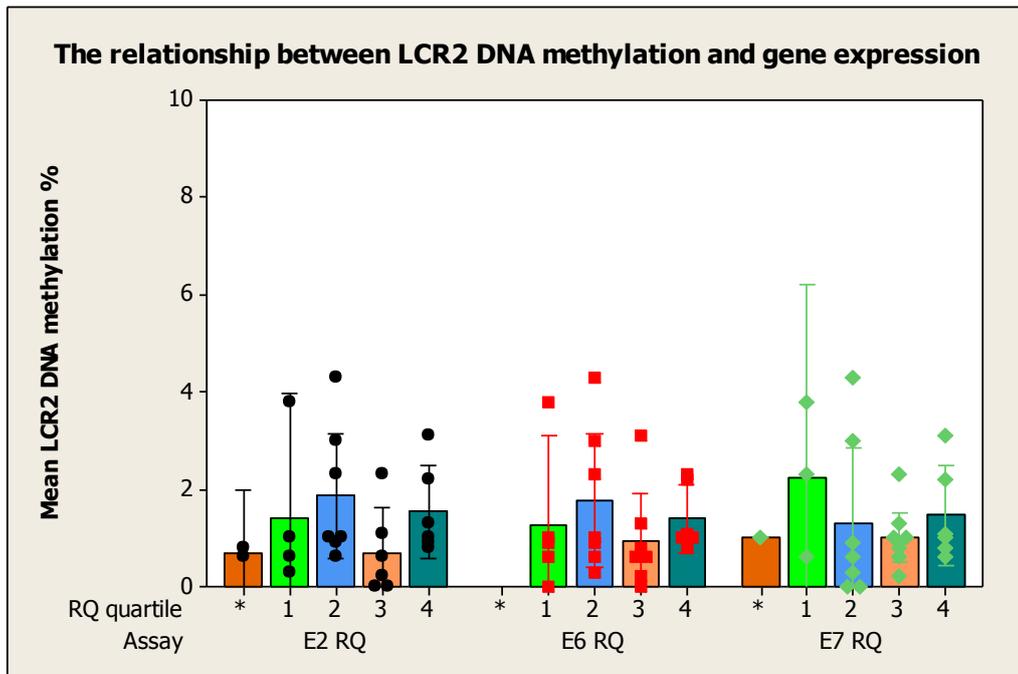
Figure 8.17: DNA methylation of the VR series LCR2 region, separated by gene expression assay and by RQ quartile.

* Represents those samples where the gene expression assays have failed to produce an RQ result. Bars represent the mean methylation and 95% confidence intervals are included.
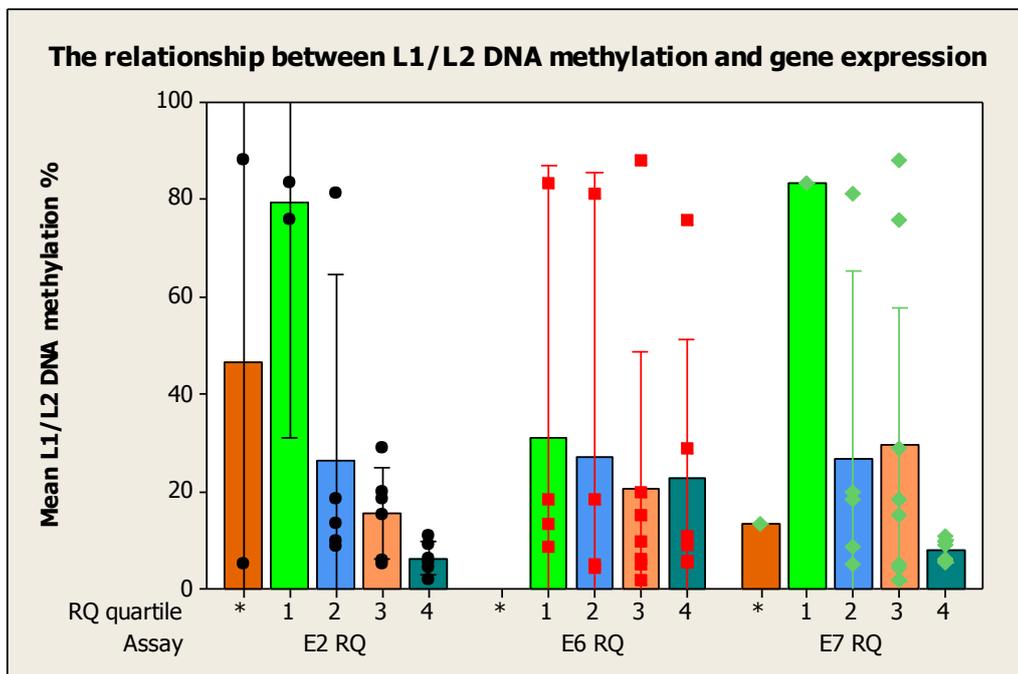
Figure 8.18: DNA methylation of the VR series *L1/L2* region, separated by gene expression assay and by RQ quartile.

* Represents those samples where the gene expression assays have failed to produce an RQ result. Bars represent the mean methylation and 95% confidence intervals are included.

8.2.3.5.1.1. E2 Methylation

For the *E2* region, there appeared to be a relationship between *E2* methylation and RQ quartile for both *E2* RQ and *E7* RQ (Figure 8.15). There was a pronounced decrease in mean methylation between Q1 and Q2/Q3/Q4 for *E2* RQ and a progressive decrease from Q1 to Q4 for *E7* RQ. There were however large confidence intervals for many of the quartile groups so these relationships needed to be confirmed statistically.

For several comparisons by one-way ANOVA, some assumptions of the test were violated. It was felt that using the one-way ANOVA instead of Kruskall-Wallis tests followed by multiple Mann-Whitney U tests would be more informative. Nonetheless, the results were interpreted with caution. Table 8.8 contains the results of one-way ANOVA tests to determine differences in *E2* DNA methylation between the *E2*, *E6* and *E7* RQ quartiles. Violations of the ANOVA are listed in the figure legend. There was a significant difference (significant at Bonferroni corrected P=0.006) in *E2* DNA methylation among the samples in the various *E2* RQ and *E7* RQ quartiles, but not the *E6* RQ quartiles. Pairwise comparisons reinforce the observation, for *E2* RQ, a significant difference was observed between Q1 and Q2/Q3/Q4. Similarly, for *E7* RQ, significant differences in *E2* DNA methylation were observed between Q1 and Q3, Q1 and Q4 and Q2 and Q4; again reinforcing the general decline in *E2* DNA methylation observed with increasing *E7* transcription.

In summary, samples expressing the least *E2* mRNA had a significantly higher level of *E2* DNA methylation which might signify an effect of intragenic methylation. Nonetheless, as *E2* DNA methylation increased, *E7* expression also decreased. The Spearman's rank correlation analysis performed previously found a significant relationship between *E2* DNA methylation and *E7* expression, but not between *E2* DNA methylation and *E2* expression. The relationship between DNA methylation and gene expression appears more complex than a simple linear correlation.

	With integrated samples included *E2* DNA methylation	Significant Tukey-Kramer comparisons (P<0.05)
vs. *E2* RQ	P=0.002[*1]	1 vs. 2, 1 vs. 3, 1 vs. 4
vs. *E6* RQ	P=0.062	None
vs. *E7* RQ	P=0.005[*2]	1 vs. 3, 1 vs. 4, 2 vs. 4

Table 8.8: The results of multiple one-way ANOVA tests with Tukey-Kramer pairwise comparisons.

Violations of the one-way ANOVA assumptions: *1, non-normally distributed residuals (Anderson Darling P=0.008) and unequal variances (Bartlett's P=0.000); *2, unequal variances (Bartlett's P=0.015). Significant results are highlighted yellow.

8.2.3.5.1.2. LCR1 Methylation

There did not appear to be a relationship between RQ quartile and LCR1 DNA methylation. The methylation of samples in different RQ quartiles was consistent and evenly distributed (Figure 8.16). Multiple one-way ANOVAs did not identify any significant results.

8.2.3.5.1.3. LCR2 Methylation

There did not appear to be a relationship between RQ quartile and LCR2 DNA methylation. As per the LCR1 region, DNA methylation of the LCR2 region was consistently low and evenly throughout the various RQ quartile groups (Figure 8.17). Multiple one-way ANOVAs did not identify any significant results.

8.2.3.5.1.4. L1/L2 Methylation

There appeared to be a relationship between *L1/L2* DNA methylation and RQ quartile group for both the *E2* and *E7* RQ (Figure 8.18). With increasing *E2* RQ quartile group, there appeared to be a progressive decline in *L1/L2* methylation. For the *E7* RQ groups, there was a less progressive decrease, with Q2 and Q3 being very similar and with large confidence intervals. Mean *L1/L2* methylation of the samples in *E7* Q4 was the lowest such that samples with the highest *E7* RQ had low level *L1/L2* methylation.

Statistical comparison was performed using multiple one-way ANOVA tests (Table 8.9), one violation of the ANOVA assumptions existed and is detailed in the table legend. There was a significant difference in *L1/L2* methylation (at Bonferroni corrected P=0.017) only among the *E2* RQ quartiles. Pairwise comparisons found a significant difference between Q1 and Q3, and Q1 and Q4. The observed relationship between *E7* RQ and *L1/L2* methylation was not confirmed statistically.

In summary, high *L1/L2* methylation was associated with low *E2* and *E7* RQ. Similarly, low *L1/L2* methylation was associated with high *E2* and *E7* RQ

	With integrated samples included	Significant Tukey-Kramer comparisons (P<0.05)
	L1/L2 DNA methylation	
vs. *E2* RQ	P=0.002	1 vs. 3, 1 vs. 4
vs. *E6* RQ	P=0.853	None
vs. *E7* RQ	P=0.198[*1]	None

Table 8.9: The results of multiple One-way ANOVA tests with Tukey-Kramer pairwise comparisons.

Violations of the one-way ANOVA assumptions: *1, unequal variances (Bartlett's P=0.013). Significant results are highlighted yellow.

8.2.3.5.2. Integrated Samples Excluded Analysis

Next, a similar set was performed for the data with the integrated samples excluded. As most comparisons were non-significant, the data are presented graphically in Appendix V and significant observations are summarised here.

8.2.3.5.2.1. E2 Methylation

For the *E2* region, DNA methylation did not appear to be associated with RQ quartile. Multiple one-way ANOVAs did not identify any significant results.

8.2.3.5.2.2. LCR1 Methylation

There did not appear to be a relationship between RQ quartile and LCR1 DNA methylation. Multiple one-way ANOVAs did not identify any significant results.

8.2.3.5.2.3. LCR2 Methylation

There was a consistently low and even distribution of DNA methylation level between the RQ quartiles. One significant result was obtained (although not when considering the Bonferroni correction), LCR2 methylation varied with *E2* RQ (P=0.039) and Tukey-Kramer identified Q1 vs. Q3 as significantly different. However there was a very low sample number in *E2* RQ Q1 and the level of DNA methylation in each group was very low increasing the likelihood of error as a result of background experimental noise. This observation must be treated with scepticism unless further evidence is produced. No other significant differences were observed.

8.2.3.5.2.4. L1/L2 Methylation

There was a varied level of *L1/L2* DNA methylation for each *E6* RQ quartile. There was considerably less variation in the *L1/L2* DNA methylation for the highest *E2* and *E7* expressing samples (i.e. Q4 for both assays). As per the data set with integrated samples included, there was a significant difference (although not when considering the Bonferroni correction) in *L1/L2* DNA methylation among *E2* RQ quartiles (P=0.025), more specifically a pairwise difference between Q3 and Q4 (Tukey-Kramer). The relationship was less convincing than when analysing all samples due to the sample number and lower significance. Nonetheless, the highest *E2* expressing samples had a lower mean *L1/L2* methylation than those with lower *E2* RQ. The *L1/L2* DNA methylation of the other *E2* RQ quartiles was quite variable. No other significant differences were observed.

8.2.3.5.3. Integrated Samples, Individual Consideration

For the integrated samples, a complex relationship appeared to exist between methylation and gene expression that necessitated interpretation independent of the non-integrated samples. The low number of integrated samples and the diverse methylation/mRNA levels made

statistical analysis difficult and comparisons of means/medians problematic. Instead, each sample was considered and compared to the other samples separately in an attempt to investigate the affects of different integration events. The following sample descriptions are compiled from the data presented in Figure 8.4, Table 8.4 and Figure 8.19.
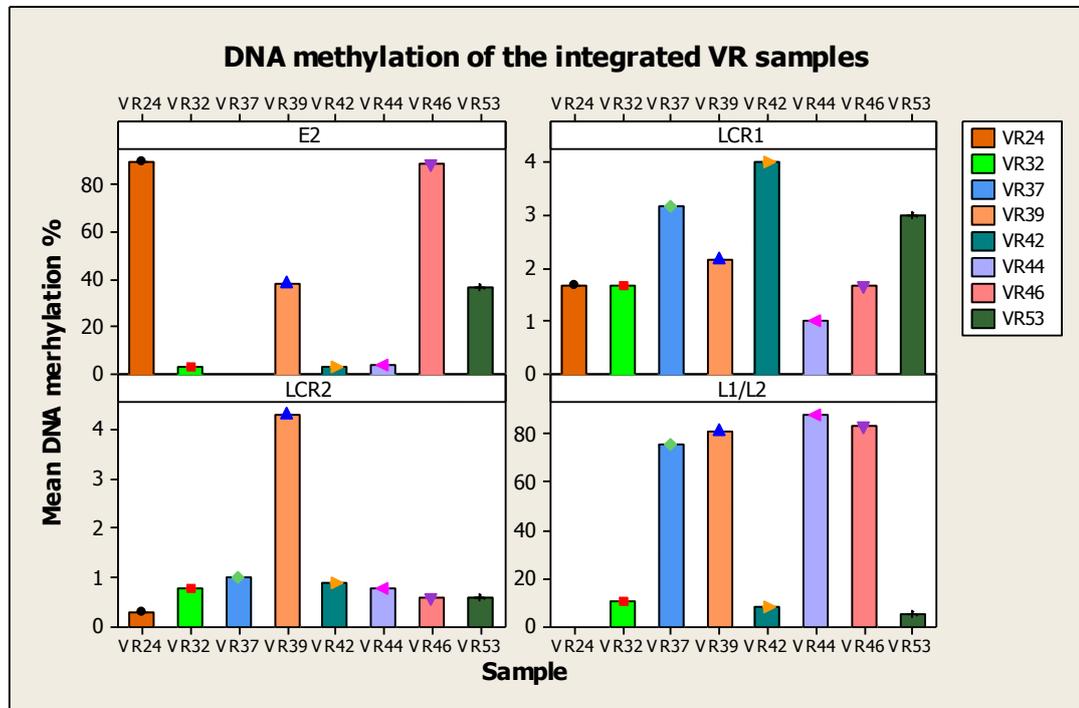
Figure 8.19: Mean DNA methylation of the VR series integration group 1 samples.

A different graph pane is present for each HPV genome region and each bar represents mean DNA methylation for that sample.

8.2.3.5.4. VR24

The VIN3 VR24 sample contained HPV16 integrated in the *E1* gene; there was hypermethylation of the *E2* region however the *L1/L2* data was not available to compare. There was hypomethylation of both the LCR1 and LCR2 regions. VR24 exhibited an average *E6* RQ, an above average *E7* RQ and a very low *E2* RQ.

8.2.3.5.5. VR32

The VIN3 VR32 sample, much like the VR42 sample was unusual in that there was no hypermethylation of *E2* or *L1/L2* observed despite a detected integration event. The VR32 sample had a detected HVP16 integration event that resulted in disruption of the *E2* gene. VR32 had very very high *E2* and *E6* RQ and a very high *E7* RQ, several Log_{10} higher than average in every instance. Both LCR1 and LCR2 were hypomethylated. A possible explanation for the integration without hypermethylation at *E2* and/or *L1/L2* was that the HPV16 genomes in this

213

sample could exist as both an integrated form, and a highly numerous episomal form. If this were the case, the small numbers of integrated sequences were detectable by DIPS but episomal forms were still present. Episomal HPV are often maintained in very high numbers and are relatively hypomethylated (Kalantari *et al.*, 2008a, Doorbar, 2006), the quantitative pyrosequencing assay would detect methylation as an average of each HPV16 genome in the sample, and an overabundance of unmethylated episomal HPV16 genomes would essentially mask any hypermethylated integrated HVP16 sequences.

8.2.3.5.6. VR37
The VIN3 VR37 sample contained HPV16 genomes with integration sites in two regions, the *E2* and *L1/L2* regions which likely represented a deletion. There was no data for the *E2* methylation assay as it failed repeatedly however methylation of the *L1/L2* region was high. The *E2* RQ was very low, several Log_{10} lower than the average whereas the *E6* and *E7* RQ were considered above average and high respectively. The LCR1 and LCR2 regions were largely unmethylated. In other samples an association was observed between site of integration and site of hypomethylation.

8.2.3.5.7. VR39
The VIN3 VR39 sample had a detected integration event in the *E1* region of the HPV16 genome. It had a relatively high level of methylation in the *E2* region and a high level of methylation in the *L1/L2* region. Both LCR1 and LCR2 were hypomethylated. The *E6* RQ was close to the average amount of expression, whereas the *E2* and *E7* RQ were above average.

8.2.3.5.8. VR42
The VIN3 VR42 sample had a detected integration event in the *L2* region. The HPV16 genome was hypomethylated in every tested region, including the *E2* and *L1/L2* regions. There was below average *E6* RQ, slightly above average *E7* RQ and high *E2* RQ. Again, a possible explanation for the integration status and the observed hypomethylation of *E2* and *L1/L2* was that both episomal and integrated HPV were present and the numerous, unmethylated episomal forms masked the methylation status of the integrated genomes.

8.2.3.5.9. VR44
The VIN3 VR44 sample was an interesting case. The *E2* qPCR assay failed repeatedly however the *E2* pyrosequencing assay worked well. The *E6* and *E7* RQ for VR46 were average and above average respectively. The HPV16 integration event into the human genome resulted in *E2* disruption. *E2* was also hypomethylated whereas interestingly, the HPV16 *L1/L2* region was highly methylated.

214

8.2.3.5.10. VR46

The VAIN3 sample VR46 had both *E2* and *L1/L2* hypermethylation. HPV16 was integrated at the *E1* region. mRNA level of the *E6* and *E7* genes were below average, however mRNA level of the *E2* gene was very very low. The down regulation of *E2* expression did not appear to be a result of E2BS (i.e. LCR2) methylation (very lowly methylated) but may be associated with *E2* and *L1/L2* hypermethylation. The HPV16 integration event occurs within the MLT1D LTR sequence (Chr 9, q21.31) which may have an effect on the methylation profile of the integrated HPV16 DNA.

8.2.3.5.11. VR53

For VIN3 sample VR53, there was a low level of *E6/E7* transcription (*E2* RQ data missing). The *L1/L2* region was hypomethylated whilst the *E2* region was more methylated than all non-integrated samples which might be related to the integration event in the *L2* region of the HPV16 genome. As per all other samples, the LCR was hypomethylated.

8.3. Discussion - HPV Distribution in VIN

This study used the Greiner Bio-One PapilloCheck DNA chip assay to identify HPV types present in histologically confirmed VIN. There were two aims; firstly to identify HPV16 positive samples that could be used for further molecular analysis, and secondly to determine the proportion of VIN potentially preventable by HPV prophylactic vaccination.

8.3.1.1. Strengths and Weaknesses

The main limitation of the study was the use of a self selected sample, i.e. women attending a specialist VIN clinic, which could bias the sample towards women with persistent or recurrent disease; this might result in some over-representation of more persistent types (e.g. HPV16). A further limitation was the use of separate biopsies for pathology and HPV analysis, in practice the biopsies were immediately adjacent and cases were only included if both biopsies were part of the same macroscopically visible lesion. It is however theoretically possible that the research biopsy may not be representative of the diagnostic biopsy.

The strengths of the study were the large sample size, the use of histologically confirmed VIN, the availability of patient age data and the use of the accurate and increasingly common PapilloCheck DNA microarray assay.

8.3.2. Meaning and Comparison to Other Studies

This is the largest UK series of VIN in which HPV type has been determined. Several studies have investigated HPV prevalence in VIN in the UK and estimated HPV positivity at 37.5–100%, with the proportion of HPV positive cases containing HPV16 varying from 37.5% to 93.3% (Abdel-Hady *et al.*, 2001, Gasco *et al.*, 2002, Todd *et al.*, 2002, Baldwin *et al.*, 2003, Tristram and Fiander, 2005, Fiander *et al.*, 2006, Woo *et al.*, 2007, Daayana *et al.*, 2010). The sample size in previous studies ranged from 8-29 cases. At n=57, the sample size for this study was much larger. HPV typing data from 57 samples were used and the findings are more consistent with the overall results of a recent international meta-analysis, in which 84.0% of VIN cases tested HPV positive (67.5% were HPV16, 7.7% were HPV33, and 4.6% were HPV18) (de Vuyst *et al.*, 2009). For this study, the corresponding proportions were overall HPV positivity of 75.4% (59.6% were positive for HPV16, 8.8% for HPV33, and 3.5% for HPV18). These proportions are also similar to another meta-analysis which indicated an overall HPV positivity of 80.4% for VIN2/3 (HPV16 present in 71.2%, HPV33 in 7.7%, and HPV18 in 5.5% (Smith *et al.*, 2009)).

HPV positive patients were significantly younger than those testing negative (mean difference of over 10 years). This is consistent with HPV-associated VIN being more common in younger

women and VIN associated with LS being more common in older women (Bonvicini *et al.*, 2005).

Three cases with VIN grade 2 or higher were associated with an LrHPV type alone (HPV6) which appears inconsistent with the suggestion that LrHPV types cause only warts and low grade lesions. However an association between LrHPV types and high-grade disease was observed in the De Vuyst meta analysis (de Vuyst *et al.*, 2009) which reported approximately 5% of VIN2/3 as linked to HPV6.

HPV-associated VIN is most common in women in their 30's and 40's (Hart, 2001). The incidence of VIN is approximately 5 per 100,000 women per year and is increasing (Joura, 2002). HPV infection is endemic in the UK, in Wales almost 30% of women in the 20–25 year age range are now infected with a HR anogenital HPV type (Hibbitts *et al.*, 2008). Hence, in the short term, the number of women affected by VIN is likely to continue increasing. HPV vaccination has been shown to be 100% effective in prevention of VIN2/3 associated with HPV16/18 in a per protocol population (Joura *et al.*, 2007). In the UK, vaccination of 12–13 year old girls against HPV16/18 infection began in 2008. Hence it is likely to be 20–30 years before the effects of this intervention become apparent in reduced incidence of VIN. Ultimately however, as HPV16/18 were present in 61.4% of VIN, our data suggest that greater than half of VIN cases could be potentially prevented by vaccination and that vaccination with a quadrivalent vaccine (protection conferred against types 16, 18, 6 and 11) might also prevent HPV6 associated VIN.

8.4. Discussion – HPV DNA Methylation, Integration and Gene Expression

8.4.1.1. Aims

This study was devised to characterise a series of VIN2/3 biopsies and to test a series of hypotheses. The primary aim was to establish the relationship between HPV methylation, integration and gene expression. We hypothesised that where integration was not present, DNA methylation might constitute an alternative transforming event, either by hypermethylation of one or more regions of the HPV genome or by specific hypermethylation directed at the E2BSs. An alternative hypothesis was that DNA methylation might be a consequence of integration.

8.4.1.2. Weaknesses

A limitation of DIPS is that when integration is not detected, it does not necessarily mean that an integration event is not present (i.e. a true negative result). A negative result could arise if there genuinely was not an integration event present, because the method was not sensitive enough or because it had failed to work properly. In addition, DIPS cannot differentiate between HPV infections where a mix of integrated and episomal HPV forms are present and infections where concatemeric HPV integrants are present.

For the qPCR assays, the biopsy dosage effect (Chapter 4: 4.1.3) could influence the RQ ratios produced. The biopsy dosage effect can essentially cause apparent differences in RQ ratio that are not the result of differences in mRNA level.

The assumptions of several statistical analyses were violated in several instances. Where violations occurred, these are detailed in the text. Sometimes these data were used on the basis that the test produced data confirmed by other tests. It is common practice to use the tests with violations provided the results are interpreted with appropriate caution. Furthermore the results of these tests were used primarily in an exploratory nature to generate hypotheses. Where the data was wholly unsuitable for an analysis (i.e. the distribution of the RQ data was unsuitable for parametric multifactorial ANOVA) less powerful and less informative non-parametric alternatives were used.

8.4.1.3. Strengths

This study is unique in being the only study to compare DNA methylation, gene expression and integration status in VIN. Histology was reviewed by a consultant pathologist specialising in gynaecological oncology. The VR series comprise a clinically relevant cohort for which both

DNA and RNA were available. The study was designed to investigate the basic biology of HPV gene expression and to assess the potential of DNA methylation as a biomarker. From this perspective, the VR series was a well designed study which allowed the assessment of several biological characteristics.

Use of pyrosequencing was a significant strength as it provides an accurate assessment of methylation status for numerous CpGs in a heterogeneous mix of DNAs, in a fully quantitative manner. The pyrosequencing assays used in this study are to be used as an adjunct to an ongoing clinical trial of topical therapy for VIN (RT3 VIN; CRUK trial number CRUK/06/024).

8.4.2. Meaning and Comparison to Other Studies

8.4.2.1. DNA Methylation of the VR Series

DNA methylation varied considerably among samples, among HPV regions and among CpGs within each region. *L1/L2* was the most heavily methylated region although *E2* hypermethylation was observed in several samples. The LCR1 (enhancer) and LCR2 (promoter) were hypomethylated in all samples. The patterns of CpG specific methylation were very similar to those observed in other sample cohorts and cell line studies (Chapter 5: 5.1.3, Chapter 6: 6.1.1, and Chapter 7: 7.1.3.3.2) and very large differences were observed between CpGs separated by only a few bp. The region and CpG specific methylation patterns observed are discussed in further detail in the general discussion (Chapter 9: 9.1.4 & 9.1.5).

DNA methylation of the VR series (predominantly VIN2/3) was similar to the level of methylation observed for the severe dyskaryosis group (Chapter 5: 5.1.1) and for CIN2/3 (Fernandez *et al.*, 2009). HPV DNA methylation of other non-cervical, anogenital sites has been assessed (Kalantari *et al.*, 2008b, Wiley *et al.*, 2005) however this is the first instance of HPV DNA methylation being determined for VIN. The results for HPV methylation studies of other anogenital sites are generally consistent with cervical lesion studies, thus expanding the potential clinical utility that a knowledge of HPV methylation might have.

8.4.2.2. Integration Status of the VR Series

Integration was detected in 8 samples with a variety of human and HPV loci disrupted. Considering only samples where the sample controls were passed, 6/15 (40%) of VIN2/3 samples had integration events; this is consistent with a previously estimated rate of 38.1% integration of HP16/HPV18 in VIN3 (Hillemanns and Wang, 2006). HPV integration in CIN has been reported in 11-60% of cases (Hafner *et al.*, 2008, Klaes *et al.*, 1999, Hopman *et al.*, 2004, Li *et al.*, 2012). Variation in the estimates may be a result of methodological differences and

differences in sample size and CIN grade. VR37 DIPS data appeared to show a deletion between the *E2* and *L2* regions. Interestingly, *E2* pyrosequencing PCRs failed whilst *E2* mRNA was detected by qPCR. Both pyrosequencing and qPCR targeted regions inside the potentially deleted zone which could hint at a low level presence of intact genomes and transcripts, and a difference in sensitivity of the pyrosequencing and qPCR assays such that qPCR but not pyrosequencing were capable of amplifying the *E2* region of these genomes.

8.4.2.3. Gene Expressions Status of the VR Series

E2, E6 and *E7* mRNA level varied considerably among samples and among the qPCR assays, the most variable mRNA was *E2*. E2 is a repressor of oncogene expression, and several factors (e.g. *E2* disruption, E2BS methylation, non-functional E2 proteins, splice variation etc) may affect its expression and account for the greater variability (Zheng and Baker, 2006). *E2/E6/E7* mRNA levels were positively correlated which was consistent with the PC08/PC09 data. The strong correlation between VR series *E2*, *E6* and *E7* RQ suggests that high/low transcription of one mRNA is accompanied by similar changes in the other mRNAs. However there were several exceptions to this observation where *E2* RQ was much lower than *E6/E7* RQ. The disparity might be explained by observations made on the W12 model, where *E2* downregulation and *E6/E7* upregulation were considered to be early events in cellular transformation and *E2/E6/E7* mRNA were present at similar levels after these early events (Gray *et al.*, 2010).

Biopsy samples are subject to the biopsy dosage effect (Chapter 4: 4.1.3). The biopsy dosage effect is caused by variation in the ratio of HPV containing versus non-HPV containing cells in each biopsy. Both cell types contain reference gene transcripts however only HPV positive cells contain HPV transcripts which leads to disparity during relative quantification. This variation could conceal differences in *E2/E6/E7* mRNA level among biopsies and promote an apparent *E2/E6/E7* correlation. However correlation between *E2/E6/E7* was also observed in PC08/PC09 which were not subject to this effect.

8.4.2.3.1. Patient Age

An association between age and integration status was observed previously (Vinokurova *et al.*, 2008) but was not apparent in our study, possibly due to the relatively small number of samples.

8.4.2.4. Gene Expression and Integration

It is interesting to note that as well as a low *E6/E7*:*E2* RQ ratio, samples VR37, VR24 and VR46 had integration events detected. Nonetheless, no significant difference in RQ was observed between the integrated and non-integrated samples when tested directly. This was consistent with one study (Hafner *et al.*, 2008), but not with another (Ho *et al.*, 2011). The lack of

correlation might correspond to the inability of DIPS to differentiate between different types of integration event which may exert distinct effects on gene expression. For instance differences in gene expression may exist between tissues containing only integrated genomes as opposed to both integrated and episomal HPV forms, and differences between concatemeric repeats and single insertions might also affect transcription.

An association between site of integration and *E2/E6/E7* level was considered but no relationship was apparent. VR32 was particularly interesting as it showed *E2* disruption but also the highest levels of *E2/E6/E7* mRNA. *E2* was disrupted just downstream of the site amplified by qPCR, hence full length E2 protein could not be produced by the integrated genome (Kim *et al.*, 2003), and *E2* transcripts would not correspond to functional E2 protein. *E6* and *E7* could be highly expressed without being repressed by E2.

8.4.2.5. DNA Methylation and Integration

Integration of HPV into the host genome with associated disruption of the *E2* ORF is commonly considered to be a mechanism of cellular transformation and carcinogenesis (Thierry, 2009). However, a substantial proportion of HPV-associated neoplasias do not contain integrated HPV or also contain episomal HPV such that functional E2 repressor may be produced (Arias-Pulido *et al.*, 2006, Klaes *et al.*, 1999, Pett and Coleman, 2007, Vinokurova *et al.*, 2008, Bhattacharjee and Sengupta, 2006b). As such, it has been suggested that integration of HPV into the host genome could be a consequence of chromosomal instability caused by aberrant HPV oncogene expression, rather than the cause of aberrant oncogene expression (Pett and Coleman, 2007, Melsheimer *et al.*, 2004). There is reason to suspect that HPV DNA methylation might be involved in initial oncogene deregulation in some instances (Vinokurova and von Knebel Doeberitz, 2011).

Our initial hypothesis was that DNA hypermethylation could be an alternative transforming event to integration. If this were true, one might expect a characteristic level or pattern of hypermethylation in non-integrated samples that was not present in integrated samples, i.e. the transforming methylation pattern and the presence of integration would be mutually exclusive.

For the VR series, this relationship was not present and hypermethylation of *E2* and *L1/L2* was strongly and significantly associated with HPV integration. We also suspected that E2BS methylation in the LCR might constitute an alternative transforming event by blocking binding of the E2 repressor proteins and enhancing P97 activity, however LCR hypermethylation was not observed in any instance (Thain *et al.*, 1996, Kim *et al.*, 2003). In fact, the highest level of

LCR methylation (still <10%) was observed for an integrated sample (VR42). LCR methylation of this sample might instead reflect the type of integration event (single insertion vs. concatemeric) as multiple insertions appear more likely to be silenced by dense LCR methylation (Kalantari *et al.*, 2009).

Rather than being mutually exclusive with integration, hypermethylation of the HPV genome could be specifically associated with integration. Integrated samples were significantly more methylated in the *E2* and/or *L1/L2* regions than the non-integrated samples. There were two exceptions to this, VR32 and VR42 (*E2* and *L1/L2* hypomethylation) may have had HPV in both integrated and episomal forms. Episomal HPV16s are typically maintained at high copy number and are hypomethylated (Kalantari *et al.*, 2008a, Doorbar, 2006). The pyrosequencing assays would detect methylation as an average of each HPV16 genome in the sample, and an overabundance of unmethylated episomal HPV16 genomes could potentially mask any hypermethylated integrated HVP16 sequences. VR42 also had the highest level of LCR1 methylation. LCR1 methylation of the VR42 integrated genomes could be considerably higher than indicated by the pyrosequencing assay but masked by the presence of episomal HPV, such that these integrated genomes were repressed (Rosl *et al.*, 1993). Alternatively, episomal HPV genomes could be effectively targeted for LCR methylation which may or may not affect P97 activity (Bechtold *et al.*, 2003, Kim *et al.*, 2003).

The association between DNA methylation and integration appeared to be complex. Some samples showed both *E2* and *L1/L2* hypermethylation, some had *E2* or *L1/L2* hypermethylation and others had *E2* and *L1/L2* hypomethylation. It appeared that an integration event in a particular HPV region could be associated with hypomethylation in adjacent regions and hypermethylation in distal regions. *E1* disruption for instance was typically associated with hypermethylation of both *E2* and *L1/L2*, *E2* disruption was associated with *L1/L2* hypermethylation and *L2* disruption with *E2* hypermethylation. As this was also observed in the cell lines study, it is discussed in further detail in the general discussion (Chapter 9: 9.1.1.1).

8.4.2.6. DNA Methylation and Gene Expression
Because DNA methylation in the integrated samples was different from the non-integrated samples, the relationship of DNA methylation with mRNA level was considered both with and without integrated samples included.

8.4.2.6.1. E2 Intragenic Methylation and E2 Expression

Initially, the *E2* pyrosequencing assay was adopted on the basis that intragenic methylation can affect gene transcription (Lorincz *et al.*, 2004, Murrell *et al.*, 2001), and so intragenic methylation might reduce *E2* transcription. The samples with the least *E2* mRNA showed significantly higher *E2* DNA methylation. A negative correlation was also observed between *E2* methylation and *E2* RQ for the integrated samples. Because *E2*/*E6*/*E7* RQ were positively correlated, the relationship between *E2* methylation and *E6*/*E7* RQ was very similar to *E2* RQ (albeit *E6* was not quite significant). This data is consistent with *E2* methylation contributing to regulation of *E2* expression. This is also discussed in further detail in the general discussion (Chapter 9: 9.1.3 & 9.1.4).

For the non-integrated samples, variation in *E2* methylation was limited and no relationship between *E2* methylation and transcription was observed. Intragenic methylation affects transcription by hindering elongation efficiency of RNA polymerase II in regions of dense methylation and condensed chromatin structure (Lorincz *et al.*, 2004). Intragenic methylation might also affect transcription from alternative start sites (Maunakea *et al.*, 2010), which has particular relevance for HPV where alternative transcription start sites are recognised (Lambert *et al.*, 1990) and various *E2* start sites are responsible for additional *E2* ORF proteins (Introduction). Whether this form of repression would be apparent for an mRNA transcript based assay (i.e. qPCR) where an early part of the *E2* ORF is amplified, and for HPV where post-transcriptional regulation is important is difficult to determine.

8.4.2.6.2. LCR Methylation and Gene Expression

We did not see a clear relationship between LCR methylation and gene expression. We had initially hypothesised a relationship between LCR hypermethylation and transcription in the absence of integration, however this was not observed and the LCR of all samples was hypomethylated. LCR hypomethylation in VIN is consistent with the PC08/09 data (Chapter 7: 7.1.3.3) and also cervical disease grade studies (Badal *et al.*, 2003, Hublarova *et al.*, 2009, Patel *et al.*, 2012, Piyathilake *et al.*, 2011).

8.4.2.6.3. L1/L2 Methylation and Gene Expression

A significant association between *L1*/*L2* methylation and *E2* RQ was observed. No relationship was observed for *E6* RQ, and the *E7* RQ relationship was deemed non-significant. Similar to *E2* methylation, the samples with the lowest levels of *E2* mRNA had the highest levels of *L1*/*L2* methylation.

That *L1*/*L2* methylation is also associated with *E2* transcription makes specific *E2* repression by intragenic methylation less likely. These data would be more consistent with a model in which

223

low levels of *E2* transcription are related to integration events and these integration events are characterised by high *E2* and/or *L1/L2* methylation. With this in mind, it is perhaps surprising that we did not observe a significant difference in *E2/E6/E7* mRNA level between the integrated and non-integrated HPV containing samples in the previous analysis. This might stem from the lack of methylation data for several of the lowest mRNA expressing samples such that they were not included in this analysis of methylation and gene expression, but were included in the previous analysis just concerning gene expression and integration. Adequate DNA was retrieved from these biopsies, however if HPV genomes were only present at very low copy number within the biopsies, the less sensitive DNA based pyrosequencing assays might fail whilst the highly sensitive qPCR assays might not. Relative quantification normalises against input RNA amount, but the biopsy dosage effect (Chapter 4: 4.1.3) could artificially lower the RQ returned if a low number of HPV genomes were present in the biopsy.

8.4.2.6.4. Summary

In summary, *L1/L2* and *E2* hypermethylation and/or HPV integration appeared to be associated with low levels of *E2/E6/E7* transcription.

Whilst there was no clear difference in mRNA level between integrated and non-integrated samples, this observation was complicated by the apparent hypermethylation of integrated HPV genomes and the observation of these hypermethylated genomes typically having lower mRNA levels. Lower level of mRNA for *E2*, *E6* and *E7* in the integrated samples was unexpected but might be explained by the requirement of *E6/E7* expression for the replication of episomal HPV genomes, but not the maintenance of integrated HPV that are replicated alongside the host genome (Pett *et al.*, 2006). Alternatively, low level gene expression for hypermethylated integrated genomes might simply be a by-product of gene dosage; episomal HPV are typically maintained in high numbers, whilst integrated genomes are frequently integrated as single insertions. Indeed, the methylation profiles of our samples are more consistent with single insertions (Kalantari *et al.*, 2009). Finally, only a small number of integrated samples were considered, there was variation in integration site (both HPV and human), potential biopsy dosage effects, and small violations of the analysis; together these factors could produce erroneous correlations.

8.5. Conclusions

In conclusion, there was considerable variation in integration state, gene transcription and DNA methylation among the VR series. There was also a pattern of CpG specific methylation reminiscent of other sample cohorts.

The hypothesis that DNA methylation is an alternative transforming event to integration was not supported by the data. Neither was E2BS methylation specifically observed in non-integrated samples, instead the LCR was hypomethylated in all instances. The biggest determinant of HPV methylation appeared to be HPV integration whereby most integrated HPV were hypermethylated in *E2* and/or *L1/L2*. There was also a potential relationship between HPV site of disruption by integration and site of hypomethylation. Whether HPV methylation is a cause or a consequence of HPV integration remains unclear.

There was not a clear difference in HPV gene transcription between samples containing integrated and non-integrated HPV genomes, however there was a suggestion of lower transcription in integrated samples which was difficult to explain. There was an association between hypermethylation and low gene transcription however the role of intragenic methylation in HPV transcription also remains unclear.

8.5.1. Further Work

In samples where integration was identified, determination of the methylation status of adjacent human DNA might provide further insight into the mechanism of methylation of the integrated HPV genome. Analysis of the type of integration event (single or concatemeric insertion) and the detection of episomal HPV in combination with integrated HPV might also help to clarify the apparent association between integration and DNA methylation.

A tissue biopsy is frequently a heterogeneous sample of cells, comprising of cells with differences in viral copy number, HPV integration, gene expression, DNA methylation, differentiation status and a plethora of other characteristics (Vinokurova and von Knebel Doeberitz, 2011, Kalantari *et al.*, 2009, Kalantari *et al.*, 2008a). When considering the limited amount of data published with regards to a potential link between HPV integration status and DNA methylation, it would be inherently interesting to study individual or small groups of cells within a tissue type. LCM allows individual and small numbers of cells to be isolated from a heterogeneous tissue by transferring cells to a transparent polymer film using a laser (Emmert-Buck *et al.*, 1996). DNA, RNA and proteins from LCM cells can then be extracted and analysed (Edwards, 2007). Whilst LCM has been used for HPV methylation studies with varying degrees of success, the limited amount of material available for analysis following LCM would likely prove problematic for DNA methylation assays, but especially for integration assays that require a large amount of nucleic acid (Vinokurova and von Knebel Doeberitz, 2011, Kalantari *et al.*, 2009).

Chapter 9. General Discussion

Several components of this investigation were common to and/or had broader relevance to multiple chapters. In order to avoid repetition and provide a more succinct discussion, these common components are discussed in this general discussion chapter.

Briefly, this discussion will consider the association between HPV methylation and integration, the potential self methylation of HPV and the observation that HPV methylation might not be associated with gene expression in some instances. There is also discussion of the potential mechanisms of methylation that result in regional and CpG specific differences and further discussion of the potential effects of HPV methylation.

Finally, several chapters determined methylation differences to be associated with clinical characteristics. This discussion therefore covers the broader relevance of these observations from the perspective of disease biomarkers.

9.1. DNA Methylation

Pyrosequencing was successfully applied to a range of sample DNAs including cell lines, LBC samples, biopsy samples and FFPE cervical cancers. It proved to be a versatile technique and benefitted significantly from an objective series of internal controls.

9.1.1. HPV Methylation was Associated with HPV Integration

For this study, integrated HPV genomes were typically hypermethylated at *E2* and/or *L1/L2* both *in vitro* and *in vivo*. HPV does not encode DNA methyltransferases (DNMTs) so methylation must be performed by human DNMTs. It is difficult to determine whether the observed levels of HPV DNA methylation are specifically a consequence of the action of HPV proteins or the result of more general DNA methylation changes in human DNA methylation that are associated with neoplasia (Henken *et al.*, 2007, Sun *et al.*, 2012, Eijsink *et al.*, 2012, Wentzensen *et al.*, 2009, Leonard *et al.*, 2012, Apostolidou *et al.*, 2009).

Human cells actively target foreign DNA elements for silencing by DNA methylation (Doerfler, 2006, Doerfler, 2008, Doerfler *et al.*, 2001). Upon integration, foreign DNAs including adenoviral, plasmid and bacteriophage DNAs are often *de novo* methylated by what might be a cellular defence mechanism to repress foreign DNA elements (Sutter *et al.*, 1978, Orend *et al.*, 1995, Heller *et al.*, 1995, Doerfler, 2008). Methylation of foreign DNAs seems to depend on the location of integration in the human genome in terms of local methylation density, methylation pattern and the strength of any foreign promoters (Doerfler, 2008). In some

226

instances DNA methylation of foreign elements could be a reflection of the DNA methylation status of regions adjacent to the integrant, where methylation could begin as an initial seed before spreading to adjacent regions (Heller *et al.*, 1995, Turker, 2002, Doerfler *et al.*, 2001). Unfortunately, we were unable to pursue assessment of the DNA methylation status of flanking human DNAs.

Methylation of HPV integrants may also depend on the type of integration event, i.e. single or concatemeric integrants. A densely methylated, single insertion integrant would be effectively repressed to the extent that no selective growth advantage of the cell would be gained, and cells containing unmethylated HPV integrants would clonally expand (Van Tine *et al.*, 2004b). For concatemeric insertions, dense methylation of most genomes could occur both as a result of targeting of foreign DNAs by host DNA methylation machinery and/or viral mechanisms that limit a copy number related gene dosage effect that is deleterious to the growth rate of the host cell (De-Castro Arce *et al.*, 2011, Doerfler, 2008). Tandem repeats may also be better recognised as genomic parasites and more efficiently targeted by DNA methylation. There are however instances of unmethylated concatemeric W12 HPV16 genomes (Kalantari *et al.*, 2008a, Yoder *et al.*, 1997).

9.1.1.1. *Site specific Hypomethylation and Site of Integration*

Whilst hypermethylation of integrated DNAs was a commonly observed event, there may also have been an association between site of disruption of the HPV genome, and the site of hypomethylation of the HPV genome. For many of the integrated VR series, but also for several of the PC08/PC09 clones, regions of the HPV genome adjacent to sites of the HPV genome disrupted by integration were frequently hypomethylated. Areas distal to the disruption site were typically hypermethylated. Furthermore in W12Ser7 we observed *L1*/*L2* but not *E2* hypermethylation.

This regional hypomethylation adjacent to the disruption site is the first report of such an observation for HPV. Nonetheless, hypomethylation of the ends of adenoviral integrants has been described previously (Hochstein *et al.*, 2007) and can be seen in Figure 9.1. Doerfler speculated that a highly specific and localised, perhaps sequence dependent genome structure might be involved in topologically denying the methylation machinery access to the ends of the integrant (Doerfler, 2011). HPV *L1* methylation has been described as a potential biomarker for integrated HPV (Kalantari *et al.*, 2010). Our data, if confirmed in a larger sample set would suggest that *L1* methylation alone would be a poor assessment of integration status

for HPV disrupted at or near to *L1,* and that assessment methylation of multiple regions would be more informative.
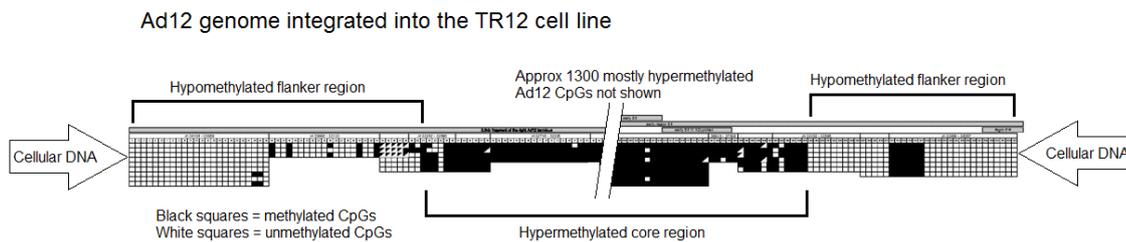
Figure 9.1: The methylation profile of integrated Ad12 adenovirus genome.

Adapted from (Hochstein *et al.*, 2007). Black squares represent methylated CpGs whilst white squares represent unmethylated CpGs. Each column of squares represents a CpG whilst each row represents an individual bisulfite sequenced cloned DNA. The large core region represents most of the Ad12 genome and apart from some exceptions is densely methylated throughout. The two flanker regions represent DNAs adjacent to host cell DNA after the integration event and are typically hypomethylated.

Hypomethylation of host DNA in the area flanking the insertion site of foreign DNAs is also commonly observed (Hejnar *et al.*, 2003, Lichtenberg *et al.*, 1988, Doerfler, 2011). In this instance, hypomethylation of the ends of integrated DNAs might represent a mechanism of a spread of methylation (Turker, 2002). Alternatively, as host DNA methylation pattern might influence integrant *de novo* methylation, hypomethylation of the flanking host sequence might represent the adoption of the adjacent (i.e. human) DNA methylation profile (Doerfler, 2011).

9.1.1.2. E2BS Methylation and Integration

We hypothesised that hypermethylation of the promoter E2BSs would be present in the absence of integration and would act to block E2 trans-acting repression as a mechanism of oncogene deregulation. There was no evidence for such a mechanism in VIN2/3 or the PC08/09 short term cultures. LCR2 hypermethylation was very rarely observed in precancerous neoplasia, and was not even commonly observed in cancer. Other studies have reported E2BS methylation in association with cervical cancers, although more rarely in high grade precancerous lesions (Bhattacharjee and Sengupta, 2006a, Fernandez *et al.*, 2009, Snellenberg *et al.*, 2012). Another study that used LCM to dissect a biopsy specifically did not see promoter E2BS methylation in regions where CIN3 histology was present, but the integration state of HPV in the biopsy was not determined (Vinokurova and von Knebel Doeberitz, 2011). There is however reason to suggest that the LCR E2BSs need not be heavily methylated in order to enhance P97 activity such that one or two methylated E2BS CpGs per genome might suffice. As

pyrosequencing considers the mean methylation of a mix of DNAs, if each DNA were variably methylated in one or two E2BS CpGs the overall level of methylation might be low, but most genomes might still be repressed by E2BS methylation (Stunkel and Bernard, 1999, Kalantari *et al.*, 2008a).

9.1.2. HPV Methylation may be Self Determined

HPV does not encode a DNMT, so DNA methylation is achieved using the cellular DNMTs (DNMT1, DNMT3A and DNMT3B) (Portela and Esteller, 2010). It is well established that oncogenic viruses can modulate DNA methylation either directly or indirectly; HBV, Hepatitis C, Kaposi's sarcoma associated herpes virus and EBV are all known to interact with and modulate the expression of DNMTs with associated changes in DNA methylation (Leonard *et al.*, 2012). HPV appears to affect the DNA methylation machinery indirectly via degradation of the p53 tumour suppressor and directly by upregulation of *DNMT1* expression (Au Yeung *et al.*, 2010). Transfection of keratinocytes with episomal HPV16 led to *DNMT1* and *DNMT3B* upregulation (Leonard *et al.*, 2012). E6 can also bind to proteins involved in chromatin remodelling and E7 binds DNMT1, enhancing its methyltransferase activity (Burgers *et al.*, 2007, Thomas and Chiang, 2005). E7 also disrupts the pRb/E2F pathway, responsible for regulating the DNMT1 promoter (McCabe *et al.*, 2005). Hence HPV transfection is associated with numerous changes in human DNA methylation (Leonard *et al.*, 2012).

E2 and E6 ORF products have long been known to modulate P97 activity (Shirasawa *et al.*, 1994, Dostatni *et al.*, 1991) and it appears that in some instances high E2, E6 and E7 activity promotes enhancer and promoter hypermethylation (De-Castro Arce *et al.*, 2011). Overexpression of HPV genes could effectively target DNA methylation towards the LCR. HPV self-methylation has been proposed as a means by which to regulate HPV gene expression in multi copy integrants where promoter and gene copy number would otherwise contribute to excessive and deleterious levels of gene expression (De-Castro Arce *et al.*, 2011).

9.1.3. HPV DNA Methylation Might not be Associated with Gene Expression

In HPV, leaky ribosome scanning and translation reinitiation are implicated in the regulation of gene expression (Zheng *et al.*, 2004, Remm *et al.*, 1999, Longworth and Laimins, 2004). Further, it is hypothesised that intragenic DNA methylation might prevent spurious initiation of transcription by preventing the binding of transcription factors to gene body binding sites (Zilberman *et al.*, 2007). We had therefore hypothesised that intragenic HPV methylation might affect the level of transcription of more 3' ORFs (i.e. *E2* hypermethylation and *E2*

transcription). Whilst we did observe a negative correlation between *E2* methylation and *E2* transcription in several instances, a similar relationship was also present for other regions of the HPV genome and for other mRNAs also. Typically, *E2* and/or *L1/L2* methylation were negatively associated with *E2*, *E6* and/or *E7* mRNA transcription. Whilst significant observations did exist, they appeared to be influenced heavily by the apparent relationship between DNA methylation and integration, with integration potentially having a negative effect on gene expression.

9.1.4. Variation in Methylation

9.1.4.1. Regional Variation

There was variation in methylation among the four HPV genome regions for every clinical and cell line cohort tested in this investigation. Typically, the *E2* and *L1/L2* regions were more methylated than the LCR, which was almost universally hypomethylated.

9.1.4.2. Permissive Infections

Whilst *E2* and *L1/L2* were methylated in several instances of precancerous neoplasia, the LCR rarely was. The most plausible explanation for universal LCR hypomethylation is that methylation of the LCR constituted a selective disadvantage. In a productive infection for instance, the primary role of E2 is in the replication of the viral genome, where E2 in association with E1 directs cell factors to the origin of replication (Kadaja *et al.*, 2009, Kim *et al.*, 2009). Methylation of the origin adjacent E2BS (E2BS$_2$) might interfere with episomal genome replication, whilst methylation of other parts of the LCR might induce conformation changes, impeding viral gene expression/replication (Rosl *et al.*, 1993).

Methylation of the promoter E2BSs might also cause deregulated expression of the viral oncogenes, a state undesirable to a productive infection and not part of the normal HPV lifecycle. Methylation of the promoter and origin adjacent E2BSs was not observed in LCM sections considered as permissive to viral replication which is consistent with this hypothesis; however enhancer (LCR1) methylation was frequently observed in these sections (Vinokurova and von Knebel Doeberitz, 2011).

9.1.4.3. Transformed Infections

Oncogene expression induces transformation, is required for the continued propagation of episomal genomes and also for the maintenance of the transformed cell state (Gray *et al.*, 2010, Van Tine *et al.*, 2004b, Hawley-Nelson *et al.*, 1989). Hypermethylation of E2BS$_1$ and the promoter E2BSs would likely block E2 binding and enhance P97 activity, whilst general LCR hypermethylation is associated with P97 repression (Thain *et al.*, 1996, De-Castro Arce *et al.*,

2011, Vinokurova and von Knebel Doeberitz, 2011, Kim *et al.*, 2003). LCR1 and LCR2 cover three CpGs of the viral enhancer and two repressive E2BSs within the promoter region respectively. It is interesting that the enhancer region was less methylated than the promoter E2BSs in our cancer samples. There were also more cancer samples with promoter hypermethylation than there were with enhancer hypermethylation. It seems that for the transformed phenotype, finding an equilibrium of methylation level at various sites of the LCR is required in order to permit oncogene expression without the potentially deleterious effects of oncogene overexpression (De-Castro Arce *et al.*, 2011, Gray *et al.*, 2010).

9.1.4.4. E2 and L1/L2

E2 and *L1/L2* hypermethylation appear to confer less of a disadvantage and in some instances were very common (i.e. integrated genomes and cancer cell HPV genomes). We could not confirm a role for intragenic methylation in modulating *E2* transcription, however possible consequences include both the enhancement and repression of P97 activity. If intragenic methylation did alter the ability of HPV to express functional E2 proteins, *E2* methylation could potentially increase oncogene expression. Alternatively, intragenic methylation might interfere with the expression of truncated E2 proteins that could act as an inhibitor of E2 function, thereby repressing P97 activity (Zilberman *et al.*, 2007, Kim *et al.*, 2003). As *L1/L2* expression is not required in non-productive infections, *L1/L2* hypermethylation might also be irrelevant to the maintenance of a transformed phenotype.

9.1.4.5. LCR

LCR hypomethylation due to selection appears likely in several instances but there was a general tendency for increased LCR methylation in more severe disease. It is possible that LCR hypermethylation/hypomethylation might be a consequence of gene expression level, rather than a cause of gene expression level; DNA methylation is hypothesised to spread from densely methylated to sparsely methylated regions and methylate regions where open chromatin formations are not present (i.e. non-transcriptionally active regions) (Turker, 2002). As such, low gene transcription in itself might facilitate promoter methylation and further downregulation (Szyf *et al.*, 2004). The relevance to HPV is that deregulated gene expression might be an undesirable state after cellular transformation where high levels of oncoprotein are sometimes detrimental to cell growth (Gray *et al.*, 2010, De-Castro Arce *et al.*, 2011). In summary, LCR methylation might both influence, and be influenced by gene expression level.

9.1.5. CpG Specific Pattern of Methylation

The pattern of CpG specific methylation, especially for the *L1/L2* and *E2* regions was remarkably well conserved across a wide range of material. The "N" shaped *L1/L2* methylation pattern for instance was seen in the VR series, most PC08 and PC09 clones, W12, CaSki, all three disease groups of the disease grade study and all outcome groups of the CRISP study. Upon closer inspection of published methylation data, the CaSki *L1/L2* "N" shaped methylation pattern is also visible (Fernandez *et al.*, 2009). There were also instances of massive methylation differences between CpGs separated by only a few bp.

Initially, this pattern was thought to be a result of either a specific mechanism of DNA methylation targeting each CpG differently, or a selective advantage conferred by each pattern of DNA methylation. The latter, seems somewhat unlikely as methylation patterns in a non/lowly transcribed region of the HPV genome (i.e. *L1/L2* in all but differentiated cells of a productive infection (Doorbar, 2006)) would need to confer a considerable selective advantage in order to be maintained and universally present. The selective benefit would need to be present for episomal and integrated HPV, HPV in precancerous lesions, cancers, normal cytology samples, samples with various levels of gene expression and methylation and even in transcriptionally inactive occluded chromatin (i.e. CaSki). Whilst hypomethylation of the LCR has a plausible basis for selection, a selective benefit of a CpG methylation pattern in *L1/L2* seems unlikely. Instead, the pattern is more likely produced as a result of enzymatic or structural differences within each region that might affect the ability of DNMTs and meCpG binding proteins to bind the DNA; these might include nucleosome position, nuclear matrix attachment regions and DNA secondary structure.

HPV DNA is organised as nucleosomes (Stunkel and Bernard, 1999). DNA methylation varies in a periodic manner with an interval of 10 bp in many forms of human DNA; genes, promoters, pericentromeric regions and euchromatic arm regions which might be due to nucleosomal DNA structure (Chodavarapu *et al.*, 2010). DNMTs access the major groove of nucleosomal DNA which would allow better access to DNA on the outside of nucleosomes (Chodavarapu *et al.*, 2010). DNA has a helical pitch of 10-10.5 bp (Wang, 1979) which is consistent with the 10 bp methylation periodicity. Figure 9.2 shows that the 10 bp methylation periodicity coincides perfectly with the period of the "N" shaped profile of the *L1/L2* region that was particularly evident in the PC08/PC09 cell lines (Chapter 7: 7.1.3.3.2).
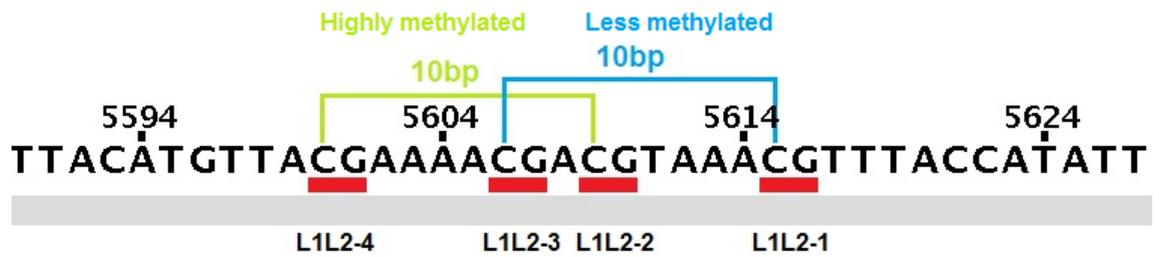
Figure 9.2: The position of CpGs within the *L1/L2* region.

Nucleotide positions are listed according to GenBank ID NC001526.1. Listed are the four CpGs of the *L1/L2* region (red boxes) and the distance (bp) between CpGs with high and low methylation.

DNA methylation is also associated with nucleosome position. Nucleosomal DNA was more methylated than flanking internucleosomal DNA and DNA methylation varied according to the distance from nucleosomes in a progressive manner (Chodavarapu *et al.*, 2010). This progressive change in methylation might explain the upwards trend of the *L1/L2* methylation profile (Chapter 7: 7.1.3.3.2). Nucleosome position analysis has not been performed for HPV16 outside of the LCR so confirmation is difficult. Further, nucleosome analysis of the LCR is typically performed on integrated HPV containing cell lines so the relevance to episomal HPV is similarly difficult to establish (Stunkel and Bernard, 1999).

For the *E2* region, a similar 10 bp periodicity is observed for the first 4/5 CpGs, after which the relationship breaks down, possibly due to additional affects on methylation such as nucleosome positioning. The HPV16 LCR has several nucleosomes, one nucleosome encompasses the whole origin and promoter region whilst the large enhancer region contains two nucleosomes, with substantial linker regions (Stunkel and Bernard, 1999). Both the LCR1 and LCR2 regions are fully encompassed within these nucleosome regions. Although some instances of the 10 bp periodicity were observed for the LCR, and present in published data (Ding *et al.*, 2009), the relationship was less well defined which might stem from the potential selectable effect of LCR methylation on oncogene expression.

9.2. Methylation and Carcinogenesis

When considering the potential contribution of HPV methylation to carcinogenesis, two main topics should be considered. The first is the effect of methylation in transformed cells, the second is the initial deregulation of the oncogenes and the induction of cellular transformation.

9.2.1. The Potential Effects of HPV DNA Methylation

The P97 promoter contains binding sites for various cellular transcription factors. Our LCR1 region represents 3 CpGs of the viral enhancer and LCR2 represents five viral promoter CpGs, four of which are contained within the two promoter E2BSs (LCR2-1 is within a Sp1 binding site, LCR2-2+3 are in $E2BS_3$ and LCR2-4+5 are in $E2BS_4$) (Tan *et al.*, 1994). E2 cannot bind methylated E2BSs (Fernandez *et al.*, 2009). One of the LCR2 CpGs (LCR2-1) is not part of an E2BS but instead is present in a binding site for the Sp1 transcription factor (Tan *et al.*, 1994). In this instance, one might expect LCR2-1 to be less methylated than the E2BS CpGs however this was not the case and LCR2-1 was methylated at a level comparable to LCR2-2 and LCR2-3. P97 activation is functionally dependent on Sp1 binding which can be displaced by E2 binding to $E2BS_3$. However, Sp1 is capable of binding methylated DNA so methylation of LCR2-1 in addition to the adjacent $E2BS_3$ CpGs might not confer a selectable disadvantage, hence its methylation at a level comparable to E2BS CpGs (Tan *et al.*, 1994, Emili *et al.*, 1994, Harrington *et al.*, 1988).

Methylation of $E2BS_1$, the so called activator E2BS is also associated with increased P97 activity which may be due to an uncharacterised complex of transcription factors binding specifically to methylated $E2BS_1$ (Vinokurova and von Knebel Doeberitz, 2011). $E2BS_1$ methylation was also associated with cervical cancer (Bhattacharjee and Sengupta, 2006a). Our efforts to produce a pyrosequencing assay targeting $E2BS_1$ were initially unsuccessful however in light of these more recent studies it seems appropriate to pursue the assessment of $E2BS_1$ methylation.

Whilst several cancers had high level LCR2 methylation, we did not frequently observe high levels of methylation of the LCR in precancerous neoplasia however it is possible that high level methylation is not required for P97 repression. Nucleosomes form over the LCR and thus mediation of HDAC activity by DNA methylation affects transcription at the promoter region as a result of topological changes (de Ruijter *et al.* 2003; Harrington *et al.* 1988; Stunkel *et al.* 2000). Multiple CpGs are present in the promoter and enhancer regions, a single methylated CpG in these regions could bind methylated DNA binding proteins and could potentially induce sufficient topological changes to repress P97 activity (Stunkel and Bernard, 1999, Kalantari *et al.*, 2008a).

9.2.2. Initial Deregulation of Oncogenes

Increased methylation of the *E2* and *L1/L2* regions was observed during and after the emergence of integrated HPV in W12Ser7 (Chapter 7: 7.1.2). Integration and the frequent

disruption of the *E2* ORF are commonly considered to be a mechanism of cellular transformation and carcinogenesis (Thierry, 2009). However, many HPV-associated cancers do not contain integrated HPV, many more do not contain *E2* disrupted integrants and many high grade lesions contain both integrated and episomal forms which should be competent at producing functional E2 repressor protein (Arias-Pulido *et al.*, 2006, Klaes *et al.*, 1999, Pett and Coleman, 2007, Vinokurova *et al.*, 2008, Bhattacharjee and Sengupta, 2006b). Instead, integration of HPV into the host genome could be a consequence of chromosomal instability caused by aberrant HPV oncogene expression, rather than the cause of initial aberrant oncogene expression (Pett and Coleman, 2007, Melsheimer *et al.*, 2004). The initial causes of aberrant oncogene expression could be varied and are considerably understudied however there is reason to suspect that HPV DNA methylation might be involved (Vinokurova and von Knebel Doeberitz, 2011). We had initially hypothesised that heavy E2BS methylation would be present in the absence of integration however this was not observed either *in vivo* or *in vitro*.

9.3. Methylation as a Biomarker

During this study, we were mindful of the processes for biomarker development recommended by CRUK (Cancer Research UK, 2012d, Cancer Research UK, 2012e). Whilst this study was not exclusively designed for biomarker development, many aspects fit well within the advised framework for the predictive/prognostic and screening biomarker roadmaps (Appendix I). The early stages of biomarker development are common to multiple CRUK roadmaps. The rationale section for instance, is identical for each type of biomarker; there must be a clinical utility for a biomarker, a reliable assay, and a sample set with which to test the biomarker. All of these requirements were at least partially met by this investigation.

9.3.1. Potential Clinical Utility for a HPV Methylation Biomarker

Liquid based cytology (LBC) replaced traditional Pap smear testing in England between 2003-2008 (Albrow *et al.*, 2012). The UK HPV vaccination programme is expected to reduce the number of HPV positive LBC samples and therefore the efficiency of cytological screening (Wentzensen *et al.*, 2009). Whilst HPV vaccination will substantially reduce the prevalence of HPV-associated disease, the current vaccines only target two of the HrHPV types. Due to the long time required for carcinogenesis, it is also likely to be many years before there is a reduction in cervical cancer rates due to vaccination (McIndoe *et al.*, 1984).

HPV testing as a primary screening test would identify a large number of positive samples, many of which would be harmless transient infections with a low risk of progression. HPV

testing alone is unlikely to achieve a sufficient level of specificity (Unger *et al.*, 2004). In order to prevent the treatment of the numerous clinically irrelevant HPV infections, many studies have attempted to develop specific CIN biomarkers to be used in conjunction with HPV testing (Wentzensen and von Knebel Doeberitz, 2007, Cuschieri and Wentzensen, 2008, Mirabello *et al.*, 2012). Using cytology as a triage test for HPV primary testing has also been proposed (Naucler *et al.*, 2009, Rijkaart *et al.*, 2012). In this instance, many patients would be identified as HPV positive with low grade cytology, hence a prognostic biomarker to predict long/short term outcome and to determine a suitable management strategy would be of substantial benefit.

9.3.2. CRUK Biomarker Roadmaps – Biomarker Discovery and Development

Several components of this investigation were relevant to CRUK "biomarker discovery and development" roadmap components (Roadmaps presented in Appendix I). We tested and developed assays for potential biomarkers (Chapter 4) and performed studies to address the early stages of biomarker development.

The components of this investigation with relevance to the biomarker roadmaps are shown in Table 9.1. We addressed several stage 1 and stage 2 biomarker development requirements. A reproducible assay was developed and applied to determine the DNA methylation status of various clinical cohorts. In these cohorts methylation was seen to vary among disease stages, between cancer and non-cancer, and even among samples of a more homogeneous VIN2/3 cohort (Chapter 5: 5.1.2 & Chapter 8: 8.2.3.1). Differences in DNA methylation were also observed among samples with low grade cytological abnormalities but different longitudinal outcomes (Chapter 6: 6.1.3).

9.3.3. Progress

Several aspects of the results were consistent with similar published studies. In the CRISP study (Chapter 6: 6.1.3), the most informative regions appeared to be LCR2 and *L1*/*L2*, and the most informative CpGs appeared to be located in the LCR (LCR1-1 and LCR2-3). We also observed differences in DNA methylation between samples with normal cytology/severe dyskaryosis and cancer (Chapter 5: 5.1.2); in this study, the most informative region appeared to be *L1*/*L2* and the most informative CpG was likely to be L1L2-2. The difference in CpGs identified as informative in the NSC and CRISP studies is likely the result of study design and sample selection. In the CRISP series, grouping was determined by eventual outcome and all samples were initially low grade cytology. For the NSC series, grouping was performed with respect to cytology grade/cancer.

Other biomarker type studies have considered multiple CpGs within the genome and identified several informative CpGs or methylation patterns indicative of disease state (Brandsma *et al.*, 2009, Mirabello *et al.*, 2012). In light of our observation of large differences among clinical groups at single CpGs, and published reports using multiple CpGs; any clinically relevant HPV methylation biomarker is likely to be comprised of several HPV CpGs considered using an algorithm (Brandsma *et al.*, 2009, Mirabello *et al.*, 2012). That different regions/CpGs of the HPV genome appear to be more discerning for different purposes (Chapter 5: 5.1.2 & Chapter 6: 6.1.3) further support the idea that considering multiple CpGs at different regions is a useful strategy.

	Biomarker discovery and development (Cancer Research UK 2012)	
	Prognostic/predictive BM	Screening BM
	Do you have an assay?	Do you have an assay?
	Yes	Yes
Requirement	Development of an accurate and reproducible assay to measure BM	
Result	Pyrosequencing assays were shown to be reproducible and accurate (Method development)	
Requirement	Define BM distribution on specimens representative of the target patient population	
Result	The NSC study and VR series achieved this	
Requirement	Does the distribution of BM values indicate a BM with potential clinical utility?	Can the biomarker distinguish between subjects with and without cancer?
Result	There is variation among disease grades (NSC study) and within a group of similar samples (VR series)	There was a significant difference between cancer and non-cancer samples (NSC study)
Requirement	Study the relationship between the BM and clinical outcome retrospectively	Can the BM detect cancer earlier than the time of clinical diagnosis?
Result	The CRISP study achieved this	-
Requirement	Is there a correlation between the BM and clinical outcome?	-
Result	There was a significant difference between outcome groups	-

Table 9.1: How relevant aspects of this investigation relate to the CRUK biomarker roadmaps.

Components of the CRUK biomarker roadmaps are detailed under the row title "requirement". Where an aspect of this investigation has relevance to this requirement, the details of the study which achieved it or the result of the study are detailed under "result". Pink and blue highlights represent the prognostic/predictive and screening biomarker requirements respectively whilst yellow highlights indicate commonalities between the biomarker roadmaps.

9.3.4. Potential Application

Only the very early stages of biomarker development have been performed but methylation of HPV DNA shows potential as a biomarker in several contexts. While cytology remains the

primary screen to prevent cervical cancer, a prognostic biomarker capable of differentiating clinical outcomes would be of particular use for mild/second borderline cytology cases where it is difficult to determine a suitable management strategy (Castanon *et al.*, 2012). HPV testing has been suggested as a triage for low grade cytological abnormalities (and is being adopted in England (Albrow *et al.*, 2012)) but is likely to substantially increase the number of patients referred to colposcopy (Moss *et al.*, 2006). HPV triage testing is also being used as a test of cure for women treated for CIN. Nonetheless, there is probably a limited scope for a CIN biomarker in these instances.

The greatest potential use of an HPV disease associated biomarker will probably arise once HPV testing is implemented as the primary screen for cervical neoplasia. Primary HPV testing has been suggested as a more cost effective alternative to primary cytology screening for European countries (Kok *et al.*, 2012) and offers greater sensitivity and negative predictive value, but at the cost of reduced specificity relative to cytology (Unger *et al.*, 2004, Katki *et al.*, 2011). As such, primary HPV testing will identify numerous patients with little risk of CIN or cancer, especially young patients where HPV infection is highly prevalent (Hibbitts *et al.*, 2006). A screening biomarker in this setting would have potential to identify those patients with clinically significant infections. HPV testing triaged with cytology has also been suggested as a suitable primary screening system, this would identify a large number of HPV positive patients with low grade cytology (Naucler *et al.*, 2009, Rijkaart *et al.*, 2012). It has previously proven difficult to determine a suitable management strategy for patients with low grade cytology and further triage of these patients using a prognostic biomarker may prove clinically useful (Castanon *et al.*, 2012).

Further, a DNA based screening system (i.e. HPV testing triaged with a DNA based biomarker) might allow self collected samples to be used (Brink *et al.*, 2006). Self collection has been shown to increase the coverage of the screening programme (Gok *et al.*, 2010). An objective DNA based primary screening technology might also find application in countries where a well established cytological screening programme is not present.

As the relationship between HPV DNA methylation and HPV-associated disease of the cervix appears to be similar for other anogenital sites (i.e. vulva, penis, anus) and other HPV types, HPV DNA methylation biomarkers could have a broad clinical utility (Kalantari *et al.*, 2008b, Turan *et al.*, 2007, Wiley *et al.*, 2005, Mirabello *et al.*, 2012). Human genes are also differentially methylated in association with HPV disease and represent potential biomarkers (Eijsink *et al.*, 2012, Wentzensen *et al.*, 2009). Our finding of *E2* and/or *L1/L2* hypermethylation also partially supports the hypothesis that *L1* methylation might be a surrogate marker for HPV

integration and a potential cervical disease biomarker (Kalantari *et al.*, 2010). Our observations would however suggest that methylation analysis of more than one HPV region would be more suitable as integrated HPV genomes appear to be hypomethylated at the HPV-human DNA boundaries.

9.4. Further Work

Further investigation into the biology of HPV is required to fully understand the apparently complex relationship between integration, methylation and gene expression. There is a potential role for DNA methylation as an initiator of deregulation that should be more thoroughly explored and also a potential relationship between HPV methylation and clearance of an infection. In a productive infection, methylation of $E2BS_2$ (the origin adjacent E2BS) could interfere with genome replication. As such, it is worth considering whether $E2BS_2$ methylation is implicated in the transition from a productive to a transformed phenotype, as a precursor to integration or is involved with viral clearance; three events where the inhibition of episomal replication might be involved.

The results of this investigation are promising in terms of a biomarker of HPV-associated disease, further development is required on a larger sample set to further explore the clinical relevance of HPV methylation.

Chapter 10.   General Conclusion

This investigation began by assessing several methods for determining HPV DNA methylation. MS-PCR and bisulfite sequencing were trialled but the limitations of these assays were quickly realised. Eventually, the relatively new technology of pyrosequencing was trialled and adopted because it is quantitative, assesses multiple CpGs and in comparison to bisulfite sequencing with cloning, is high throughput. Pyrosequencing was shown to be both sensitive and reproducible and could be applied to a diverse range of sample DNAs including clinically relevant LBC samples.

Pyrosequencing was applied to various DNAs with two main objectives, to better understand the basic biology of HPV DNA methylation and to evaluate the clinical utility of HPV DNA methylation. The cell lines and VR series studies allowed several conclusions to be drawn. First, we did not see evidence that methylation was an alternative transforming event to HPV integration. LCR and E2BS methylation was very rarely observed and the largest determinant of HPV methylation state was the presence of integrated HPV. Integrants were typically hypermethylated in *E2* or *L1/L2*, but not in regions of the genome adjacent to human DNA, similar to the methylation profile of other foreign DNAs inserted into the human genome. In general, *E2* and/or *L1/L2* methylation were negatively associated with *E2*, *E6* and/or *E7* mRNA transcription and a relationship between integration and gene expression was not successfully established.

To better understand the potential clinical utility of HPV DNA methylation, the relationship between HPV DNA methylation and disease state/clinical outcome was evaluated in LBC material from the CRISP clinical series and from routine screening, and in biopsy material from invasive cervical cancers. These studies demonstrated that there were significant differences in DNA methylation among samples with normal cytology, severe dyskaryosis and invasive cervical cancer. There were also considerable differences in methylation of the cohort of low grade cytology samples with a variety of 6 month outcomes. Methylation differences were apparent for samples that were cleared, and those that were persistent and associated with either low or high grade disease. Several regions and CpGs of the HPV genome were further identified that were the most significantly different among the various disease and outcome groups.

Throughout the investigation, consistent patterns of CpG specific methylation were observed which might be explained by subtle positioning differences in reference to the nucleosomal

structure, and the subsequent level of access that this structure provides for the cellular methylation machinery.

In summary, the relationship between methylation, integration and gene expression is complex and assessment of HPV DNA methylation status does appear to hold potential as a triage test for HPV positive women.

References

ABDEL-HADY, E. S., MARTIN-HIRSCH, P., DUGGAN-KEEN, M., STERN, P. L., MOORE, J. V., CORBITT, G., KITCHENER, H. C. & HAMPSON, I. N. 2001. Immunological and viral factors associated with the response of vulval intraepithelial neoplasia to photodynamic therapy. *Cancer Res,* 61**,** 192-6.

ABDUL, H. N., BROWN, C. & GASTON, K. 2009. The regulation of cell proliferation by the papillomavirus early proteins. *Cellular and Molecular Life Sciences***,** In Print.

AI, W., NARAHARI, J. & ROMAN, A. 2000. Yin yang 1 negatively regulates the differentiation-specific E1 promoter of human papillomavirus type 6. *J Virol,* 74**,** 5198-205.

ALBROW, R., KITCHENER, H., GUPTA, N. & DESAI, M. 2012. Cervical screening in England: The past, present, and future. *Cancer Cytopathology,* 120**,** 87-96.

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

ANTEQUERA, F. & BIRD, A. 1993. Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A,* 90**,** 11995-11999.

APOSTOLIDOU, S., HADWIN, R., BURNELL, M., JONES, A., BAFF, D., PYNDIAH, N., MOULD, T., JACOBS, I. J., BEDDOWS, S., KOCJAN, G. & WIDSCHWENDTER, M. 2009. DNA methylation analysis in liquid-based cytology for cervical cancer screening. *Int J Cancer,* 125**,** 2995-3002.

APT, D., CHONG, T., LIU, Y. & BERNARD, H. U. 1993. Nuclear factor I and epithelial cell-specific transcription of human papillomavirus type 16. *J Virol,* 67**,** 4455-4463.

ARAIN, S., WALTS, A., THOMAS, P. & BOSE, S. 2005. The Anal Pap Smear: Cytomorphology of squamous intraepithelial lesions. *CytoJournal,* 2**,** 4.

ARAN, D., TOPEROFF, G., ROSENBERG, M. & HELLMAN, A. 2011. Replication timing-related and gene body-specific methylation of active human genes. *Hum Mol Genet,* 20**,** 670-80.

ARIAS-PULIDO, H., PEYTON, C. L., JOSTE, N. E., VARGAS, H. & WHEELER, C. M. 2006. Human Papillomavirus Type 16 Integration in Cervical Carcinoma In Situ and in Invasive Cervical Cancer. *Journal of Clinical Microbiology,* 44**,** 1755-1762.

ARICO, J. K., KATZ, D. J., VAN DER VLAG, J. & KELLY, W. G. 2011. Epigenetic Patterns Maintained in Early *Caenorhabditis elegans* Embryos Can Be Established by Gene Activity in the Parental Germ Cells. *PLoS Genet,* 7**,** e1001391.

AU YEUNG, C. L., TSANG, W. P., TSANG, T. Y., CO, N. N., YAU, P. L. & KWOK, T. T. 2010. HPV-16 E6 upregulation of DNMT1 through repression of tumor suppressor p53. *Oncol Rep,* 24**,** 1599-604.

BACHMAN, K. E., PARK, B. H., RHEE, I., RAJAGOPALAN, H., HERMAN, J. G., BAYLIN, S. B., KINZLER, K. W. & VOGELSTEIN, B. 2003. Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. *Cancer Cell,* 3**,** 89-95.

BADAL, S., BADAL, V., CALLEJA-MACIAS, I. E., KALANTARI, M., CHUANG, L. S. H., LI, B. F. L. & BERNARD, H.-U. 2004. The human papillomavirus-18 genome is efficiently targeted by cellular DNA methylation. *Virology,* 324**,** 483-492.

BADAL, V., CHUANG, L. S. H., TAN, E. H.-H., BADAL, S., VILLA, L. L., WHEELER, C. M., LI, B. F. L. & BERNARD, H.-U. 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**,** 6227-6234.

BAKER, C. C., PHELPS, W. C., LINDGREN, V., BRAUN, M. J., GONDA, M. A. & HOWLEY, P. M. 1987. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *J Virol,* 61**,** 962-971.

BALDERAS-LOAEZA, A., ANAYA-SAAVEDRA, G., RAMIREZ-AMADOR, V. A., GUIDO-JIMENEZ, M. C., KALANTARI, M., CALLEJA-MACIAS, I. E., BERNARD, H. U. & GARCIA-CARRANCA, A. 2007. Human papillomavirus-16 DNA methylation patterns support a causal association of the virus with oral squamous cell carcinomas. *Int J Cancer,* 120**,** 2165-9.

BALDWIN, P. J., VAN DER BURG, S. H., BOSWELL, C. M., OFFRINGA, R., HICKLING, J. K., DOBSON, J., ROBERTS, J. S., LATIMER, J. A., MOSELEY, R. P., COLEMAN, N., STANLEY, M. A. & STERLING, J. C. 2003. Vaccinia-expressed human papillomavirus 16 and 18 e6 and e7 as a therapeutic vaccination for vulval and vaginal intraepithelial neoplasia. *Clin Cancer Res,* 9**,** 5205-13.

BARTON, S. C., ARNEY, K. L., SHI, W., NIVELEAU, A., FUNDELE, R., SURANI, M. A. & HAAF, T. 2001. Genome-wide methylation patterns in normal and uniparental early mouse embryos. *Hum Mol Genet,* 10**,** 2983-7.

BEAUJEAN, N., TAYLOR, J. E., MCGARRY, M., GARDNER, J. O., WILMUT, I., LOI, P., PTAK, G., GALLI, C., LAZZARI, G., BIRD, A., YOUNG, L. E. & MEEHAN, R. R. 2004. The effect of interspecific oocytes on demethylation of sperm DNA. *Proc Natl Acad Sci U S A,* 101**,** 7636-40.

BECHTOLD, V., BEARD, P. & RAJ, K. 2003. Human papillomavirus type 16 E2 protein has no effect on transcription from episomal viral DNA. *J Virol,* 77**,** 2021-8.

BELINSKY, S. A., NIKULA, K. J., BAYLIN, S. B. & ISSA, J. P. 1996. Increased cytosine DNA-methyltransferase activity is target-cell-specific and an early event in lung cancer. *Proc Natl Acad Sci U S A,* 93**,** 4045-50.

BERDASCO, M. & ESTELLER, M. 2010. Aberrant Epigenetic Landscape in Cancer: How Cellular Identity Goes Awry. *Developmental Cell,* 19**,** 698-711.

BERG, M. & STENLUND, A. 1997. Functional interactions between papillomavirus E1 and E2 proteins. *Journal of Virology,* 71**,** 3853-63.

BERNARD, H.-U., BURK, R. D., CHEN, Z., VAN DOORSLAER, K., HAUSEN, H. Z. & DE VILLIERS, E.-M. 2010. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology,* 401**,** 70-79.

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

BHATTACHARJEE, B. & SENGUPTA, S. 2006b. HPV16 E2 gene disruption and polymorphisms of E2 and LCR: some significant associations with cervical cancer in Indian women. *Gynecol Oncol,* 100**,** 372-8.

BHATTACHARYA, S. K., RAMCHANDANI, S., CERVONI, N. & SZYF, M. 1999. A mammalian protein with specific demethylase activity for mCpG DNA. *Nature,* 397**,** 579-83.

BIRD, A. P. 1986. CpG-rich islands and the function of DNA methylation. *Nature,* 321**,** 209-13.

BIRD, A. P. & WOLFFE, A. P. 1999. Methylation-induced repression--belts, braces, and chromatin. *Cell,* 99**,** 451-4.

BLACHON, S. & DEMERET, C. 2003. The regulatory E2 proteins of human genital papillomaviruses are pro-apoptotic. *Biochimie,* 85**,** 813-819.

BONVICINI, F., VENTUROLI, S., AMBRETTI, S., PATERINI, P., SANTINI, D., CECCARELLI, C., ZERBINI, M. & MUSIANI, M. 2005. Presence and type of oncogenic human papillomavirus in classic and in differentiated vulvar intraepithelial neoplasia and keratinizing vulvar squamous cell carcinoma. *J Med Virol,* 77**,** 102-6.

BOWKER, D. W. R., P. F. 2007. *Practical Data Analysis Workbook for Biosciences (3rd ed.)* Harlow, Essex: Pearson Education Ltd.

BOYER, S. N., WAZER, D. E. & BAND, V. 1996. E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. *Cancer Res,* 56**,** 4620-4.

BRAATEN, K. P. & LAUFER, M. R. 2008. Human Papillomavirus (HPV), HPV-Related Disease, and the HPV Vaccine. *Reviews in Obstetrics & Gynecology,* 1**,** 2-10.

BRANDSMA, J. L., SUN, Y., LIZARDI, P. M., TUCK, D. P., ZELTERMAN, D., HAINES, G. K., 3RD, MARTEL, M., HARIGOPAL, M., SCHOFIELD, K. & NEAPOLITANO, M. 2009. Distinct human papillomavirus type 16 methylomes in cervical cells at different stages of premalignancy. *Virology,* 389**,** 100-7.

BREITBURD, F., KIRNBAUER, R., HUBBERT, N. L., NONNENMACHER, B., TRIN-DINH-DESMARQUET, C., ORTH, G., SCHILLER, J. T. & LOWY, D. R. 1995. Immunization with viruslike particles from cottontail rabbit papillomavirus (CRPV) can protect against experimental CRPV infection. *J Virol,* 69**,** 3959-63.

BRINK, A. A., MEIJER, C. J., WIEGERINCK, M. A., NIEBOER, T. E., KRUITWAGEN, R. F., VAN KEMENADE, F., FRANSEN DAALMEIJER, N., HESSELINK, A. T., BERKHOF, J. & SNIJDERS, P. J. 2006. High concordance of results of testing for human papillomavirus in cervicovaginal samples collected by two methods, with comparison of a novel self-sampling device to a conventional endocervical brush. *J Clin Microbiol,* 44**,** 2518-23.

BRYAN, J. T. & BROWN, D. R. 2000. Association of the Human Papillomavirus Type 11 E1[logical and]E4 Protein with Cornified Cell Envelopes Derived from Infected Genital Epithelium. *Virology,* 277**,** 262-269.

BRYANT, D., RAI, N., ROWLANDS, G., HIBBITTS, S., JONES, J., TRISTRAM, A., FIANDER, A. & POWELL, N. 2011. Human papillomavirus type distribution in vulval intraepithelial neoplasia determined using PapilloCheck DNA Microarray. *J Med Virol,* 83**,** 1358-61.

BUNZ, F., DUTRIAUX, A., LENGAUER, C., WALDMAN, T., ZHOU, S., BROWN, J. P., SEDIVY, J. M., KINZLER, K. W. & VOGELSTEIN, B. 1998. Requirement for p53 and p21 to Sustain G2 Arrest After DNA Damage. *Science,* 282**,** 1497-1501.

BURD, E. M. 2003. Human Papillomavirus and Cervical Cancer. *Clinical Microbiology reviews,* 16**,** 1-17.

BURGERS, W. A., BLANCHON, L., PRADHAN, S., DE LAUNOIT, Y., KOUZARIDES, T. & FUKS, F. 2007. Viral oncoproteins target the DNA methyltransferases. *Oncogene,* 26**,** 1650-5.

BURGERS, W. A., FUKS, F. & KOUZARIDES, T. 2002. DNA methyltransferases get connected to chromatin. *Trends Genet,* 18**,** 275-277.

CAI, F. F., KOHLER, C., ZHANG, B., WANG, M. H., CHEN, W. J. & ZHONG, X. Y. 2011. Epigenetic therapy for breast cancer. *Int J Mol Sci,* 12**,** 4465-87.

CANCER RESEARCH UK. 2012a. *Cervical cancer risks and causes* [Online]. Available: http://cancerhelp.cancerresearchuk.org/type/cervical-cancer/about/cervical-cancer-risks-and-causes [Accessed 25/07/2012 2012].

CANCER RESEARCH UK. 2012b. *Definitions/glossary* [Online]. Available: http://science.cancerresearchuk.org/funding/find-grant/all-funding-schemes/biomarkers-imaging-discovery-development-project-grants/definitions-glossary/ [Accessed 08/08/2012 2012].

CANCER RESEARCH UK. 2012c. *Diagnostic biomarker (BM) roadmap* [Online]. Available: http://science.cancerresearchuk.org/prod_consump/groups/cr_common/@fre/@fun/documents/generalcontent/cr_027484.pdf [Accessed 08/08/2012 2012].

CANCER RESEARCH UK. 2012d. *Prognostic/predictive biomarker (BM) roadmap* [Online]. Available: http://science.cancerresearchuk.org/prod_consump/groups/cr_common/@fre/@fun/documents/generalcontent/cr_027481.pdf [Accessed 06/08/2012 2012].

CANCER RESEARCH UK. 2012e. *Screening biomarker (BM) roadmap* [Online]. Available: http://science.cancerresearchuk.org/prod_consump/groups/cr_common/@fre/@fun/documents/generalcontent/cr_027487.pdf [Accessed 08/08/2012 2012].

CARLI, P., CATTANEO, A., DE MAGNIS, A., BIGGERI, A., TADDEI, G. & GIANNOTTI, B. 1995. Squamous cell carcinoma arising in vulval lichen sclerosus: a longitudinal cohort study. *Eur J Cancer Prev,* 4**,** 491-5.

CASTANON, A., TRISTRAM, A., MESHER, D., POWELL, N., BEER, H., ASHMAN, S., RIECK, G., FIELDER, H., FIANDER, A. & SASIENI, P. 2012. Effect of diindolylmethane supplementation on low-grade cervical cytological abnormalities: double-blind, randomised, controlled trial. *Br J Cancer,* 106**,** 45-52.

CDC. 2011. *Sexually Transmitted Diseases (STDs): Treatment Guidelines, 2010* [Online]. Available: http://www.cdc.gov/std/treatment/2010/genital-warts.htm [Accessed 19/09/2012 2012].

CERVICAL SCREENING WALES 2004. KC53/61/65 Statistical Report - Adroddiad Ystadegol 2002/03 data. Cardiff: Cervical Screening Wales.

CERVICAL SCREENING WALES 2011. KC53/61/65 Statistical Report - Adroddiad Ystadegol 2010/2011 data. Cardiff: Cervical Screening Wales.

CHEN, G. & STENLUND, A. 2000. Two patches of amino acids on the E2 DNA binding domain define the surface for interaction with E1. *J Virol,* 74**,** 1506-12.

CHEN, T. & LI, E. 2004. Structure and function of eukaryotic DNA methyltransferases. *Curr Top Dev Biol,* 60**,** 55-89.

CHEN, Z. X. & RIGGS, A. D. 2011. DNA methylation and demethylation in mammals. *J Biol Chem*.

CHENG, Y.-W., WU, T.-C., CHEN, C.-Y., CHOU, M.-C., KO, J.-L. & LEE, H. 2008. Human Telomerase Reverse Transcriptase Activated by E6 Oncoprotein Is Required for Human Papillomavirus-16/18-Infected Lung Tumorigenesis. *Clin Cancer Res,* 14**,** 7173-7179.

CHIANG, C. M., USTAV, M., STENLUND, A., HO, T. F., BROKER, T. R. & CHOW, L. T. 1992. Viral E1 and E2 proteins support replication of homologous and heterologous papillomaviral origins. *Proc Natl Acad Sci U S A,* 89**,** 5799-803.

CHIK, F. & SZYF, M. 2011. Effects of specific DNMT gene depletion on cancer cell transformation and breast cancer cell invasion; toward selective DNMT inhibitors. *Carcinogenesis,* 32**,** 224-32.

CHODAVARAPU, R. K., FENG, S., BERNATAVICHUTE, Y. V., CHEN, P.-Y., STROUD, H., YU, Y., HETZEL, J. A., KUO, F., KIM, J., COKUS, S. J., CASERO, D., BERNAL, M., HUIJSER, P., CLARK, A. T., KRAMER, U., MERCHANT, S. S., ZHANG, X., JACOBSEN, S. E. & PELLEGRINI, M. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature,* 466**,** 388-392.

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

CHRISTMAN, J. K. 2002. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene,* 21**,** 5483-95.

CHUANG, J. C., YOO, C. B., KWAN, J. M., LI, T. W. H., LIANG, G., YANG, A. S. & JONES, P. A. 2005. Comparison of biological effects of non-nucleoside DNA methylation inhibitors versus 5-aza-2'-deoxycytidine. *Molecular Cancer Therapeutics,* 4**,** 1515-1520.

CIRELLI, R. & TYRING, S. K. 1994. Interferons in human papillomavirus infections. *Antiviral Res,* 24**,** 191-204.

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

CONGER, K. L., LIU, J.-S., KUO, S.-R., CHOW, L. T. & WANG, T. S. F. 1999. Human Papillomavirus DNA Replication. Interactions between the viral E1 protein and two subunits of human DNA polymerase alpha/primase. *J. Biol. Chem.,* 274**,** 2696-2705.

CONRAD, M., BUBB, V. J. & SCHLEGEL, R. 1993. The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein. *J Virol,* 67**,** 6170-8.

CORDEN, S. A., SANT-CASSIA, L. J., EASTON, A. J. & MORRIS, A. G. 1999. The integration of HPV-18 DNA in cervical carcinoma. *Mol Pathol,* 52**,** 275-82.

CRICCA, M., VENTUROLI, S., LEO, E., COSTA, S., MUSIANI, M. & ZERBINI, M. 2009. Disruption of HPV 16 E1 and E2 genes in precancerous cervical lesions. *Journal of Virological Methods,* In Press, Corrected Proof.

CULLEN, A. P., REID, R., CAMPION, M. & LORINCZ, A. T. 1991. Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive cervical neoplasm. *J Virol,* 65**,** 606-12.

CUSCHIERI, K. & WENTZENSEN, N. 2008. Human Papillomavirus mRNA and p16 Detection as Biomarkers for the Improved Diagnosis of Cervical Neoplasia. *Cancer Epidemiology Biomarkers & Prevention,* 17**,** 2536-2545.

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

DAAYANA, S., ELKORD, E., WINTERS, U., PAWLITA, M., RODEN, R., STERN, P. L. & KITCHENER, H. C. 2010. Phase II trial of imiquimod and HPV therapeutic vaccination in patients with vulval intraepithelial neoplasia. *Br J Cancer,* 102**,** 1129-36.

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

DAUJAT, S., ZEISSLER, U., WALDMANN, T., HAPPEL, N. & SCHNEIDER, R. 2005. HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *J Biol Chem,* 280**,** 38090-5.

DE-CASTRO ARCE, J., GÖCKEL-KRZIKALLA, E. & RÖSL, F. 2011. Silencing of multi-copy HPV16 by viral self-methylation and chromatin occlusion: a model for epigenetic virus–host interaction. *Human Molecular Genetics*.

DE RUIJTER, A. J., VAN GENNIP, A. H., CARON, H. N., KEMP, S. & VAN KUILENBURG, A. B. 2003. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J,* 370**,** 737-49.

DE SANJOSE, S., DIAZ, M., CASTELLSAGUE, X., CLIFFORD, G., BRUNI, L., MUNOZ, N. & BOSCH, F. X. 2007. Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. *Lancet Infect Dis,* 7**,** 453-9.

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

DE VUYST, H., CLIFFORD, G. M., NASCIMENTO, M. C., MADELEINE, M. M. & FRANCESCHI, S. 2009. Prevalence and type distribution of human papillomavirus in carcinoma and intraepithelial neoplasia of the vulva, vagina and anus: A meta-analysis. *Int J Cancer,* 124**,** 1626-1636.

DEFILIPPIS, R. A., GOODWIN, E. C., WU, L. & DIMAIO, D. 2003. Endogenous human papillomavirus E6 and E7 proteins differentially regulate proliferation, senescence, and apoptosis in HeLa cervical carcinoma cells. *J Virol,* 77**,** 1551-63.

DEMERET, C., DESAINTES, C., YANIV, M. & THIERRY, F. 1997. Different mechanisms contribute to the E2-mediated transcriptional repression of human papillomavirus type 18 viral oncogenes. *J Virol,* 71**,** 9343-9.

DEMERET, C., GOYAT, S., YANIV, M. & THIERRY, F. 1998. The Human Papillomavirus Type 18 (HPV18) Replication Protein E1 Is a Transcriptional Activator When Interacting with HPV18 E2. *Virology,* 242**,** 378-386.

DENG, G., NGUYEN, A., TANAKA, H., MATSUZAKI, K., BELL, I., MEHTA, K. R., TERDIMAN, J. P., WALDMAN, F. M., KAKAR, S., GUM, J., CRAWLEY, S., SLEISENGER, M. H. & KIM, Y. S. 2006. Regional hypermethylation and global hypomethylation are associated with

altered chromatin conformation and histone acetylation in colorectal cancer. *Int J Cancer,* 118**,** 2999-3005.

DING, D. C., CHIANG, M. H., LAI, H. C., HSIUNG, C. A., HSIEH, C. Y. & CHU, T. Y. 2009. Methylation of the long control region of HPV16 is related to the severity of cervical neoplasia. *Eur J Obstet Gynecol Reprod Biol,* 147**,** 215-20.

DOERFLER, W. 2006. De novo methylation, long-term promoter silencing, methylation patterns in the human genome, and consequences of foreign DNA insertion. *Curr Top Microbiol Immunol,* 301**,** 125-75.

DOERFLER, W. 2008. In pursuit of the first recognized epigenetic signal--DNA methylation: a 1976 to 2008 synopsis. *Epigenetics,* 3**,** 125-33.

DOERFLER, W. 2011. Epigenetic consequences of foreign DNA insertions: de novo methylation and global alterations of methylation patterns in recipient genomes. *Reviews in Medical Virology,* 21**,** 336-346.

DOERFLER, W., REMUS, R., MULLER, K., HELLER, H., HOHLWEG, U. & SCHUBBERT, R. 2001. The fate of foreign DNA in mammalian cells and organisms. *Dev Biol (Basel),* 106**,** 89-97; discussion 143-60.

DOLLARD, S. C., BROKER, T. R. & CHOW, L. T. 1993. Regulation of the human papillomavirus type 11 E6 promoter by viral and host transcription factors in primary human keratinocytes. *J Virol,* 67**,** 1721-6.

DONG, G., BROKER, T. R. & CHOW, L. T. 1994. Human papillomavirus type 11 E2 proteins repress the homologous E6 promoter by interfering with the binding of host transcription factors to adjacent elements. *J Virol,* 68**,** 1115-27.

DOORBAR, J. 2006. Molecular biology of human papillomavirus infection and cervical cancer. *Clin. Sci.,* 110**,** 525-541.

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

DOSTATNI, N., LAMBERT, P. F., SOUSA, R., HAM, J., HOWLEY, P. M. & YANIV, M. 1991. The functional BPV-1 E2 trans-activating protein can act as a repressor by preventing formation of the initiation complex. *Genes Dev,* 5**,** 1657-71.

DUENSING, S., LEE, L. Y., DUENSING, A., BASILE, J., PIBOONNIYOM, S.-O., GONZALEZ, S., CRUM, C. P. & MÃ¼NGER, K. 2000. The human papillomavirus type 16 E6 and E7 oncoproteins cooperate to induce mitotic defects and genomic instability by uncoupling centrosome duplication from the cell division cycle. *Proc Natl Acad Sci U S A,* 97**,** 10002-10007.

DUENSING, S. & MÜNGER, K. 2004. Mechanisms of genomic instability in human cancer: insights from studies with human papillomavirus oncoproteins. *Int J Cancer,* 109**,** 157-162.

DYSON, N., HOWLEY, P., MUNGER, K. & HARLOW, E. 1989. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science,* 243**,** 934-937.

EADS, C. A., DANENBERG, K. D., KAWAKAMI, K., SALTZ, L. B., BLAKE, C., SHIBATA, D., DANENBERG, P. V. & LAIRD, P. W. 2000. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res,* 28**,** E32.

ECKHARDT, F., LEWIN, J., CORTESE, R., RAKYAN, V. K., ATTWOOD, J., BURGER, M., BURTON, J., COX, T. V., DAVIES, R., DOWN, T. A., HAEFLIGER, C., HORTON, R., HOWE, K., JACKSON, D. K., KUNDE, J., KOENIG, C., LIDDLE, J., NIBLETT, D., OTTO, T., PETTETT, R., SEEMANN, S., THOMPSON, C., WEST, T., ROGERS, J., OLEK, A., BERLIN, K. & BECK, S. 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet,* 38**,** 1378-1385.

EDEN, S. & CEDAR, H. 1994. Role of DNA methylation in the regulation of transcription. *Curr Opin Genet Dev,* 4**,** 255-9.

EDWARDS, R. A. 2007. Laser capture microdissection of mammalian tissue. *J Vis Exp***,** 309.

EHRICH, M., NELSON, M. R., STANSSENS, P., ZABEAU, M., LILOGLOU, T., XINARIANOS, G., CANTOR, C. R., FIELD, J. K. & VAN DEN BOOM, D. 2005. Quantitative high-throughput

analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A,* 102**,** 15785-15790.

EIJSINK, J. J., LENDVAI, A., DEREGOWSKI, V., KLIP, H. G., VERPOOTEN, G., DEHASPE, L., DE BOCK, G. H., HOLLEMA, H., VAN CRIEKINGE, W., SCHUURING, E., VAN DER ZEE, A. G. & WISMAN, G. B. 2012. A four-gene methylation marker panel as triage test in high-risk human papillomavirus positive patients. *Int J Cancer,* 130**,** 1861-9.

EL-DEIRY, W. S., NELKIN, B. D., CELANO, P., YEN, R. W., FALCO, J. P., HAMILTON, S. R. & BAYLIN, S. B. 1991. High expression of the DNA methyltransferase gene characterizes human neoplastic cells and progression stages of colon cancer. *Proc Natl Acad Sci U S A,* 88**,** 3470-4.

EMILI, A., GREENBLATT, J. & INGLES, C. J. 1994. Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. *Mol Cell Biol,* 14**,** 1582-1593.

EMMERT-BUCK, M. R., BONNER, R. F., SMITH, P. D., CHUAQUI, R. F., ZHUANG, Z., GOLDSTEIN, S. R., WEISS, R. A. & LIOTTA, L. A. 1996. Laser Capture Microdissection. *Science,* 274**,** 998-1001.

ERNST, J. & KELLIS, M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotech,* 28**,** 817-825.

ESTELLER, M. 2002. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene,* 21**,** 5427-40.

ESTELLER, M. 2007a. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet,* 8**,** 286-98.

ESTELLER, M. 2007b. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum Mol Genet,* 16 Spec No 1**,** R50-9.

ESTELLER, M., CORN, P. G., BAYLIN, S. B. & HERMAN, J. G. 2001. A Gene Hypermethylation Profile of Human Cancer. *Cancer Research,* 61**,** 3225-3229.

ESTELLER, M., HAMILTON, S. R., BURGER, P. C., BAYLIN, S. B. & HERMAN, J. G. 1999. Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. *Cancer Res,* 59**,** 793-7.

ESTELLER, M. & HERMAN, J. G. 2004. Generating mutations but providing chemosensitivity: the role of O6-methylguanine DNA methyltransferase in human cancer. *Oncogene,* 23**,** 1-8.

ETZIONI, R., URBAN, N., RAMSEY, S., MCINTOSH, M., SCHWARTZ, S., REID, B., RADICH, J., ANDERSON, G. & HARTWELL, L. 2003. The case for early detection. *Nat Rev Cancer,* 3**,** 243-252.

FALLS, J. G., PULFORD, D. J., WYLIE, A. A. & JIRTLE, R. L. 1999. Genomic Imprinting: Implications for Human Disease. *The American Journal of Pathology,* 154**,** 635-647.

FEHRMANN, F., KLUMPP, D. J. & LAIMINS, L. A. 2003. Human papillomavirus type 31 E5 protein supports cell cycle progression and activates late viral functions upon epithelial differentiation. *J Virol,* 77**,** 2819-31.

FEINBERG, A. P. & TYCKO, B. 2004. The history of cancer epigenetics. *Nat Rev Cancer,* 4**,** 143-153.

FERNANDEZ, A. F. & ESTELLER, M. 2010. Viral epigenomes in human tumorigenesis. *Oncogene,* 29**,** 1405-20.

FERNANDEZ, A. F., ROSALES, C., LOPEZ-NIEVA, P., GRANA, O., BALLESTAR, E., ROPERO, S., ESPADA, J., MELO, S. A., LUJAMBIO, A., FRAGA, M. F., PINO, I., JAVIERRE, B., CARMONA, F. J., ACQUADRO, F., STEENBERGEN, R. D., SNIJDERS, P. J., MEIJER, C. J., PINEAU, P., DEJEAN, A., LLOVERAS, B., CAPELLA, G., QUER, J., BUTI, M., ESTEBAN, J. I., ALLENDE, H., RODRIGUEZ-FRIAS, F., CASTELLSAGUE, X., MINAROVITS, J., PONCE, J., CAPELLO, D., GAIDANO, G., CIGUDOSA, J. C., GOMEZ-LOPEZ, G., PISANO, D. G., VALENCIA, A., PIRIS, M. A., BOSCH, F. X., CAHIR-MCFARLAND, E., KIEFF, E. & ESTELLER,

M. 2009. The dynamic DNA methylomes of double-stranded DNA viruses associated with human cancer. *Genome Res,* 19**,** 438-51.

FIANDER, A. N., TRISTRAM, A. J., DAVIDSON, E. J., TOMLINSON, A. E., MAN, S., BALDWIN, P. J., STERLING, J. C. & KITCHENER, H. C. 2006. Prime-boost vaccination strategy in women with high-grade, noncervical anogenital intraepithelial neoplasia: clinical results from a multicenter phase II trial. *Int J Gynecol Cancer,* 16**,** 1075-81.

FLORES, E. R., ALLEN-HOFFMANN, B. L., LEE, D. & LAMBERT, P. F. 2000. The Human Papillomavirus Type 16 E7 Oncogene Is Required for the Productive Stage of the Viral Life Cycle. *J Virol,* 74**,** 6622-6631.

FLORES, E. R., ALLEN-HOFFMANN, B. L., LEE, D., SATTLER, C. A. & LAMBERT, P. F. 1999. Establishment of the Human Papillomavirus Type 16 (HPV-16) Life Cycle in an Immortalized Human Foreskin Keratinocyte Cell Line. *Virology,* 262**,** 344-354.

FLORES, E. R. & LAMBERT, P. F. 1997. Evidence for a switch in the mode of human papillomavirus type 16 DNA replication during the viral life cycle. *J Virol,* 71**,** 7167-79.

FRAGA, M. F., BALLESTAR, E., VILLAR-GAREA, A., BOIX-CHORNET, M., ESPADA, J., SCHOTTA, G., BONALDI, T., HAYDON, C., ROPERO, S., PETRIE, K., IYER, N. G., PEREZ-ROSADO, A., CALVO, E., LOPEZ, J. A., CANO, A., CALASANZ, M. J., COLOMER, D., PIRIS, M. A., AHN, N., IMHOF, A., CALDAS, C., JENUWEIN, T. & ESTELLER, M. 2005. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet,* 37**,** 391-400.

FRAZER, I. H. 2009. Interaction of human papillomaviruses with the host immune system: A well evolved relationship. *Virology,* 384**,** 410-414.

FROMMER, M., MCDONALD, L. E., MILLAR, D. S., COLLIS, C. M., WATT, F., GRIGG, G. W., MOLLOY, P. L. & PAUL, C. L. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A,* 89**,** 1827-1831.

GAGE, J. R., MEYERS, C. & WETTSTEIN, F. O. 1990. The E7 proteins of the nononcogenic human papillomavirus type 6b (HPV-6b) and of the oncogenic HPV-16 differ in retinoblastoma protein binding and other properties. *J Virol,* 64**,** 723-30.

GARTEL, A. L. & RADHAKRISHNAN, S. K. 2005. Lost in Transcription: p21 Repression, Mechanisms, and Consequences. *Cancer Res,* 65**,** 3980-3985.

GASCO, M., SULLIVAN, A., REPELLIN, C., BROOKS, L., FARRELL, P. J., TIDY, J. A., DUNNE, B., GUSTERSON, B., EVANS, D. J. & CROOK, T. 2002. Coincident inactivation of 14-3-3sigma and p16INK4a is an early event in vulval squamous neoplasia. *Oncogene,* 21**,** 1876-81.

GENTHER, S. M., STERLING, S., DUENSING, S., MUNGER, K., SATTLER, C. & LAMBERT, P. F. 2003. Quantitative Role of the Human Papillomavirus Type 16 E5 Gene during the Productive Stage of the Viral Life Cycle. *J Virol,* 77**,** 2832-2842.

GIARRE, M., CALDEIRA, S., MALANCHI, I., CICCOLINI, F., LEAO, M. J. & TOMMASINO, M. 2001. Induction of pRb degradation by the human papillomavirus type 16 E7 protein is essential to efficiently overcome p16INK4a-imposed G1 cell cycle Arrest. *J Virol,* 75**,** 4705-12.

GJERSET, R. A. & MARTIN, D. W., JR. 1982. Presence of a DNA demethylating activity in the nucleus of murine erythroleukemic cells. *J Biol Chem,* 257**,** 8581-3.

GLAXOSMITHKLINE. 2008. *Summary of Product Characteristics (SPC)* [Online]. Available: http://emc.medicines.org.uk/medicine/20204/SPC/Cervarix/ [Accessed 27/02/2009 2009].

GLOSS, B. & BERNARD, H. U. 1990. The E6/E7 promoter of human papillomavirus type 16 is activated in the absence of E2 proteins by a sequence-aberrant Sp1 distal element. *J Virol,* 64**,** 5577-5584.

GOK, M., HEIDEMAN, D. A., VAN KEMENADE, F. J., BERKHOF, J., ROZENDAAL, L., SPRUYT, J. W., VOORHORST, F., BELIEN, J. A., BABOVIC, M., SNIJDERS, P. J. & MEIJER, C. J. 2010. HPV

testing on self collected cervicovaginal lavage specimens as screening method for women who do not attend cervical screening: cohort study. *BMJ,* 340**,** c1040.

GOLL, M. G., KIRPEKAR, F., MAGGERT, K. A., YODER, J. A., HSIEH, C. L., ZHANG, X., GOLIC, K. G., JACOBSEN, S. E. & BESTOR, T. H. 2006. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science,* 311**,** 395-8.

GOODWIN, E. C., YANG, E., LEE, C.-J., LEE, H.-W., DIMAIO, D. & HWANG, E.-S. 2000. Rapid induction of senescence in human cervical carcinoma cells. *Proc Natl Acad Sci U S A,* 97**,** 10978-10983.

GRASSMANN, K., RAPP, B., MASCHEK, H., PETRY, K. U. & IFTNER, T. 1996. Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA. *Journal of Virology,* 70**,** 2339-49.

GRAY, E., PETT, M. R., WARD, D., WINDER, D. M., STANLEY, M. A., ROBERTS, I., SCARPINI, C. G. & COLEMAN, N. 2010. In vitro Progression of Human Papillomavirus 16 Episome-Associated Cervical Neoplasia Displays Fundamental Similarities to Integrant-Associated Carcinogenesis. *Cancer Research,* 70**,** 4081-4091.

HAFNER, N., DRIESCH, C., GAJDA, M., JANSEN, L., KIRCHMAYR, R., RUNNEBAUM, I. B. & DURST, M. 2008. Integration of the HPV16 genome does not invariably result in high levels of viral oncogene transcripts. *Oncogene,* 27**,** 1610-7.

HAGENSEE, M. E., OLSON, N. H., BAKER, T. S. & GALLOWAY, D. A. 1994. Three-dimensional structure of vaccinia virus-produced human papillomavirus type 1 capsids. *J Virol,* 68**,** 4503-5.

HAMM, S., JUST, G., LACOSTE, N., MOITESSIER, N., SZYF, M. & MAMER, O. 2008. On the mechanism of demethylation of 5-methylcytosine in DNA. *Bioorg Med Chem Lett,* 18**,** 1046-9.

HARRINGTON, M. A., JONES, P. A., IMAGAWA, M. & KARIN, M. 1988. Cytosine methylation does not affect binding of transcription factor Sp1. *Proc Natl Acad Sci U S A,* 85**,** 2066-70.

HART, W. R. 2001. Vulvar intraepithelial neoplasia: historical aspects and current status. *Int J Gynecol Pathol,* 20**,** 16-30.

HATTORI, N., IMAO, Y., NISHINO, K., OHGANE, J., YAGI, S., TANAKA, S. & SHIOTA, K. 2007. Epigenetic regulation of Nanog gene in embryonic stem and trophoblast stem cells. *Genes Cells,* 12**,** 387-96.

HATTORI, N., NISHINO, K., KO, Y. G., OHGANE, J., TANAKA, S. & SHIOTA, K. 2004. Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. *J Biol Chem,* 279**,** 17063-9.

HAWLEY-NELSON, P., VOUSDEN, K. H., HUBBERT, N. L., LOWY, D. R. & SCHILLER, J. T. 1989. HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *Embo J,* 8**,** 3905-10.

HEALTH PROTECTION AGENCY. 2012. Sexually transmitted infections in England, 2011. In Health Protection Agency Report. Vol 6, No.22. Available: http://www.hpa.org.uk/hpr/archives/2012/hpr2212.pdf [Accessed 01/06/2012].

HEGDE, R. S., GROSSMAN, S. R., LAIMINS, L. A. & SIGLER, P. B. 1992. Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DMA-binding domain bound to its DNA target. *Nature,* 359**,** 505-512.

HEJNAR, J., ELLEDER, D., HÁJKOVÁ, P., WALTER, J., BLAŽKOVÁ, J. & SVOBODA, J. 2003. Demethylation of host-cell DNA at the site of avian retrovirus integration. *Biochemical and Biophysical Research Communications,* 311**,** 641-648.

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

HELLEMANS, J., PREOBRAZHENSKA, O., WILLAERT, A., DEBEER, P., VERDONK, P. C., COSTA, T., JANSSENS, K., MENTEN, B., VAN ROY, N., VERMEULEN, S. J., SAVARIRAYAN, R., VAN HUL, W., VANHOENACKER, F., HUYLEBROECK, D., DE PAEPE, A., NAEYAERT, J. M., VANDESOMPELE, J., SPELEMAN, F., VERSCHUEREN, K., COUCKE, P. J. & MORTIER, G. R. 2004. Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. *Nat Genet,* 36**,** 1213-8.

HELLER, H., KAMMER, C., WILGENBUS, P. & DOERFLER, W. 1995. Chromosomal insertion of foreign (adenovirus type 12, plasmid, or bacteriophage lambda) DNA is associated with enhanced methylation of cellular DNA segments. *Proc Natl Acad Sci U S A,* 92**,** 5515-9.

HELLMAN, A. & CHESS, A. 2007. Gene Body-Specific Methylation on the Active X Chromosome. *Science,* 315**,** 1141-1143.

HENKEN, F. E., WILTING, S. M., OVERMEER, R. M., VAN RIETSCHOTEN, J. G., NYGREN, A. O., ERRAMI, A., SCHOUTEN, J. P., MEIJER, C. J., SNIJDERS, P. J. & STEENBERGEN, R. D. 2007. Sequential gene promoter methylation during HPV-induced cervical carcinogenesis. *Br J Cancer,* 97**,** 1457-64.

HERDMAN, M. T., PETT, M. R., ROBERTS, I., ALAZAWI, W. O., TESCHENDORFF, A. E., ZHANG, X. Y., STANLEY, M. A. & COLEMAN, N. 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**,** 2341-53.

HERMAN, J. G., GRAFF, J. R., MYÖHÄNEN, S., NELKIN, B. D. & BAYLIN, S. B. 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A,* 93**,** 9821-9826.

HERRICK, J., CONTI, C., TEISSIER, S., THIERRY, F., COUTURIER, J., SASTRE-GARAU, X., FAVRE, M., ORTH, G. & BENSIMON, A. 2005. Genomic organization of amplified MYC genes suggests distinct mechanisms of amplification in tumorigenesis. *Cancer Res,* 65**,** 1174-9.

HIBBITTS, S., JONES, J., POWELL, N., DALLIMORE, N., MCREA, J., BEER, H., TRISTRAM, A., FIELDER, H. & FIANDER, A. N. 2008. Human papillomavirus prevalence in women attending routine cervical screening in South Wales, UK: a cross-sectional study. *Br J Cancer,* 99**,** 1929-33.

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

HILDESHEIM A, H. R. W. S. & *et al.* 2007. Effect of human papillomavirus 16/18 l1 viruslike particle vaccine among young women with preexisting infection: A randomized trial. *JAMA: The Journal of the American Medical Association,* 298**,** 743-753.

HILLEMANNS, P. & WANG, X. 2006. Integration of HPV-16 and HPV-18 DNA in vulvar intraepithelial neoplasia. *Gynecol Oncol,* 100**,** 276-82.

HO, C. M., LEE, B. H., CHANG, S. F., CHIEN, T. Y., HUANG, S. H., YAN, C. C. & CHENG, W. F. 2011. Integration of human papillomavirus correlates with high levels of viral oncogene transcripts in cervical carcinogenesis. *Virus Res,* 161**,** 124-30.

HO, G. Y., BIERMAN, R., BEARDSLEY, L., CHANG, C. J. & BURK, R. D. 1998. Natural history of cervicovaginal papillomavirus infection in young women. *N Engl J Med,* 338**,** 423-8.

HO, G. Y. F., BURK, R. D., KLEIN, S., KADISH, A. S., CHANG, C. J., PALAN, P., BASU, J., TACHEZY, R., LEWIS, R. & ROMNEY, S. 1995. Persistent Genital Human Papillomavirus Infection as a Risk Factor for Persistent Cervical Dysplasia. *Journal of the National Cancer Institute,* 87**,** 1365-1371.

HOCHSTEIN, N., MUIZNIEKS, I., MANGEL, L., BRONDKE, H. & DOERFLER, W. 2007. Epigenetic Status of an Adenovirus Type 12 Transgenome upon Long-Term Cultivation in Hamster Cells. *Journal of Virology,* 81**,** 5349-5361.

HOEBEECK, J., VAN DER LUIJT, R., POPPE, B., DE SMET, E., YIGIT, N., CLAES, K., ZEWALD, R., DE JONG, G. J., DE PAEPE, A., SPELEMAN, F. & VANDESOMPELE, J. 2005. Rapid detection of VHL exon deletions using real-time quantitative PCR. *Lab Invest,* 85**,** 24-33.

HOHN, T., CORSTEN, S., RIEKE, S., MULLER, M. & ROTHNIE, H. 1996. Methylation of coding region alone inhibits gene expression in plant protoplasts. *Proc Natl Acad Sci U S A,* 93**,** 8334-9.

HOLLAND, P. M., ABRAMSON, R. D., WATSON, R. & GELFAND, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci U S A,* 88**,** 7276-80.

HONG, D., YE, F., LU, W., HU, Y., WAN, X., CHEN, Y. & XIE, X. 2008. Methylation status of the long control region of HPV 16 in clinical cervical specimens. *Molecular Medicine Reports,* 1**,** 555-560.

HOPMAN, A. H. N., SMEDTS, F., DIGNEF, W., UMMELEN, M., SONKE, G., MRAVUNAC, M., VOOIJS, G. P., SPEEL, E.-J. M. & RAMAEKERS, F. C. S. 2004. Transition of high-grade cervical intraepithelial neoplasia to micro-invasive carcinoma is characterized by integration of HPV 16/18 and numerical chromosome abnormalities. *The Journal of Pathology,* 202**,** 23-33.

HOU, S. Y., WU, S. Y., ZHOU, T., THOMAS, M. C. & CHIANG, C. M. 2000. Alleviation of human papillomavirus E2-mediated transcriptional repression via formation of a TATA binding protein (or TFIID)-TFIIB-RNA polymerase II-TFIIF preinitiation complex. *Mol Cell Biol,* 20**,** 113-25.

HOWLEY, P. M. & LOWY, D. R. 2006. *Papillomaviruses. In Field's Virology*, Lippincott Williams and Wilkins.

HSU, C. H., PENG, K. L., JHANG, H. C., LIN, C. H., WU, S. Y., CHIANG, C. M., LEE, S. C., YU, W. C. & JUAN, L. J. 2012. The HPV E6 oncoprotein targets histone methyltransferases for modulating specific gene transcription. *Oncogene,* 31**,** 2335-49.

HU, J. F., PHAM, J., DEY, I., LI, T., VU, T. H. & HOFFMAN, A. R. 2000. Allele-specific histone acetylation accompanies genomic imprinting of the insulin-like growth factor II receptor gene. *Endocrinology,* 141**,** 4428-35.

HUBLAROVA, P., HRSTKA, R., ROTTEROVA, P., ROTTER, L., COUPKOVA, M., BADAL, V., NENUTIL, R. & VOJTESEK, B. 2009. Prediction of human papillomavirus 16 e6 gene expression and cervical intraepithelial neoplasia progression by methylation status. *Int J Gynecol Cancer,* 19**,** 321-5.

HUGHES, F. J. & ROMANOS, M. A. 1993. E1 protein of human papillomavirus is a DNA helicase/ATPase. *Nucl. Acids Res.,* 21**,** 5817-5823.

HWANG, E. S., NOTTOLI, T. & DIMAIO, D. 1995. The HPV16 E5 protein: expression, detection, and stable complex formation with transmembrane proteins in COS cells. *Virology,* 211**,** 227-33.

IVANOVA, T. A., GOLOVINA, D. A., ZAVALISHINA, L. E., VOLGAREVA, G. M., KATARGIN, A. N., ANDREEVA, Y. Y., FRANK, G. A., KISSELJOV, F. L. & KISSELJOVA, N. P. 2007. Up-regulation of expression and lack of 5' CpG island hypermethylation of p16 INK4a in HPV-positive cervical carcinomas. *BMC Cancer,* 7**,** 47.

JACOBS, M. V., SNIJDERS, P. J., VAN DEN BRULE, A. J., HELMERHORST, T. J., MEIJER, C. J. & WALBOOMERS, J. M. 1997. A general primer GP5+/GP6(+)-mediated PCR-enzyme immunoassay method for rapid detection of 14 high-risk and 6 low-risk human papillomavirus genotypes in cervical scrapings. *J Clin Microbiol,* 35**,** 791-5.

JCNI 2010. StatBite Proportion of Specific Cancers Caused by HPV. *Journal of the National Cancer Institute,* 102**,** 839.

JEON, S., ALLEN-HOFFMANN, B. & LAMBERT, P. 1995. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J Virol,* 69**,** 2989-2997.

JEON, S. & 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**,** 1654-1658.

JIANG, Y. L., RIGOLET, M., BOURC'HIS, D., NIGON, F., BOKESOY, I., FRYNS, J. P., HULTEN, M., JONVEAUX, P., MARASCHIO, P., MEGARBANE, A., MONCLA, A. & VIEGAS-PEQUIGNOT, E. 2005. DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. *Hum Mutat,* 25**,** 56-63.

JIRICNY, J. 1996. Mismatch repair and cancer. *Cancer Surveys: Genetic Instability in Cancer. Imperial Cancer Research Fund,* 28**,** 47-68.

JONES, P. A. 1996. DNA Methylation Errors and Cancer. *Cancer Res,* 56**,** 2463-2467.

JONES, P. A. 1999. The DNA methylation paradox. *Trends Genet,* 15**,** 34-7.

JONES, P. A. & LIANG, G. 2009. Rethinking how DNA methylation patterns are maintained. *Nature reviews. Genetics,* 10**,** 805-11.

JONES, P. L., WADE, P. A. & WOLFFE, A. P. 2001. Purification of the MeCP2/Histone Deacetylase Complex from Xenopus laevis. *Methods Mol Biol,* 181**,** 297-307.

JOURA, E. A. 2002. Epidemiology, diagnosis and treatment of vulvar intraepithelial neoplasia. *Curr Opin Obstet Gynecol,* 14**,** 39-43.

JOURA, E. A., LEODOLTER, S., HERNANDEZ-AVILA, M., WHEELER, C. M., PEREZ, G., KOUTSKY, L. A., GARLAND, S. M., HARPER, D. M., TANG, G. W., FERRIS, D. G., STEBEN, M., JONES, R. W., BRYAN, J., TADDEO, F. J., BAUTISTA, O. M., ESSER, M. T., SINGS, H. L., NELSON, M., BOSLEGO, J. W., SATTLER, C., BARR, E. & PAAVONEN, J. 2007. Efficacy of a quadrivalent prophylactic human papillomavirus (types 6, 11, 16, and 18) L1 virus-like-particle vaccine against high-grade vulval and vaginal lesions: a combined analysis of three randomised clinical trials. *Lancet,* 369**,** 1693-702.

JÜTTERMANN, R., LI, E. & JAENISCH, R. 1994. Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proceedings of the National Academy of Sciences,* 91**,** 11797-11801.

KADAJA, M., SILLA, T., USTAV, E. & USTAV, M. 2009. Papillomavirus DNA replication -- From initiation to genomic instability. *Virology,* 384**,** 360-368.

KADAJA, M., SUMERINA, A., VERST, T., OJARAND, M., USTAV, E. & USTAV, M. 2007. Genomic instability of the host cell induced by the human papillomavirus replication machinery. *Embo J,* 26**,** 2180-91.

KAGIE, M. J., KENTER, G. G., HERMANS, J., TRIMBOS, J. B. & FLEUREN, G. J. 1997. The relevance of various vulvar epithelial changes in the early detection of squamous cell carcinoma of the vulva. *International Journal of Gynecological Cancer,* 7**,** 50-57.

KAHN, J. A. & BURK, R. D. 2007. Papillomavirus vaccines in perspective. *The Lancet,* 369**,** 2135-2137.

KALANTARI, M., CALLEJA-MACIAS, I. E., TEWARI, D., HAGMAR, B., LIE, K., BARRERA-SALDANA, H. A., WILEY, D. J. & BERNARD, H. U. 2004. Conserved methylation patterns of human papillomavirus type 16 DNA in asymptomatic infection and cervical neoplasia. *J Virol,* 78**,** 12762-72.

KALANTARI, M., CHASE, D. M., TEWARI, K. S. & BERNARD, H. U. 2010. Recombination of human papillomavirus-16 and host DNA in exfoliated cervical cells: a pilot study of L1 gene methylation and chromosomal integration as biomarkers of carcinogenic progression. *J Med Virol,* 82**,** 311-20.

KALANTARI, M., GARCIA-CARRANCA, A., MORALES-VAZQUEZ, C. D., ZUNA, R., MONTIEL, D. P., CALLEJA-MACIAS, I. E., JOHANSSON, B., ANDERSSON, S. & BERNARD, H. U. 2009. Laser capture microdissection of cervical human papillomavirus infections: copy number of the virus in cancerous and normal tissue and heterogeneous DNA methylation. *Virology,* 390**,** 261-7.

KALANTARI, M., LEE, D., CALLEJA-MACIAS, I. E., LAMBERT, P. F. & BERNARD, H.-U. 2008a. Effects of cellular differentiation, chromosomal integration and 5-aza-2'-deoxycytidine treatment on human papillomavirus-16 DNA methylation in cultured cell lines. *Virology,* 374**,** 292-303.

KALANTARI, M., VILLA, L. L., CALLEJA-MACIAS, I. E. & BERNARD, H. U. 2008b. Human papillomavirus-16 and -18 in penile carcinomas: DNA methylation, chromosomal recombination and genomic variation. *Int J Cancer,* 123**,** 1832-40.

KARLIC, R., CHUNG, H. R., LASSERRE, J., VLAHOVICEK, K. & VINGRON, M. 2010. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A,* 107**,** 2926-31.

KATKI, H. A., KINNEY, W. K., FETTERMAN, B., LOREY, T., POITRAS, N. E., CHEUNG, L., DEMUTH, F., SCHIFFMAN, M., WACHOLDER, S. & CASTLE, P. E. 2011. Cervical cancer risk for women undergoing concurrent testing for human papillomavirus and cervical cytology: a population-based study in routine clinical practice. *Lancet Oncol,* 12**,** 663-72.

KAWAKAMI, K., BRABENDER, J., LORD, R. V., GROSHEN, S., GREENWALD, B. D., KRASNA, M. J., YIN, J., FLEISHER, A. S., ABRAHAM, J. M., BEER, D. G., SIDRANSKY, D., HUSS, H. T., DEMEESTER, T. R., EADS, C., LAIRD, P. W., ILSON, D. H., KELSEN, D. P., HARPOLE, D., MOORE, M. B., DANENBERG, K. D., DANENBERG, P. V. & MELTZER, S. J. 2000. Hypermethylated APC DNA in plasma and prognosis of patients with esophageal adenocarcinoma. *J Natl Cancer Inst,* 92**,** 1805-11.

KIM, K., GARNER-HAMRICK, P. A., FISHER, C., LEE, D. & LAMBERT, P. F. 2003. Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection. *J Virol,* 77**,** 12450-9.

KIM, K. Y., BLATT, L. & TAYLOR, M. W. 2000. The effects of interferon on the expression of human papillomavirus oncogenes. *J Gen Virol,* 81**,** 695-700.

KIM, M.-S., KONDO, T., TAKADA, I., YOUN, M.-Y., YAMAMOTO, Y., TAKAHASHI, S., MATSUMOTO, T., FUJIYAMA, S., SHIRODE, Y., YAMAOKA, I., KITAGAWA, H., TAKEYAMA, K.-I., SHIBUYA, H., OHTAKE, F. & KATO, S. 2009. DNA demethylation in hormone-induced transcriptional derepression. *Nature,* 461**,** 1007-1012.

KINES, R. C., THOMPSON, C. D., LOWY, D. R., SCHILLER, J. T. & DAY, P. M. 2009. The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding. *Proc Natl Acad Sci U S A,* 106**,** 20458-63.

KIYONO, T., FOSTER, S. A., KOOP, J. I., MCDOUGALL, J. K., GALLOWAY, D. A. & KLINGELHUTZ, A. J. 1998. Both Rb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells. *Nature,* 396**,** 84-8.

KLAES, R., WOERNER, S. M., RIDDER, R., WENTZENSEN, N., DUERST, M., SCHNEIDER, A., LOTZ, B., MELSHEIMER, P. & DOEBERITZ, M. V. K. 1999. Detection of High-Risk Cervical Intraepithelial Neoplasia and Cervical Cancer by Amplification of Transcripts Derived from Integrated Papillomavirus Oncogenes. *Cancer Res,* 59**,** 6132-6136.

KLAUSNER, R. D. 2002. The fabric of cancer cell biology—Weaving together the strands. *Cancer Cell,* 1**,** 3-10.

KLUMPP, D. J. & LAIMINS, L. A. 1999. Differentiation-induced changes in promoter usage for transcripts encoding the human papillomavirus type 31 replication protein E1. *Virology,* 257**,** 239-46.

KOFFA, M. D., GRAHAM, S. V., TAKAGAKI, Y., MANLEY, J. L. & CLEMENTS, J. B. 2000. The human papillomavirus type 16 negative regulatory RNA element interacts with three proteins that act at different posttranscriptional levels. *Proc Natl Acad Sci U S A,* 97**,** 4677-82.

KOK, I. M. C. M. D., ROSMALEN, J. V., DILLNER, J., ARBYN, M., SASIENI, P., IFTNER, T. & BALLEGOOIJEN, M. V. 2012. Primary screening for human papillomavirus compared with cytology screening for cervical cancer in European settings: cost effectiveness analysis based on a Dutch microsimulation model. *BMJ,* 344.

KOROMILAS, A. E., LI, S. & MATLASHEWSKI, G. 2001. Control of interferon signaling in human papillomavirus infection. *Cytokine Growth Factor Rev,* 12**,** 157-70.

KURMAN RJ, H. D. E. H. A. L. & *et al.* 1994. Interim guidelines for management of abnormal cervical cytology. *JAMA: The Journal of the American Medical Association,* 271**,** 1866-1869.

KURODA, A., RAUCH, T. A., TODOROV, I., KU, H. T., AL-ABDULLAH, I. H., KANDEEL, F., MULLEN, Y., PFEIFER, G. P. & FERRERI, K. 2009. Insulin Gene Expression Is Regulated by DNA Methylation. *PLoS ONE,* 4**,** e6953.

LACEY, C. J. N., LOWNDES, C. M. & SHAH, K. V. 2006. Chapter 4: Burden and management of non-cancerous HPV-related conditions: HPV-6/11 disease. *Vaccine,* 24, Supplement 3**,** S35-S41.

LAIRD, P. W., JACKSON-GRUSBY, L., FAZELI, A., DICKINSON, S. L., JUNG, W. E., LI, E., WEINBERG, R. A. & JAENISCH, R. 1995. Suppression of intestinal neoplasia by DNA hypomethylation. *Cell,* 81**,** 197-205.

LAMBERT, P. F., MONK, B. C. & HOWLEY, P. M. 1990. Phenotypic analysis of bovine papillomavirus type 1 E2 repressor mutants. *J Virol,* 64**,** 950-956.

LANE, N., DEAN, W., ERHARDT, S., HAJKOVA, P., SURANI, A., WALTER, J. & REIK, W. 2003. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *genesis,* 35**,** 88-93.

LAW, J. A. & JACOBSEN, S. E. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet,* 11**,** 204-20.

LEHMAN, C. W. & BOTCHAN, M. R. 1998. Segregation of viral plasmids depends on tethering to chromosomes and is regulated by phosphorylation. *Proc Natl Acad Sci U S A,* 95**,** 4338-43.

LEIBOWITCH, M., NEILL, S., PELISSE, M. & MOYAL-BARACCO, M. 1990. The epithelial changes associated with squamous cell carcinoma of the vulva: a review of the clinical, histological and viral findings in 78 women. *Br J Obstet Gynaecol,* 97**,** 1135-9.

LEONARD, S. M., WEI, W., COLLINS, S. I., PEREIRA, M., DIYAF, A., CONSTANDINOU-WILLIAMS, C., YOUNG, L. S., ROBERTS, S. & WOODMAN, C. B. 2012. Oncogenic human papillomavirus imposes an instructive pattern of DNA methylation changes which parallel the natural history of cervical HPV infection in young women. *Carcinogenesis*.

LESNIKOVA, I., LIDANG, M., HAMILTON-DUTOIT, S. & KOCH, J. 2009. p16 as a diagnostic marker of cervical neoplasia: a tissue microarray study of 796 archival specimens. *Diagn Pathol,* 4**,** 22.

LEY, T. J., DING, L., WALTER, M. J., MCLELLAN, M. D., LAMPRECHT, T., LARSON, D. E., KANDOTH, C., PAYTON, J. E., BATY, J., WELCH, J., HARRIS, C. C., LICHTI, C. F., TOWNSEND, R. R., FULTON, R. S., DOOLING, D. J., KOBOLDT, D. C., SCHMIDT, H., ZHANG, Q., OSBORNE, J. R., LIN, L., O'LAUGHLIN, M., MCMICHAEL, J. F., DELEHAUNTY, K. D., MCGRATH, S. D., FULTON, L. A., MAGRINI, V. J., VICKERY, T. L., HUNDAL, J., COOK, L. L., CONYERS, J. J., SWIFT, G. W., REED, J. P., ALLDREDGE, P. A., WYLIE, T., WALKER, J., KALICKI, J., WATSON, M. A., HEATH, S., SHANNON, W. D., VARGHESE, N., NAGARAJAN, R., WESTERVELT, P., TOMASSON, M. H., LINK, D. C., GRAUBERT, T. A., DIPERSIO, J. F., MARDIS, E. R. & WILSON, R. K. 2010. DNMT3A mutations in acute myeloid leukemia. *The New England journal of medicine,* 363**,** 2424-33.

LI, E. 2002. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet,* 3**,** 662-673.

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

LICHTENBERG, U., ZOCK, C. & DOERFLER, W. 1988. Integration of foreign DNA into mammalian genome can be associated with hypomethylation at site of insertion. *Virus Research,* 11**,** 335-342.

LILOGLOU, T., BEDIAGA, N. G., BROWN, B. R. B., FIELD, J. K. & DAVIES, M. P. A. 2012. Epigenetic biomarkers in lung cancer. *Cancer Letters*.

LIN, B. Y., MAKHOV, A. M., GRIFFITH, J. D., BROKER, T. R. & CHOW, L. T. 2002. Chaperone Proteins Abrogate Inhibition of the Human Papillomavirus (HPV) E1 Replicative Helicase by the HPV E2 Protein. *Mol Cell Biol,* 22**,** 6592-6604.

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q.-M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462**,** 315-322.

LIU, W. & SAINT, D. A. 2002. A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Anal Biochem,* 302**,** 52-9.

LIU, X., DAKIC, A., ZHANG, Y., DAI, Y., CHEN, R. & SCHLEGEL, R. 2009. HPV E6 protein interacts physically and functionally with the cellular telomerase complex. *Proceedings of the National Academy of Sciences,* 106**,** 18780-18785.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods,* 25**,** 402-408.

LOEYS, B. L., CHEN, J., NEPTUNE, E. R., JUDGE, D. P., PODOWSKI, M., HOLM, T., MEYERS, J., LEITCH, C. C., KATSANIS, N., SHARIFI, N., XU, F. L., MYERS, L. A., SPEVAK, P. J., CAMERON, D. E., DE BACKER, J., HELLEMANS, J., CHEN, Y., DAVIS, E. C., WEBB, C. L., KRESS, W., COUCKE, P., RIFKIN, D. B., DE PAEPE, A. M. & DIETZ, H. C. 2005. A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. *Nat Genet,* 37**,** 275-81.

LONGWORTH, M. S. & LAIMINS, L. A. 2004. Pathogenesis of Human Papillomaviruses in Differentiating Epithelia. *Microbiol. Mol. Biol. Rev.,* 68**,** 362-372.

LORINCZ, M. C., DICKERSON, D. R., SCHMITT, M. & GROUDINE, M. 2004. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat Struct Mol Biol,* 11**,** 1068-75.

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

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8[thinsp]A resolution. *Nature,* 389**,** 251-260.

LUONG, D. 2009. *Basic Principles of Genetics: Lecture 2, Genes are mostly located on chromosomes* [Online]. Available: http://cnx.org/content/m26565/latest/ [Accessed 12/09/2012 2012].

MADISON, K. C. 2003. Barrier function of the skin: "la raison d'etre" of the epidermis. *J Invest Dermatol,* 121**,** 231-41.

MANNIK, A., RUNKORG, K., JAANSON, N., USTAV, M. & USTAV, E. 2002. Induction of the bovine papillomavirus origin "onion skin"-type DNA replication at high E1 protein concentrations in vivo. *J Virol,* 76**,** 5835-45.

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

MAUNAKEA, A. K., NAGARAJAN, R. P., BILENKY, M., BALLINGER, T. J., D/'SOUZA, C., FOUSE, S. D., JOHNSON, B. E., HONG, C., NIELSEN, C., ZHAO, Y., TURECKI, G., DELANEY, A.,

VARHOL, R., THIESSEN, N., SHCHORS, K., HEINE, V. M., ROWITCH, D. H., XING, X., FIORE, C., SCHILLEBEECKX, M., JONES, S. J. M., HAUSSLER, D., MARRA, M. A., HIRST, M., WANG, T. & COSTELLO, J. F. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature,* 466**,** 253-257.

MAYER, W., NIVELEAU, A., WALTER, J., FUNDELE, R. & HAAF, T. 2000. Demethylation of the zygotic paternal genome. *Nature,* 403**,** 501-2.

MCCABE, M. T., DAVIS, J. N. & DAY, M. L. 2005. Regulation of DNA methyltransferase 1 by the pRb/E2F1 pathway. *Cancer Res,* 65**,** 3624-32.

MCCLUGGAGE, W. G. 2009. Recent developments in vulvovaginal pathology. *Histopathology,* 54**,** 156-73.

MCGRATH, J. & SOLTER, D. 1984. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell,* 37**,** 179-183.

MCINDOE, W., MCLEAN, M., JONES, R. W. & MULLINS, P. 1984. The invasive potential of carcinoma in situ of the cervix. *Joural of The American College of Obstetricians and Gynecologists,* 64**,** 451-458.

MCPHILLIPS, M. G., VEERAPRADITSIN, T., CUMMING, S. A., KARALI, D., MILLIGAN, S. G., BONER, W., MORGAN, I. M. & GRAHAM, S. V. 2004. SF2/ASF Binds the Human Papillomavirus Type 16 Late RNA Control Element and Is Regulated during Differentiation of Virus-Infected Epithelial Cells. *J Virol,* 78**,** 10598-10605.

MEIJER, C. J. L. M., SNIJDERS, P. J. F. & BRULE, A. 2000. Screening for cervical cancer: Should we test for infection with high-risk HPV? *Canadian Medical Association Journal,* 163**,** 535-538.

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)**,** 1725-33.

MELSHEIMER, P., VINOKUROVA, S., WENTZENSEN, N., BASTERT, G. & VON KNEBEL DOEBERITZ, M. 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**,** 3059-63.

METIVIER, R., GALLAIS, R., TIFFOCHE, C., LE PERON, C., JURKOWSKA, R. Z., CARMOUCHE, R. P., IBBERSON, D., BARATH, P., DEMAY, F., REID, G., BENES, V., JELTSCH, A., GANNON, F. & SALBERT, G. 2008. Cyclical DNA methylation of a transcriptionally active promoter. *Nature,* 452**,** 45-50.

MIDDLETON, K., PEH, W., SOUTHERN, S., GRIFFIN, H., SOTLAR, K., NAKAHARA, T., EL-SHERIF, A., MORRIS, L., SETH, R., HIBMA, M., JENKINS, D., LAMBERT, P., COLEMAN, N. & DOORBAR, J. 2003. Organization of Human Papillomavirus Productive Cycle during Neoplastic Progression Provides a Basis for Selection of Diagnostic Markers. *J Virol,* 77**,** 10186-10201.

MINCHEVA, A., GISSMANN, L. & ZUR HAUSEN, H. 1987. Chromosomal integration sites of human papillomavirus DNA in three cervical cancer cell lines mapped by in situ hybridization. *Med Microbiol Immunol,* 176**,** 245-56.

MIRABELLO, L., SUN, C., GHOSH, A., RODRIGUEZ, A. C., SCHIFFMAN, M., WENTZENSEN, N., HILDESHEIM, A., HERRERO, R., WACHOLDER, S., LORINCZ, A. & BURK, R. D. 2012. Methylation of Human Papillomavirus Type 16 Genome and Risk of Cervical Precancer in a Costa Rican Population. *Journal of the National Cancer Institute,* 104**,** 556-565.

MIRZA, S., SHARMA, G., PANDYA, P. & RALHAN, R. 2010. Demethylating agent 5-aza-2-deoxycytidine enhances susceptibility of breast cancer cells to anticancer agents. *Molecular and Cellular Biochemistry,* 342**,** 101-109.

MODIS, Y., TRUS, B. L. & HARRISON, S. C. 2002. Atomic model of the papillomavirus capsid. *Embo J,* 21**,** 4754-62.

MOHN, F., WEBER, M., REBHAN, M., ROLOFF, T. C., RICHTER, J., STADLER, M. B., BIBEL, M. & SCHUBELER, D. 2008. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell,* 30**,** 755-66.

MOMPARLER, R. 2005. Pharmacology of 5-Aza-2'-deoxycytidine (decitabine). *Seminars in Hematology,* 42, Supplement 2**,** S9-S16.

MONK, M., BOUBELIK, M. & LEHNERT, S. 1987. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development,* 99**,** 371-382.

MORI, T., O'DAY, S. J., UMETANI, N., MARTINEZ, S. R., KITAGO, M., KOYANAGI, K., KUO, C., TAKESHIMA, T. L., MILFORD, R., WANG, H. J., VU, V. D., NGUYEN, S. L. & HOON, D. S. 2005. Predictive utility of circulating methylated DNA in serum of melanoma patients receiving biochemotherapy. *J Clin Oncol,* 23**,** 9351-8.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods,* 5**,** 621-8.

MOSS, S., GRAY, A., LEGOOD, R., VESSEY, M., PATNICK, J. & KITCHENER, H. 2006. Effect of testing for human papillomavirus as a triage during screening for cervical cancer: observational before and after study. *BMJ,* 332**,** 83-85.

MULLER, H. M., WIDSCHWENDTER, A., FIEGL, H., IVARSSON, L., GOEBEL, G., PERKMANN, E., MARTH, C. & WIDSCHWENDTER, M. 2003. DNA methylation in serum of breast cancer patients: an independent prognostic marker. *Cancer Res,* 63**,** 7641-5.

MÜNGER, K. & HOWLEY, P. M. 2002. Human papillomavirus immortalization and transformation functions. *Virus Research,* 89**,** 213-228.

MUÑOZ, N., BOSCH, F. X., CASTELLSAGUÉ, X., DÍAZ, M., DE SANJOSE, S., HAMMOUDA, D., SHAH, K. V. & MEIJER, C. J. L. M. 2004. Against which human papillomavirus types shall we vaccinate and screen? the international perspective. *International Journal of Cancer,* 111**,** 278-285.

MURRELL, A., HEESON, S., BOWDEN, L., CONSTANCIA, M., DEAN, W., KELSEY, G. & REIK, W. 2001. An intragenic methylated region in the imprinted Igf2 gene augments transcription. *EMBO Rep,* 2**,** 1101-6.

NAKAHARA, T., PEH, W. L., DOORBAR, J., LEE, D. & LAMBERT, P. F. 2005. Human Papillomavirus Type 16 E1{wedge}E4 Contributes to Multiple Facets of the Papillomavirus Life Cycle. *J Virol,* 79**,** 13150-13165.

NATURE BIOTECHOLOGY 2010. Glossary. *Nat Biotech,* 28**,** 434-435.

NAUCLER, P., RYD, W., TÖRNBERG, S., STRAND, A., WADELL, G., ELFGREN, K., RÅDBERG, T., STRANDER, B., FORSLUND, O., HANSSON, B.-G., HAGMAR, B., JOHANSSON, B., RYLANDER, E. & DILLNER, J. 2009. Efficacy of HPV DNA Testing With Cytology Triage and/or Repeat HPV DNA Testing in Primary Cervical Cancer Screening. *Journal of the National Cancer Institute,* 101**,** 88-99.

NCBI. 1993. *Human papillomavirus - 16, complete genome* [Online]. Available: http://www.ncbi.nlm.nih.gov/sites/entrez?Db=genome&Cmd=ShowDetailView&TermToSearch=10337 [Accessed 30/01/2009 2009].

NCBI. 2009. *Entrez Gene: TERT telomerase reverse transcriptase [Homo sapiens]* [Online]. Available: http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=7015&ordinalpos=1&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum [Accessed 06/03/2009 2009].

NEPHEW, K. P. 2009. What will it take to obtain DNA methylation markers for early cervical cancer detection? *Gynecologic Oncology,* 112**,** 291-292.

NHS CERVICAL SCREENING PROGRAMME. 2012. *Modules from Remedica cervical screening DVD* [Online]. Available: http://www.cancerscreening.nhs.uk/cervical/publications/video/index.html [Accessed 30/08/2012 2012].

NORDGARD, O., KVALOY, J. T., FARMEN, R. K. & HEIKKILA, R. 2006. Error propagation in relative real-time reverse transcription polymerase chain reaction quantification models: the balance between accuracy and precision. *Anal Biochem,* 356**,** 182-93.

O'BRIEN, P. M. & SAVERIA CAMPO, M. 2002. Evasion of host immunity directed by papillomavirus-encoded proteins. *Virus Research,* 88**,** 103-117.

O'CONNOR, M. J., STUNKEL, W., KOH, C. H., ZIMMERMANN, H. & BERNARD, H. U. 2000. The differentiation-specific factor CDP/Cut represses transcription and replication of human papillomaviruses through a conserved silencing element. *J Virol,* 74**,** 401-10.

O'CONNOR, M. J., TAN, S. H., TAN, C. H. & BERNARD, H. U. 1996. YY1 represses human papillomavirus type 16 transcription by quenching AP-1 activity. *J Virol,* 70**,** 6529-39.

OH, S. T., KYO, S. & LAIMINS, L. A. 2001. Telomerase Activation by Human Papillomavirus Type 16 E6 Protein: Induction of Human Telomerase Reverse Transcriptase Expression through Myc and GC-Rich Sp1 Binding Sites. *Journal of Virology,* 75**,** 5559-5566.

OKANO, M., BELL, D. W., HABER, D. A. & LI, E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* 99**,** 247-57.

OLEK, A. & WALTER, J. 1997. The pre-implantation ontogeny of the H19 methylation imprint. *Nat Genet,* 17**,** 275-276.

OOI, S. K. & BESTOR, T. H. 2008. The colorful history of active DNA demethylation. *Cell,* 133**,** 1145-8.

OOI, S. K., QIU, C., BERNSTEIN, E., LI, K., JIA, D., YANG, Z., ERDJUMENT-BROMAGE, H., TEMPST, P., LIN, S. P., ALLIS, C. D., CHENG, X. & BESTOR, T. H. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature,* 448**,** 714-7.

OREND, G., KNOBLAUCH, M., KAMMER, C., TJIA, S. T., SCHMITZ, B., LINKWITZ, A., MEYER, G., MAAS, J. & DOERFLER, W. 1995. The initiation of de novo methylation of foreign DNA integrated into a mammalian genome is not exclusively targeted by nucleotide sequence. *J Virol,* 69**,** 1226-42.

OVERMEER, R. M., HENKEN, F. E., BIERKENS, M., WILTING, S. M., TIMMERMAN, I., MEIJER, C. J., SNIJDERS, P. J. & STEENBERGEN, R. D. 2009. Repression of MAL tumour suppressor activity by promoter methylation during cervical carcinogenesis. *J Pathol,* 219**,** 327-36.

PATEL, D. A., ROZEK, L. S., COLACINO, J. A., VAN ZOMEREN-DOHM, A., RUFFIN, M. T., UNGER, E. R., DOLINOY, D. C., SWAN, D. C., ONYEKWULUJE, J., DEGRAFFINREID, C. R. & PASKETT, E. D. 2012. Patterns of cellular and HPV 16 methylation as biomarkers for cervical neoplasia. *Journal of Virological Methods*.

PAULSON, E. J. & SPECK, S. H. 1999. Differential methylation of Epstein-Barr virus latency promoters facilitates viral persistence in healthy seropositive individuals. *J Virol,* 73**,** 9959-68.

PETER, M., ROSTY, C., COUTURIER, J., RADVANYI, F., TESHIMA, H. & SASTRE-GARAU, X. 2006. MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors. *Oncogene,* 25**,** 5985-93.

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

PETT, M. R., ALAZAWI, W. O., ROBERTS, I., DOWEN, S., SMITH, D. I., STANLEY, M. A. & COLEMAN, N. 2004. Acquisition of high-level chromosomal instability is associated with integration of human papillomavirus type 16 in cervical keratinocytes. *Cancer Res,* 64**,** 1359-68.

PETT, M. R., HERDMAN, M. T., PALMER, R. D., YEO, G. S., SHIVJI, M. K., STANLEY, M. A. & COLEMAN, N. 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**,** 3822-7.

PFAFFL, M. W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res,* 29**,** e45.

PFAFFL, M. W., HORGAN, G. W. & DEMPFLE, L. 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res,* 30**,** e36.

PIRAMI, L., GIACHE, V. & BECCIOLINI, A. 1997. Analysis of HPV16, 18, 31, and 35 DNA in pre-invasive and invasive lesions of the uterine cervix. *J Clin Pathol,* 50**,** 600-4.

PIYATHILAKE, C. J., MACALUSO, M., ALVAREZ, R. D., CHEN, M., BADIGA, S., EDBERG, J. C., PARTRIDGE, E. E. & JOHANNING, G. L. 2011. A higher degree of methylation of the HPV 16 E6 gene is associated with a lower likelihood of being diagnosed with cervical intraepithelial neoplasia. *Cancer,* 117**,** 957-63.

POPPE, B., VANDESOMPELE, J., SCHOCH, C., LINDVALL, C., MROZEK, K., BLOOMFIELD, C. D., BEVERLOO, H. B., MICHAUX, L., DASTUGUE, N., HERENS, C., YIGIT, N., DE PAEPE, A., HAGEMEIJER, A. & SPELEMAN, F. 2004. Expression analyses identify MLL as a prominent target of 11q23 amplification and support an etiologic role for MLL gain of function in myeloid malignancies. *Blood,* 103**,** 229-35.

PORTELA, A. & ESTELLER, M. 2010. Epigenetic modifications and human disease. *Nat Biotechnol,* 28**,** 1057-68.

POWELL, N., BOYDE, A., TRISTRAM, A., HIBBITTS, S. & FIANDER, A. 2009. The potential impact of human papillomavirus vaccination in contemporary cytologically screened populations may be underestimated: an observational retrospective analysis of invasive cervical cancers. *Int J Cancer,* 125**,** 2425-7.

POWELL, N. G., HIBBITTS, S. J., BOYDE, A. M., NEWCOMBE, R. G., TRISTRAM, A. J. & FIANDER, A. N. 2011. The risk of cervical cancer associated with specific types of human papillomavirus: a case-control study in a UK population. *Int J Cancer,* 128**,** 1676-82.

PRADHAN, S., BACOLLA, A., WELLS, R. D. & ROBERTS, R. J. 1999. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J Biol Chem,* 274**,** 33002-10.

QBASEPLUS. 2007. *qBasePlus Home Page* [Online]. Biogazelle. Available: http://www.biogazelle.com/products [Accessed 26/04/2012 2012].

QIAGEN. 2009. *REST 2009 Software* [Online]. Qiagen. Available: http://www.qiagen.com/Products/REST2009Software.aspx?r=8042#Tabs=t1 [Accessed 05/06/2012 2012].

QU, Z., FU, J., YAN, P., HU, J., CHENG, S. Y. & XIAO, G. 2010. Epigenetic repression of PDZ-LIM domain-containing protein 2: implications for the biology and treatment of breast cancer. *J Biol Chem,* 285**,** 11786-92.

QUINT, W., JENKINS, D., MOLIJN, A., STRUIJK, L., VAN DE SANDT, M., DOORBAR, J., MOLS, J., VAN HOOF, C., HARDT, K., STRUYF, F. & COLAU, B. 2012. One virus, one lesion—individual components of CIN lesions contain a specific HPV type. *The Journal of Pathology,* 227**,** 62-71.

RAGIN, C. C., RESHMI, S. C. & GOLLIN, S. M. 2004. Mapping and analysis of HPV16 integration sites in a head and neck cancer cell line. *Int J Cancer,* 110**,** 701-9.

RAJ, K., BERGUERAND, S., SOUTHERN, S., DOORBAR, J. & BEARD, P. 2004. E1^E4 Protein of Human Papillomavirus Type 16 Associates with Mitochondria. *J Virol,* 78**,** 7199-7207.

RAJEEVAN, M. S., SWAN, D. C., DUNCAN, K., LEE, D. R., LIMOR, J. R. & UNGER, E. R. 2006. Quantitation of site-specific HPV 16 DNA methylation by pyrosequencing. *Journal of Virological Methods,* 138**,** 170-176.

RAMAKERS, C., RUIJTER, J. M., DEPREZ, R. H. L. & MOORMAN, A. F. M. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters,* 339**,** 62-66.

RAMCHANDANI, S., BHATTACHARYA, S. K., CERVONI, N. & SZYF, M. 1999. DNA methylation is a reversible biological signal. *Proc Natl Acad Sci U S A,* 96**,** 6107-12.

RASSMUSSEN, R. 2001. Quantification on the LightCycler. *In:* MEUER, S., WITTWER,C. AND NAKAGAWARA,K (ed.) *Rapid Cycle Real-time PCR, Methods and Applications.* Heidelberg: Springer Press.

RAUCH, T. A., WU, X., ZHONG, X., RIGGS, A. D. & PFEIFER, G. P. 2009. A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci U S A,* 106**,** 671-8.

REGIER, N. & FREY, B. 2010. Experimental comparison of relative RT-qPCR quantification approaches for gene expression studies in poplar. *BMC Mol Biol,* 11**,** 57.

REIK, W. & WALTER, J. 2001. Genomic imprinting: parental influence on the genome. *Nat Rev Genet,* 2**,** 21-32.

REMM, M., REMM, A. & USTAV, M. 1999. Human papillomavirus type 18 E1 protein is translated from polycistronic mRNA by a discontinuous scanning mechanism. *J Virol,* 73**,** 3062-70.

RIJKAART, D. C., BERKHOF, J., VAN KEMENADE, F. J., COUPE, V. M. H., HESSELINK, A. T., ROZENDAAL, L., HEIDEMAN, D. A. M., VERHEIJEN, R. H., BULK, S., VERWEIJ, W. M., SNIJDERS, P. J. F. & MEIJER, C. J. L. M. 2012. Evaluation of 14 triage strategies for HPV DNA-positive women in population-based cervical screening. *International Journal of Cancer,* 130**,** 602-610.

ROBERTS, I., NG, G., FOSTER, N., STANLEY, M., HERDMAN, M. T., PETT, M. R., TESCHENDORFF, A. & COLEMAN, N. 2008a. Critical evaluation of HPV16 gene copy number quantification by SYBR green PCR. *BMC Biotechnol,* 8**,** 57.

ROBERTS, S., KINGSBURY, S. R., STOEBER, K., KNIGHT, G. L., GALLIMORE, P. H. & WILLIAMS, G. H. 2008b. Identification of an Arginine-Rich Motif in Human Papillomavirus Type 1 E1^E4 Protein Necessary for E4-Mediated Inhibition of Cellular DNA Synthesis In Vitro and in Cells. *J Virol,* 82**,** 9056-9064.

ROBERTSON, K. D. 2001. DNA methylation, methyltransferases, and cancer. *Oncogene,* 20**,** 3139-55.

ROBERTSON, K. D. & JONES, P. A. 2000. DNA methylation: past, present and future directions. *Carcinogenesis,* 21**,** 461-467.

ROBERTSON, K. D. & WOLFFE, A. P. 2000. DNA methylation in health and disease. *Nat Rev Genet,* 1**,** 11-19.

RODRIGUEZ-OSORIO, N., WANG, H., RUPINSKI, J., BRIDGES, S. M. & MEMILI, E. 2010. Comparative functional genomics of mammalian DNA methyltransferases. *Reprod Biomed Online,* 20**,** 243-55.

RODRIGUEZ-PAREDES, M. & ESTELLER, M. 2011. Cancer epigenetics reaches mainstream oncology. *Nat Med,* 17**,** 330-9.

ROMANCZUK, H. & 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**,** 3159-3163.

ROMANCZUK, H., THIERRY, F. & HOWLEY, P. M. 1990. Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters. *J Virol,* 64**,** 2849-59.

RONAGHI, M. 2001. Pyrosequencing Sheds Light on DNA Sequencing. *Genome Research,* 11**,** 3-11.

RONAGHI, M., UHLÉN, M. & NYRÉN, P. 1998. A Sequencing Method Based on Real-Time Pyrophosphate. *Science,* 281**,** 363-365.

ROSL, F., ARAB, A., KLEVENZ, B. & ZUR HAUSEN, H. 1993. The Effect of DNA Methylation on Gene Regulation of Human Papillomaviruses. *J Gen Virol,* 74**,** 791-801.

ROUGIER, N., BOURC'HIS, D., GOMES, D. M., NIVELEAU, A., PLACHOT, M., PALDI, A. & VIEGAS-PEQUIGNOT, E. 1998. Chromosome methylation patterns during mammalian preimplantation development. *Genes Dev,* 12**,** 2108-13.

RUESCH, M. N. & LAIMINS, L. A. 1997. Initiation of DNA synthesis by human papillomavirus E7 oncoproteins is resistant to p21-mediated inhibition of cyclin E-cdk2 activity. *J Virol,* 71**,** 5570-5578.

RUIJTER, J. M., RAMAKERS, C., HOOGAARS, W. M., KARLEN, Y., BAKKER, O., VAN DEN HOFF, M. J. & MOORMAN, A. F. 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res,* 37**,** e45.

SANDERS, C. M. & STENLUND, A. 1998. Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor. *Embo J,* 17**,** 7044-55.

SANDERS, G. D. & TAIRA, A. V. 2003. Cost Effectiveness of a Potential Vaccine for Human papillomavirus. *Emerging Infectious Diseases,* 9**,** 37-48.

SANTOS, F., HENDRICH, B., REIK, W. & DEAN, W. 2002. Dynamic Reprogramming of DNA Methylation in the Early Mouse Embryo. *Developmental Biology,* 241**,** 172-182.

SAXONOV, S., BERG, P. & BRUTLAG, D. L. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A,* 103**,** 1412-7.

SCHEFFNER, M., MUNGER, K., BYRNE, J. C. & HOWLEY, P. M. 1991. The state of the p53 and retinoblastoma genes in human cervical carcinoma cell lines. *Proc Natl Acad Sci U S A,* 88**,** 5523-7.

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

SCHEURER, M. E., TORTOLERO-LUNA, G. & ADLER-STORTHZ, K. 2005. Human papillomavirus infection: biology, epidemiology, and prevention. *International Journal of Gynecological Cancer,* 15**,** 727-746.

SCHILLER, J. T., DAY, P. M. & KINES, R. C. 2010. Current understanding of the mechanism of HPV infection. *Gynecol Oncol,* 118**,** S12-7.

SCHILLER, J. T. & LOWY, D. R. 2000. Papillomavirus-Like Particle Vaccines. *J Natl Cancer Inst Monogr,* 2000**,** 50-54.

SCHILLING, E. & REHLI, M. 2007. Global, comparative analysis of tissue-specific promoter CpG methylation. *Genomics,* 90**,** 314-323.

SCHMITTGEN, T. D. & LIVAK, K. J. 2008. Analyzing real-time PCR data by the comparative CT method. *Nat. Protocols,* 3**,** 1101-1108.

SCHORDERET, D. F. & GARTLER, S. M. 1992. Analysis of CpG suppression in methylated and nonmethylated species. *Proc Natl Acad Sci U S A,* 89**,** 957-961.

SCHUEBEL, K. & BAYLIN, S. 2005. In living color: DNA methyltransferase caught in the act. *Nat Meth,* 2**,** 736-738.

SCHWARZ, E., FREESE, U. K., GISSMANN, L., MAYER, W., ROGGENBUCK, B., STREMLAU, A. & ZUR HAUSEN, H. 1985. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature,* 314**,** 111-4.

SEDMAN, J. & STENLUND, A. 1998. The Papillomavirus E1 Protein Forms a DNA-Dependent Hexameric Complex with ATPase and DNA Helicase Activities. *J Virol,* 72**,** 6893-6897.

SELKER, E. U. 1998. Trichostatin A causes selective loss of DNA methylation in Neurospora. *Proc Natl Acad Sci U S A,* 95**,** 9430-5.

SEN, E., ALAM, S. & MEYERS, C. 2004. Genetic and biochemical analysis of cis regulatory elements within the keratinocyte enhancer region of the human papillomavirus type 31 upstream regulatory region during different stages of the viral life cycle. *J Virol,* 78**,** 612-29.

SHENKER, N. & FLANAGAN, J. M. 2012. Intragenic DNA methylation: implications of this epigenetic mechanism for cancer research. *Br J Cancer,* 106**,** 248-253.

SHERMAN, M. E. & KELLY, D. 1992. High-grade squamous intraepithelial lesions and invasive carcinoma following the report of three negative Papanicolaou smears: screening failures or rapid progression? *Mod Pathol,* 5**,** 337-42.

SHIRASAWA, H., JIN, M. H., SHIMIZU, K., AKUTSU, N., SHINO, Y. & SIMIZU, B. 1994. Transcription-modulatory activity of full-length E6 and E6*I proteins of human papillomavirus type 16. *Virology,* 203**,** 36-42.

SIDERI, M., JONES, R. W., WILKINSON, E. J., PRETI, M., HELLER, D. S., SCURRY, J., HAEFNER, H. & NEILL, S. 2005. Squamous vulvar intraepithelial neoplasia: 2004 modified terminology, ISSVD Vulvar Oncology Subcommittee. *J Reprod Med,* 50**,** 807-10.

SMITH, J. S., BACKES, D. M., HOOTS, B. E., KURMAN, R. J. & PIMENTA, J. M. 2009. Human papillomavirus type-distribution in vulvar and vaginal cancers and their associated precursors. *Obstet Gynecol,* 113**,** 917-24.

SMITH, T. M., HICKS-BERGER, C. A., KIM, S. & KIRLEY, T. L. 2002. Cloning, expression, and characterization of a soluble calcium-activated nucleotidase, a human enzyme belonging to a new family of extracellular nucleotidases. *Archives of Biochemistry and Biophysics,* 406**,** 105-115.

SNELLENBERG, S., SCHÜTZE, D. M., CLAASSEN-KRAMER, D., MEIJER, C. J. L. M., SNIJDERS, P. J. F. & STEENBERGEN, R. D. M. 2012. Methylation status of the E2 binding sites of HPV16 in cervical lesions determined with the Luminex® xMAP™ system. *Virology,* 422**,** 357-365.

SNIJDERS, P. J., STEENBERGEN, R. D. M., HEIDEMAN, D. A. M. & MEIJER, C. J. L. M. 2006. HPV-mediated cervical carcinogenesis: concepts and clinical implications. *The Journal of Pathology,* 208**,** 152-164.

SPINK, K. M. & LAIMINS, L. A. 2005. Induction of the Human Papillomavirus Type 31 Late Promoter Requires Differentiation but Not DNA Amplification. *J Virol,* 79**,** 4918-4926.

STACEY, S. N., JORDAN, D., WILLIAMSON, A. J., BROWN, M., COOTE, J. H. & ARRAND, J. R. 2000. Leaky scanning is the predominant mechanism for translation of human papillomavirus type 16 E7 oncoprotein from E6/E7 bicistronic mRNA. *J Virol,* 74**,** 7284-97.

STANLEY, M. 2007. Prophylactic HPV vaccines. *J Clin Pathol,* 60**,** 961-5.

STANLEY, M. A. 2002. Imiquimod and the imidazoquinolones: mechanism of action and therapeutic potential. *Clin Exp Dermatol,* 27**,** 571-7.

STANLEY, M. A. 2006a. HPV and pathogenesis of cervical cancer. *European Journal of Obstetrics & Gynaecology and Reproductive Biology,* Article In Press.

STANLEY, M. A. 2006b. Human papillomavirus vaccines. *Rev Med Virol,* 16**,** 139-49.

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

STANLEY, M. A., PETT, M. R. & COLEMAN, N. 2007. HPV: from infection to cancer. *Biochemical Society Transactions,* 35**,** 1456-1460.

STEENBERGEN, R. D., KRAMER, D., BRAAKHUIS, B. J., STERN, P. L., VERHEIJEN, R. H., MEIJER, C. J. & SNIJDERS, P. J. 2004. TSLC1 gene silencing in cervical cancer cell lines and cervical neoplasia. *J Natl Cancer Inst,* 96**,** 294-305.

STEGER, G. & CORBACH, S. 1997. Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein. *J Virol,* 71**,** 50-58.

STOLER, M. H., WOLINSKY, S. M., WHITBECK, A., BROKER, T. R. & CHOW, L. T. 1989. Differentiation-linked human papillomavirus types 6 and 11 transcription in genital condylomata revealed by in situ hybridization with message-specific RNA probes. *Virology,* 172**,** 331-40.

STRAIGHT, S. W., HERMAN, B. & MCCANCE, D. J. 1995. The E5 oncoprotein of human papillomavirus type 16 inhibits the acidification of endosomes in human keratinocytes. *J Virol,* 69**,** 3185-92.

STRAIGHT, S. W., HINKLE, P. M., JEWERS, R. J. & MCCANCE, D. J. 1993. The E5 oncoprotein of human papillomavirus type 16 transforms fibroblasts and effects the downregulation of the epidermal growth factor receptor in keratinocytes. *J Virol,* 67**,** 4521-32.

STUBENRAUCH, F., ZOBEL, T. & IFTNER, T. 2001. The E8 domain confers a novel long-distance transcriptional repression activity on the E8E2C protein of high-risk human papillomavirus type 31. *J Virol,* 75**,** 4139-49.

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

STUNKEL, W., HUANG, Z., TAN, S. H., O'CONNOR, M. J. & BERNARD, H. U. 2000. Nuclear matrix attachment regions of human papillomavirus type 16 repress or activate the E6 promoter, depending on the physical state of the viral DNA. *J Virol,* 74**,** 2489-501.

SUN, C., REIMERS, L. L. & BURK, R. D. 2011. Methylation of HPV16 genome CpG sites is associated with cervix precancer and cancer. *Gynecologic Oncology,* 121**,** 59-63.

SUN, L. L., CAO, D. Y., YANG, J. X., LI, H., ZHOU, X. R., SONG, Z. Q., CHENG, X. M., CHEN, J. & SHEN, K. 2012. Population-based case-control study on DAPK1, RAR-beta2 and MGMT methylation in liquid-based cytology. *Arch Gynecol Obstet,* 285**,** 1433-9.

SURANI, M. A. H., BARTON, S. C. & NORRIS, M. L. 1984. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature,* 308**,** 548-550.

SUTTER, D., WESTPHAL, M. & DOERFLER, W. 1978. Patterns of integration of viral DNA sequences in the genomes of adenovirus type 12-transformed hamster cells. *Cell,* 14**,** 569-85.

SZYF, M., ELIASSON, L., MANN, V., KLEIN, G. & RAZIN, A. 1985. Cellular and viral DNA hypomethylation associated with induction of Epstein-Barr virus lytic cycle. *Proc Natl Acad Sci U S A,* 82**,** 8090-4.

SZYF, M., PAKNESHAN, P. & RABBANI, S. A. 2004. DNA demethylation and cancer: therapeutic implications. *Cancer Lett,* 211**,** 133-43.

TAKAI, D. & JONES, P. A. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A,* 99**,** 3740-5.

TAKIZAWA, T., NAKASHIMA, K., NAMIHIRA, M., OCHIAI, W., UEMURA, A., YANAGISAWA, M., FUJITA, N., NAKAO, M. & TAGA, T. 2001. DNA Methylation Is a Critical Cell-Intrinsic Determinant of Astrocyte Differentiation in the Fetal Brain. *Developmental Cell,* 1**,** 749-758.

TAN, S. H., BAKER, C. C., STUNKEL, W. & BERNARD, H. U. 2003. A transcriptional initiator overlaps with a conserved YY1 binding site in the long control region of human papillomavirus type 16. *Virology,* 305**,** 486-501.

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

TAO, Q. & ROBERTSON, K. D. 2003. Stealth technology: how Epstein-Barr virus utilizes DNA methylation to cloak itself from immune detection. *Clin Immunol,* 109**,** 53-63.

TASKER, G. L. & WOJNAROWSKA, F. 2003. Lichen sclerosus. *Clinical and Experimental Dermatology,* 28**,** 128-133.

TAZI, J. & BIRD, A. 1990. Alternative chromatin structure at CpG islands. *Cell,* 60**,** 909-920.

THAIN, A., JENKINS, O., CLARKE, A. R. & GASTON, K. 1996. CpG methylation directly inhibits binding of the human papillomavirus type 16 E2 protein to specific DNA sequences. *J Virol,* 70**,** 7233-5.

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

THIERRY, F., SPYROU, G., YANIV, M. & HOWLEY, P. 1992. Two AP1 sites binding JunB are essential for human papillomavirus type 18 transcription in keratinocytes. *J Virol,* 66**,** 3740-8.

THOMAS, M. C. & CHIANG, C. M. 2005. E6 oncoprotein represses p53-dependent gene activation via inhibition of protein acetylation independently of inducing p53 degradation. *Mol Cell,* 17**,** 251-64.

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

THORLAND, E. C., MYERS, S. L., PERSING, D. H., SARKAR, G., MCGOVERN, R. M., GOSTOUT, B. S. & SMITH, D. I. 2000. Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. *Cancer Res,* 60**,** 5916-21.

TODD, R. W., ETHERINGTON, I. J. & LUESLEY, D. M. 2002. The effects of 5% imiquimod cream on high-grade vulval intraepithelial neoplasia. *Gynecol Oncol,* 85**,** 67-70.

TRISTRAM, A. & FIANDER, A. 2005. Clinical responses to Cidofovir applied topically to women with high grade vulval intraepithelial neoplasia. *Gynecol Oncol,* 99**,** 652-5.

TURAN, T., KALANTARI, M., CALLEJA-MACIAS, I. E., CUBIE, H. A., CUSCHIERI, K., VILLA, L. L., SKOMEDAL, H., BARRERA-SALDAÑA, H. A. & BERNARD, H.-U. 2006. Methylation of the human papillomavirus-18 L1 gene: A biomarker of neoplastic progression? *Virology,* 349**,** 175-183.

TURAN, T., KALANTARI, M., CUSCHIERI, K., CUBIE, H. A., SKOMEDAL, H. & BERNARD, H. U. 2007. High-throughput detection of human papillomavirus-18 L1 gene methylation, a candidate biomarker for the progression of cervical neoplasia. *Virology,* 361**,** 185-93.

TURKER, M. S. 2002. Gene silencing in mammalian cells and the spread of DNA methylation. *Oncogene,* 21**,** 5388-93.

UNGER, E. R., STEINAU, M., RAJEEVAN, M. S., SWAN, D., LEE, D. R. & VERNON, S. D. 2004. Molecular markers for early detection of cervical neoplasia. *Dis Markers,* 20**,** 103-16.

VAN DE NIEUWENHOF, H. P., BULTEN, J., HOLLEMA, H., DOMMERHOLT, R. G., MASSUGER, L. F. A. G., VAN DER ZEE, A. G. J., DE HULLU, J. A. & VAN KEMPEN, L. C. L. T. 2011. Differentiated vulvar intraepithelial neoplasia is often found in lesions, previously diagnosed as lichen sclerosus, which have progressed to vulvar squamous cell carcinoma. *Mod Pathol,* 24**,** 297-305.

VAN DE NIEUWENHOF, H. P., VAN DER AVOORT, I. A. M. & DE HULLU, J. A. 2008. Review of squamous premalignant vulvar lesions. *Critical Reviews in Oncology/Hematology,* 68**,** 131-156.

VAN DE NIEUWENHOF, H. P., VAN KEMPEN, L. C. L. T., DE HULLU, J. A., BEKKERS, R. L. M., BULTEN, J., MELCHERS, W. J. G. & MASSUGER, L. F. A. G. 2009. The Etiologic Role of HPV in Vulvar Squamous Cell Carcinoma Fine Tuned. *Cancer Epidemiology Biomarkers & Prevention,* 18**,** 2061-2067.

VAN TINE, B. A., DAO, L. D., WU, S.-Y., SONBUCHNER, T. M., LIN, B. Y., ZOU, N., CHIANG, C.-M., BROKER, T. R. & CHOW, L. T. 2004a. Human papillomavirus (HPV) origin-binding protein associates with mitotic spindles to enable viral DNA partitioning. *Proc Natl Acad Sci U S A,* 101**,** 4030-4035.

VAN TINE, B. A., KAPPES, J. C., BANERJEE, N. S., KNOPS, J., LAI, L., STEENBERGEN, R. D. M., MEIJER, C. L. J. M., SNIJDERS, P. J. F., CHATIS, P., BROKER, T. R., MOEN, P. T., JR. & CHOW, L. T. 2004b. Clonal Selection for Transcriptionally Active Viral Oncogenes during Progression to Cancer. *J Virol,* 78**,** 11172-11186.

VAN TINE, B. A., KNOPS, J., BROKER, T. R., CHOW, L. T. & MOEN, P. T., JR. 2001. In situ analysis of the transcriptional activity of integrated viral DNA using tyramide-FISH. *Dev Biol (Basel),* 106**,** 381-5.

VANDESOMPELE, J., DE PRETER, K., PATTYN, F., POPPE, B., VAN ROY, N., DE PAEPE, A. & SPELEMAN, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol,* 3**,** RESEARCH0034.

VELDMAN, T., LIU, X., YUAN, H. & SCHLEGEL, R. 2003. Human papillomavirus E6 and Myc proteins associate in vivo and bind to and cooperatively activate the telomerase reverse transcriptase promoter. *Proc Natl Acad Sci U S A,* 100**,** 8211-6.

VENUTI, A., PAOLINI, F., NASIR, L., CORTEGGIO, A., ROPERTO, S., CAMPO, M. S. & BORZACCHIELLO, G. 2011. Papillomavirus E5: the smallest oncoprotein with many functions. *Mol Cancer,* 10**,** 140.

VINOKUROVA, S. & VON KNEBEL DOEBERITZ, M. 2011. Differential Methylation of the HPV 16 Upstream Regulatory Region during Epithelial Differentiation and Neoplastic Transformation. *PloS one,* 6**,** e24451.

VINOKUROVA, S., WENTZENSEN, N., KRAUS, I., KLAES, R., DRIESCH, C., MELSHEIMER, P., KISSELJOV, F., DURST, M., SCHNEIDER, A. & VON KNEBEL DOEBERITZ, M. 2008. Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. *Cancer Res,* 68**,** 307-13.

W.H.O. 2010. *Immunization, Vaccines and Biologicals: Human papillomavirus (HPV)* [Online]. World Health Organisation. Available: http://www.who.int/immunization/topics/hpv/en/ [Accessed 5/09/2011 2011].

WALDMAN, T., KINZLER, K. W. & VOGELSTEIN, B. 1995. p21 Is Necessary for the p53-mediated G1 Arrest in Human Cancer Cells. *Cancer Res,* 55**,** 5187-5190.

WANG-JOHANNING, F., LU, D. W., WANG, Y., JOHNSON, M. R. & JOHANNING, G. L. 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**,** 2199-2210.

WANG, J. C. 1979. Helical repeat of DNA in solution. *Proceedings of the National Academy of Sciences,* 76**,** 200-203.

WATT, P. M., KUMAR, R. & KEES, U. R. 2000. Promoter demethylation accompanies reactivation of the HOX11 proto-oncogene in leukemia. *Genes Chromosomes Cancer,* 29**,** 371-7.

WEBER, M., DAVIES, J. J., WITTIG, D., OAKELEY, E. J., HAASE, M., LAM, W. L. & SCHUBELER, D. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet,* 37**,** 853-62.

WELLS, S. I., FRANCIS, D. A., KARPOVA, A. Y., DOWHANICK, J. J., BENSON, J. D. & HOWLEY, P. M. 2000. Papillomavirus E2 induces senescence in HPV-positive cells via pRB- and p21(CIP)-dependent pathways. *Embo J,* 19**,** 5762-71.

WENTZENSEN, N., BERGERON, C., CAS, F., ESCHENBACH, D., VINOKUROVA, S. & VON KNEBEL DOEBERITZ, M. 2005. Evaluation of a nuclear score for p16INK4a-stained cervical squamous cells in liquid-based cytology samples. *Cancer,* 105**,** 461-7.

WENTZENSEN, N., SHERMAN, M. E., SCHIFFMAN, M. & WANG, S. S. 2009. Utility of methylation markers in cervical cancer early detection: Appraisal of the state-of-the-science. *Gynecologic Oncology,* 112**,** 293-299.

WENTZENSEN, N. & VON KNEBEL DOEBERITZ, M. 2007. Biomarkers in cervical cancer screening. *Dis Markers,* 23**,** 315-30.

WILEY, D. J., HUH, J., RAO, J. Y., CHANG, C., GOETZ, M., POULTER, M., MASONGSONG, E., CHANG, C. I. & BERNARD, H. U. 2005. Methylation of human papillomavirus genomes in cells of anal epithelia of HIV-infected men. *J Acquir Immune Defic Syndr,* 39**,** 143-51.

WILSON, A. S., POWER, B. E. & MOLLOY, P. L. 2007. DNA hypomethylation and human diseases. *Biochimica et biophysica acta,* 1775**,** 138-62.

WILSON, V. G., WEST, M., WOYTEK, K. & RANGASAMY, D. 2002. Papillomavirus E1 Proteins: Form, Function, and Features. *Virus Genes,* 24**,** 275-290.

WOJDACZ, T. K. & DOBROVIC, A. 2007. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucl. Acids Res.,* gkm013.

WOO, Y. L., DAMAY, I., STANLEY, M., CRAWFORD, R. & STERLING, J. 2007. The use of HPV Linear Array Assay for multiple HPV typing on archival frozen tissue and DNA specimens. *J Virol Methods,* 142**,** 226-30.

WOODMAN, C. B., COLLINS, S., WINTER, H., BAILEY, A., ELLIS, J., PRIOR, P., YATES, M., ROLLASON, T. P. & YOUNG, L. S. 2001. Natural history of cervical human papillomavirus infection in young women: a longitudinal cohort study. *Lancet,* 357**,** 1831-6.

WOODMAN, C. B. J., COLLINS, S. I. & YOUNG, L. S. 2007. The natural history of cervical HPV infection: unresolved issues. *Nat Rev Cancer,* 7**,** 11-22.

WU, J., ISSA, J. P., HERMAN, J., BASSETT, D. E., JR., NELKIN, B. D. & BAYLIN, S. B. 1993. Expression of an exogenous eukaryotic DNA methyltransferase gene induces transformation of NIH 3T3 cells. *Proc Natl Acad Sci U S A,* 90**,** 8891-5.

WU, S. C. & ZHANG, Y. 2010. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol,* 11**,** 607-20.

XU, M., LUO, W., ELZI, D. J., GRANDORI, C. & GALLOWAY, D. A. 2008. NFX1 interacts with mSin3A/histone deacetylase to repress hTERT transcription in keratinocytes. *Mol Cell Biol,* 28**,** 4819-28.

YEE, C., KRISHNAN-HEWLETT, I., BAKER, C. C., SCHLEGEL, R. & HOWLEY, P. M. 1985. Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. *Am J Pathol,* 119**,** 361-6.

YODER, J. A., WALSH, C. P. & BESTOR, T. H. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet,* 13**,** 335-40.

YOU, J., CROYLE, J. L., NISHIMURA, A., OZATO, K. & HOWLEY, P. M. 2004. Interaction of the Bovine Papillomavirus E2 Protein with Brd4 Tethers the Viral DNA to Host Mitotic Chromosomes. *Cell,* 117**,** 349-360.

YUGAWA, T. & KIYONO, T. 2009. Molecular mechanisms of cervical carcinogenesis by high-risk human papillomaviruses: novel functions of E6 and E7 oncoproteins. *Rev Med Virol,* 19**,** 97-113.

ZHENG, Z. M. & BAKER, C. C. 2006. Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front Biosci,* 11**,** 2286-302.

ZHENG, Z. M., TAO, M., YAMANEGI, K., BODAGHI, S. & XIAO, W. 2004. Splicing of a cap-proximal human Papillomavirus 16 E6E7 intron promotes E7 expression, but can be restrained by distance of the intron from its RNA 5' cap. *J Mol Biol,* 337**,** 1091-108.

ZHOU, J., GISSMANN, L., ZENTGRAF, H., MULLER, H., PICKEN, M. & MULLER, M. 1995. Early phase in the infection of cultured cells with papillomavirus virions. *Virology,* 214**,** 167-176.

ZHOU, J., LIU, W. J., PENG, S. W., SUN, X. Y. & FRAZER, I. 1999. Papillomavirus Capsid Protein Expression Level Depends on the Match between Codon Usage and tRNA Availability. *J Virol,* 73**,** 4972-4982.

ZILBERMAN, D., GEHRING, M., TRAN, R. K., BALLINGER, T. & HENIKOFF, S. 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet,* 39**,** 61-69.

ZILBERMAN, D. & HENIKOFF, S. 2007. Genome-wide analysis of DNA methylation patterns. *Development,* 134**,** 3959-3965.

ZUR HAUSEN, H. 1976. Condylomata acuminata and human genital cancer. *Cancer Res,* 36**,** 794.

ZUR HAUSEN, H. 1977. Human papillomaviruses and their possible role in squamous cell carcinomas. *Curr Top Microbiol Immunol,* 78**,** 1-30.

ZUR HAUSEN, H. 1991. Viruses in human cancers. *Science,* 254**,** 1167-1173.

ZUR HAUSEN, H., MEINHOF, W., SCHEIBER, W. & BORNKAMM, G. 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**,** 650-656.

Appendices

Appendix I – Cancer Research UK Biomarker Roadmaps
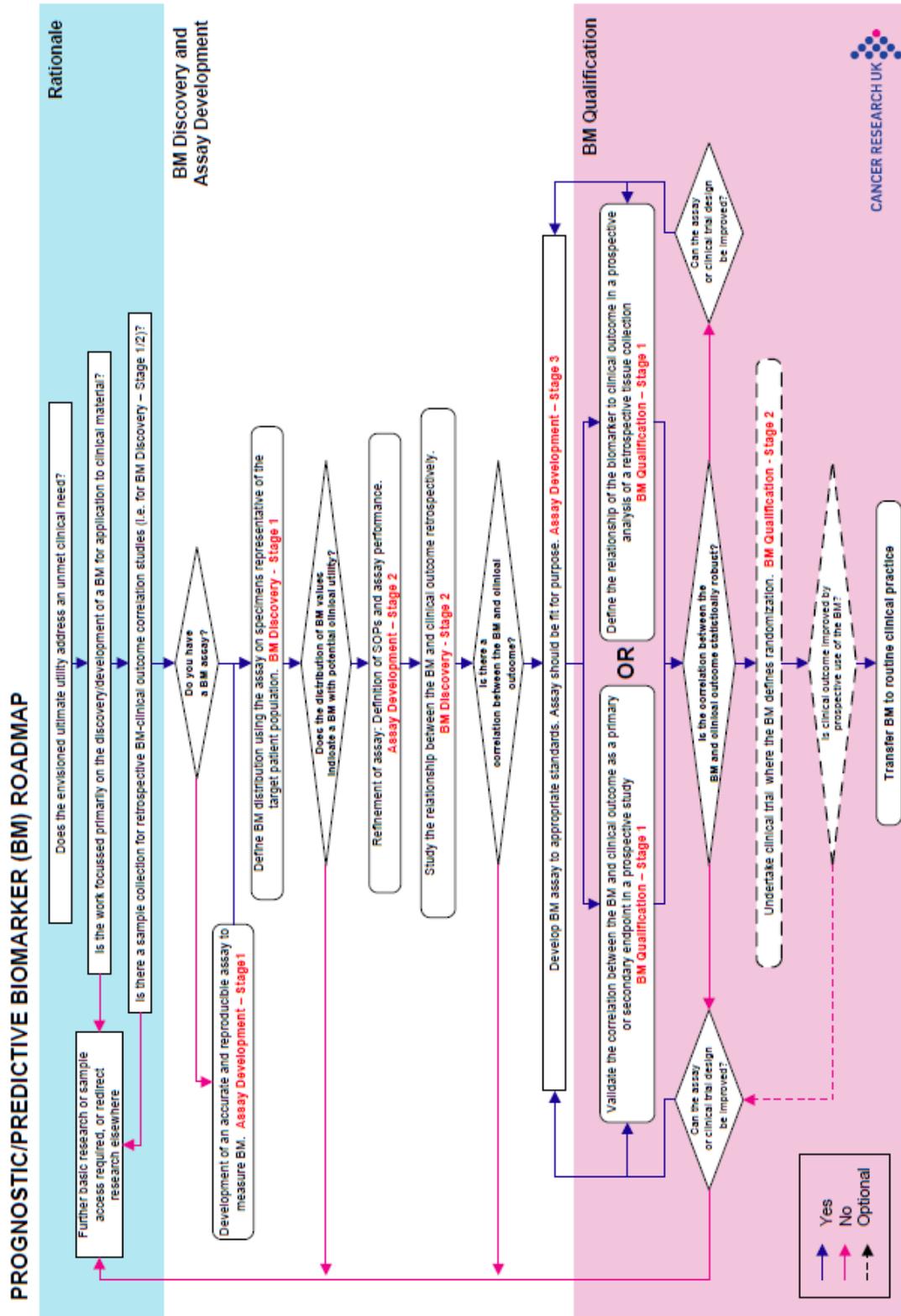
Figure I.a: The CRUK prognostic/predictive biomarker roadmap.

SCREENING BIOMARKER (BM) ROADMAP
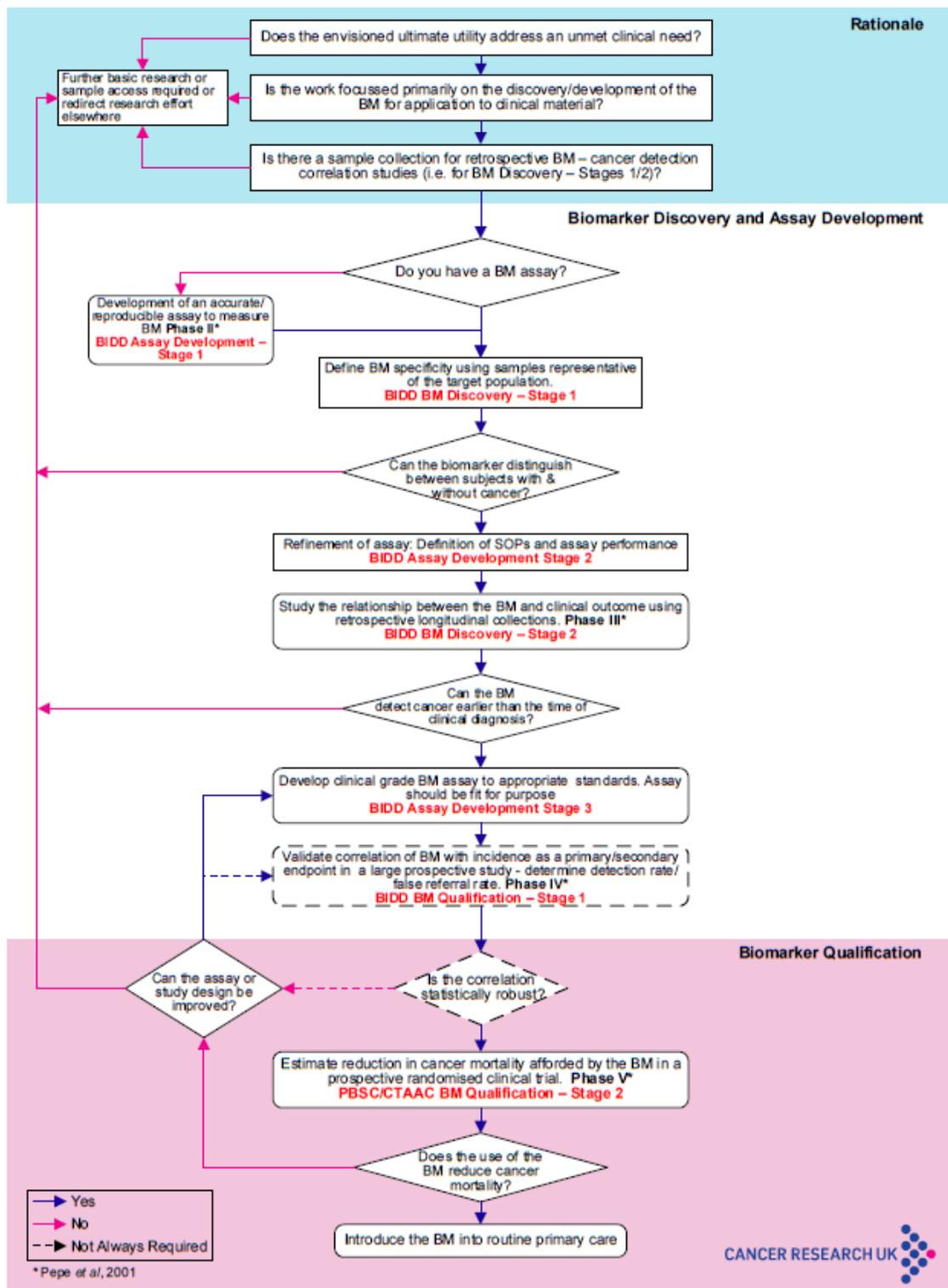
Figure I.b: The CRUK screening biomarker roadmap.
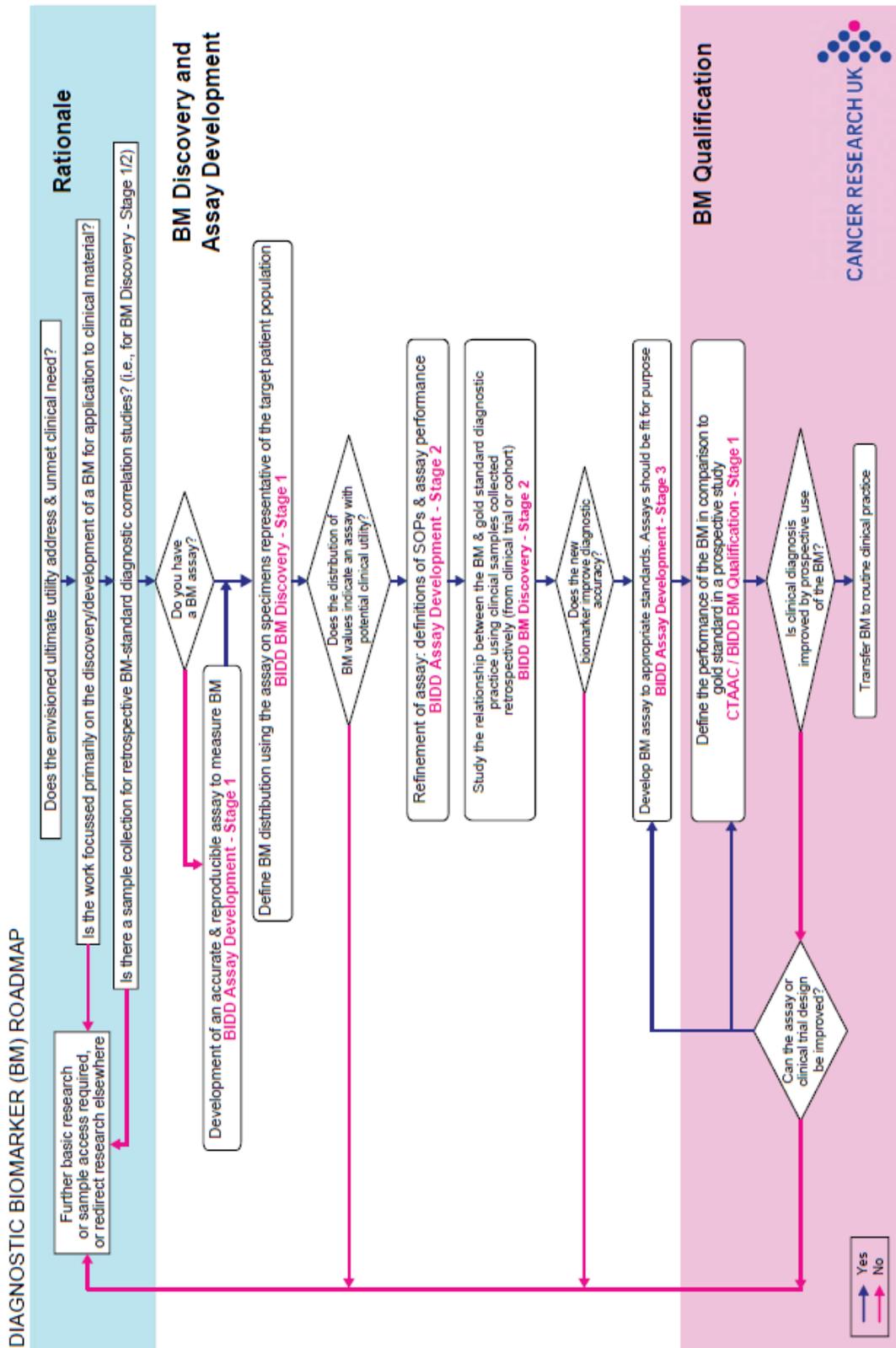
Figure I.c: The CRUK diagnostic biomarker roadmap.

Appendix II – Chapter 4 qPCR Method Validation
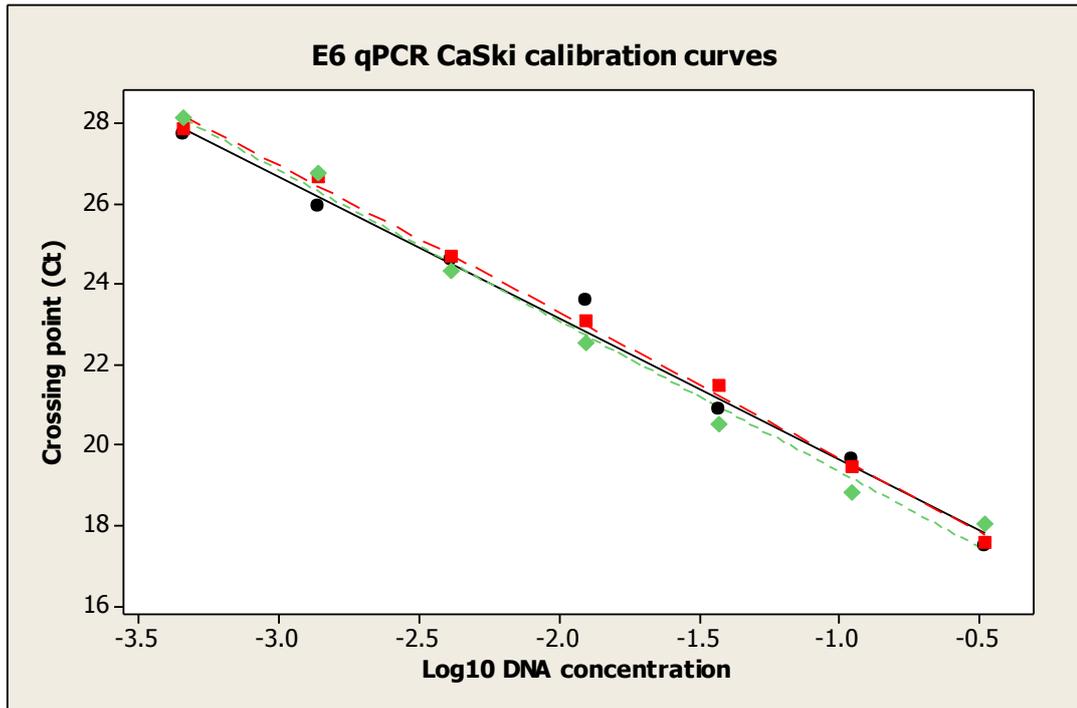
Figure II.a: The relationship between Ct value and DNA concentration for the *E6* qPCR primer pair.

Included are regression lines for each of the three repeats. Each colour and point shape represents an experimental repeat.
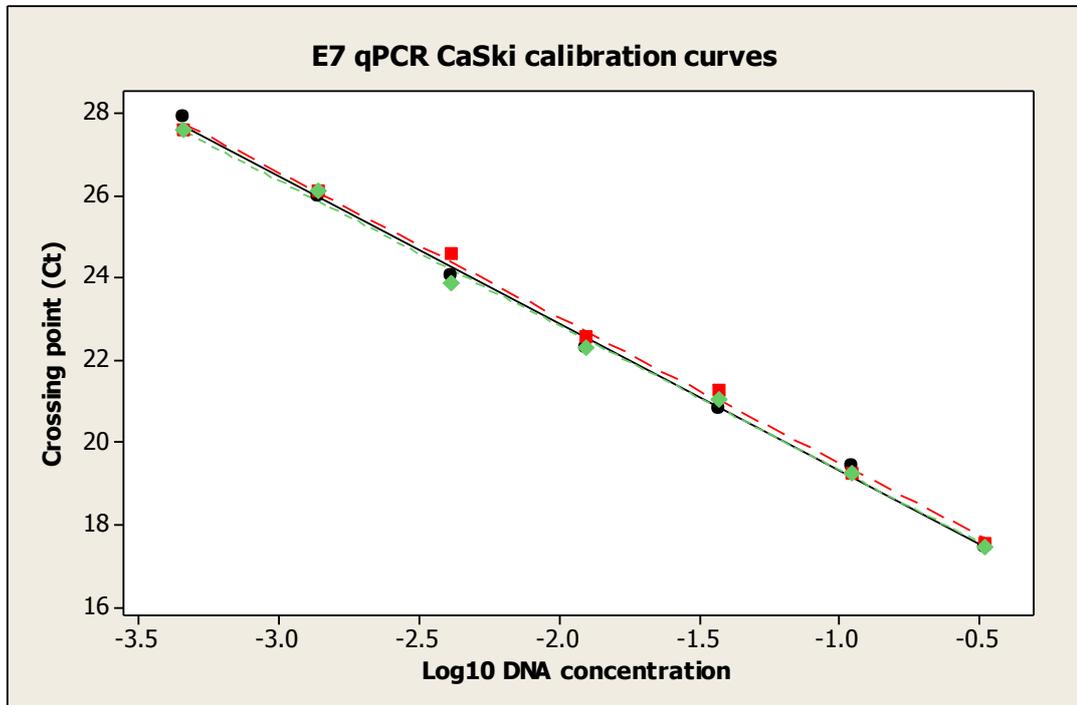
Figure II.b: The relationship between Ct value and DNA concentration for the *E7* qPCR primer pair.

Included are regression lines for each of the three repeats. Each colour and point shape represents an experimental repeat.
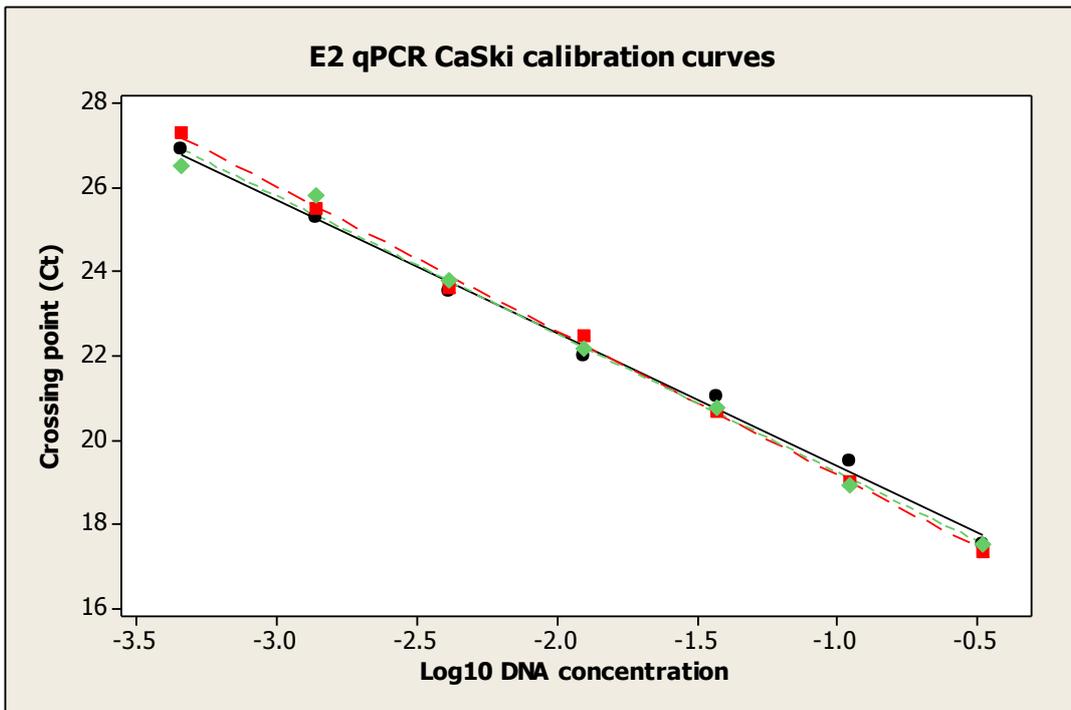
Figure II.c: The relationship between Ct value and DNA concentration for the _E2_ qPCR primer pair.

Included are regression lines for each of the three repeats. Each colour and point shape represents an experimental repeat.
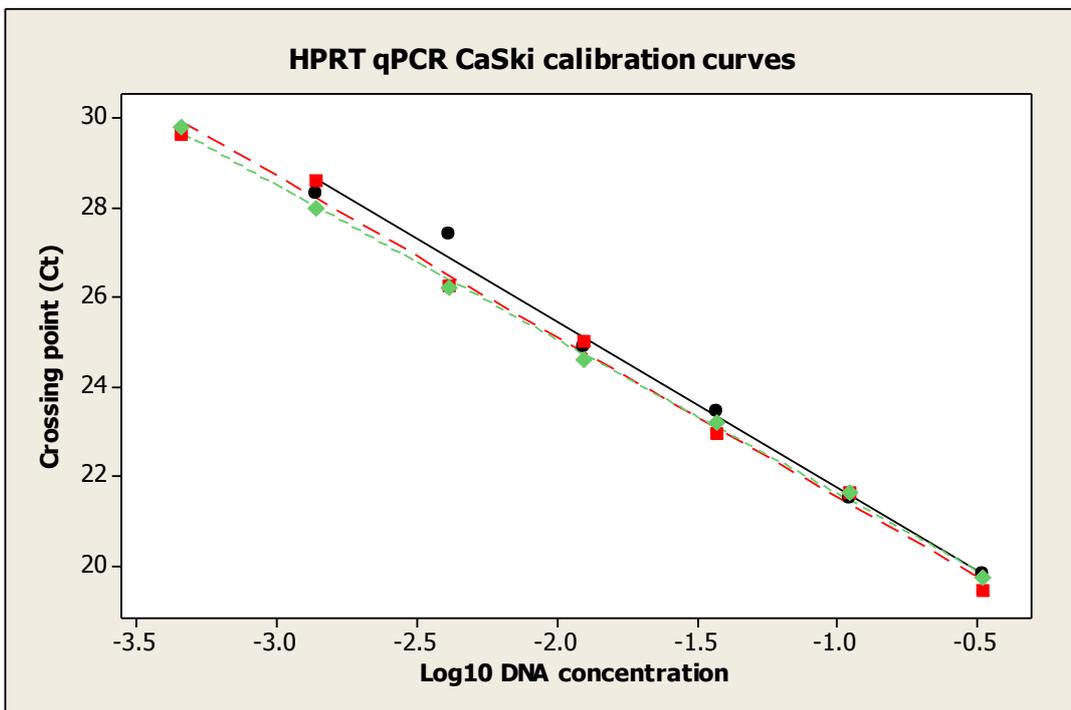
Figure II.d: The relationship between Ct value and DNA concentration for the _HPRT_ qPCR primer pair.

Included are regression lines for each of the three repeats. Each colour and point shape represents an experimental repeat. One data point was missing from one of experimental repeats.
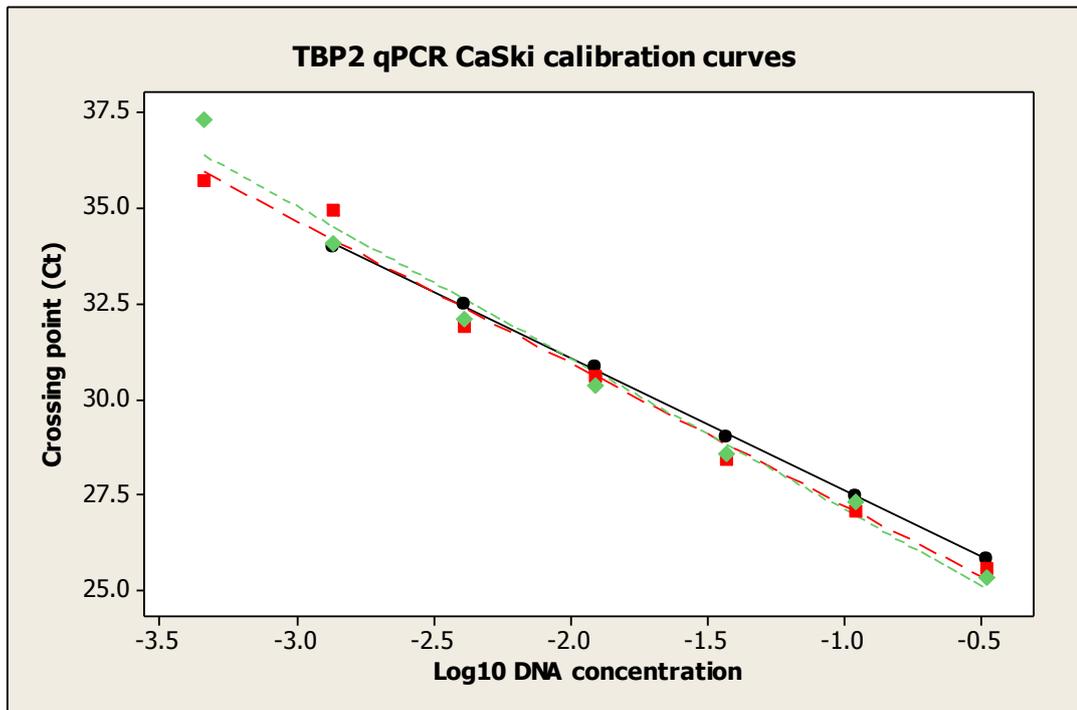
Figure II.e: The relationship between Ct value and DNA concentration for the *TBP*2 qPCR primer pair.

Included are regression lines for each of the three repeats. Each colour and point shape represents an experimental repeat. One data point was missing from one of experimental repeats.

Appendix III – Chapter 4 Pyrosequencing Method Validation

Appendix III contains the results of 6x CaSki Pyrosequencing analyses used to determine variability among runs.
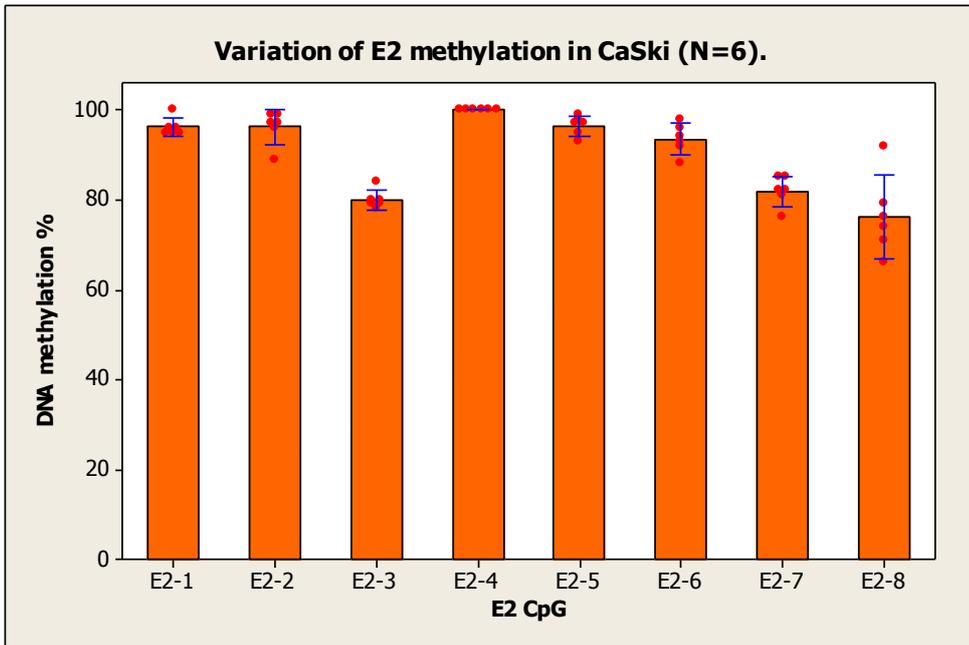
Figure III.a: Variation in CaSki *E2* methylation for each CpG.

Bars represent mean values, red circles represent individual values and error bars represent 95% confidence intervals.
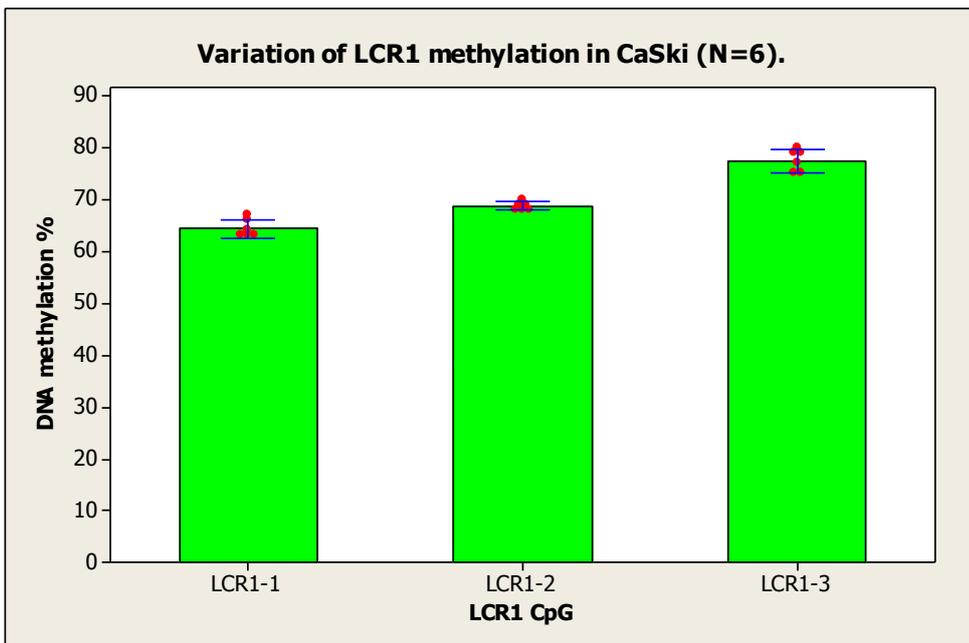
Figure III.b: Variation in CaSki LCR1 methylation for each CpG.

Bars represent mean values, red circles represent individual values and error bars represent 95% confidence intervals
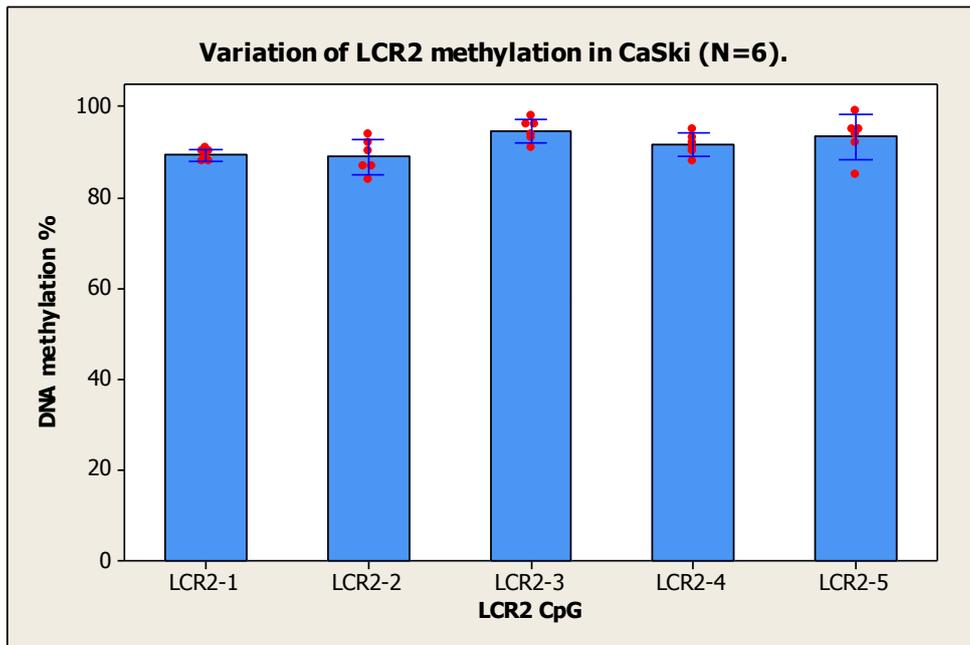
Figure III.c: Variation in CaSki LCR2 methylation for each CpG.

Bars represent mean values, red circles represent individual values and error bars represent 95% confidence intervals
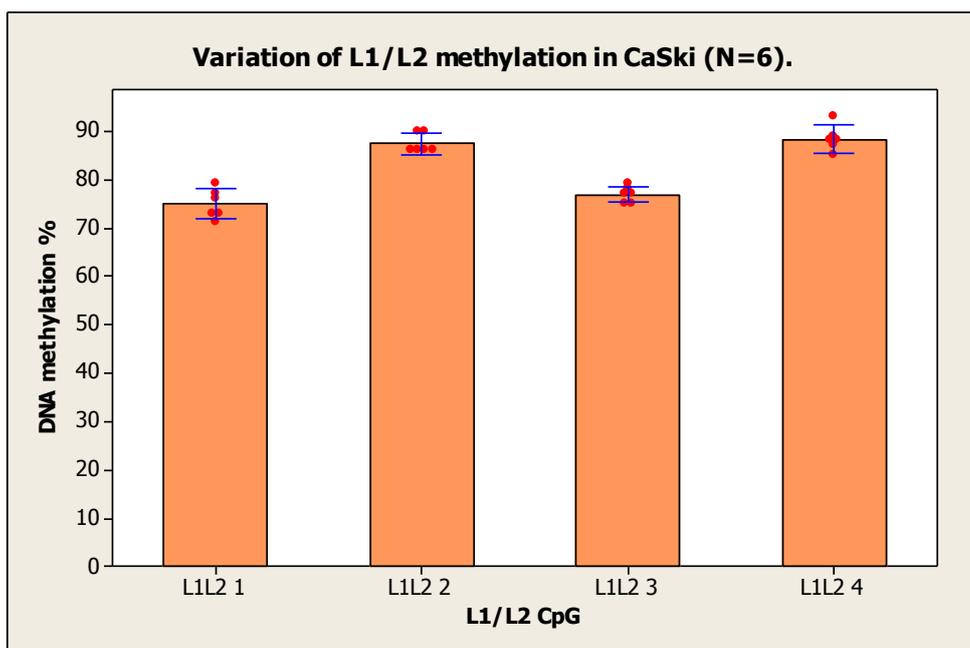
Figure III.d: Variation in CaSki *L1/L2* methylation for each CpG.

Bars represent mean values, red circles represent individual values and error bars represent 95% confidence intervals

Appendix IV – Chapter 6 Supplementary Statistics

Appendix IV contains the statistical analyses that were performed for Chapter 6, section 6.1.3.1. The so that the chapter was more concise, the mostly non-significant results were briefly summarised in the main body of the text. These tables represent supplementary information.

Each region was considered separately with the aim of identifying which CpG was the most significantly differently methylated among the three outcome groups. The Bonferroni corrected P-value of significance was used (*E2* CpG, P=0.006; LCR1 CpG, P=0.017; LCR2 CpG, P=0.010; *L1/L2* CpG, P=0.013) however comparisons significant at P=0.05 were considered.

	E2-1	E2-2	E2-3	E2-4	E2-5	E2-6	E2-7	E2-8
Kruskall-Wallis P-value	0.463	0.122	0.321	0.012	0.700	0.179	0.123	0.518
H-value (DF)	1.54 (2)	4.21 (2)	2.28 (2)	8.85 (2)	0.71 (2)	3.44 (2)	4.19 (2)	1.34 (2)
ANOVA P-value	0.696	0.112	0.341	0.034	0.674	0.099	0.049	0.336
F-value (DF)	0.37 (2)	2.40 (2)	1.13 (2)	3.90 (2)	0.40 (2)	2.55 (2)	3.45 (2)	1.14 (2)
Equal variances	Yes	Yes	Yes	Close	Yes	Yes	Yes	Yes
Normal residuals	No	No	No	No	No	Yes	Yes	Yes
	Significant Tukey-Kramer Pairwise Comparison?							
	E2-1	E2-2	E2-3	E2-4	E2-5	E2-6	E2-7	E2-8
C vs. PL	No	No	No	No	No	No	No	No
C vs. PH	No	No	No	Yes	No	No	No	No
PL vs. PH	No	No	No	No	No	No	Yes	No

Table IV.a: The results of multiple statistical analyses on E2 CpG methylation data.

The non-parametric Kruskall-Wallis P-values and H-values represent the significant differences among the medians of each disease group, higher significance is represented by a lower P-value and a larger H-value. ANOVA P-values and F-values are accompanied by degrees of freedom (DF) and two statements for violations of the ANOVA assumptions (equal variances and normally distributed residuals), here three degrees of success are present, No; a violation (P=<0.01), Close; P=0.01-0.05 and Yes; P>0.05. Finally, the results of C vs. Pl, C vs. PH and PL vs. PH Tukey-Kramer pairwise comparisons are listed where No = no significant difference in means at P=0.05 and Yes = a significant difference in means at P=0.05.

	LCR1-1	LCR1-2	LCR1-3
Kruskall-Wallis P-value	0.017	0.639	0.579
H-value (DF)	8.13 (2)	0.89 (2)	1.09 (2)
ANOVA P-value	0.014	0.581	0.650
F-value (DF)	5.02 (2)	0.55 (2)	0.44 (2)
Equal variances	Close	Yes	Yes
Normal residuals	No	Yes	No
	Significant Tukey-Kramer Pairwise Comparison?		
	LCR1-1	LCR1-2	LCR1-3
C vs. PL	Yes	No	No
C vs. PH	No	No	No
PL vs. PH	Yes	No	No

Table IV.b: The results of multiple statistical analyses on LCR1 CpG methylation data.

The non-parametric Kruskall-Wallis P-values and H-values represent the significant differences among the medians of each disease group, higher significance is represented by a lower P-value and a larger H-value. ANOVA P-values and F-values are accompanied by degrees of freedom (DF) and two statements for violations of the ANOVA assumptions (equal variances and normally distributed residuals), here three degrees of success are present, No; a violation (P=<0.01), Close; P=0.01-0.05 and Yes; P>0.05. Finally, the results of C vs. Pl, C vs. PH and PL vs. PH Tukey-Kramer pairwise comparisons are listed where No = no significant difference in means at P=0.05 and Yes = a significant difference in means at P=0.05.

	LCR2-1	LCR2-2	LCR2-3	LCR2-4	LCR2-5
Kruskall-Wallis P-value	0.364	0.069	0.000	0.486	0.269
H-value (DF)	2.02 (2)	5.36 (2)	18.53 (2)	1.44 (2)	2.63 (2)
ANOVA P-value	0.338	0.058	0.000	0.381	0.278
F-value (DF)	1.13 (2)	3.18 (2)	23.14 (2)	1.00 (2)	1.34 (2)
Equal variances	Yes	Yes	Yes	Close	N/A[*1]
Normal residuals	No	Close	No	No	No
	Significant Tukey-Kramer Pairwise Comparison?				
	LCR2-1	LCR2-2	LCR2-3	LCR2-4	LCR2-5
C vs. PL	No	No	No	No	No
C vs. PH	No	Yes	Yes	No	No
PL vs. PH	No	No	Yes	No	No

Table IV.c: The results of multiple statistical analyses on LCR2 CpG methylation data.

The non-parametric Kruskall-Wallis P-values and H-values represent the significant differences among the medians of each disease group, higher significance is represented by a lower P-value and a larger H-value. ANOVA P-values and F-values are accompanied by degrees of freedom (DF) and two statements for violations of the ANOVA assumptions (equal variances and normally distributed residuals), here three degrees of success are present, No; a violation (P=<0.01), Close; P=0.01-0.05 and Yes; P>0.05. Finally, the results of C vs. Pl, C vs. PH and PL vs. PH Tukey-Kramer pairwise comparisons are listed where No = no significant difference in means at P=0.05 and Yes = a significant difference in means at P=0.05.

	L1L2-1	L1L2-2	L1L2-3	L1L2-4
Kruskall-Wallis P-value	0.122	0.202	0.318	0.493
H-value (DF)	4.21 (2)	3.20 (2)	2.29 (2)	1.42 (2)
ANOVA P-value	0.107	0.291	0.277	0.300
F-value (DF)	2.42 (2)	1.29 (2)	1.34 (2)	1.27 (2)
Equal variances	Yes	No	No	Yes
Normal residuals	Yes	No	No	No
	Significant Tukey-Kramer Pairwise Comparison?			
	L1L2-1	L1L2-2	L1L2-3	L1L2-4
C vs. PL	No	No	No	No
C vs. PH	No	No	No	No
PL vs. PH	No	No	No	No

Table IV.d: The results of multiple statistical analyses on *L1/L2* CpG methylation data.

The non-parametric Kruskall-Wallis P-values and H-values represent the significant differences among the medians of each disease group, higher significance is represented by a lower P-value and a larger H-value. ANOVA P-values and F-values are accompanied by degrees of freedom (DF) and two statements for violations of the ANOVA assumptions (equal variances and normally distributed residuals), here three degrees of success are present, No; a violation (P=<0.01), Close; P=0.01-0.05 and Yes; P>0.05. Finally, the results of C vs. Pl, C vs. PH and PL vs. PH Tukey-Kramer pairwise comparisons are listed where No = no significant difference in means at P=0.05 and Yes = a significant difference in means at P=0.05.

Appendix V – Chapter 8 Supplementary Tables

Appendix V contains tables with data summarised in the main body of text. The tables are supplementary to the analysis performed in Chapter 8.
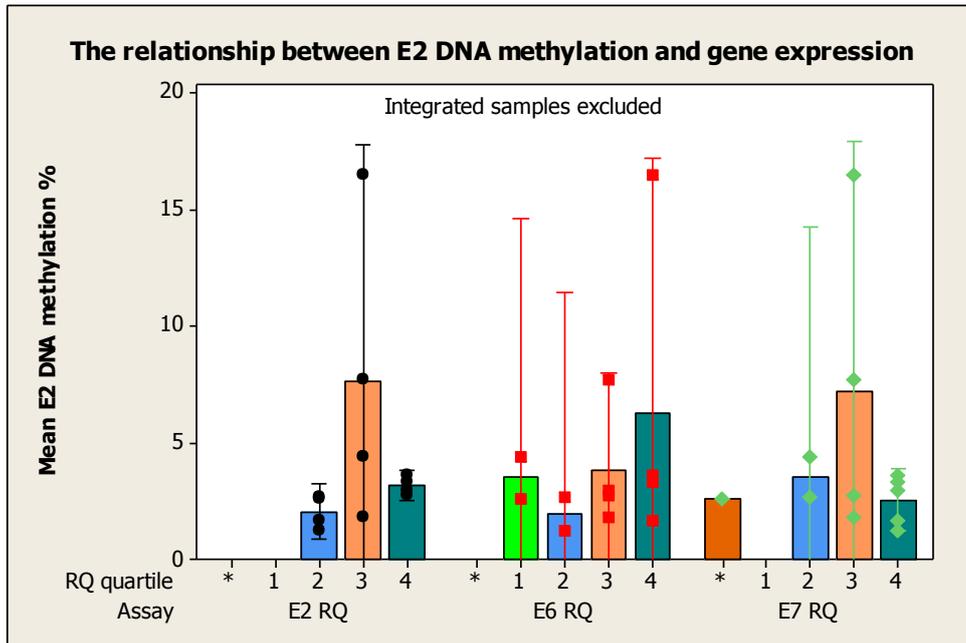
Figure V.a: DNA methylation of the VR series integration groups 0 and X *E2* region, separated by gene expression assay (RQ assay) and by RQ quartile within each assay.

* Represents those samples whereby the gene expression assays have failed to produce an RQ result.
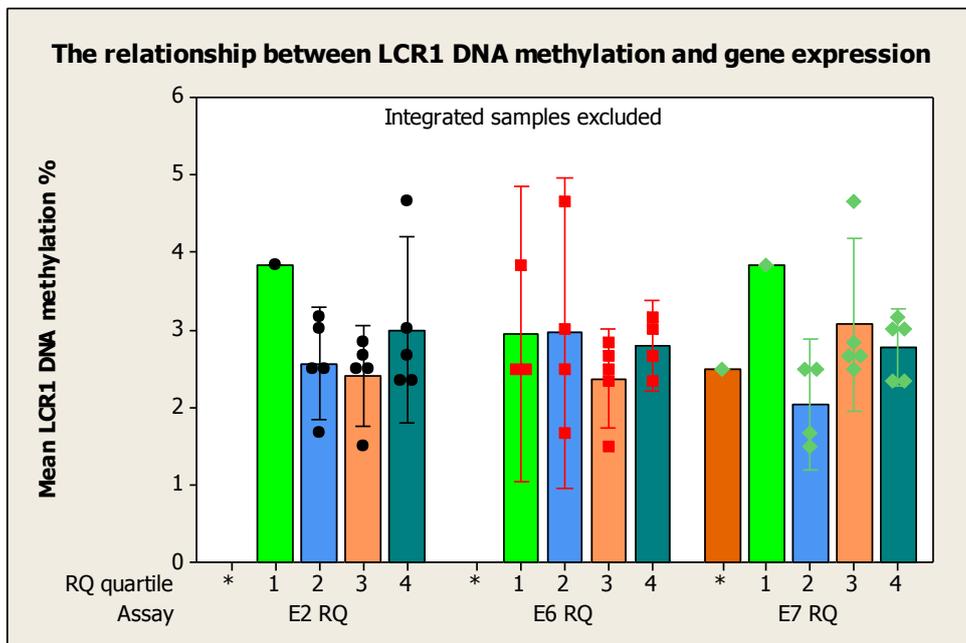
Figure V.b: DNA methylation of the VR series integration groups 0 and X LCR1 region, separated by gene expression assay (RQ assay) and by RQ quartile within each assay.

* Represents those samples whereby the gene expression assays have failed to produce an RQ result.
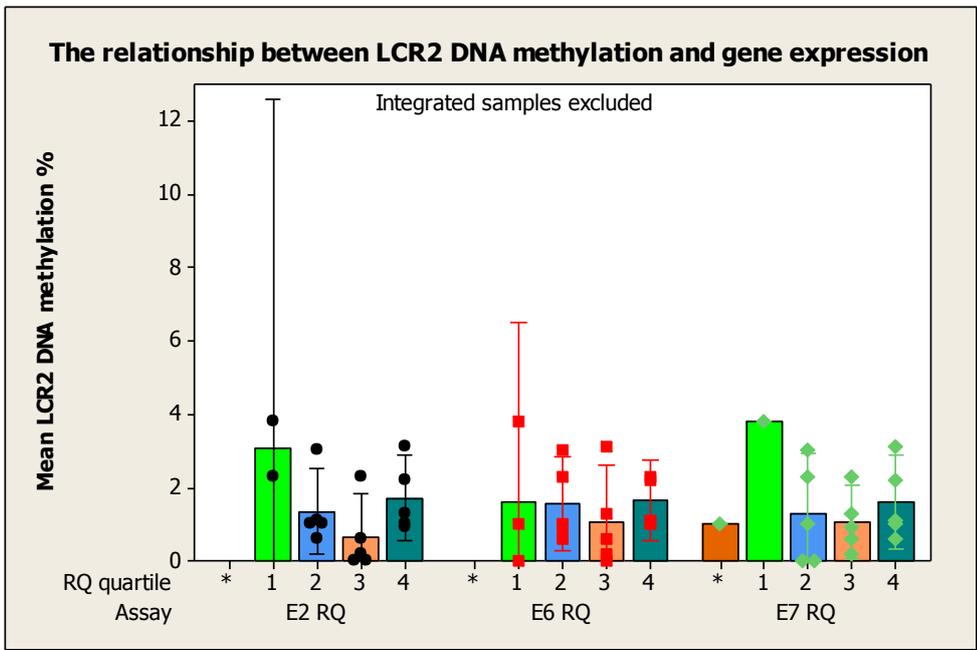
Figure V.c1: DNA methylation of the VR series integration groups 0 and X LCR2 region, separated by gene expression assay (RQ assay) and by RQ quartile within each assay.

* Represents those samples whereby the gene expression assays have failed to produce an RQ result.
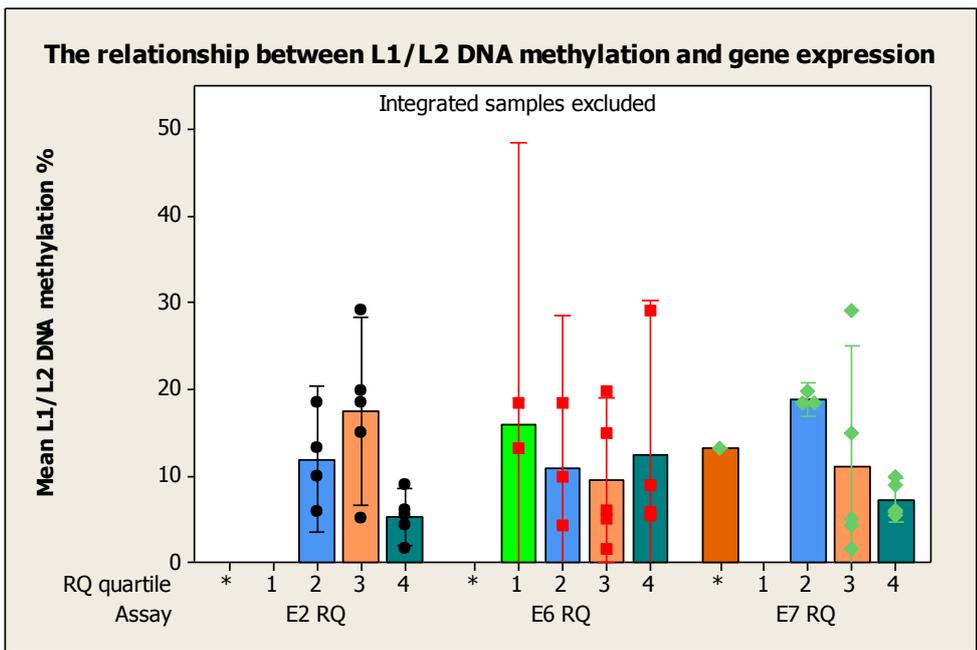
Figure V.d: DNA methylation of the VR series integration groups 0 and X *L1/L2* region, separated by gene expression assay (RQ assay) and by RQ quartile within each assay.

* Represents those samples whereby the gene expression assays have failed to produce an RQ result.

Appendix VI – HPV Methylation Studies, Methodology Summary

Appendix VI contains a summary of the methodologies used by a number of HPV16 DNA methylation studies.

Summary of HPV16 methylation studies					
Authors	Methylation assay used	Size (n)	Quantitative	regions	Material
Badal *et al* 2003	Restriction enzyme/Bisulfite sequencing	81	No	LCR, *E6*	Biopsy and exfoliated
Kalantari *et al* 2004	Bisulfite sequencing	115	Yes	LCR, *L1*	Biopsy
Bhattacharjee *et al* 2006	Restriction enzyme/Bisulfite sequencing	72	No	LCR	Biopsy and exfoliated
Hong *et al* 2008	Pyrosequencing	70	Yes	LCR	Exfoliated
Kalantari *et al* 2008	Bisulfite sequencing	19	Yes	LCR, *L1*	Biopsy*1
Brandsma *et al* 2009	Bisulfite sequencing	13	Yes	Genome	Exfoliated
Ding *et al* 2009	Bisulfite sequencing	53	Yes	LCR	Biopsy and exfoliated
Fernandez *et al* 2009	Bisulfite sequencing, MS-PCR	18	Yes, No	Genome	*2
Hublarova *et al* 2009	Restriction enzyme	121	No	LCR	Exfoliated
Kalantari *et al* 2010	Bisulfite sequencing	21	Yes	*L1*	Exfoliated
Piyathilake *et al* 2011	Pyrosequencing	75	Yes	LCR	Exfoliated
Vinokurova *et al* 2011	Bisulfite sequencing	3	Yes	LCR	LCM biopsy*3
Mirabello *et al* 2012	Pyrosequencing	108	yes	Genome	Exfoliated
Patel *et al* 2012	Pyrosequencing	89	Yes	LCR	Exfoliated
Snellenberg *et al* 2012	Luminex xMAP	84	Yes	LCR	FFPE and exfoliated*4
Sun *et al* 2012	EpiTyper/Pyrosequencing	85	Yes	LCR, *L1*	Exfoliated

TableVI.a: A summary of the methods of a number of HPV16 DNA methylation studies.

Here, each study is listed along with the main methylation assays used, the sample size of the study (n), whether the techniques produced quantitative data, the regions of the HPV genome studied and finally the type of clinical material used. Under material, biopsy and exfoliated represent cervical biopsies and exfoliated cervical cells as obtained for cytological testing respectively. Notes: *1, material was penile biopsy; *2 insufficient detail included in the paper to determine origin of samples; *3 cervical biopsies were microdissected into several components; *4 FFPE represents formalin fixed paraffin embedded tissue samples.

Appendix VII – Published HPV Typing Study

Appendix VII contains the published VIN HPV typing study (Bryant *et al.*, 2011).

Human Papillomavirus Type Distribution in Vulval Intraepithelial Neoplasia Determined Using PapilloCheck DNA Microarray

Dean Bryant, Nirmala Rai, Gareth Rowlands, Sam Hibbitts, Joanne Jones, Amanda Tristram, Alison Fiander, and Ned Powell*

HPV Research Group, Obstetrics and Gynecology, School of Medicine, Cardiff University, Cardiff, United Kingdom

Vulval intraepithelial neoplasia is a precursor of vulval carcinoma, and is frequently associated with human papillomavirus (HPV) infection. Estimates of HPV prevalence in vulval intraepithelial neoplasia vary widely in the UK. The objective of this study was to assess HPV infection in a sample of women with vulval intraepithelial neoplasia, confirmed histologically, and determine the proportion of disease associated with HPV types targeted by prophylactic HPV vaccines. HPV infection was assessed in biopsies from 59 patients using the Greiner Bio-One PapilloCheck® DNA chip assay. Valid results were obtained for 54 cases. HPV infection was present in 43 of the 54 cases (79.6%: 95% CI 67.1–88.2%). The most common HPV types were HPV 16 (33/54: 61.1%), HPV 33 (8/54: 14.8%), HPV 6 (5/54: 9.3%), and HPV 42 (3/54: 5.6%). The mean age of HPV positive women was significantly less than the mean age of HPV negative women. This is the largest UK series of vulval intraepithelial neoplasia in which HPV type has been investigated, and 34/54 (63.0%, 95% CI: 49.6–78.6%) cases were associated with HPV 16/18, which are targeted by current prophylactic HPV vaccines. *J. Med. Virol. 83:1358–1361, 2011.*
© 2011 Wiley-Liss, Inc.

KEY WORDS: human papillomavirus; vulval intraepithelial neoplasia; papillocheck

INTRODUCTION

Vulval intraepithelial neoplasia is frequently a painful, distressing condition, and a precursor lesion of invasive vulval carcinoma. In recent decades the incidence of vulval intraepithelial neoplasia has increased, while the mean age of patients has decreased. The incidence of vulval cancer rose by 17% between 1979 and 2001 [MacLean, 2004], with 996 new cases registered in England in 2000 [MacLean, 2006].

Human papillomavirus (HPV) positivity in vulval intraepithelial neoplasia lesions is reported in international meta-analyses as 84% with most cases attributable to HPV type 16 [De Vuyst et al., 2009]. The most likely explanation for the increased incidence of vulval intraepithelial neoplasia is the substantial rise in HPV infections of the lower genital tract [Peto et al., 2004]; the lifetime risk of HPV infection being estimated at up to 79% [Syrjanen et al., 1990].

The aim of this investigation was to identify HPV types present in vulval intraepithelial neoplasia, confirmed histologically, using the Greiner Bio-One PapilloCheck® DNA chip assay and determine the proportion of vulval intraepithelial neoplasia potentially preventable by HPV prophylactic vaccination.

MATERIALS AND METHODS

Study Population

The study population comprised women attending a specialist vulval intraepithelial neoplasia clinic at the University Hospital of Wales, Cardiff, UK between 2003 and 2009, with vulval intraepithelial neoplasia confirmed histologically, who gave written informed consent for a biopsy to be taken for research purposes. The study was approved by South Wales LREC (SMKW/EL/03/5178). Fifty-nine cases were investigated, with an age range of 22–82 years, and a mean age of 45.8 years. There were 6 cases of vulval

282

intraepithelial neoplasia grade 1, 2 of vulval intraepithelial neoplasia grade 2, and 51 cases of vulval intraepithelial neoplasia grade 3.

Sample Collection and Processing

Separate biopsies were taken for histology and research use. The research biopsy was immediately placed in liquid based cytology medium (which fixed cells and preserved DNA). DNA was extracted from the research biopsy using the Qiagen QIAamp kit (QIAGEN GmbH, Hilden, Germany) according to the manufacturer's instructions.

Pathology Review

A recent review of vulval intraepithelial neoplasia histopathology recommended replacement of the vulval intraepithelial neoplasia 1/2/3 subdivisions with two categories reflecting the putative aetiology of the disease: usual vulval intraepithelial neoplasia (HPV-associated); and differentiated vulval intraepithelial neoplasia (associated with lichen sclerosis) [Sideri et al., 2005]. However, the existence of differentiated vulval intraepithelial neoplasia is disputed, and in practice this diagnosis is rarely made [McCluggage, 2009]. The histopathology of lesions in this study is reported using both the vulval intraepithelial neoplasia 1/2/3 and usual/differentiated classification. Pathology review was undertaken blind to HPV status by an experienced Consultant Histopathologist with a special interest in gynaecological pathology (GR).

PapilloCheck® Assay

HPV infection was identified using the Greiner Bio-One Papillocheck® assay (Greiner Bio-One GmbH, Frickenhausen, Germany), which enables simultaneous detection and typing of 24 different HPV types in a single reaction (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73, 82, 6, 11, 40, 42, 43, and 44). The assay uses multiplex PCR with fluorescent primers to amplify a 350 bp fragment of the E1 gene. The PCR product is then hybridized to a PapilloCheck® DNA array, comprising 28 probes, each in 5 replicate spots. To identify false negatives

the assay includes amplification of the human ADAT1 gene. Hybridization is assessed using the CheckScanner™ array reader. Amplification, hybridization, detection, and interpretation were performed according to the manufacturer's instructions. A valid result was defined as one which passed all of the PapilloCheck® internal controls (including DNA adequacy, sample inhibition, and hybridization).

Previous studies have confirmed the analytical [Dalstein et al., 2009; Jones et al., 2009] and clinical validity [Halfon et al., 2009; Hesselink et al., 2010] of this assay in large sample sets as compared to GP5+/6+, HC2, and Roche Linear Array. The assay is certified within the European Union as an in vitro diagnostic for qualitative detection of HPV in clinical specimens.

RESULTS

HPV Type Distribution

HPV type distribution, stratified by histology, is shown in Table I. Valid results were obtained for 54 of the 59 cases. HPV infection was present in 43/54 cases (79.6%: 95% CI 67.1–88.2%). HPV 16 and/or 18 were present in 34/54 cases (63.0%, 95% CI: 49.6–78.6%). HPV 16 and/or 18 were the only types present in 26/54 cases (48%, 95% CI: 35.4–61.2%).

High-risk (HR) HPV types were present in 38 cases and low risk (LR) types in 9 cases; 4 cases contained both HR and LR types. Multiple HPV types were present in 10 cases; in these cases the distribution of types was: 1 vulval intraepithelial neoplasia grade 1 sample contained HPV 18, 42, and 44/55; all other cases with multiple types were vulval intraepithelial neoplasia 3 and the combinations of types were: 3 cases with HPV 16 and 33; 2 cases with HPV 16 and 6; and 1 case each of (16,18), (33, 42), (16, 31, 33), (16, 51). In the five cases where only LR types were present, three were HPV 6 (histology: two cases of vulval intraepithelial neoplasia grade 3, and one vulval intraepithelial neoplasia grade 3 with well-differentiated carcinoma and viral warts), one was HPV 40 (vulval intraepithelial neoplasia 1), and one contained HPV 42 (vulval intraepithelial neoplasia 1).

TABLE I. HPV Type Distribution Within Histological Grade

Histology	HPV type								
	16	18	31	33	51	6	40	42	44/55[a]
VIN 1	0	1	0	0	0	0	1	2	1
VIN 2	1	0	0	0	0	1	0	0	0
VIN 3	32	1	1	8	1	4	0	1	0
Total	33	2	1	8	1	5	1	3	1
%	61.1	3.7	1.9	14.8	1.9	9.3	1.9	5.6	1.9

There were 10 cases showing multiple infection, hence the number of types identified is greater that the number of cases. Percentages indicate the proportion of cases that tested positive for a given type.
[a]The HPV 44 probe cross-hybridizes to HPV 55, hence these types are reported together.

HPV and Age

HPV positive cases were aged 22–81 with a mean of 43.4 years. HPV negative cases were aged 33–82 years with a mean of 54.9 years. The means of these groups were significantly different (t-test $P = 0.019$).

HPV and Histology

There was no significant correlation between HPV positivity and grade of lesion. Of the HPV positive cases, 40/43 (93.0%) showed vulval intraepithelial neoplasia grade 2 or worse. This was slightly lower in HPV negative lesions, where 9/11 (81.8%) showed vulval intraepithelial neoplasia grade 2 or worse, but this difference was not significant (Fisher's $P = 0.266$).

There was no significant correlation between HPV type and histological grade; however it was noticeable that all of the HPV 16 positive cases were associated with vulval intraepithelial neoplasia 2 or worse, while a HR type (HPV 18) was identified in only 1/5 vulval intraepithelial neoplasia grade 1 cases with valid HPV results.

Slides were available for pathology review for 54/59 cases. 52/54 cases were "Usual vulval intraepithelial neoplasia"; 2/54 cases were "Probable Differentiated vulval intraepithelial neoplasia." Both cases of probable differentiated vulval intraepithelial neoplasia were in HPV negative patients: one an 82 year old woman with vulval intraepithelial neoplasia 3, the other a 61 year old woman with vulval intraepithelial neoplasia 1 (but with a history of vulval cancer). Among the cases with valid HPV results, 42/49 cases of usual vulval intraepithelial neoplasia were HPV positive (85.7% 95%CI: 73.3–92.9%).

DISCUSSION

This is the largest UK series of vulval intraepithelial neoplasia in which HPV type has been determined. The main limitation of this study was the use of a self selected sample, i.e., women attending a specialist vulval intraepithelial neoplasia clinic. This might bias the sample toward women with persistent or recurrent disease, which might possibly result in some over-representation of more persistent types (HPV 16). A further limitation was the use of separate biopsies for pathology and HPV analysis; in practice both biopsies were immediately adjacent and cases were only included if both biopsies were part of the same macroscopically visible lesion. It is however theoretically possible that the research biopsy may not be representative of the diagnostic biopsy.

Several studies have investigated HPV prevalence in vulval intraepithelial neoplasia in the UK and estimate HPV positivity at 37.5–100%, with the proportion of HPV positive cases containing HPV 16 varying from 37.5% to 93.3% [Abdel-Hady et al., 2001; Gasco et al., 2002; Todd et al., 2002; Baldwin et al., 2003; Tristram and Fiander, 2005; Fiander et al.,

2006; Woo et al., 2007; Daayana et al., 2010]. The sample size in these studies ranged from 8 to 29 cases. The sample size for the current study is larger, and the findings are more consistent with the overall results of a recent international meta-analysis, in which 84.0% of vulval intraepithelial neoplasia cases tested HPV positive (67.5% were HPV 16, 7.7% were HPV 33, and 4.6% were HPV 18) [De Vuyst et al., 2009]. The corresponding proportions in the current study were overall HPV positivity of 79.6% (61.1% were positive for HPV 16, 14.8% for HPV 33, and 3.7% for HPV 18). These are also similar to a separate meta-analysis which indicated an overall HPV positivity of 80.4% for vulval intraepithelial neoplasia 2/3 (HPV 16 present in 71.2%, HPV33 in 7.7%, and HPV 18 in 5.5% [Smith et al., 2009]).

Patients who tested positive for HPV infection were significantly younger (by over 11 years on average) than those testing negative. This is consistent with HPV associated disease being more common in younger women and vulval intraepithelial neoplasia associated with lichen sclerosis being more common in older patients [Bonvicini et al., 2005].

The finding that 85.7% of usual vulval intraepithelial neoplasia were HPV positive is in accord with data previously reviewed [van de Nieuwenhof et al., 2008]. The seven HPV negative usual vulval intraepithelial neoplasia patients represent an interesting group; it is not clear whether these patients are accounted for by mis-diagnosis of differentiated vulval intraepithelial neoplasia (possibly due to sampling error), or whether there is a subset of usual vulval intraepithelial neoplasia that is truly HPV negative.

Three cases showing vulval intraepithelial neoplasia grade 2 or worse were associated with a LR HPV type alone (HPV6). One of these cases showed vulval intraepithelial neoplasia grade 3 with well-differentiated squamous cell carcinoma and viral warts; it is possible that the warts were associated with the HPV 6 but the vulval intraepithelial neoplasia and carcinoma were not. This apparent association between high-grade vulval intraepithelial neoplasia and LR HPV types runs contrary to the suggestion that LR types cause only warts and low grade lesions. However an association between LR HPV types and high-grade disease was observed in the De Vuyst meta analysis [De Vuyst et al., 2009] which reported approximately 5% of vulval intraepithelial neoplasia 2/3 as linked to HPV 6.

HPV associated vulval intraepithelial neoplasia is most common in women in their 30's and 40's [Hart, 2001]. The incidence of VIN is approximately 5 per 100,000 women per year and is increasing [Joura, 2002]. HPV infection is endemic in the UK with almost 30% of women in the 20–25 year age range now infected with a HR anogenital HPV type [Hibbitts et al., 2008]. Hence, in the short term, the number of women affected by vulval intraepithelial neoplasia is likely to continue increasing. HPV vaccination has been shown to be 100% effective in

prevention of vulval epithelial neoplasia 2/3 associated with HPV 16/18 in a per protocol population [Joura et al., 2007]. In the UK vaccination of 12–13 year old girls against HPV 16/18 infection began in 2008. Hence it is likely to be 20–30 years before the effects of this intervention become apparent in reduced incidence of vulval intraepithelial neoplasia. Ultimately however, as HPV 16/18 were present in 63.0% of vulval intraepithelial neoplasia, our data suggest that greater than half of vulval intraepithelial neoplasia cases could be potentially prevented by vaccination.

ACKNOWLEDGMENTS

REFERENCES

Abdel-Hady ES, Martin-Hirsch P, Duggan-Keen M, Stern PL, Moore JV, Corbitt G, Kitchener HC, Hampson IN. 2001. Immunological and viral factors associated with the response of vulval intraepithelial neoplasia to photodynamic therapy. Cancer Res 61:192–196.

Baldwin PJ, van der Burg SH, Boswell CM, Offringa R, Hickling JK, Dobson J, Roberts JS, Latimer JA, Moseley RP, Coleman N, Stanley MA, Sterling JC. 2003. Vaccinia-expressed human papillomavirus 16 and 18 e6 and e7 as a therapeutic vaccination for vulval and vaginal intraepithelial neoplasia. Clin Cancer Res 9:5205–5213.

Bonvicini F, Venturoli S, Ambretti S, Paterini P, Santini D, Ceccarelli C, Zerbini M, Musiani M. 2005. Presence and type of oncogenic human papillomavirus in classic and in differentiated vulvar intraepithelial neoplasia and keratinizing vulvar squamous cell carcinoma. J Med Virol 77:102–106.

Daayana S, Elkord E, Winters U, Pawlita M, Roden R, Stern PL, Kitchener HC. 2010. Phase II trial of imiquimod and HPV therapeutic vaccination in patients with vulval intraepithelial neoplasia. Br J Cancer 102:1129–1136.

Dalstein V, Merlin S, Bali C, Saunier M, Dachez R, Ronsin C. 2009. Analytical evaluation of the PapilloCheck test, a new commercial DNA chip for detection and genotyping of human papillomavirus. J Virol Methods 156:77–83.

De Vuyst H, Clifford GM, Nascimento MC, Madeleine MM, Franceschi S. 2009. Prevalence and type distribution of human papillomavirus in carcinoma and intraepithelial neoplasia of the vulva, vagina and anus: A meta-analysis. Int J Cancer 124:1626–1636.

Fiander AN, Tristram AJ, Davidson EJ, Tomlinson AE, Man S, Baldwin PJ, Sterling JC, Kitchener HC. 2006. Prime-boost vaccination strategy in women with high-grade, noncervical anogenital intraepithelial neoplasia: Clinical results from a multicenter phase II trial. Int J Gynecol Cancer 16:1075–1081.

Gasco M, Sullivan A, Repellin C, Brooks L, Farrell PJ, Tidy JA, Dunne B, Gusterson B, Evans DJ, Crook T. 2002. Coincident inactivation of 14-3-3sigma and p16INK4a is an early event in vulval squamous neoplasia. Oncogene 21:1876–1881.

Halfon P, Benmoura D, Khiri H, Penaranda G, Blanc B, Riggio D, Sandri MT. 2009. Comparison of the clinical performance of carcinogenic HPV typing of the Linear Array and Papillocheck((R)) HPV-screening assay. J Clin Virol 47:38–42.

Hart WR. 2001. Vulvar intraepithelial neoplasia: Historical aspects and current status. Int J Gynecol Pathol 20:16–30.

Hesselink AT, Heideman DA, Berkhof J, Topal F, Pol RP, Meijer CJ, Snijders PJ. 2010. Comparison of the clinical performance of the PapilloCheck HPV detection with GP5+/6+-PCR-EIA in population-based cervical screening. J Clin Microbiol 48:797–801.

Hibbitts S, Jones J, Powell N, Dallimore N, McRea J, Beer H, Tristram A, Fielder H, Fiander AN. 2008. Human papillomavirus prevalence in women attending routine cervical screening in South Wales, UK: A cross-sectional study. Br J Cancer 99:1929–1933.

Jones J, Powell NG, Tristram A, Fiander AN, Hibbitts S. 2009. Comparison of the PapilloCheck DNA microarray human papillomavirus detection assay with Hybrid Capture II and PCR-enzyme immunoassay using the GP5/6+ primer set. J Clin Virol 45:100–104.

Joura EA. 2002. Epidemiology, diagnosis and treatment of vulvar intraepithelial neoplasia. Curr Opin Obstet Gynecol 14:39–43.

Joura EA, Leodolter S, Hernandez-Avila M, Wheeler CM, Perez G, Koutsky LA, Garland SM, Harper DM, Tang GWK, Ferris DG, Steben M, Jones RW, Bryan J, Taddeo FJ, Bautista OM, Esser MT, Sings HL, Nelson M, Boslego JW, Sattler C, Barr E, Paavonen J. 2007. Efficacy of a quadrivalent prophylactic human papillomavirus (types 6, 11 16, and 18) L1 virus-like-particle vaccine against high-grade vulval and vaginal lesions: A combined analysis of three randomised clinical trials. The Lancet 369: 1693–1702.

MacLean AB. 2004. Vulval cancer: The past 100 years, and into the next century: The Presidential Oration, British Society for the Study of Vulval Disease, 2002. J Obstet Gynaecol 24:491–497.

MacLean AB. 2006. Vulval cancer: Prevention and screening. Best Pract Res Clin Obstet Gynaecol 20:379–395.

McCluggage WG. 2009. Recent developments in vulvovaginal pathology. Histopathology 54:156–173.

Peto J, Gilham C, Fletcher O, Matthews FE. 2004. The cervical cancer epidemic that screening has prevented in the UK. Lancet 364:249–256.

Sideri M, Jones RW, Wilkinson EJ, Preti M, Heller DS, Scurry J, Haefner H, Neill S. 2005. Squamous vulvar intraepithelial neoplasia: 2004 modified terminology, ISSVD Vulvar Oncology Subcommittee. J Reprod Med 50:807–810.

Smith JS, Backes DM, Hoots BE, Kurman RJ, Pimenta JM. 2009. Human papillomavirus type-distribution in vulvar and vaginal cancers and their associated precursors. Obstet Gynecol 113: 917–924.

Syrjanen K, Hakama M, Saarikoski S, Vayrynen M, Yliskoski M, Syrjanen S, Kataja V, Castren O. 1990. Prevalence, incidence, and estimated life-time risk of cervical human papillomavirus infections in a nonselected Finnish female population. Sex Transm Dis 17:15–19.

Todd RW, Etherington IJ, Luesley DM. 2002. The effects of 5% imiquimod cream on high-grade vulval intraepithelial neoplasia. Gynecol Oncol 85:67–70.

Tristram A, Fiander A. 2005. Clinical responses to Cidofovir applied topically to women with high grade vulval intraepithelial neoplasia. Gynecol Oncol 99:652–655.

van de Nieuwenhof HP, van der Avoort IA, de Hullu JA. 2008. Review of squamous premalignant vulvar lesions. Crit Rev Oncol Hematol 68:131–156.

Woo YL, Damay I, Stanley M, Crawford R, Sterling J. 2007. The use of HPV Linear Array Assay for multiple HPV typing on archival frozen tissue and DNA specimens. J Virol Methods 142:226–230.