AN INVESTIGATION OF THE CONTRIBUTION OF CIGARETTE SMOKING AND HUMAN PAPILLOMAVIRUS INFECTION TO THE EPIGENETIC MODULATION OF CELLULAR GENES IN CERVICAL EPITHELIUM

by

YUK TING MA

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Cancer Sciences
College of Medical and Dental Sciences
The University of Birmingham
July 2010

ABSTRACT

In this thesis, I examine the contribution of cigarette smoking and human papillomavirus infection, two independent risk factors for cervical neoplasia, to the epigenetic modulation of cellular genes in cervical epithelium.

To determine the temporal relationship between cigarette smoking and the detection of *CDKN2A* methylation in cervical cytological samples, I used a unique cohort of 2011 women aged 15-19 who were recruited soon after they first had sexual intercourse. I have shown that compared with never-smokers, women who first started to smoke during follow-up had an increased risk of acquiring methylation of the tumour suppressor gene (TSG), *CDKN2A* (odds ratio=3.67; 95% confidence interval: 1.09 to 12.33; p=0.04). I was also able to show that this epigenetic change was often reversible following smoking cessation.

To determine the spectrum of epigenetic changes associated with HPV infection I performed CpG island and gene expression arrays on two *in vitro* models, a HPV replication model and a cervical disease progression model. I went on to show that HPV infection is followed by the up-regulation of the DNA methyltransferase DNMT1 which binds to the TSG, *RARB*, resulting in its *de novo* methylation. Specific CpG loci in the *RARB* promoter appeared to be targeted for *de novo* methylation, rather than methylation of all the CpGs, and may represent the methylation pattern seen at the earliest stages of HPV-induced carcinogenesis.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors, Professor Ciaran Woodman and Professor Paul Murray for making this thesis possible. In particular, I am extremely grateful for the immense support they have given me over the past few years, for their constant guidance and dedication, and for always believing in me. Above all I would like to thank them for making this an enjoyable and rewarding experience.

I am also extremely grateful to the many people in the Institute who have offered me help and advice during my time here, particularly the Woodman and Murray Groups, Dr Wenbin Wei and Dr Sally Roberts.

I would also like to thank Cancer Research UK for funding my Clinical Research Fellowship.

Finally, I would like to thank my family for their constant support over the years, without whom this would not have been achieved.

<u>RESULTS 2</u>

RESULTS 2

<u>**APPENDIX**</u>

Abbreviation	Full description
°C	degrees Celsius
μ	micro-
AIS	adenocarcinoma *in situ*
AP-1	activator protein-1
ATP	adenosine triphosphate
BGS	bisulphite genomic sequencing
bp	base pairs
BSA	bovine serum albumin
CAB	cytological abnormality
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
CI	confidence interval
CIN	cervical intraepithelial neoplasia
cRNA	complementary RNA
Da	daltons
DMEM	Dulbeccco's modified Eagle's medium
DMH	differential methylation hybridisation
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	deoxyribonucleoside triphosphate
DTT	1,4-dithiothreitol
EBV	Epstein-barr virus
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EST	expressed sequence tag
F	forward primer
FBS	foetal bovine serum
FC	fold change
FDA	Food and Drug Administration
g	grams
GAPDH	glyceraldehydephosphate dehydrogenase
GCOS	GeneChip Operating Software
H3	histone 3
HCl	hydrochloric acid
HDAC	histone deacetylases
HEEC	human oesophageal squamous epithelial cells
HES	human embryonic stem cells
HGCIN	high-grade cervical intraepithelial neoplasia
HPV	human papillomavirus

IFN	Interferon
l	litres
LCR	long control region
LMP1	latent membrane protein 1
LOH	loss of heterozygosity
m	milli-
M	molar
M	methylated
MBD	methyl CpG binding domain
MDS	myelodysplastic syndrome
MECP	methyl CpG binding proteins
MSP	methylation specific polymerase chain reaction
mwt	molecular weight
n	nano-
NHFK	normal human foreskin keratinocyte
NNK	4-methylnitrosamino-1-(3-pyridyl)-1-butanone
ORF	open reading frame
p	pico-
P	passage
PARP	poly (ADP-ribose) polymerase
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PHFK	primary human foreskin keratinocytes
PMSF	phenylmethylsulfonyl fluoride
Q-RT-PCR	quantitative reverse transcription polymerase chain reaction
R	reverse primer
R base	Adenine or guanine
RNA	ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
SCC	squamous cell carcinoma
SDS	sodium dodecyl sulphate
SLR	signal log ratio
SSC	sodium chloride-sodium citrate
TBE	Tris-Borate-EDTA
TEMED	tetramethylethylenediamine
Tm	annealing temperature
TSG	tumour suppressor gene
U	unmethylated
UCSC	The University of California, Santa Cruz

UHN	University Health Network
URR	upstream regulatory region
UV	ultraviolet
V	voltage
Y base	Cytosine or thymine

Chapter 1

INTRODUCTION

This chapter is divided into 3 sections. In the first section, I provide a general introduction to epigenetics focusing in particular on DNA methylation and its contribution to the regulation of gene expression.

1.1 Introduction to epigenetics

Epigenetics refers to the study of heritable changes in gene expression that are mediated by changes in DNA other than the nucleotide sequence (Holliday, 1987). Four inter-related types of epigenetic modifications are recognised including DNA methylation, histone modification, chromatin structure, and RNA interference (Bird, 2002; Feinberg & Tycko, 2004; Esteller, 2008).

1.1.1 DNA methylation

DNA methylation involves the enzymatic transfer of a methyl group from the universal methyl donor S-adenosylmethionine to the carbon-5 position of the cytosine ring, producing the modified base 5-methylcytosine (Jones & Baylin, 2002; Herman & Baylin, 2003). This covalent modification occurs after DNA synthesis and is catalysed by the DNA methyltransferase (DNMT) family of enzymes, which I describe in more detail later (Jones & Baylin, 2002; Herman & Baylin, 2003). In the human genome, cytosine methylation occurs almost exclusively at cytosine bases that are located 5' to a guanosine base (the CpG dinucleotide); in embryonic stem cells, cytosine methylation has also been reported at CpA and to a lesser extent CpT dinucleotides, but such non-CpG methylation is barely detectable in adult somatic cells (Ramsahoye *et al*, 2000).

The frequency of the CpG dinucleotide is actually substantially lower than expected in most of the human genome, present at approximately one-fifth of its expected frequency, and about 80% of these are methylated in normal cells (Bird, 1980). 5-methylcytosine undergoes spontaneous deamination to thymine at a relatively high rate, and the failure to recognise and repair this mutation is thought to be the mechanism by which the CpG dinucleotide has been progressively depleted from the bulk of the genome over evolution (Bird, 1980; Sved & Bird, 1990). In contrast, CpG islands contain CpG dinucleotides at the expected frequency, and in normal cells these are usually unmethylated, regardless of the transcriptional state of the gene (Bird, 1986; Gardiner-Garden & Frommer, 1987; Antequera & Bird, 1993; Takai & Jones, 2002). CpG islands may be defined as short regions of DNA (longer than 500bp) with a GC content greater than 55% and an observed/expected CpG ratio of 0.65 (Gardiner-Garden & Frommer, 1987; Takai & Jones, 2002). There are approximately 29 000 CpG islands in the genome predicted by computational analysis (Lander *et al*, 2001; Venter *et al*, 2001), and the majority of these are located in the proximal promoter regions of 40-50% of human genes (Antequera & Bird, 1993; Bird, 2002).

1.1.2 DNA methyltransferases

The DNA methyltransferase (DNMT) family of enzymes catalyse the transfer of a methyl group from the methyl donor S-adenosylmethionine to the carbon-5 position of the cytosine ring (Jones & Baylin, 2002; Herman & Baylin, 2003). There are 2 distinct methyltransferase activities, *de novo* and maintenance. *De novo* methyltransferases methylate unmethylated DNA, thus establishing new DNA methylation patterns. In contrast, maintenance methyltransferases recognise

hemimethylated DNA, and copy the pre-existing methylation pattern onto the nascent DNA strand after DNA replication, thus maintaining the methylation pattern in proliferating cells (Chen & Li, 2004). There are 3 known biologically active DNMTs in mammalian cells, DNMT1, DNMT3A and DNMT3B. DNMT3L is an additional member of the DNMT3 family, but has no methyltransferase activity (Bourc'his *et al*, 2001).

DNMT1 was the first mammalian DNA methyltransferase to be identified (Bestor *et al*, 1988). *In vitro*, DNMT1 exhibited a 5- to 30-fold preference for hemimethylated DNA, and it was therefore originally classified as a maintenance methyltransferase (Yoder *et al*, 1997). DNMT3A and DNMT3B were identified shortly after from searching expressed sequence tag (EST) databases; both enzymes were found to methylate CpG dinucleotides without a preference for hemimethylated DNA and were thus classified as *de novo* methyltransferases (Okano *et al*, 1998). DNMT3A and DNMT3B were also found to be essential for *de novo* methylation in mice, further supporting this classification (Okano *et al*, 1999).

Currently, several lines of evidence suggest that all 3 DNMTs possess both *de novo* and maintenance methyltransferase activities, and that they cooperate *in vivo*. Following disruption of both alleles of DNMT1 in the HCT116 colorectal carcinoma cell line, only a 20% decrease in the overall genomic methylation was observed, and most of the loci examined remained fully methylated and silenced, suggesting that other methyltransferases can maintain the methylation of most of the genome (Rhee *et al*, 2000). Similarly, following disruption of both alleles of DNMT3B in the HCT116

cell line, only a 3% decrease in the global methylation level was observed, whereas disruption of both DNMT1 and DNMT3B in the same cells nearly eliminated methyltransferase activity and decreased the overall genomic methylation by over 95%, despite the presence of DNMT3A, suggesting that DNMT1 and DNMT3B cooperate to maintain DNA methylation in human cancer cells (Rhee *et al*, 2002). Cooperation between the maintenance and *de novo* methyltransferases was also demonstrated in a separate study using *Micrococcus luteus* genomic DNA. Very little methylation was observed in the presence of DNMT3A or DNMT3B alone or in combination. However combinations of *de novo* and maintenance enzymes (DNMT1 + DNMT3A, DNMT1 + DNMT3B, DNMT1 + DNMT3A + DNMT3B) showed increased rates of methylation, which was much higher than with any of the individual enzymes alone (Kim *et al*, 2002).

1.1.3 DNA methylation: normal role

DNA methylation is essential for normal embryonic development, and plays an important role in regulating gene expression (see below), maintaining genome stability, genomic imprinting and X-chromosome inactivation.

<u>1.1.3.1 Embryonic development</u>

Properly established and maintained DNA methylation patterns are essential for embryonic development, and require both *de novo* and maintenance methyltransferase activities (Li, 2002; Jaenisch & Bird, 2003). Consistent with this, inactivation of DNMT1 or DNMT3A and DNMT3B in mouse embryos resulted in global

demethylation or inhibition of *de novo* methylation respectively, both leading to embryonic lethality (Okano *et al*, 1999; Li *et al*, 1992).

In the early embryo an initial wave of genome-wide demethylation of parental DNA occurs in the first cleavage divisions after fertilisation. This erases most, but not all, of the methylation patterns that are inherited from the gametes; methylation of imprinted genes is protected from demethylation at this time (see below). Demethylation is followed by a wave of global *de novo* methylation between blastocyst implantation and gastrulation, re-establishing the overall methylation pattern. This methylation pattern is then maintained throughout life in the somatic cells (Jaenisch *et al*, 1997; Li *et al*, 2002; Jaenisch & Bird, 2003). In the paternal genome, demethylation is an active process, although the demethylase responsible for this has not yet been identified (Mayer *et al*, 2000; Oswald *et al*, 2000). In contrast, the maternal genome is passively demethylated during successive rounds of replication in the absence of nuclear DNMT1 (Rougier *et al*, 1998; Mayer *et al*, 2000).

1.1.3.2 Maintaining genome stability

DNA methylation also helps to maintain transcriptional silence in the noncoding regions of the genome. These regions contain repeat elements, inserted viral sequences and transposons, the transcription of which could be harmful to cells (Yoder *et al*, 1997; Walsh *et al*, 1998; Remus *et al*, 1999; Herman & Baylin 2005).

1.1.3.3 Genomic imprinting

Genomic imprinting refers to the conditioning of maternal and paternal genomes during gametogenesis which leads to the preferential expression of a specific parental allele in the somatic cells of the offspring (Feinberg & Tycko, 2004). The majority of imprinted genes show differences in DNA methylation between the parental alleles, and also differences with respect to bulk chromatin structure and histone modifications (Figure 1.1) (Reik & Walter, 2001).

The imprints are reprogrammed in the gametes to ensure that they reflect the sex of the germ line. This first involves erasure of all the imprints by genome-wide demethylation in the germ cells, followed by *de novo* methylation during oogenesis in females and spermatogenesis in males, in order to establish the correct pattern of allele-specific methylation. This pattern is then maintained throughout development (Reik & Walter, 2001). Evidence suggests that establishment of maternal imprints requires cooperation of DNMT3L, which itself has no methyltransferase activity, with DNMT3A and DNMT3B (Bourc'his *et al*, 2001; Hata *et al*, 2002).

The continuous expression of DNMT1 is essential for maintenance of imprinting. In studies of DNMT1-deficient mice embryos, the *H19* gene, which is normally maternally expressed, was found to be transcribed from both alleles, and the normally active paternal allele of the *Igf2* gene and the normally active maternal allele of the *Igfr2* gene were both silenced (Li *et al*, 1993). These findings indicate that in the absence of DNA methylation, other epigenetic mechanisms such as histone

modifications are insufficient to maintain the mono-allelic expression of imprinted

genes (Li *et al*, 1993).
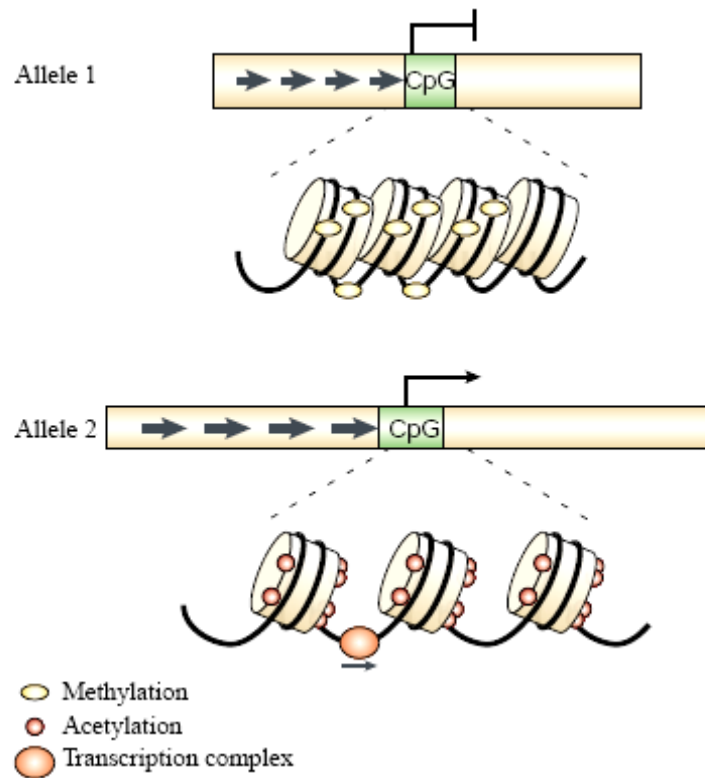
Figure 1.1 Characteristics of imprinted genes. The figure shows a schematic pair of imprinted alleles. The enlarged region below the chromosomes highlights the allele-specific epigenetic changes, such as nucleosomal condensation through deacetylation, and methylation (allele 1) and opening of the chromatin by acetylation and demethylation (allele 2). The transcriptional competence of allele 2 is indicated by the binding of a transcription complex (from Reik & Walter, 2001).

1.1.3.4 X-chromosome inactivation

X-chromosome inactivation refers to transcriptional silencing of one of the 2 X chromosomes in the female, in order to ensure equivalent levels of gene expression from the X chromosome in both males and females (Avner & Heard, 2001; Li, 2002). Inactivation of the X chromosome occurs early in embryogenesis, both X chromosomes have an equal probability of being silenced, and once established the inactive state is stably maintained during subsequent cell divisions (Heard, 2005).

Inactivation of the X chromosome is initiated by transcription of the noncoding *Xist* (X-inactive-specific transcript) RNA, which is expressed exclusively from the inactive X-chromosome. The *Xist* transcripts spread from their site of synthesis to coat the entire X chromosome, triggering transcriptional silencing of the entire chromosome (Avner & Heard, 2001; Heard, 2005). Several mechanisms then act synergistically to maintain the inactive state: continuous expression of *Xist* transcript; histone modifications, such as hypoacetylation of histone H3 and H4, di-methylation of H3 lysine 9 (H3K9me2) and tri-methylation of H3 lysine 27 (H3K27me3); and DNA methylation (Csankovski *et al*, 2001; Heard *et al*, 2001; Plath *et al*, 2003; Heard, 2005). In particular, the recruitment of DNA methylation appears to be crucial for the stable silencing of the X chromosome in somatic cells; in DNMT1-knockout embryos, reactivation of an X-linked transgene was observed when the levels of DNA methylation decreased below a certain threshold (Sado *et al*, 2000).

1.1.4 Epigenetic regulation of gene expression

The local chromatin structure is central to the epigenetic regulation of gene expression and determines whether a gene is transcribed or repressed (Figure 1.2). The chromatin structure is regulated by both nucleosome structure and histone acetylation: Nucleosomes which are in a more open configuration enable regulatory proteins to access chromatin and are thus associated with active gene transcription, while more tightly compacted nucleosomes are associated with gene silencing. Similarly, acetylated histones help to maintain the chromatin in an open and transcriptionally active state, enabling binding of transcription factors and other regulatory proteins that promote gene transcription, while histone deacetylases (HDAC) keep the histones deacetylated and thus transcriptionally silent (Baylin, 2005).
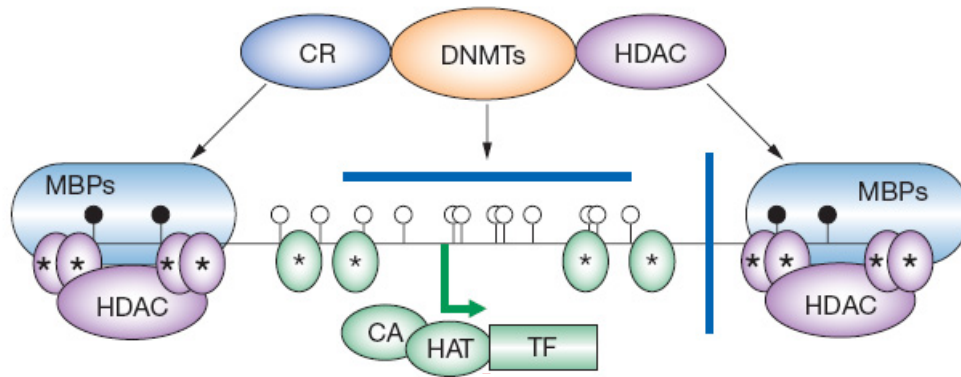
Figure 1.2 Effects of chromatin structure and DNA methylation on gene transcription in normal cells. The chromatin around the transcriptionally active (green arrow), unmethylated promoter (white circles) is occupied by widely spaced nucleosomes composed of histone complexes in which key residues in the tails of histone H3 are in the acetylated state (green ovals). This region is accessible to key components of the gene-transcription apparatus – primary transcription factors (TF), proteins with histone acetyltransferase activity (HAT) and co-activators (CA). The flanking regions to either side of the unmethylated CpG island contain methylated cytosines (black circles). The chromatin structure here is characteristic of transcriptionally silenced regions: nucleosomes are more tightly compacted, with deacetylated histones (purple ovals), and methylcytosine binding proteins (MBP) are bound to methylated DNA. The MBP are also part of complexes that contain histone deacetylases (HDAC), which facilitate histone deacetylation. The apparatus for DNA methylation, consisting of DNMTs, transcriptional co-repressors and HDAC, have access to the flanking regions but not to the unmethylated CpG island in the promoter (from Herman & Baylin, 2003).

1.1.4.1 DNA methylation and regulation of gene expression

DNA methylation has been reported to repress gene transcription through several distinct mechanisms. One mechanism involves directly inhibiting the binding of activating transcription factors to their target sequences. Many transcription factors are targeted to GC-rich sequences and methylation has been shown to prevent the binding of several transcription factors to these sites (Comb *et al*, 1990; Prendergast & Ziff, 1991; Gaston & Fried, 1995; Clark *et al*, 1997). A second mechanism involves the methyl CpG binding proteins (MeCPs), which specifically recognise and bind to methylated DNA (Li, 2002; Klose & Bird, 2006). A family of 5 MeCPs have been identified (MeCP2, MBD1, MBD2, MBD3, MBD4), each containing a homologous methyl CpG binding domain (MBD) (Nan *et al*, 1993; Nan *et al*, 1997; Hendrich & Bird, 1998). In addition, the Kaiso family of proteins, which lack a MBD domain, have also been shown to bind to methylated DNA through their zinc-finger domains (Prokhortchouk *et al*, 2001; Filion *et al*, 2006). The MeCPs mediate repression of gene transcription by recruiting chromatin remodelling co-repressor complexes to regions containing methylated DNA (Jones *et al*, 1998; Nan *et al*, 1998; Ng *et al*, 1999; Wade *et al*, 1999; Zhang *et al*, 1999; Feng *et al*, 2001; Ng *et al*, 2000; Fujita *et al*, 2003; Sarraf & Stancheva, 2004). The DNMT family of proteins can also directly repress transcription, independent of their methyltransferase activity. DNMT1, DNMT3A and DNMT3B contain transcriptional repression domains that function by recruiting histone deacetylases and other co-repressor proteins, similar to the MeCPs, also resulting in chromatin remodelling and transcriptional silencing (Fuks *et al*, 2000; Robertson *et al*, 2000; Rountree *et al*, 2000; Bachman *et al*, 2001; Fuks *et al*, 2001).

DNA methylation can also repress gene expression in association with covalent histone modifications (Bird and Wolffe, 1999; Fuks, 2005). The 2 most studied histone modifications associated with gene silencing are histone deacetylation and methylation of histone H3 lysine 9 (H3K9). Some studies have demonstrated that histone modifications are a prerequisite for DNA methylation (Lehnertz *et al*, 2003; Bachman *et al*, 2003). For example, H3K9 methylation was shown to initiate silencing of *CDKN2A* independently of DNA methylation, with DNA methylation acting subsequently to lock the chromatin in a repressed state (Bachman *et al*, 2003). This is also a feature of X-chromosome inactivation (see above). Other investigators have revealed that DNA methylation may be the dominant epigenetic mechanism regulating gene transcription, with subsequent recruitment of repressive histone modifications to gene promoters (Cameron *et al*, 1999; Fahrner *et al*, 2002; Suzuki *et al*, 2002). For example, treatment of malignant cells with the DNA demethylating agent 5-aza-2'-deoxycytidine, resulted first in demethylation of the hypermethylated *hLMH1* promoter, second to reactivation of gene expression, and finally to reversal of the histone code (enrichment of acetylated H3 and H3K4 methylation with loss of H3K9 methylation), while treatment with the histone deacetylase inhibitor, trichostatin A, failed to reactivate the hypermethylated gene or to dramatically alter histone modifications (Fahrner *et al*, 2002). It seems likely that both DNA methylation and histone modifications play a role at the promoters of different genes.

Another epigenetic mark of transcriptional repression is trimethylation of H3K27 (H3K27me3). While enrichment of H3K27me3 has recently been reported at the promoters of many genes that are hypermethylated and silenced in cancer (McGarvey

et al, 2006), the functional relationship between these 2 repressive mechanisms remains unclear. One study demonstrated a direct interaction between EZH2 (which catalyses trimethylation of H3K27) and the DNMTs both *in vitro* and *in vivo* (Vire *et al*, 2005), but several subsequent studies have provided compelling evidence that H3K27me3 and DNA methylation function independently (McGarvey *et al*, 2007; Kondo *et al*, 2008).

1.1.5 DNA methylation: cancer

A link between DNA methylation and cancer was first demonstrated in 1983 (Feinberg & Vogelstein, 1983). Since that time, nearly every tumour type examined has been characterised by both the global loss of methylation in the CpG-poor regions of the genome where the CpG dinucleotides are usually methylated, and by hypermethylation in the CpG islands located in the promoters of genes where the CpG dinucleotides are usually not methylated (Figure 1.3) (Herman & Baylin, 2003).

<u>1.1.5.1 Hypomethylation and cancer</u>

Hypomethylation at CpG dinucleotides was the first epigenetic abnormality to be identified in cancer cells; using methylation-sensitive restriction enzymes, Feinberg and Vogelstein (1983) observed that a substantial proportion of CpGs that were methylated in normal tissues, were unmethylated in cancer cells.
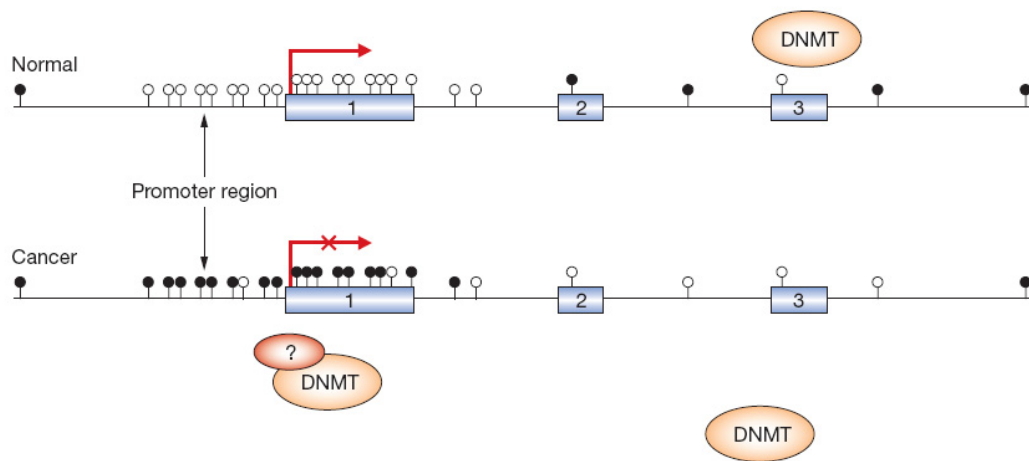
Figure 1.3 Distribution of DNA methylation in normal and cancer cells. In normal cells, most CpG sites outside of CpG islands are methylated (black circles), whereas most of the CpG sites within CpG islands located in the promoter regions are unmethylated (white circles). In cancer cells, many of the CpG sites in the bulk of the genome become unmethylated, with aberrant methylation of CpG islands located in the promoter regions, which is associated with transcriptional silencing (from Herman & Baylin, 2003).

Several mechanisms have been proposed to explain the contribution of hypomethylation to the development of cancer (Esteller, 2008). Firstly, hypomethylation may contribute to carcinogenesis by inducing chromosomal instability. Aneuploidy and a large increase in the number of novel chromosomal translocations have been reported following disruption of the DNMTs in the HCT116 colorectal carcinoma cell line (Karpf & Matsui, 2005). In sporadic colorectal cancers, tumours with global hypomethylation were also reported to have a significantly increased number of loci with loss of heterozygosity (LOH) (Matsuzaki *et al*, 2005). Secondly, loss of methylation may allow the reactivation of transposable elements which can then be translocated to other genomic regions, disrupting normal cellular genes (Bestor *et al*, 2005; Esteller, 2008). Thirdly, loss of methylation may also allow the expression of normally silenced genes such as imprinted genes. The first gene to display loss of imprinting in cancers was *IGF2* (insulin-like growth factor gene), and loss of imprinting of this gene has been associated with the development of 50% of Wilm's tumours in children and with an increased risk of colorectal cancer in adults (Wilkins, 1988; Ravenel *et al*, 2003; Cui *et al*, 2003). In addition hypomethylation may also lead to the activation of oncogenes. Mechanistically, while genome-wide hypomethylation has been associated with the activation of normally silent genes in some cancers eg *MAGE-1*, for other genes, hypomethylation appears to gene-specific rather than random by-products of global hypomethylation (De Smet *et al*, 1996; Sato *et al*, 2003).

1.1.5.2 Hypermethylation and cancer

The first tumour suppressor gene reported to be hypermethylated was the retinoblastoma gene *(Rb)* (Greger *et al*, 1989; Sakai *et al*, 1991). Since then, hypermethylation of CpG islands in the proximal promoter region of tumour suppressor genes has been reported in virtually every tumour type and is frequently associated with aberrant silencing of gene expression. Indeed, the number of tumour suppressor genes inactivated by aberrant promoter hypermethylation equals or exceeds the number that are inactivated by mutations or deletions (Herman & Baylin, 2003).

However the causes of aberrant hypermethylation in cancer remains poorly understood. In particular, it is not known why some CpG islands become hypermethylated in cancer while others do not. Increased DNMT1 and DNMT3B mRNA levels have been observed in a number of solid and haematological malignancies, and the DNA methyltransferases are considered to be central to the process of aberrant methylation (Esteller, 2005). However, it is clear that methylation is not a random process; different tumour suppressor genes are inactivated by methylation in different cancers, and there has been no TSG which has been found to be hypermethylated in every cancer type (Costello *et al*, 2000; Esteller *et al*, 2001; Issa *et al*, 2004; Keshet *et al*, 2006). Studies investigating methylation following exposure to carcinogens such as smoking and viruses are beginning to increase our understanding of the possible underlying mechanisms and these aspects will be discussed in more detail later; smoking and HPV-induced DNA methylation are the main focus of this thesis.

Recently, DNA methylation has been reported to be a function of promoter CpG content (Weber *et al*, 2007). In this study, promoters were classified into 3 categories based on CpG ratio, GC content and the length of the CpG-rich region. Promoters with a high CpG content represented 'strong' CpG islands, promoters with an intermediate CpG content represented 'weak' CpG islands and the promoters with low CpG content represented promoters with no local enrichment of CpG dinucleotides. The promoters with intermediate CpG content were found to be methylated in somatic cells but not in the germline, suggesting that 'weak' CpG islands are the preferential targets for *de novo* methylation during development. Conversely, the promoters with high CpG content were usually not methylated even when inactive, and were possibly protected from *de novo* methylation by the high levels of H3K4 di-methylation present (Weber *et al*, 2007).

Another recent mechanism that has been reported to predispose genes to DNA methylation in cancer is the presence of a stem cell-like chromatin pattern. Embryonic stem cells are pluripotent cells which rely on the Polycomb-mediated tri-methylation of H3K27 to reversibly repress genes that are essential for differentiation (Ringrose *et al*, 2004; Lee *et al*, 2006). Several studies have demonstrated that genes that become *de novo* methylated in cancer are frequently marked by H3K27me3 in embryonic stem cells (Widschwendter *et al*, 2007; Schlesinger *et al*, 2007). Subsequent studies have revealed that the promoters of key developmental genes in embryonic stem cells are actually marked by a 'bivalent' chromatin pattern, the active mark H3K4 tri-methylation (H3K4me3) and the repressive mark H3K27me3, which keeps these genes in a 'transcription-ready' state (Berstein *et al*, 2006; Ohm *et al*, 2007; Zhao *et*

al, 2007). These bivalent domains are usually lost during embryonic stem cell differentiation, and differentiated cells are typically marked by either H3K27me3 or H3K4me3, but not by both (Berstein *et al*, 2006). This bivalent chromatin pattern characteristic of stem cells has been shown to predispose genes to DNA hypermethylation during the initiation and progression of malignancy (Ohm *et al*, 2007).

In Chapter 5 of this thesis, I perform an analysis based on these 2 determinants of DNA methylation, promoter CpG content and stem cell-like chromatin pattern.

1.1.5.3 Epigenetic switching

It is clear the relationship between H3K27me3 and DNA methylation is complex. The recently described process of epigenetic switching adds yet another layer to this complicated relationship. In embryonic stem cells, genes essential for differentiation are repressed by H3K27me3 in the absence of methylation. In prostate cancer cells it has been shown that these marks can be removed and replaced with DNA methylation, which changes the reversible gene repression to one of permanent silencing (Figure 1.4) (Gal-Yam *et al*, 2008). In Chapter 5 of this thesis, I also investigate possible epigenetic switching for one candidate gene.
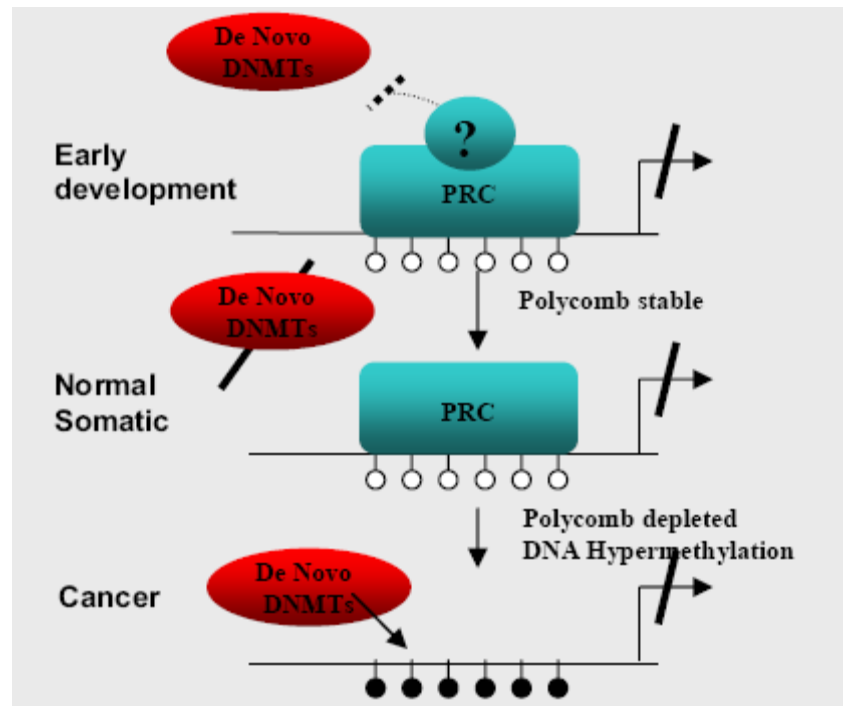
Figure 1.4 Epigenetic switching. Schematic promoters are shown. Arrows represent transcription. Black circles represent methylated CpG loci. White circles represent unmethylated CpG loci. In embryonic stem cells, genes can be repressed by polycomb repressive complexes (PRC) in the absence of DNA methylation. In cancer, loss of the PRC occurs and this is replaced with DNA methylation (from Gal-Yam *et al*, 2008).

1.1.6 Clinical implications: Reversal of epigenetic silencing

Unlike mutations, epigenetic gene silencing is potentially reversible, and there is therefore great interest in developing new therapeutic strategies using epigenetic agents for the management of patients with cancer.

In experimental settings, inhibitors of DNA methylation rapidly reactivate genes that have become epigenetically silenced. The prototype inhibitors of DNA methylation, 5-azacytidine (azacitidine), and 5-aza-2'-deoxycytidine (decitabine) were initially developed as classical cytotoxic agents over 40 years ago (Sorm *et al*, 1964). Subsequently it was found that these drugs could inhibit DNA methylation in human cell lines, inducing gene expression and differentiation (Constantinides *et al*, 1977; Jones & Taylor, 1980). Both drugs are nucleoside analogues and are converted to deoxynucleoside tri-phosphates which are then incorporated into DNA in place of cytosine. They are therefore only active during S-phase. These 'false' bases then form covalent bonds with the DNA methyltransferases, trap these enzymes and target them for degradation, resulting in demethylation of DNA during subsequent rounds of cell division (Jones & Taylor, 1980; Egger *et al*, 2004). In patients with high-risk myelodysplastic syndrome (MDS), a pre-leukaemic bone marrow disorder, azacitidine and decitabine have been tested in phase 3 clinical trials and have both demonstrated high response rates, reduced risk of leukaemic transformation and improved median survival (Silverman *et al*, 2002; Kantarjian *et al*, 2006). *In vivo*, demethylation and gene reactivation have also been demonstrated (Daskalakis *et al*, 2002; Mund *et al*, 2005; Yang *et al*, 2006). Consequently, both drugs have now been approved by the FDA for treatment of MDS.

However decitabine was found to be ineffective when initially used in patients with various solid tumours (Abele *et al*, 1987). It is now thought that this lack of activity may be related to the use of too high doses, limited number of days of drug exposure and assessment of response after only one cycle (Issa, 2007). These drugs are also limited by being S-phase dependent, thus they will be inactive in non-cycling cells. They cause significant myelosuppresion and they are also unstable in aqueous solutions. As they are incorporated into DNA, this is also thought to contribute to their cytotoxicity (Herman & Baylin, 2003; Egger *et al*, 2004). Consequently the development of new agents is an active area of research. Rationally designed inhibitors of the DNMTs are emerging, but as the DNMTs are known to cooperate, it is likely that simultaneous inhibition of several DNMTs will be required. Rational design of combination therapies, such as DNA methylation and histone acetylation inhibitors, which are known to act synergistically *in vitro*, are also being considered (Issa, 2007).

1.2 Cervical cancer

In this section I provide an introduction to cervical cancer, focusing in particular on the role of the 2 main risk factors, human papillomavirus (HPV) and smoking.

1.2.1 Epidemiology of cervical cancer

Cervical cancer is the second most common cancer among women worldwide. In 2007, there were an estimated 555,000 new cases and 309,000 deaths from cervical cancer worldwide (American Cancer Society website, May 2010). The majority of cases occur in developing countries, where they account for 15% of all female cancers. In comparison, the incidence rate is low in developed countries, accounting for less than 5% of all female cancers; in the UK, there were 2,828 new cases in 2007, making it only the eleventh most common cancer in women in the UK (Parkin & Bray, 2006; CRUK website, May 2010). The incidence rate is highest among those aged 30-39, with a second peak seen in older women in some countries (Bosch & de Sanjose, 2003; Franceschi *et al*, 2006).

Since the introduction of population-based screening programmes in the 1960s and the 1970s, the incidence and mortality from cervical cancer has declined dramatically in many developed countries. In some populations, the decline in mortality predated the introduction of screening, and thus factors such as improved socioeconomic levels, earlier stage at diagnosis and more effective treatments may have also contributed to this improvement. In developing countries, the incidence and mortality rates have remained relatively stable, and cervical cancer remains the most common cause of cancer death in females in these countries (Parkin & Bray, 2006).

Squamous cell carcinoma (SCC) is the most common histological type, accounting for approximately 80% of all cervical carcinomas, while adenocarcinomas account for most of the remaining cases (International Collaboration of Epidemiological Studies of Cervical Cancer 2006a). In many developed countries the incidence of cervical SCC has fallen considerably as a consequence of cervical screening. Conversely, numerous studies in the past 2 decades have reported increasing rates of adenocarcinoma, particularly in younger women (age <40 years), although since the mid- to late-1990s declining rates of adenocarcinoma are now being observed in some countries (Schwartz & Weiss, 1986; Antilla *et al*, 1999; Smith *et al*, 2000; Hemminki *et al*, 2002; Bulk *et al*, 2005; Howlett *et al*, 2007; Andrae *et al*, 2008; Sasieni *et al*, 2009; Mathew & George, 2009). This initial increase in the incidence of adenocarcinoma is thought to be related to generational changes in sexual behaviour leading to increased transmission of HPV in younger generations; younger age at first intercourse, increased number of sexual partners, and increasing risk that each sexual partner is HPV positive (Bray *et al*, 2005). The poorer detection of adenocarcinomas by screening programmes is also thought to have contributed to the increasing rates of adenocarcinoma: as adenocarcinomas and their precursor lesions arise from glandular epithelium within the endocervical canal, they are therefore less amenable to detection by cytological screening which involves scraping cells from the epithelium of the ectocervix (Bray *et al*, 2005; Parkin & Bray, 2006; Sasieni *et al*, 2009). In turn, an increased awareness of adenocarcinoma precursor lesions among pathologists and clinicians and improved specimen collection are both thought to have contributed to the more recent decline in the incidence of adenocarcinoma seen in some countries

(Howlett *et al*, 2007; Andrae *et al*, 2008; Sasieni *et al*, 2009; Mathew & George, 2009).

1.2.2 Histology of the cervix

The epithelial lining of the cervix varies; the endocervix is lined by a single layer of columnar epithelium, while the ectocervix is covered by non-keratinising, stratified squamous epithelium. The junction between the endocervix and the ectocervix is known as the original squamo-columnar junction (Figure 1.5a). This junctional region undergoes continuous dynamic modification which is greatest at puberty and during a woman's first pregnancy, when hormonal changes result in extension of the columnar epithelium onto the ectocervix forming an ectropion or cervical erosion (Figure 1.5b). Exposure of the columnar epithelium to the acidic environment of the vagina (post-puberty) then induces squamous metaplasia, where the columnar epithelium becomes replaced by new squamous epithelium, and this area is known as the transformation zone (Figure 1.5c) (Stevens & Lowe, 1992; Critchlow *et al*, 1995; Jacobson *et al*, 1999). The transformation zone is the most common site for the development of cervical neoplasia (Arends *et al*, 1998).
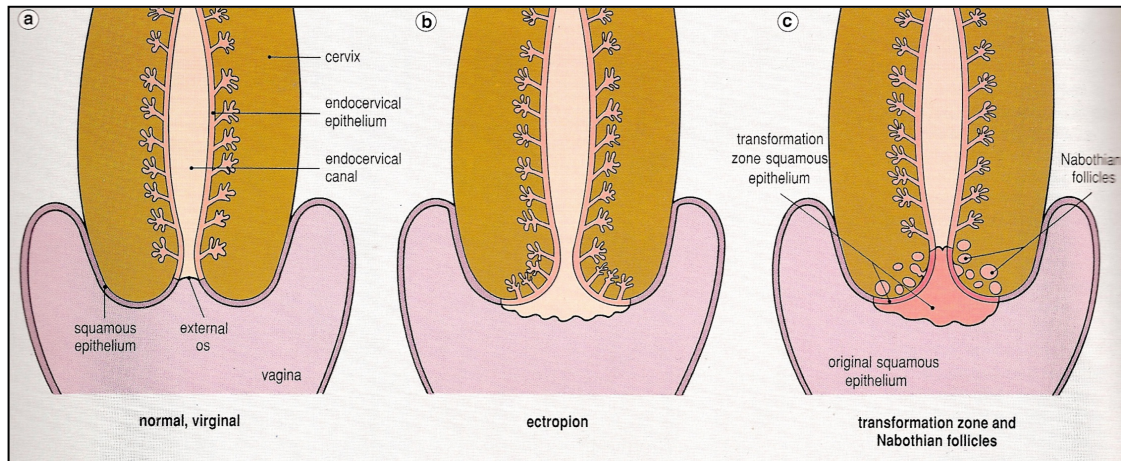
Figure 1.5 Squamo-columnar junction of the cervix. a) The squamo-columnar junction is originally situated in the region of the external os. b) At puberty, the endocervical epithelium extends distally into the acid environment of the vagina, forming an ectropion. c) A transformation zone forms as squamous epithelium grows over the ectropion (from Stevens & Lowe, 1992).

1.2.3 Pre-invasive lesions

Pre-invasive lesions of the cervix are asymptomatic and are usually detected during routine cervical cytological screening.

1.2.3.1 Cervical intraepithelial neoplasia

Cervical intraepithelial neoplasia (CIN) represents a spectrum of histological abnormalities. The grading of CIN refers to the level in the epithelium at which cytoplasmic maturation occurs (Wells, 1996):

- CIN grade 1 (CIN1) represents mild dysplasia; mild nuclear abnormalities are present throughout the epithelium, and cytoplasmic differentiation is present in the upper two-thirds of the epithelium.

- CIN grade 2 (CIN2) represents moderate dysplasia; nuclear abnormalities are more marked and are again present throughout the epithelium, and cytoplasmic differentiation is limited to the upper third of the epithelium.

- CIN grade 3 (CIN3) represents severe dysplasia and carcinoma *in situ*; severe nuclear abnormalities are present throughout the epithelium, and there is virtually no normal cytoplasmic differentiation visible.

A modification of the original CIN classification was introduced in 1990, where CIN1 is now designated low-grade CIN, and CIN2 and CIN3 are classified together as high-grade CIN (Richart, 1990). In the UK, women with low-grade CIN may be kept under colposcopic surveillance, but all women with high-grade CIN are treated immediately (NHS Cervical Screening Programme website, May 2010). Treatment of CIN lesions

involves either removal of the transformation zone (where the abnormalities arise) by large loop excision of the transformation zone (LLETZ), cone biopsy or rarely, hysterectomy, or by destroying the cells in the abnormal area with laser therapy, cold coagulation, or cryotherapy (Schiffmann *et al*, 2007). In the UK, LLETZ is the most common method used.

1.2.3.2 Adenocarcinoma *in situ*

The pre-invasive component of adenocarcinoma is known as adenocarcinoma *in situ* (AIS). The management of women with AIS is challenging: The colposcopic changes associated with AIS may be minimal, which makes it difficult to determine the full extent of the lesion. AIS also tends to extend a considerable distance into the endocervical canal, making complete excision difficult. In addition, AIS is often multifocal and skip lesions are commonly found, thus negative margins from a diagnostic excision specimen do not necessarily mean that the lesion has been completely excised. Consequently hysterectomy is the treatment of choice in women who have completed childbearing (Wright *et al*, 2007).

1.2.4 Invasive disease

Eventually the abnormal epithelial cells may breach the basement membrane and invade into the cervical stroma.

1.2.4.1 Squamous cell carcinoma

This is the most common histological type of invasive cervical cancer. The earliest sign of SCC is early stromal invasion, when small foci (<1mm) in the basal

epithelium breach the basement membrane. The tumour can then spread both locally and via the lymphatic vessels. The 2 most important prognostic factors in SCC are the depth of the invasion of the primary tumour and the presence and extent of lymph node metastases (Wells, 1996).

1.2.4.2 Adenocarcinoma

Adenocarcinoma of the cervix is the second most common histological type of cervical cancer, after SCC. Adenocarcinoma has a pattern of dissemination which is different to that of SCC. For example adenocarcinoma more frequently metastasises to the lymph nodes and ovary. The most important prognostic factors for adenocarcinoma are clinical stage and the presence of lymph node metastases (Gien *et al*, 2010).

1.2.5 Natural history of cervical cancer

The natural history of cervical cancer is thought to be characterised by a well-defined pre-malignant phase. However, an accurate assessment of the rates of progression or regression of cervical epithelial abnormalities in the modern era is difficult to attain because all women with high-grade CIN and many women with low-grade CIN are now treated. Based on a review of the early natural history studies (1950s onwards), it was estimated that the majority of low-grade CIN lesions will regress spontaneously (60%), while 30% will persist, 10% will progress to CIN3 and only 1% will progress to invasive cancer. High-grade CIN lesions may still spontaneously regress in up to 33% of cases, however the probability of an atypical epithelium becoming invasive increased with the severity of the atypia (Ostör, 1993). Currently cytological and

histological examination is not able to accurately predict the minority of women with CIN who will progress from the vast majority whose abnormalities will spontaneously regress, consequently all high grade lesions are treated (Woodman *et al*, 2007). More useful prognostic markers than those provided by morphological examination are clearly needed.

The time to progression from pre-invasive disease to invasive cervical cancer is also thought to be a slow process; the mean age of a woman with CIN is approximately 15 years younger than that of a woman with invasive cervical cancer (zur Hausen, 2002).

1.2.6 Risk factors for cervical neoplasia

The single most important risk factor for the development of cervical neoplasia is infection with high-risk types of human papillomavirus (HPV) and this will be discussed in more detail below. In addition, the number of sexual partners, younger age at first sexual intercourse, high parity, younger age at first full term pregnancy, increasing duration of oral contraceptive use and smoking are all associated with an increased risk of developing cervical neoplasia (International Collaboration of Epidemiological Studies of Cervical Cancer 2006b, 2006c, 2007, 2009). Recently a re-analysis of 12 epidemiological studies involving 8,097 women with SCC and 1,374 women with adenocarcinoma, demonstrated that all the risk factors listed above are common to both cervical SCC and adenocarcinoma, with the exception of smoking. Current smoking was associated with a significantly increased risk of SCC (RR=1.50, 95% CI: 1.35-1.66) but not of adenocarcinoma (RR=0.86, 95% CI: 0.70-1.05), and the difference between these 2 histological types was statistically significant

(International Collaboration of Epidemiological Studies of Cervical Cancer 2006a). This was also demonstrated in the larger pooled analysis of individual data from 13,541 women with cervical cancer from 23 epidemiological studies (International Collaboration of Epidemiological Studies of Cervical Cancer 2006b). With the exception of infection with HPV, the magnitude of the risk imparted by all the other risk factors is only modest, with relative risks between 1.5-2.5 (International Collaboration of Epidemiological Studies of Cervical Cancer 2006b, 2006c, 2007, 2009).

1.3 Human papillomavirus

Human papillomaviruses (HPVs) are small double-stranded DNA viruses that infect basal epithelial cells, causing benign and malignant lesions in the cutaneous and mucosal epithelia. More than 100 HPV types have been identified, and over 40 of these are known to infect the anogenital tract. The genital HPV types are further subdivided into low-risk and high-risk types (Table 1.1). The low-risk types usually cause benign lesions such as warts and condyloma acuminata, and are rarely associated with malignancy. In contrast, the high-risk types are frequently associated with cervical cancer. Fifteen high-risk HPV types have been identified and 5 of these (types 16, 18, 45, 31 and 33) account for 90% of all cases of cervical cancer worldwide (Muñoz *et al*, 2003).

Classification	HPV types
High-risk	16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, 82
Probable high-risk	26, 53, 66
Low-risk	6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81

Table 1.1 Classification of the major HPV types

1.3.1 Organisation of the HPV genome

The HPV virion consists of a non-enveloped, double-stranded circular DNA within an icosahedral capsid (Doorbar, 2006). The HPV genome is approximately 8kb in length and can be divided into 3 major regions, an early coding region, a late coding region and an 850bp long control region (LCR) (Figure 1.6). The early region contains 6 open reading frames (ORF) that encode the early (E) proteins, E1, E2, E4, E5, E6 and E7. The late region contains 2 ORF which encode the late (L) proteins, L1 and L2. The LCR (also known as the upstream regulatory region, URR) is a non-coding region, which contains enhancer elements, a promoter and binding sites for many transcription factors (Zheng & Baker, 2006).

The high-risk HPV types contain 2 major viral promoters. The early promoter lies upstream of the E6 ORF and is responsible for nearly all early gene expression; in HPV16 the early promoter is referred to as P97, and in HPV18 it is referred to as P105. The second, or late, promoter lies within the E7 ORF and is responsible for late gene expression; in HPV16 this is referred to as P670 and there is accumulating evidence that a late promoter is also present in HPV18 (Hirochika *et al*, 1987; Gloss & Bernard, 1990; Romanczuk *et al*, 1990; Grassman *et al*, 1996; Parker *et al*, 1997; Zheng & Baker, 2006).
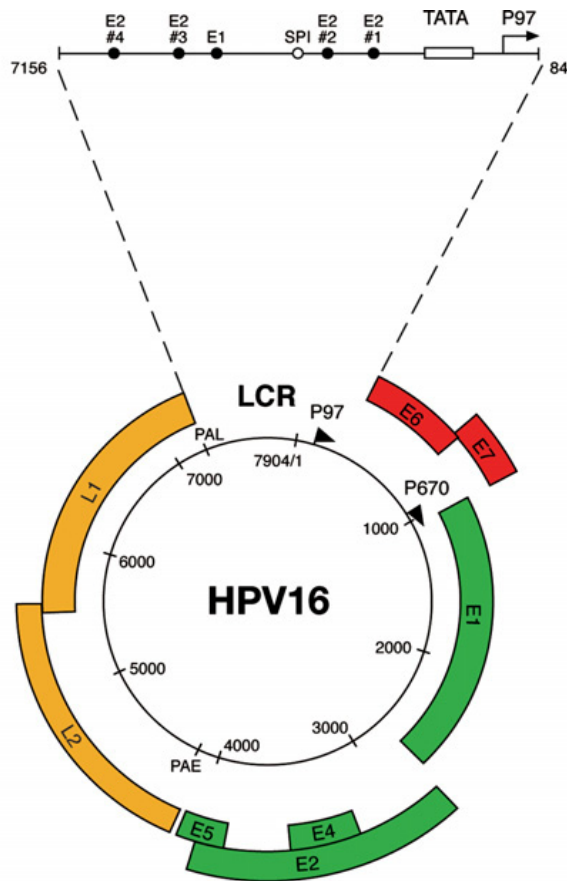
Figure 1.6 Organisation of the HPV genome. The HPV16 genome (7904bp) is shown as a black circle with the early (P97) and late (P670) promoters marked by arrows. The locations of the 8 open reading frames (E1, E2, E4, E5, E6, E7, L1 and L2) are shown. The long control region (LCR) is enlarged to show the E1 and E2 binding sites, the SP-1 binding sites, the TATA box and the P97 promoter (from Doorbar, 2006).

<u>1.3.1.1 Early proteins</u>

The E1 and E2 proteins are essential for viral replication (Hebner & Laimins, 2006).

The E1 protein functions as a helicase, catalysing the unwinding of DNA, and also

recruits the DNA polymerase alpha-primase to the viral origin of replication

(Masterson *et al*, 1998; Conger *et al*, 1999). The E2 protein helps to recruit E1 to the

origin of replication, and also functions as a transcription factor, regulating

transcription of E6 and E7 from the early promoter. The LCR of the high-risk HPVs

contains four E2-binding sites (see Figure 1.6); at low levels E2 activates viral

transcription, while at high levels E2 represses expression of the viral oncogenes, by

displacing basal transcription factors that are essential for promoter activation (Cripe

et al, 1987; Steger *et al*, 1997; Hebner & Laimins, 2006).

The E1^E4 and E5 proteins are differentiation-dependent proteins and may play a role

in viral genome amplification by inducing cell cycle arrest at G2, or by stimulating

cell cycle progression, respectively (Doorbar *et al*, 1997; Fehrmann *et al*, 2003;

Knight *et al*, 2006; Conway & Meyers, 2009). In addition, in the stratified epithelium,

the E1^E4 proteins are believed to cause collapse of the cytoskeleton networks, thus

allowing release of the virions from these cells (Bryan & Brown, 2000).

The E6 and E7 proteins from high-risk, but not low-risk, HPV types are oncoproteins.

The high-risk E6 protein binds to the p53 tumour suppressor protein and promotes its

degradation by forming a trimeric complex with the cellular ubiquitin ligase E6AP

(Scheffner *et al*, 1990; Werness *et al*, 1990). The E6 oncoprotein can also down-

regulate p53 activity by targeting the transcriptional activator CBP/p300 (Patel *et al*,

1999; Zimmermann *et al*, 1999). The resulting rapid reduction in the steady-state levels of p53 alleviates the restraint on cell cycle progression, enabling viral replication. Activation of telomerase is another important function of the E6 oncoprotein (Klingelhutz *et al*, 1996). The high-risk E7 protein binds to and degrades the Rb tumour suppressor protein and other members of the pocket protein family, which releases the E2F transcription factor from Rb inhibition, leading to transcription of S-phase genes (Dyson *et al*, 1989).

1.3.1.2 Late proteins

The L1 and L2 ORF encode the major (L1) and minor (L2) capsid proteins, respectively, and once expressed these proteins assemble into viral capsids. The HPV icosahedral capsid is composed of 360 copies of the 55 kDa L1 capsid protein arranged as 72 pentameric capsomeres, with 12 copies of the 74 kDa L2 protein, possibly associated with the 12 pentavalent capsomeres. (Chen *et al*, 2000; Modis *et al*, 2002; Finnen *et al*, 2003).

1.3.2 HPV life cycle

The human papillomavirus life cycle is tightly linked to differentiation of the host epithelial cell, and can be divided into 2 stages, nonproductive and productive (Figure 1.7) (Doorbar, 2005). HPV can only replicate in basal epithelial cells (Egawa, 2003), and the virus is thought to gain access to these cells through micro-abrasions in the epithelial tissue (Doorbar, 2005). In the basal cells the virus expresses the early proteins E1, E2, E5, E6, and E7 and replicates with the cellular DNA during S-phase, maintaining its genome as a low copy number episome (20-100 copies per basal cell)

(Stanley *et al*, 1989; De Geest *et al*, 1993). This is the non-productive phase of the life cycle.
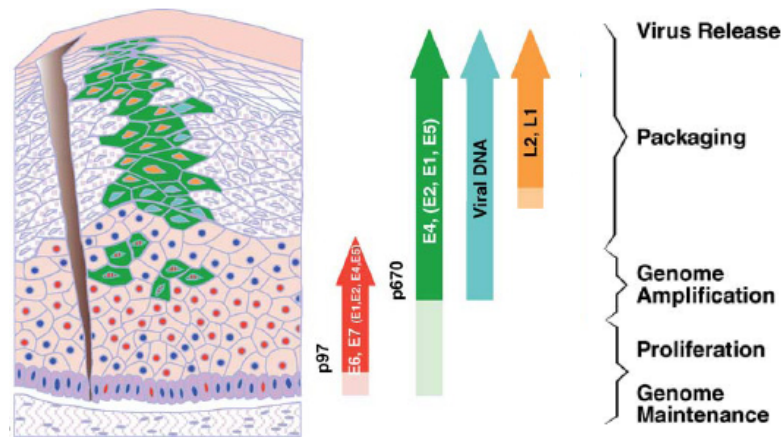
Figure 1.7 HPV life cycle. Diagrammatic representation of the skin to reveal the pattern of HPV16 gene expression as the infected cell migrates towards the epithelial surface. After infection the viral genome is maintained as a low copy number episome. During epithelial differentiation the P97 promoter directs expression of E6 and E7 genes necessary for S-phase entry (red). The P670 promoter is up-regulated in the higher epithelial layers, and viral replication proteins (E1, E2, E4, E5) increase in abundance (green), facilitating amplification of viral genomes (blue). In the upper epithelial layers the viral capsid proteins are found (yellow) (from Doorbar, 2005).

As the HPV-infected basal cells divide, the viral episomes are replicated and are distributed evenly among the daughter cells. Following mitosis, one cell remains attached to the basal layer, while the other cell migrates up through the suprabasal layers and begins to differentiate (Lehman & Botchan, 1998; Bastien & McBride, 2000; Doorbar, 2005). Non-infected cells normally exit the cell cycle once they become detached from the basal layer. In contrast, detached HPV-infected epithelial cells remain mitotically active due to the expression of the E6 and E7 proteins (Cheng *et al*, 1995; Ruesch & Laimins, 1998; Flores *et al*, 2000). Hence, HPV infection with consequential expression of the E6 and E7 proteins, allows the HPV-infected cell to override the normal cell cycle exit that occurs on differentiation, in order to enable production of viral progeny (Laimins, 1996).

In the suprabasal cells, the late promoter is activated, the virus expresses the E4, L1 and L2 proteins and the viral genomes are amplified to a high copy number per cell (Laimins, 1996; Doorbar, 2005). This is the productive phase of the life cycle. In the uppermost layers of the epithelium, the viral genomes are packaged into the newly formed capsids, and the progeny virions are then released from the cell (Doorbar, 2005). The time from HPV infection to subsequent release of virions is approximately 3 weeks, which is the time required for the basal cells to undergo complete differentiation and desquamation (Stanley, 2006).

1.3.3 Persistence of HPV infection

Cervical HPV infection is a very common sexually transmitted infection and most young women are infected soon after the onset of sexual activity (Collins *et al*, 2002;

Bosch & de Sanjose, 2003). However, the majority of women will spontaneously clear their infection within a short period of time, and persistent infection is seen in only approximately 15% of women (Ho *et al*, 1998; Woodman *et al*, 2001; Stanley *et al*, 2007). Persistent infection with high-risk HPV types is generally considered to be a major risk factor for the development of high-grade CIN and cervical cancer (Ho *et al*, 1995; Remmink *et al*, 1995). However, there is no accepted definition of persistent infection (Schiffman *et al*, 2007). A review of the literature demonstrated that in many studies women were considered to have persistent infection if they tested positive for HPV on 2 or more occasions, but in these studies the duration of the interval between tests ranged from 2 months to 7 years (Woodman *et al*, 2007). Clearly, the shorter the interval between tests, the more likely it is for an infection to be defined as persistent (Figure 1.8). Nevertheless, it is now clear that persistent viral infection is not always necessary for disease progression; high-grade CIN has been detected soon after infection with high-risk HPV (Woodman *et al*, 2001).
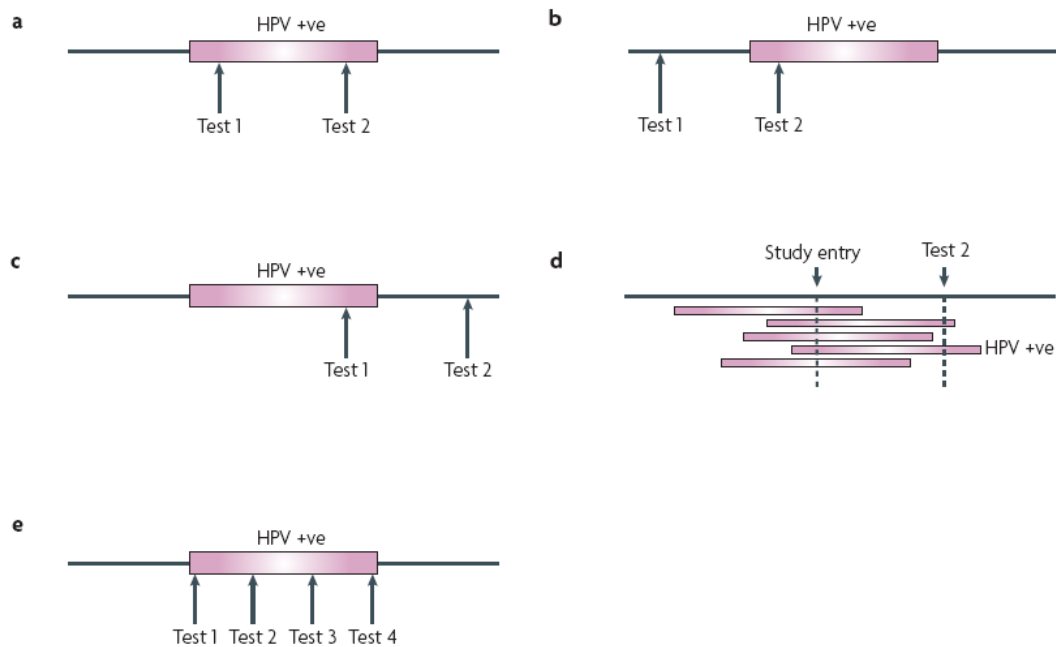
Figure 1.8 Persistent HPV infection. The rectangles denote a period of time during which HPV DNA sequences could be detected. In many studies, a woman is considered to have a persistent infection if she is HPV DNA-positive in two or more consecutive tests, and a transient infection if she is positive only once. Panels a) - c) show how the same infection might be considered persistent or transient, depending on when the samples were taken. a) The infection is characterized as persistent. b), c) The same infection is now characterized as transient merely by changing the sampling times. d) These episodes of HPV infection are of identical duration, but began at different, undetermined times before entry to the study. Two of these infections would have been considered persistent, and three transient, on the basis of when the second test was performed. e) With a definition based on the number of positive tests, an infection is more likely to be considered persistent if a woman is tested more frequently (from Woodman, 2007).

1.3.4 HPV and cervical neoplasia

Infection with the high-risk types of human papillomavirus is the major cause of cervical cancer, and high-risk HPV DNA can be detected in virtually all cervical cancers (International Agency for Research on Cancer 1995 and 2007; Walboomers *et al*, 1999). In squamous cell carcinoma HPV16 is the most common type followed by HPV18, while in adenocarcinomas HPV18 is the most common type detected. Approximately 85% of high-grade CIN lesions also contain high-risk HPV DNA; HPV16 is again the most common type detected, but HPV18 is significantly under-represented (Smith *et al*, 2007). Infection with HPV18 is associated with an increased risk of high-grade CIN, but it has been found that there is a lack of moderate or severe cytological changes after an incident HPV18 infection, suggesting that the cytological changes detected after infection with HPV18 underestimates the severity of the underlying histological abnormality (Woodman *et al*, 2003; Kovacic *et al*, 2006). This may explain why HPV18 is significantly under-represented in high-grade CIN, as women will be referred for colposcopic assessment only if the smear shows moderate or severe dyskaryosis. The recent finding that HPV18 tends to integrate early, may explain the predominance of minor cytological changes following HPV18 infection; early integration of HPV18 was associated with decreased viral load, which in turn is known to be associated with less severe cytological abnormalities (Collins *et al*, 2009).

1.3.4.1 Integration of HPV

HPV integration occurs in up to 80% of HPV16-positive and 100% of HPV18-positive carcinomas (Bosch & de Sanjose, 2003). Viral integration usually leads to the

disruption of the E1 and/or E2 open reading frames, with subsequent release of the repression of E6 and E7 oncogene expression (Romanczuk *et al*, 1992). Integration is not a normal part of the HPV life cycle - it results in the deletions of genes that are essential for the viral synthesis – but it provides cells with a selective growth advantage (Jeon *et al*, 1995). These cells also show increased genomic instability which can also contribute to malignant progression (Pett *et al*, 2004).

Integration of high-risk HPV is known to occur near to common fragile sites (Thorland *et al*, 2000; Wentzensen *et al*, 2004; Yu *et al*, 2005), and recently it has been shown that this is because these regions are permissive sites for insertion of foreign DNA, rather than because it confers a selective growth advantage when disrupted (Dall *et al*, 2008). It was also found that although the majority of integration events occur in chromosome bands containing a cancer-associated gene, these genes were generally a considerable distance from the integration site, and there was no evidence of altered expression of selected candidate genes. This suggests that insertional mutagenesis is not a common consequence of HPV16 integration (Dall *et al*, 2008).

Persistent viral infection with constitutive expression of the E6 and E7 genes is necessary to maintain the malignant phenotype (Hu *et al*, 1995). However, *in vitro*, the expression of HPV16 E6 and E7 is only sufficient to immortalise human keratinocytes (Hawley-Nelson *et al*, 1989; Münger *et al*, 1989; Sedman *et al*, 1991; Dürst *et al*, 1995). In addition, only a small percentage of women infected with high-risk HPV will progress to invasive disease, and this is usually after a long latency

period (zur Hausen, 2002). Thus infection with HPV is necessary but not sufficient for the development of cervical cancer and additional alterations are required for progression to cervical cancer. One alteration may be HPV-induced methylation of cellular genes and this will be discussed in the next section.

1.3.5 *In vitro* models to study the role of HPV in cervical carcinogenesis

There are no animal models available for the study of the interaction between HPV and cervical cells in the pathogenesis of cervical neoplasia. However, transfection of high-risk HPV types into either primary human keratinocytes derived from neonatal foreskins or normal human cervical keratinocytes results in immortalised and transformed cell lines, providing useful models for studying the molecular biology of HPV in normal epithelia of human origin, and the role of HPV in the initiation of cervical neoplasia (Woodworth *et al*, 1988). For example, cDNA and oligonucleotide microarrays have been used to examine the global changes in gene expression in both primary human ectocervical keratinocytes and primary human foreskin keratinocytes after infection with retroviruses expressing the HPV16 E6 and E7 genes (Nees *et al*, 2001; Duffy *et al*, 2003). Using methylation specific-multiplex ligation-dependent probe amplification, a number of candidate genes have also been shown to become methylated in primary human foreskin keratinocytes following transfection with HPV16 and HPV18 (Henken *et al*, 2007). However, an assessment of the global changes in the methylation of cellular genes that follows infection with high-risk HPV has not previously been undertaken.

W12

Another *in vitro* model that has been used to study the role of HPV during progression of early disease is the W12 cell line. W12 is a HPV16-positive cervical keratinocyte cell line isolated originally from a young woman with low-grade CIN (Stanley *et al*, 1989). In these cells the HPV16 genome is stably maintained, predominantly as viral episomes, at a high copy number of 100 copies per cell. During long-term *in vitro* cultivation, loss of episomes and emergence of integrated HPV16 occurs, accompanied by an increasing atypia and an accumulation of host cytogenetic abnormalities that are similar to those seen in cervical neoplasia *in vivo* (Stanley *et al*, 1989; Alazawi *et al*, 2002). Thus the W12 cell line is considered to be a model of progression of HPV16-associated neoplasia.

Integration of HPV16 in the W12 cell line is associated with loss of transcription of the HPV16 E2 gene and increased expression of the HPV16 E7 protein. In turn this is associated with both a selective growth advantage, compared to cells harbouring episomal forms of HPV16 only, and with the acquisition of numerical and structural chromosomal instability (Jeon *et al*, 1995; Pett *et al*, 2004). The W12 cell line has also been used to assess the changes in cervical keratinocyte gene expression associated with integration of HPV16 (Alazawi *et al*, 2002). More recently, studies using the W12 cell line have demonstrated that cervical carcinogenesis requires not only integration of the high risk HPV genomes, but also the specific loss of inhibitory episomes; in cells containing a mixture of integrated and episomal HPV16, the expression of E2 from the episomal forms was sufficient to partially inhibit transcription from the coexisting integrated HPV16, and full deregulation of E6 and

E7 gene expression was not observed until there was complete loss of E2-expressing episomes (Pett *et al*, 2006). The spontaneous loss of episomes was also found to be associated with increased expression of antiviral genes that are inducible by type I interferon (IFN), and in a subsequent study, exogenous IFN-β treatment of W12 cells resulted in a rapid reduction in the numbers of HPV16 episomes, hastening the transition from episomal to integrated HPV16 in these keratinocytes (Herdman *et al*, 2006; Pett *et al*, 2006). This latter study therefore provides strong evidence against the routine use of type I interferons in the treatment of HPV16-positive cervical lesions.

An assessment of the changes in the methylation of cellular genes during long term *in vitro* cultivation of the W12 cell line has not previously been undertaken.

1.3.6 HPV vaccination

Because HPV infection is a necessary cause of cervical neoplasia, prophylactic HPV vaccination has thus been investigated as a potential approach to reduce the burden of cervical cancer and its precursor lesions. In a proof of principle study, over 2000 young women aged 16-23, were randomised to 3 doses of a HPV16 L1 virus-like-particle vaccine or placebo (Koutsky *et al*, 2002). After a median follow-up of 17 months a decreased incidence of HPV16 infection and HPV16 related CIN lesions was observed among women who tested negative for HPV16 at study entry (Koutsky *et al*, 2002). Several randomised clinical trials have since been performed using either a bivalent (HPV types 16 and 18) or quadrivalent (HPV types 6, 11, 16 and 18) L1 virus-like particle vaccine (Harper *et al*, 2004; Villa *et al*, 2005; FUTURE II Study Group, 2007; Paavonen *et al*, 2007). In women who were not infected with HPV16 or

HPV18, these studies have all demonstrated that vaccination can reduce the incidence of HPV16 or HPV18-associated CIN lesions (~95% efficacy). Additionally vaccination may also offer cross-protection against non-vaccine oncogenic HPV types. The duration of follow-up of these initial studies was only short (2-3 years), however it is encouraging that a sustained effect has been observed after longer follow-up (5-6 years) (Harper *et al*, 2006; Villa *et al*, 2006; Paavonen *et al*, 2009). Based on early reports, the quadrivalent vaccine (Gardasil, Merck and Company) was fast-tracked for approval by the FDA in 2006. In 2009, the bivalent vaccine (Cervarix, GlaxoSmithKline Biologicals) was also approved by the FDA, following completion of the associated phase 3 randomised clinical trial (Paavonen *et al*, 2009). As these are prophylactic vaccines, cervical screening remains important for the detection of and treatment of CIN lesions in women who are already HPV positive and also for lesions caused by other oncogenic HPV types.

1.4 Smoking and cervical neoplasia

In 1950, the first large-scale epidemiological studies linking smoking with lung cancer were published (Doll & Hill, 1950; Wynder & Graham, 1950). Since that time, cigarette smoking has been causally associated with cancer at most organ sites (International Agency for Research on Cancer, 2004).

In the cervix, HPV is known to be necessary but not sufficient for the development of cervical cancer; cigarette smoking may be one potential co-factor that modulates progression to high-grade CIN and invasive disease. Because smoking is strongly correlated with various aspects of sexual behaviour, an association between smoking

and cervical neoplasia has actually been difficult to prove. However, smoking is now known to be an independent risk factor for cervical neoplasia (International Collaboration of Epidemiological Studies of Cervical Cancer, 2006b; Collins *et al*, 2010). A recent pooled analysis of 23 case-control and longitudinal studies including 13,541 women with cervical cancer and 23,017 women without cervical cancer revealed that current smokers had a significantly increased risk of cervical SCC compared to non-smokers (RR=1.60 95% CI: 1.48-1.73), and this risk increased with an increasing number of cigarettes smoked per day and with a younger age at smoking initiation (International Collaboration of Epidemiological Studies of Cervical Cancer, 2006b). This risk persisted when the analysis was restricted to women who tested positive for HPV DNA. As previously discussed, smoking was not associated with an increased risk of adenocarcinoma (International Collaboration of Epidemiological Studies of Cervical Cancer 2006a, 2006b). A longitudinal study has also demonstrated that smoking is an independent risk factor for CIN; after controlling for HPV status, current smokers were twice as likely to be diagnosed with high-grade CIN compared with non-smokers (Collins *et al*, 2010).

Cigarette smoke contains over 60 known carcinogens including the polycyclic aromatic hydrocarbons, N-nitrosamines and aromatic amines (Hecht, 2003; International Agency for Research on Cancer, 2004). Nicotine is the major addictive agent in tobacco but it is not generally considered to be carcinogenic; however with every inhalation, more than 60 carcinogens are simultaneously delivered (International Agency for Research on Cancer, 2004). The components of tobacco smoke can reach the cervical cells via the bloodstream or by diffusion through tissue.

Nicotine and its major metabolite cotinine have been detected in the cervical mucus of smokers; the mean nicotine levels were, significantly higher in the smokers compared to the passive smokers (p=0.0001), and the non-smokers (p=0.0002) (Sasson *et al*, 1985; Jones *et al*, 1991; McCann *et al*, 1992). In addition, characteristic smoking-related DNA adducts have also been detected in normal cervical tissue taken from smokers, and a highly significant trend was observed when examining specimens taken from normal cervices, compared to histologically normal regions of cervices with CIN and invasive cancer (Philips & Ni Shé, 1993; Ali *et al*, 1994). These studies support the biological plausibility of the now proven association between tobacco smoke exposure and risk of cervical cancer.

However, it has been found that smoking does not increase the risk of acquiring HPV infection, and it does not prolong the duration of HPV infection, thus other mechanisms must be involved to account for the association of smoking with cervical neoplasia (Collins *et al*, 2010). One possible mechanism may be smoking-induced methylation changes, and this will be discussed in more detail in the next section.

In this section I discuss methylation changes following exposure to 2 carcinogens, HPV and cigarette smoke, particularly in relation to cervical neoplasia.

1.5 Viruses and methylation

In 1964, the first human tumour virus - Epstein-Barr virus (EBV) - was discovered (Epstein *et al*, 1964). Since that time, it is now known that approximately 15-20% of all human cancers are associated with viral infections (Parkin, 2006; McLaughlin-Drubin & Munger, 2008). Viral infection may result in activation of the host DNA methylation machinery; there are two possible reasons for this. On the one hand, methylation of the viral genome may be a host defence mechanism to prevent uptake, integration and expression of foreign DNA, such as viruses (Doerfler *et al*, 2001). On the other hand, many oncogenic viruses cause persistent infections, and methylation of the viral genome may be the masking mechanism which allows the virus to evade detection by the immune system (Fernandez & Esteller, 2010).

In addition to methylation of the viral genome, methylation of cellular genes has also frequently been reported in association with viral infection (Li *et al*, 2005; Flanagan, 2007). The reason for this remains incompletely understood, but it has been proposed that the viral-induced activation of the host DNA methylation machinery may, as a by-product, also cause methylation of cellular genes (Li *et al*, 2005). Recent reports that viral oncogenes can activate the DNA methyltransferases and induce methylation of tumour suppressor genes provides direct and compelling evidence linking viral infection with aberrant methylation of cellular genes (Tsai *et al*, 2002; Hino *et al*, 2009; Lee *et al*, 2005; Liu *et al*, 2006; Arora *et al*, 2008). For example, the EBV

oncogene latent membrane protein 1 (LMP1) induces expression of DNMT1, 3A and 3B, resulting in hypermethylation of the *E-cadherin* (*CDH1*) promoter (Tsai *et al*, 2002); in EBV-associated gastric cancer, the EBV oncogene latent membrane 2A induces phosphorylation of STAT3 which activates DNMT1, leading to methylation of the *PTEN* promoter (Hino *et al*, 2009); in hepatocellular carcinoma, methylation of the *CDH1* promoter follows activation of DNMT1 by the hepatitis B virus oncogenic X protein (Lee *et al*, 2005; Liu *et al*, 2006) or activation of both DNMT1 and DNMT3b by the hepatitis C virus Core protein (Arora *et al*, 2008).

1.5.1 Methylation of the HPV genome

Methylation of the human papillomavirus type 1a and the cottontail rabbit papillomavirus genomes was reported over 25 years ago, providing the first evidence that the papillomaviruses could be targeted by cellular DNA methylation (Danos *et al*, 1980; Sugawara *et al*, 1983; Wettstein & Stevens 1983; Burnett & Sleeman, 1984). Despite this, methylation of the high risk HPV16 and HPV18 genomes has only recently been reported. The HPV genomes do not code for any gene of the DNA methylation machinery and must therefore be methylated by the host DNA methylation machinery (Fernandez & Esteller, 2010). The current evidence suggests that the genomes of HPV16 and HPV18 can both become methylated but in slightly different patterns and in different pathological contexts. However, the causes and the consequences of methylation of the HPV genome have yet to be determined in detail.

1.5.1.1 HPV16

The early studies reporting methylation of the HPV16 genome were contradictory. Badal *et al* (2003) initially reported methylation of the long control region (LCR) and E6 gene in 52% of normal smears, 22% of precursor lesions and only 6.1% of cervical cancers examined. Conversely, in subsequent studies, methylation of the HPV16 LCR and part of the L1 gene was found to be highest in invasive cancers and lowest in CIN lesions (Kalantari *et al*, 2004; Ding *et al*, 2009). This discrepancy may reflect the differences in technology used; methylation-sensitive enzyme restriction and bisulphite modification followed by direct sequencing, 2 techniques with limited resolution, were used in the first study, while bisulphite genomic sequencing (BGS) and sequencing of individual clones, was used in the later studies. Indeed BGS has demonstrated that even in individual samples, mixtures of differentially methylated HPV16 genomes are commonly found (Kalantari *et al*, 2004).

It has also been shown that the methylation status of the HPV16 genome varies with the life cycle; in undifferentiated W12 cells, the episomal HPV16 LCR is heavily methylated, but upon differentiation the LCR region becomes hypomethylated (Kim *et al*, 2003; Kalantari *et al*, 2008). The mechanism for such differentiation linked methylation and demethylation remains unknown. Interestingly, differentiation does not alter the methylation status of HPV16 genomes that are integrated into chromosomal DNA, suggesting that the mechanisms regulating the methylation of episomal and integrated HPV16 genomes are different (Kalantari *et al*, 2008).

Recently, it has also been demonstrated that methylation of the E2 binding sites in the HPV16 LCR can prevent E2 binding (Thain *et al*, 1996; Kim *et al*, 2003; Bhattacharjee & Sengupta, 2006). Although viral integration will frequently result in disruption of the E2 gene and the consequent loss of E2 protein, in malignant lesions, both episomal and integrated HPV16 genomes can co-exist, and in some lesions only episomal HPV16 genomes are found (Matsukara *et al*, 1989). Thus, in the setting of intact E2 (ie in presence of episomal HPV16), methylation of the E2 binding sites provides an alternative mechanism for the loss of repression of E2, which in turn may lead to sustained expression of the E6 and E7 oncogenes (Bhattacharjee & Sengupta, 2006).

1.5.1.2 HPV18

The pattern of methylation of the HPV18 genome is slightly different. In asymptomatic infections and in low grade CIN lesions the HPV18 L1 gene, the LCR and the E6 gene are generally unmethylated. In contrast, in all the cervical carcinomas investigated, heavy methylation of the L1 gene was seen while the LCR and E6 gene remained unmethylated (Badal *et al*, 2004; Turan *et al*, 2006).

Hence in carcinomas, heavy methylation of the L1 region is common to both the HPV16 and the HPV18 genomes. There are 2 possible reasons for this. Firstly, as the high-risk HPV genomes are usually integrated in cervical carcinoma, the high level of methylation observed may be part of the genome defence of the host cell against a chromosomally integrating virus (Doerfler *et al*, 2001). Secondly, following integration of HPV16 or 18 into the host genome, the L1 gene now lies upstream of

the viral LCR and is therefore unlikely to be transcribed, and a lack of transcription is also known to trigger *de novo* methylation of any DNA (Fuks, 2005). In fact, the finding of increased methylation of parts of the HPV genome in carcinomas would initially appear counterintuitive – methylation is generally a repressive modification while continued transcription of the HPV genome is essential for carcinogenesis. However, it has now been demonstrated that the transforming process is maintained by a single unmethylated HPV genome (Van Tine *et al*, 2004); consistent with this most copies of HPV16 are methylated in CaSki, while no methylation was observed in SiHa, which contains one copy of HPV16 (Badal *et al*, 2003; Kalantari *et al*, 2004). It has also been suggested that L1 may become preferentially methylated during carcinogenesis, because its product, the L1 capsid protein is not required during carcinogenesis (Kalantari *et al*, 2010).

The marked difference in the methylation of the L1 gene between asymptomatic and low grade CIN lesions compared to carcinomas has led to it being investigated as a potential biomarker predictive of progression of both HPV16 and HPV18 infections (Turan *et al*, 2007; Kalantari *et al*, 2010).

1.5.2 HPV-associated methylation of cellular genes

Following infection of cervical cells with high-risk HPV types, aberrant methylation of tumour suppressor genes may be one factor that modulates progression to invasive disease. Cross-sectional studies have revealed that methylation of TSGs varies between HPV-positive and -negative vulval cancers, HPV-positive and -negative penile cancers and HPV-positive and -negative head and neck cancers (O'Nions *et al*,

2001; Gasco *et al*, 2002; Dong *et al*, 2003; Ferreux *et al*, 2003; Marsit *et al*, 2006; Yanagawa *et al*, 2008). However these studies are not informative to cervical cancer almost all of which are HPV-positive at diagnosis. Although aberrant methylation of a number of TSG has been commonly reported in women with CIN and those with invasive disease (Table 4.2, Woodman *et al*, 2007), the contribution of TSG methylation to the initiation or progression of cervical neoplasia has not been defined.

HPV-induced upregulation of the DNA methyltransferases has been reported in a few studies, and this may contribute to the aberrant methylation of TSG. Gene expression profiling has revealed increased DNMT1 expression in organotypic raft cultures of human foreskin keratinocytes transfected with HPV18 E6 and E7 genes, and in short term primary cervical cancer cultures (HPV16- or HPV18-positive) compared with normal cervical keratinocytes cultures (Garner-Hamrick *et al*, 2004; Santin *et al*, 2005). Genome-wide microarray-based comparative genomic hybridisation, followed by fine mapping analysis of 20q, has demonstrated increased DNMT3B copy number in all four human HPV-immortalised keratinocytes cell lines and in most of the cervical cancers analysed, and this was associated with increased mRNA expression in 79% of cases (Wilting *et al*, 2006). In addition, it has also been shown that HPV16 E7 can bind directly to DNMT1, stabilising the protein and increasing its methyltransferase activity (Burgers *et al*, 2006). So far for HPV, activation of the DNA methyltransferases by a HPV oncogene which subsequently induces methylation of a tumour suppressor gene has not been demonstrated. Although the HPV16 E7 protein has recently been reported to downregulate *CDH1* gene

expression, through upregulation of DNMT1 activity, this was independent of DNA methylation (Laurson *et al*, 2008).

To determine the spectrum of epigenetic changes associated with HPV infection, I have investigated the methylation and transcriptional changes in both primary human foreskin keratinocytes following transfection with HPV18, and those which occur during long term cultivation of the W12 cell line, a disease progression model.

1.6 Smoking-associated methylation changes

Currently, there is no direct evidence that smoking induces CpG methylation. However, such DNA methylation has been observed following exposure to tobacco carcinogens in both *in vitro* and animal studies. For example, in human oesophageal squamous epithelial cells (HEEC), methylation of the *fragile histidine triad* (*FHIT*) gene promoter was observed following short-term exposure to high dose nicotine (100 μM), in association with *de novo* expression of DNMT3A (Soma *et al*, 2006). Similarly, in the human lung carcinoma cell line, A549, treatment with cigarette smoke extract for 6 days increased methylation of the promoters of 2 TSGs, *RASSF1A* and *MGMT* (Liu *et al*, 2007).

Many animal studies have also demonstrated an association between DNA methylation and exposure to tobacco carcinogens (Swafford *et al*, 1997; Belinsky *et al*, 1998; Vuillemenot *et al*, 2004; Pulling *et al*, 2004; Watson *et al*, 2004; Hutt *et al*, 2005; Bachman *et al*, 2006; Vuillemenot *et al*, 2006). In F344/N rats, a high frequency of *de novo* methylation of CDKN2A was observed in primary lung tumours induced following exposure to several carcinogens including cigarette smoke, and the tobacco specific carcinogen 4-methylnitrosamino-1-(3-pyridyl)-1-butanone (NNK) (Swafford *et al*, 1997; Belinsky *et al*, 1998). Similarly, using mice models, aberrant methylation of the retinoic acid receptor beta (*RARB*) gene has been detected in virtually all (≥85%) primary lung tumours induced by cigarette smoke or NNK, and methylation of the death-associated protein kinase (*DAPK*) gene has been reported in 43% and 52% of primary lung tumours induced by cigarette smoke or NNK respectively (Pulling *et al*, 2004; Vuillemenot *et al*, 2004). Methylated forms have also been

detected in precursor lesions suggesting that smoking-induced methylation of these genes is an early event in lung carcinogenesis (Belinsky *et al*, 1998; Pulling *et al*, 2004; Vuillemenot *et al*, 2004).

In humans, cross-sectional surveys restricted to patients with cancer have revealed that aberrant methylation of TSGs is associated with various aspects of smoking exposure (summarised in Table 1.2). However such studies cannot distinguish those epigenetic changes which are a consequence of the disease process from those which are directly attributable to smoking.

More compelling evidence of smoking-induced epigenetic changes is dependent upon demonstrating, in disease-free individuals, that the onset of smoking *precedes* the detection of such changes. To determine the temporal relationship between cigarette smoking and the detection of *CDKN2A* methylation in cervical cytological samples, I have used a unique cohort of young women who were recruited soon after they first had sexual intercourse.

Site of cancer	Gene	Smoking-related variable	Reference
Bladder	CDKN2A	current and ex-smokers compared with never smokers	Marsit 2007
Bladder	RUNX3	smokers compared with non-smokers	Wolff 2008
Cervix	CDKN2A	current smokers compared with non-smokers	Lea 2004
Gastric	MLH1	smokers compared with non-smokers, pack-years smoked	Nan 2005
Head &Neck	CDH1	pack-years smoked	Hasegawa 2002
Head & Neck	CDKN2A	younger age at initiation	Hasegawa 2002
Lung	APC, CDKN2A	ever smokers compared with never smokers	Toyooka 2003
Lung	CDKN2A	smoking history of >20 pack-years compared with smoking history <20 pack-years	Nakata 2006
Lung	CDKN2A, ESR1	smoking history	Suga 2008
Lung	CDKN2A	former smokers compared with current smokers	Jarmalaite 2003
Lung	CDKN2A	current smokers, pack-years smoked, duration of smoking, and inversely with time since quitting smoking	Kim 2001
Lung	CDKN2A	pack-years smoked, duration	Kim 2003
Lung	CDKN2A	smokers compared with non-smokers (male)	Wu 2005
Lung	CDKN2A, MGMT	ever smokers compared with never smokers	Liu 2006
Lung	FANCF	former smokers compared to current smokers, inversely with number of years smoked	Marsit 2004
Lung	FHIT	current smokers compared with never smokers, pack-years smoked	Kim 2004
Lung	HIC1	current smokers compared with never smokers	Eguchi 1997
Lung	MTHFR	current and ex-smokers compared with never smokers	Vaissiere 2009
Lung	RARB	ever smokers compared with non-smokers	Tomizawa 2004
Lung	RASSF1A	duration of smoking	Dammann 2005
Lung	RASSF1A	younger age at smoking initiation	Kim 2003
Lung	RASSF1A	younger age at smoking initiation	Marsit 2005
Lung	TSLC1	pack-years smoked and number of cigarettes per day	Kikuchi 2006
Oesophageal	FHIT	current smokers compared with non-smokers, pack-years smoked	Lee 2006
Prostate	"M-score" APC, GSTP1, MDR1	current smokers compared with never smokers, pack-years smoked	Enokida 2006

Table 1.2 TSG methylation associated with smoking exposure: Studies which report a significant positive association between the prevalence of methylated TSG and some aspect of smoking exposure in patients with invasive disease. (Note: There are also many studies which have found no significant association between the prevalence of methylated TSG and smoking exposure in patients with invasive disease (data not shown)).

Chapter 2

THESIS AIM AND OBJECTIVES

2.1 Aim

To assess the contribution of cigarette smoking and human papillomavirus infection to the epigenetic modulation of cellular genes in cervical epithelium.

2.2 Objectives

- To investigate the association between smoking status and *CDKN2A* methylation in cervical smears from disease-free women.

  o To determine the prevalence and incidence of *CDKN2A* methylation in smokers and never smokers.

  o To determine if *CDKN2A* methylation becomes undetectable following smoking cessation.

- To investigate the spectrum of methylation changes associated with human papillomavirus (HPV) infection.

  o To describe and validate the changes in DNA methylation which follow the transfection of primary human foreskin keratinocytes (isolated from a single donor) with episomal HPV18, and which follow long term cultivation of the W12 disease progression model.

o To determine how often changes in DNA methylation observed using these models are recapitulated in cervical neoplasia.

o To investigate possible determinants of virus-associated methylation changes.

o To describe and validate the transcriptional changes which follow the transfection of PHFK (isolated from a single donor) with episomal HPV18 which follow long-term cultivation of the W12 disease progression model.

o To compare these transcriptional changes with those reported in published gene expression arrays using similar *in vitro* models, and to determine how often these changes are recapitulated in cervical neoplasia.

o To examine the relationship between methylation and transcriptional changes in these models.

o To investigate possible HPV18-induced 'epigenetic switching', using a candidate gene approach.

Chapter 3

MATERIALS AND METHODS

3.1 Cell culture

All the cell cultures were incubated at 37°C in 5% CO_2.

3.1.1 Cervical cancer cell lines

The cervical cancer cell lines CaSki (HPV16 positive), SiHa (HPV16 positive), HeLa (HPV18 positive) and C33A (HPV negative) were grown in Dulbeccco's modified Eagle's medium (DMEM) (Sigma-Aldrich) supplemented with 10% foetal bovine serum (FBS), 2% L-glutamine and 1% penicillin-streptomycin. The medium was changed every 2-3 days. When the cells reached 90% confluence, they were either passaged on or harvested.

3.1.2 W12 cell line

W12 keratinocytes were grown on mitomycin C-treated J2 3T3 fibroblast feeder cells in E-medium (Table 3.1) supplemented with 5 ng/ml mouse epidermal growth factor (EGF) and 2% L-glutamine, in our laboratory. The medium was changed every 2-3 days, with fresh mitomcyin C-treated feeder cells added every 5-6 days. When the cells reached 80% confluence they were either passaged onto a fresh feeder layer or harvested. Prior to harvesting, 2 ml of ethylenediaminetetraacetic acid (EDTA) were added per dish to remove the J2 3T3 feeder cells. DNA, RNA and protein has been extracted from every passage (P) between P11-P56; I have performed some of the extractions.

Reagent	Volume		100x Cocktail:	
DMEM	300 ml		Adenine (0.18 M)	20 ml
F12 medium	160 ml		Insulin (5 mg/ml)	20 ml
Penicillin/streptomycin	10 ml		Transferrin (5mg/ml)	20 ml
100x Cocktail	5 ml		Tri-iodothyronine ($2x10^{-8}$ M)	20 ml
Hydrocortisone (0.5 μg/ml)	0.5 ml		PBS	120 ml
Cholera Toxin (0.1 nM)	0.5 ml		**Total volume**	**200ml**
FBS	25 ml			
Total volume	**500 ml**			

Table 3.1 Composition of E-medium

3.1.3 Primary human foreskin keratinocytes transfected with HPV18

Primary human foreskin keratinocytes (PHFK) transfected with HPV18 have previously been established in the laboratory of Dr Sally Roberts (Senior Lecturer, our Institute) (Wilson *et al*, 2007). In brief, PHFK have been isolated from the foreskin of a single neonatal donor (#1) and were maintained in keratinocyte-SFM growth medium (Invitrogen). Prior to transfection, a plasmid containing the wildtype HPV18 genome was digested with *Eco*RI to release the HPV18 genomes, and the genomes were re-circularised in the presence of T4 DNA ligase. At passage 3 (see Figure 3.1) PHFK from donor (#1) were co-transfected with the re-circularised wildtype HPV18 genome and a plasmid containing a neomycin resistance gene. The transfected cells were then selected in G418- and serum containing medium (E-medium) for 8 days. The selected clones were pooled (passage 4/2, Figure 3.1) and were grown on mitomycin C-treated J2 3T3 fibroblast feeder cells in E-medium.

For the purposes of experiments performed in this thesis, I have taken frozen aliquots of cells from untransfected PHFK (passage 4), PHFK transfected with HPV18 at an early (passage 4/3) and a later (passage 4/8) (Figure 3.1). The untransfected keratinocytes were grown in keratinocyte-SFM growth medium, while PHFK transfected with HPV18 were grown on mitomycin C-treated J2 3T3 fibroblast feeder

cells in E-medium supplemented with 5 ng/ml EGF and 2% L-glutamine. The medium was changed every 2 days. All the cells were cultivated for one passage only, and when the cells reached 80% confluence they were harvested, once J2 3T3 feeder cells had been removed with EDTA treatment. DNA, RNA and protein extractions were performed immediately.

3.1.4 J2 3T3 feeder cells

J2 3T3 fibroblast feeder cells were grown in DMEM supplemented with 10% bovine serum. When the cells reached 80% confluence they were either passaged on or harvested for mitomycin C treatment.

3.1.4.1 Mitomycin C treatment

The medium was first removed from each dish and the cells washed with 10 ml phosphate-buffered saline (PBS). Five ml of DMEM supplemented with 5% bovine serum were then added per dish. One hundred μl of mitomycin C (0.4 mg/ml) (Roche) were then added and the cells were incubated for 2-4 hours at 37°C; the medium was mixed at 30 minutes intervals. Following mitomycin C treatment the medium was removed and the cells were washed 3 times with 10 ml PBS. The cells were then incubated with 1 ml of trypsin-EDTA at 37°C for 3-5 minutes. Once the cells were separated from the dish, the trypsin was neutralised with 1 ml of E-medium and all the suspended cells were transferred to a universal container. Ten μl of the cell suspension were removed for cell counting (section 3.1.6). The suspended cells were centrifuged at 1000 rpm for 5 minutes to pellet the cells, and then re-suspended in 1ml of E-medium for every 1.5×10^6 cells. 1 ml of re-suspended cells (1.5×10^6 cells) was

then added drop wise to each new culture dish containing 10 ml of E-medium. The mitomycin C treated J2 3T3 feeder cells were incubated at 37°C for at least one hour before keratinocytes were added.

3.1.5 Cultivation of cell lines

Once the cells reached 80-90% confluence, the medium was removed and the adherent cells were washed with 10 ml of PBS. Two-three ml of trypsin-EDTA were added and the cells were incubated at 37°C for 3-5 minutes. Once the cells were separated from the dish, the trypsin was rapidly neutralised using an equal volume of serum-containing medium and all the suspended cells were transferred to a universal container. The suspended cells were centrifuged at 1500 rpm for 5 minutes to pellet the cells, and then re-suspended in fresh medium. Cells were then plated out onto new culture dishes in a total volume of 10ml, for further cultivation.

3.1.6 Cell counting

Ten μl of cell suspension were introduced into a counting chamber containing 10 large squares. Light microscopy was used to count the number of cells in 5 large squares, which was then divided by 5 to give the average number of cells x 10^4 per ml. The average number of cells per ml was then multiplied by the total volume the cells were suspended in to provide the total number of cells present.

3.1.7 Harvesting of cell lines

Once the cells reached 80-90% confluence, the medium was removed and the adherent cells were washed with 10 ml of PBS. Two-three ml of trypsin-EDTA were

added and the cells were incubated at 37°C for 3-5 minutes. Once the cells were separated from the dish, the trypsin was rapidly neutralised using an equal volume of serum-containing medium and all the suspended cells were transferred to a universal container. Ten µl of the cell suspension were removed for cell counting (section 3.1.6). The suspended cells were centrifuged at 1500 rpm for 5 minutes to pellet the cells. Ten ml of PBS were added to wash the cells, and the cells were centrifuged at 1500 rpm for a further 5 minutes. This was repeated twice. The cells were then re-suspended in PBS and divided into 3 aliquots for DNA, RNA and protein extraction.
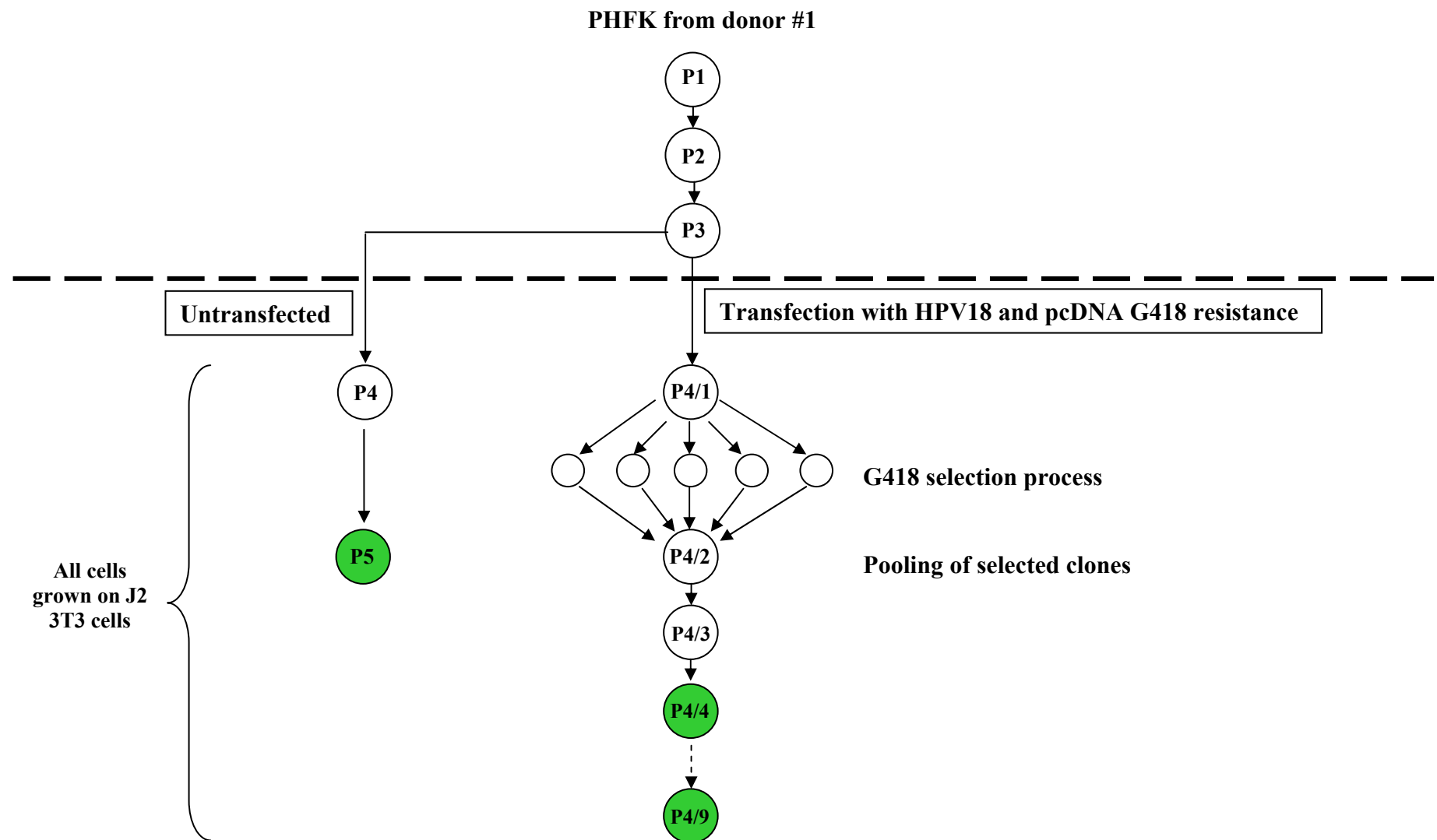
Figure 3.1 Transfection of primary human foreskin keratinocytes with HPV18. PHFK were isolated from a single neonatal donor. At passage 3, PHFK were either co-transfected with re-circularised wildtype HPV18 and a plasmid containing a neomycin resistance gene and grown on J2 3T3 fibroblast feeder cells, or maintained in culture as untransfected cells again on J2 3T3 feeder cells. Following transfection with HPV18, the cells underwent G418 selection and the selected cells were then pooled at passage 4/2. DNA, RNA and protein were extracted from the 3 passages highlighted in green.

3.2 *In vivo* model (Longitudinal cohort)

Cytological samples are available from a completed prospective cohort study that was conducted to determine the natural history of early cervical neoplasia in young women (Woodman *et al*, 2001).

3.2.1 Study population

2,011 women aged 15 to 19 years were recruited from a single family planning clinic (a Brook Advisory Centre) in Birmingham, United Kingdom, between 1988 and 1992. Women were asked to re-attend at intervals of six months: follow-up ended on 31st August 1997. The study was approved by the appropriate ethical committee.

At recruitment, a standardised interview questionnaire was used to construct a detailed social, sexual, and behavioural risk-factor profile, including smoking. Sexual and smoking histories were updated at each follow-up visit. In addition, at each visit two cytological samples were taken using the same Ayres spatula: the first was used to prepare a smear for cytological evaluation; the second was placed into 10 ml of PBS and stored at -80°C for subsequent virological examination.

All women in whom a cytological abnormality was identified were immediately referred to a dedicated research clinic for histological examination, irrespective of the severity of that abnormality. Colposcopic and cytological surveillance were maintained in these women and treatment postponed until there was histological evidence of high-grade CIN (CIN2 or CIN3), at which point women left the study.

3.2.2 Analysis of HPV DNA

After all clinical follow-up had ended, cervical samples have been tested for the presence of HPV DNA using a general primer (GP5+/GP6+) mediated polymerase chain reaction (PCR) and further PCR tests have been performed with type-specific primers on samples that were HPV positive (Woodman *et al*, 2001).

3.2.3 Smoking history

Smoking quantity was recorded as a categorical variable, with categories 0 cigarettes smoked per day, 1-9, 10-19, 20-29, 30-39, and 40 or more. Cumulative smoking exposure was measured using pack-years, estimated as the midpoint of each smoking category (with the final category set to 45), multiplied by the length of the interval during which this quantity applied, and accumulated over the lifetime of the woman.

3.2.4 Analysis of smoking and CDKN2A methylation

The study populations for all analyses performed in this thesis comprise subsets of the 1,075 women who were cytologically normal and tested negative for HPV DNA in the cervical smear taken at study entry (incident cohort). Three analyses were performed (Chapter 4):

- For the analysis comparing the prevalence of CDKN2A methylation in smokers with that in never-smokers, the study population was restricted to women who had normal cervical smears, and who tested negative for HPV DNA of any type in cervical samples, throughout follow-up.

- For the analysis comparing the risk of CDKN2A methylation in incident smokers with that in never-smokers, all women who began to smoke during follow-up were potentially available for inclusion, but only follow-up time prior to the first detection of HPV DNA of any type, or of cytological abnormality, in cervical samples, was included.

- For the analysis examining the time to disappearance of methylated forms of CDKN2A in ex-smokers, all women who quit smoking during follow-up were potentially available for inclusion, but only follow-up time prior to the first detection of HPV DNA of any type, or of cytological abnormality, in cervical samples, was included.

3.3 DNA ANALYSIS

3.3.1 DNA extraction

DNA was isolated from cell lines and cervical samples.

3.3.1.1 Lysis

Four hundred μl of freshly prepared lysis buffer (40 μl 10x PCR buffer, 2 μl Tween 20, 358 μl nuclease-free water and 13 μl of proteinase K) were added to each sample. Each sample was vortexed and incubated overnight at 55°C. Phenol-chloroform extraction was then performed.

3.3.1.2 Phenol-chloroform extraction

To remove the protein, an equal volume of phenol (400 µl) was added to each sample, vortexed and centrifuged at 13,000 rpm for 5 minutes. The supernatant was then transferred to a fresh tube, avoiding aspiration of the interlayer and the organic phase, and one volume (400 µl) each of phenol and chloroform was added. The sample was again vortexed and then centrifuged at 13,000 rpm for 5 minutes. The supernatant was then transferred to a fresh tube, again avoiding aspiration of the interlayer and the organic phase, and one volume of chloroform (400 µl) was added, to remove phenol. The sample was vortexed and centrifuged again at 13,000 rpm for 5 minutes. The supernatant was then transferred to a fresh tube, again avoiding aspiration of the interlayer and the organic phase. Two-and-a half volumes of 100% ethanol (1000 µl), 0.1 volume of 3M sodium acetate (40 µl) and 1 µl of glycogen were added, and the sample was vortexed and then precipitated at -80°C overnight.

3.3.1.3 Clean up

The precipitated sample was centrifuged at 13,000 rpm at 0°C for 30 minutes. The supernatant was discarded and 1000 µl of cold 70% ethanol were added, and the sample was centrifuged at 13,000 rpm at 4°C for 30 minutes. The supernatant was again discarded and 500 µl of 70% ethanol were added, and the sample was centrifuged at 13,000 rpm at 4°C for 10 minutes. The pellet was allowed to air-dry at room temperature for 30 minutes and then re-suspended in 50 µl of nuclease-free water. Extracted DNA samples were then stored at -20°C until use.

The NanoDrop ND-100 Spectrophotometer (Thermo Scientific) was used to measure DNA concentration, according to the manufacturer's instructions.

3.3.2 Bisulphite modification

Genomic DNA was modified by treatment with sodium bisulphite. In this reaction, all unmethylated cytosines will be converted to uracil, while methylated cytosines will be resistant to this modification (Wang *et al*, 1980). The methylated cytosines can then be detected using techniques such as methylation specific polymerase chain reaction (MSP), and bisulphite genomic sequencing (BGS).

3.3.2.1 Bisulphite modification (cell lines – traditional method)

Ten micrograms of cell line DNA in a volume of 10 μl were denatured by adding 1 μl of freshly prepared 3M sodium hydroxide, and incubating at 37°C for 15 minutes. Five hundred μl of freshly prepared 3.12M sodium bisuphite/5mM hydroquinone (pH 5.0) were added to the denatured DNA and the samples were placed in a thermocycler for 20 cycles of 99°C for 15 seconds and 50°C for 15 minutes. Bisulphite modified DNA was then purified using the Wizard DNA Clean-up System (Promega), according to the manufacturer's protocol and eluted using 50 μl of nuclease-free water. Five μl of freshly prepared 3M sodium hydroxide were added for desulphonation and the mixture was incubated at room temperature for 10 minutes. Two-and-a half volumes of 100% ethanol (125 μl), 0.1 volume of 3M sodium acetate (5 μl) and 1 μl of glycogen were added, the sample was vortexed and precipitated at -80°C for at least 1 hour.

The precipitated bisulphite modified DNA was centrifuged at 13,000 rpm at 4°C for 30 minutes. The supernatant was discarded and 1000 µl of cold 70% ethanol were added, and the sample was centrifuged at 13,000 rpm at 4°C for 5 minutes. The supernatant was discarded and the pellet was allowed to air-dry at room temperature for 30 minutes and then re-suspended in 50 µl of nuclease-free water. Bisulphite modified DNA samples were stored at -20°C until use.

3.3.2.2 Bisulphite modification using the EZ DNA Methylation-Gold Kit™

The EZ DNA Methylation Gold Kit™ (Zymo Research) allows bisulphite modification to be completed in less than 3 hours. The kit has also been designed to minimise template degradation, loss of DNA during treatment and clean-up, and to provide complete conversion of unmethylated cytosines; all of which are potential limitations of standard bisulphite modification. In addition the kit is designed to bisulphite modify much smaller quantities of DNA from 500 pg to 2 µg, with the optimal amount of DNA being 200-500 ng. This is an important consideration with our cervical samples where the total amount of DNA extracted was small

Cervical samples

Two hundred nanograms of DNA isolated from each cervical cytological sample were subjected to bisulphite modification using the EZ DNA Methylation-Gold Kit™ according to the manufacturer's protocol. The converted DNA was eluted in a final volume of 50 µl and stored at -20°C until use.

Cell Lines

Five hundred nanograms of DNA isolated from each cell line were subjected to bisulphite modification using the EZ DNA Methylation-Gold Kit^{TM} according to the manufacturer's protocol. The converted DNA was eluted in a final volume of 50 μl and stored at -20°C until use.

3.3.3 Methylation specific PCR

Methylation specific PCR (MSP) is a sensitive and specific PCR-based technique which can detect the methylation of a block of CpG sites. It takes advantage of the sequence differences following bisulphite modification, using primers designed to recognise methylated or unmethylated DNA in a bisulphite modified sample (Herman *et al*, 1996). The sensitivity of MSP has been reported to be 1:1000; 0.1% of methylated DNA can be detected in an otherwise unmethylated sample (Herman *et al*, 1996).

Five μl of bisulphite modified DNA solution (section 3.3.2.1) were amplified in a 25 μl volume using 12.5 μl of GoTaq Green Master Mix (Promega) and 0.6 μmol/L primer mix. Each sample was amplified using primers that recognise methylated DNA, and primers that recognise unmethylated DNA, in parallel. Amplification was performed using a Px2 Thermal Cycler (Thermo Scientific) and the following cycle conditions: 95ºC for 5 minutes, followed by 40 cycles of 95ºC for 45 seconds, annealing temperature for 1 minute, 72ºC for 2 minutes, and a final extension of 72ºC for 10 minutes. Table 3.2 lists the primer sequences and the corresponding annealing temperatures used. CpGenome Universal methylated and unmethylated DNA

(Millipore) were used as positive controls of the methylated and unmethylated reactions, respectively. Unmodified DNA was used as a negative control of the completeness of bisulphite modification. Water was used as a template-free control. All the PCR products were analysed using 2% agarose gel electrophoresis and ethidium bromide staining (section 3.3.5).

Name	Primer sequence 5'-3'	Product size (bp)	Annealing temperature (°C)
CDKN2A MF	TTATTAGAGGGTGGGGCGGATCGC	150	62
CDKN2A MR	GACCCCGAACCGCGACCGTAA		
CDKN2A UF	TTATTAGAGGGTGGGGTGGATTGT	151	62
CDKN2A UR	CAACCCCAAACCACAACCATAA		
DAPK1 MF	GGATAGTCGGATCGAGTTAACGTC	99	60
DAPK1 MR	CCCTCCCAAACGCCGA		
DAPK1 UF	GGAGGATAGTTGGATTGAGTTAATGTT	106	60
DAPK1 UR	CAAATCCCTCCCAAACACCAA		

Table 3.2 Primer sequences used for MSP analysis.

3.3.3.1 Designing primers for MSP

The Methprimer primer design program was used to design primers for MSP (http://www.urogene.org/methprimer/index1.html). The UCSC genome browser website (http://genome.ucsc.edu/index.html) was used to obtain the genomic sequence of the gene of interest; the upstream genomic sequence (by 1000 base pairs, bp) and the first exon were selected and copied into the Methprimer design program. The options 'pick MSP primers' and 'use CpG island prediction' were selected and the following parameters for primer design were used: product size 100-300 bp, primer annealing temperature 55-60°C, primer size between 15-30 bp. A pair of methylated and unmethylated primers was selected with annealing temperatures within 1-2°C of each other. Primers were ordered from Alta Bioscience, UK.

3.3.4 Nested MSP

Nested methylation-specific PCR (MSP) is a 2-stage PCR approach that has been reported to have a 50-fold improved sensitivity to detect methylated alleles compared to standard MSP; one methylated allele can be detected among 50,000 unmethylated alleles (Palmisano *et al*, 2000). The stage 1 primers recognise the bisulphite-modified template but do not discriminate between methylated and unmethylated alleles; the stage 2 primers are standard MSP primers. Nested MSP allows the investigation of low numbers of cells without needing to use excess PCR cycles (which can generate false positive results). Nested MSP was thus used to detect methylated forms of the *CDKN2A* gene in cervical cytological samples. Table 3.3 lists the primer sequences and the corresponding annealing temperatures used.

Name	Primer sequence 5'-3'	Product size (bp)	Annealing temperature (°C)
CDKN2A Ext F	GAAGAAAGAGGAGGGGTTGG	273bp	56
CDKN2A Ext R	CTACAAACCCTCTACCCACC		
CDKN2A MF	TTATTAGAGGGTGGGGCGGATCGC	150bp	See Table 3.2
CDKN2A MR	GACCCCGAACCGCGACCGTAA		
CDKN2A UF	TTATTAGAGGGTGGGGTGGATTGT	151bp	62
CDKN2A UR	CAACCCCAAACCACAACCATAA		
DAPK1 Ext F	GAGGAGGATAGTYGGATYGAGTTAA	141bp	56
DAPK1 Ext R	ACRAAAACACAACTAAAAAATAAATAAAAAAC		
DAPK1 MF	GGATAGTCGGATCGAGTTAACGTC	99bp	See Table 3.2
DAPK1 MR	CCCTCCCAAACGCCGA		
DAPK1 UF	GGAGGATAGTTGGATTGAGTTAATGTT	106bp	60
DAPK1 UR	CAAATCCCTCCCAAACACCAA		

Table 3.3 Primer sequences used for nested MSP analysis.

3.3.4.1 Stage 1 PCR

Five μl of bisulphite-modified DNA solution were amplified in a 25 μl volume using 12.5 μl of GoTaq Green Master Mix (Promega), and 0.6 μmol/L of primer mix. Amplification was performed using a Px2 Thermal Cycler and the following cycle conditions: 95ºC for 10 min, followed by 40 cycles of 95ºC for 30 seconds, annealing

temperature for 30 seconds, 72ºC for 30 seconds, and a final extension of 72ºC for 10 minutes.

3.3.4.2 Stage 2 PCR

The stage 2 PCR was performed on 5 μl of diluted stage 1 PCR product using primers specific for methylated and unmethylated template. Amplification was performed in a 25 μl volume, using 12.5 μl of GoTaq Green Master Mix and 0.6 μmol/L of primer mix. Amplifications were performed as in section 3.3.4.1 but all cycling times were reduced to 15 seconds. All stage 2 amplifications were performed in duplicate. CpGenome Universal methylated and unmethylated DNA (Millipore) were used as positive controls of the methylated and unmethylated reactions, respectively. Water was used as a template-free control. Stage 2 PCR products were analysed using 2% agarose gel electrophoresis and ethidium bromide staining (section 3.3.5).

3.3.4.3 Optimising nested MSP

Prior to performing nested MSP in cervical cytological samples from our *in vivo* model, nested MSP was repeated using a series of different PCR conditions in order to optimise the sensitivity and specificity of this method:

a) **Increasing the number of PCR cycles.** The sensitivity of the technique will increase with the number of PCR cycles used in the nested MSP reaction. However false positives can arise when a high number of PCR cycles are utilised due to 2 reasons. Firstly, mispriming of the methylation specific primers can lead to false positive results either from mispriming in the same region but with just a

C/T mismatch, or in a different region with some areas of similarity to the primer sequence. Secondly, if there is a low level of methylation in the "normal" counterparts of a mixed cell population, this will amplify up with a high number of PCR cycles, but the positive result is not actually arising from the "abnormal" counterpart. This is particularly a problem if a large amount of DNA is taken into the stage 2 PCR step. Thus to maintain a high sensitivity but to minimise the occurrence of false positives two further adjustments were made:

b) **Increasing the annealing temperature of the methylated primer in the stage 2 PCR.** Mispriming can occur more readily at lower annealing temperatures. By increasing the annealing temperature for the methylated primer, this will ensure the highest specificity of amplification of only methylated alleles in the sample.

c) **Diluting the stage 1 PCR product.** False positives can occur more readily if too much template is taken forward to the stage 2 PCR. The stage 1 PCR product was thus diluted 1:50 and 5 μl of this was used in the stage 2 PCR.

3.3.5 Agarose gel electrophoresis

Agarose gel electrophoresis is a technique used to separate DNA and RNA molecules based on their size. DNA and RNA are negatively charged molecules and will migrate towards the cathode when an electrical current is applied. Shorter molecules will migrate more easily through the pores in the agarose gel compared to longer molecules. Adjusting the concentration of the agarose gel enables visualisation of molecules of varying sizes, for example, increasing the concentration of the agarose

gel will reduce the migration speed and enables separation of smaller DNA molecules. Ethidium bromide is commonly added to the agarose gel; this intercalates double-stranded DNA and RNA, fluoresces under ultraviolet (UV) light and thus enables the separated nucleic acids to be visualised.

Agarose powder was dissolved in 1x Tris-Borate-EDTA (TBE) buffer (Sigma) by heating in a microwave oven on high power for 2-3 minutes. A 2% gel (for example, 2 g of agarose in 100 ml of TBE) was used to analyse DNA fragments below 1 kb in length. Ethidium bromide was added for a final concentration of 0.5 μg/ml, and the solution was poured into a gel tray and allowed to set. A suitable comb was inserted before setting. For the electrophoresis, the agarose gel was placed into an electrophoresis tank which was filled with 1x TBE buffer. Loading buffer (30% glycerol, 0.25% bromphemol blue) was mixed with every sample (1:5 dilution); a molecular weight marker (100 bp DNA ladder, Promega) was loaded into the first well and the PCR samples were loaded into the remaining wells. An electric voltage of 100V was applied and once separated the nucleic acids were visualised under UV light using a transilluminator (Bio-Rad).

3.3.6 Bisulphite genomic sequencing

Bisulphite genomic sequencing (BGS) involves amplification of bisulphite modified DNA followed by cloning of the amplified fragments then sequencing analysis of individual clones, and this provides detailed information of the methylation status of all the CpG sites within an amplified region (Frommer *et al*, 1992).

3.3.6.1 PCR amplification

Five µl of bisulphite modified DNA were amplified in a 25 µl volume using 12.5 µl of PCR Master Mix (Promega) and 0.6 µmol/L primer mix. The primers for BGS were designed to recognise the bisulphite-modified template but did not discriminate between methylated and unmethylated alleles ie there were no CpG sites located in the primer sequence. Amplification was performed using a Px2 Thermal Cycler and the following cycle conditions: 95ºC for 10 minutes, followed by 40 cycles of 95ºC for 1 minute, annealing temperature for 45 seconds, 72ºC for 2 minutes, and a final extension of 72ºC for 10 minutes. Five µl of the PCR product were analysed using 2% agarose gel electrophoresis and ethidium bromide staining (section 3.3.5). If a single band of the expected size was confirmed, the remaining PCR product (20 µl) was purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol. The purified PCR product was eluted in 50 µl of EB buffer.

Name	Primer sequence 5'-3'	Product size	Annealing temperature
SOCS3 BGSF	GTTTAAGGGATTTAGGAGATTTTTGTTT	339bp	60°C
SOCS3 BGSR	CAAATACAACCACCAACAATAACAAC		
MARK1 BGSF	AAGTGTGAGGAAGTTATTTGGAAGAG	331bp	60°C
MARK1 BGSR	TCAAAATCACCCTCAAAATTTAACA		
RARB BGSF	TTAGATTAGTTGGGTTATTTGAAGGTTAGT	451bp	60°C
RARB BGSR	AAATAATAATCTCTTTCCCAACCCC		
WT1 BGSF	AGTGGTTATGATTTTGGGGTTATT	250bp	60°C
WT1 BGSR	AAAAAATTAAATATTCCTCACTAAAAAAAA		
CDKN2A BGSF	TTGTGGTTTTATTTTTTTAATTAGGTTATT	463bp	60°C
CDKN2A BGSR	CCTTCCTCCTTTTCCAACTAAAA		
ATM BGSF	GTTTTTTATTGTGGTTTTTGTTGTG	147bp	60°C
ATM BGSR	AAAACCTTTAACCTCAAAAATCCTT		
CDH13 BGSF	AAGGGAGGTTAGAGAAATTGGATAG	331bp	60°C
CDH13 BGSR	ACTTAACAACAAAAACACAAATATAAATCA		
ROBO3 BGSF	TGTGTGTTTATGTGTGGGGTATATTA	283bp	60°C
ROBO3 BGSR	ACCTCCTAAAAATCTAAAAAAATCAAAA		

Table 3.4 Primer sequences used for bisulphite genomic sequencing

3.3.6.2 Cloning

Cloning of the PCR products was performed using the pGem T Easy Vector system II (Promega), according to the manufacturer's protocol. To analyse the transformants, 10 white colonies were selected and each one amplified in a 25 µl volume using 12.5 µl of PCR Master Mix (Promega) and 0.6 µmol/L M13 primer mix. Amplification was performed as in section 3.3.6.1 but for 30 cycles only. Five µl of the PCR product were analysed using 2% agarose gel electrophoresis and ethidium, bromide staining (section 3.3.5). If a single band of the expected size was confirmed, the remaining PCR product (20 µl) was purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol, prior to sequencing.

3.3.6.3 Sequencing

Only clones with the correct insert were selected for sequencing. The sequencing reaction was performed in a 20 µl volume and included 10 ng of purified PCR product, 3.2 pmol of M13 primer (forward or reverse), 1.5 µl BigDye® Terminator v3.1 Ready Reaction Mix (Applied Biosystems), and 3.5 µl Sequencing Buffer (Applied Biosystems). Cycle sequencing was performed using the following cycle conditions: Rapid thermal ramp (1°C per second) to 96°C, then 96°C for 1 minute, followed by 25 cycles of rapid thermal ramp to 96°C, 96°C for 10 seconds, rapid thermal ramp to 50°C, 50°C for 5 seconds, rapid thermal ramp to 60°C and 60°C for 4 minutes. The extension products were then purified using ethanol/EDTA precipitation: One µl of 0.5M EDTA and 64 µl of 95% ethanol were added to each reaction, and incubated at room temperature for 15 minutes, followed by centrifugation at 13,000 rpm for 20 minutes. The supernatant was discarded and 250

µl of 70% ethanol were added, followed by centrifugation at 13,000 rpm for 10 minutes. The supernatant was again discarded and the pellet was allowed to air-dry at 37°C for 20 minutes. Ten µl of Hi-Di™ Formamide (Applied Biosystems) were added to re-suspend the DNA. The samples were denatured at 100°C for 5 minutes, and then immediately placed on ice, prior to loading onto a 96-well plate. Automated sequencing was performed using the ABI 3100 prism™ capillary sequencer (Applied Biosystems), according to the manufacturer's protocol. Following sequencing, the extracted data files were analysed using the ABI sequencing analysis program (version 3.6.1).

3.3.7 Gel band extraction and direct sequencing

A subset of methylated products visualised using 2% agarose gel electrophoresis and ethidium bromide staining were excised from the gel using the GFX PCR DNA and Gel Band Purification Kit (Amersham), according to the manufacturer's protocol. Ten ng of the purified PCR product were then sequenced directly. The sequencing reaction was performed as described in section 3.3.6.3 but the reverse primer of the gene of interest was used instead of the M13 primer. The remainder of the purification and sequencing steps were performed as described in section 3.3.6.3.

3.3.8 Pyrosequencing methylation analysis

Pyrosequencing is a sequencing-by-synthesis method based on the luminometric detection of pyrophosphate release during nucleotide incorporation (Ronaghi *et al*, 1998). The DNA fragment of interest is firstly immobilised onto a solid support and converted into single-stranded form. A sequencing primer is hybridised onto the

single-stranded DNA and incubated with 4 enzymes, DNA polymerase, ATP sulphurylase, firefly luciferase and apyrase (a nucleotide-degrading enzyme). Repeated cycles of deoxynucleotide addition are performed. A deoxynucleotide will only be incorporated into the growing DNA strand if it's base is complimentary to the base on the template strand. The release of inorganic pyrophosphate during DNA synthesis is detected by ATP sulphurylase and luciferase in a coupled reaction; the inorganic pyrophosphate is converted to ATP by ATP sulphurylase, and the amount of ATP is then determined by the luciferase assay. The amount of light generated by the luciferase-catalysed reaction is proportional to the number of pyrophosphate molecules released, and can be detected using a suitable light-sensitive device such as a luminometer. Unincorporated deoxynucleotides and the produced ATP are degraded between each cycle by apyrase.

The technology has been applied to genotyping and haplotyping of single nucleotide polymorphisms and microbial typing. More recently it has been extensively used to assess DNA methylation. Pyrosequencing methylation analysis (PMA) offers the following advantages. It is a high-throughput, quantitative method which allows for the simultaneous examination of up to 25 CpG sites. It also provides an inbuilt internal control for the adequacy of bisulphite modification, and its reproducibility is reported to be close to 100%. (Colella *et al*, 2003; Tost *et al*, 2003).

3.3.8.1 Designing primers for pyrosequencing

The PSQ Assay Design software v1.0 (Qiagen) was used to design primers for pyrosequencing. The genomic sequence of the gene of interest was obtained from the

UCSC genome browser website (http://genome.ucsc.edu/index.html), and the sequence was then bisulphite modified *in silico*: CG dinculeotides were converted to YG, and the remaining cytosines (C) bases were then converted to thymidine (T). This modified sequence was then copied into the PSQ Assay Design software. Primers were designed using the "allele quantification" assay type setting and the following parameters: product size 50-200 bp, primer annealing temperature 55-60°C and primer size between 18-24 bp. For each gene 3 primers were generated; a biotinylated forward primer, a non-biotinylated reverse primer (or vice versa) and a sequencing primer. In order to enable the quantitative detection of the level of methylation in a given sample, pyrosequencing primers were designed to avoid CpG sites. If this was not possible in the region of interest then a maximum of 2 "Y" or "R" wobble bases were allowed in the primer sequence. Primers were ordered from Biomers.net.

3.3.8.2 Pyrosequencing PCR

For pyrosequencing analysis, all bisulphite modifications were performed using the EZ DNA Methylation Gold Kit[TM] (Zymo Research) (see section 3.3.2.2). Five μl of bisulphite modified DNA were amplified in a 50 μl volume using 25 μl of 2x Thermo-Start® High Performance PCR Master Mix (ABgene), 10 pmol non-biotinylated primer, and 5 pmol biotinylated primer. Amplifications were performed using a Px2 Thermal Cycler and the following cycle conditions: 95ºC for 15 minutes, followed by 50 cycles of 95ºC for 30 seconds, annealing temperature for 45 seconds, 72ºC for 30 seconds, and a final extension of 72ºC for 10 minutes. Ten μl of the PCR product were analysed using 2% agarose gel electrophoresis and ethidium bromide staining

(section 3.3.5). If a single band of the expected size was confirmed, the remaining PCR product (40 µl) was taken forward for pyrosequencing.

3.3.8.3 Pyrosequencing

Forty µl of PCR product were used for each sequencing reaction. The PCR products were loaded into 96 well plates, and purification using Streptavidin Sepaharose beads (Amersham) followed by co-denaturation of the biotinylated PCR product with the sequencing primer (15 pmol/reaction) were performed using the PyroMark Q96 Vacuum Prep Workstation (Qiagen), according to the manufacturer's protocol. Pyrosequencing was then performed using PyroGold SQA reagents (Qiagen) on the PyroMark™Q96 ID system (Qiagen), according to the manufacturer's protocol. Following pyrosequencing, data were analysed using the Pyro Q-CpG software (Qiagen).

3.3.9 Differential methylation hybridisation

Differential methylation hybridisation (DMH) is an array-based method that enables genome-wide assessments of changes in DNA methylation patterns (Huang *et al*, 1999). The technique uses a methylation-sensitive restriction endonuclease to identify differentially methylated DNA present in a test sample compared to the reference sample.

Genomic DNA

Cut with *Mse*I

Ligate linker primers

Divide into 2 equal aliquots

Test sample
McrBC cut

Reference sample
McrBC mock-cut

Linker-PCR

No product

Label Cy5

Label Cy3

Co-hybridise both samples to microarray
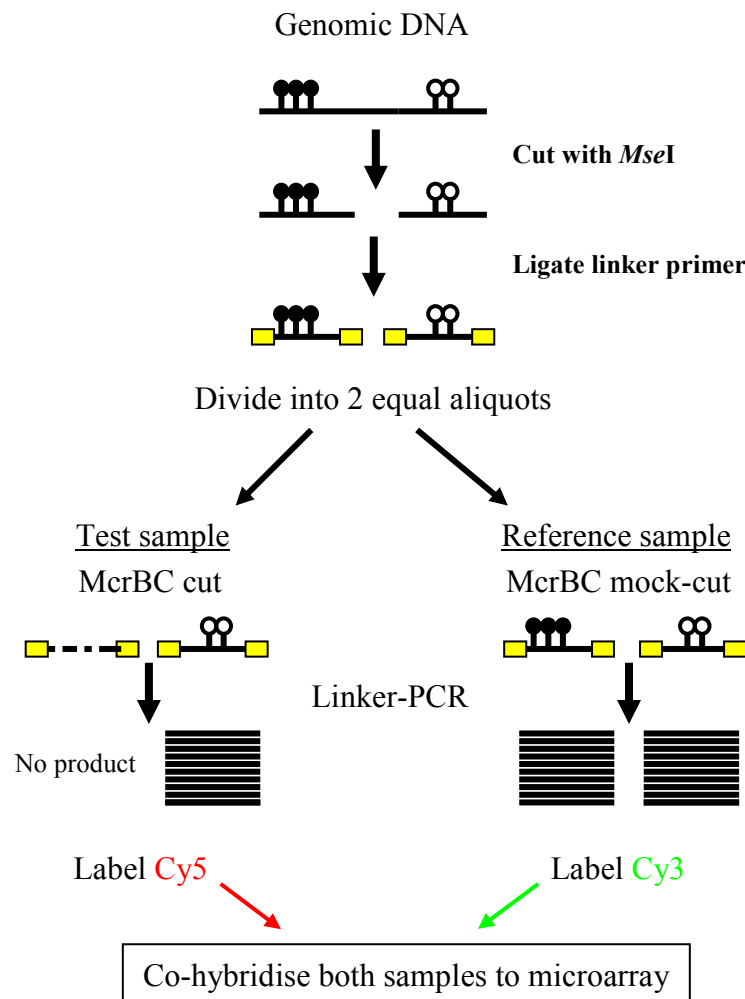
Figure 3.2 Differential methylation hybridisation. Schematic flowchart for the preparation of fluorescent labelled amplicons (see text for details, adapted from Yan *et al*, 2001).

3.3.9.1 Preparation of fluorescently labelled targets (Figure 3.2)

Fluorescently labelled targets were first prepared.

3.3.9.1.1 MseI restriction

Firstly, 2 μg of genomic DNA were digested with the restriction enzyme *Mse*I (New England Biolabs). This endonuclease restricts genomic DNA into small fragments (<200 bp), but its recognition sequence TTAA rarely occurs within GC-rich regions, leaving most CpG islands intact following digestion (Huang 1999). The digestion mix included 5 μl 10x NE Buffer 2, 5 μl *Mse*I (10 U/μl) and water to a total volume of 50 μl. The digestion was allowed to proceed overnight at 37°C, using a Px2 Thermal Cycler. The digested product was then purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol; the final elution step was repeated to produce a final volume of 80 μl. The purified digested product was then concentrated in a SpeedVac concentrator to a volume of 10 μl.

3.3.9.1.2 Ligation to linkers

The cleaved ends of DNA were then ligated to unphosphorylated linkers. The linker sequences used were: H12 5'-TAATCCCTCGGA-3' and H24 5'-AGGCAACTGTGCTATCCGAGGGAT-3'. The use of universal linkers enables the subsequent amplification of all methylated fragments in ligated DNA samples (Yan *et al*, 2002). The purified digested DNA (10 μl) was mixed with 5 μl each of H12 and H24 linkers (100 μM) and incubated at 50°C for 10 minutes. To allow the linkers to anneal, the mixture was then allowed to cool gradually to room temperature (25°C), by a 1°C drop every 2 minutes using a Px2 Thermal Cycler. The annealed linker-

primers were then ligated to the purified DNA with the addition of 1 µl of T4 DNA ligase (400 U/µl), 3 µl of T4 DNA ligase buffer and 6 µl of nuclease-free water to bring the reaction volume to 30 µl. This ligation reaction was allowed to proceed overnight at 16°C using a Px2 Thermal Cycler.

To test the efficiency of ligation a test PCR was performed. Ten µl of water were added to each ligated sample to bring the volume up to 40 µl. One µl of ligated DNA was then amplified in a 25 µl volume using 12.5 µl of GoTaq Green Master Mix (Promega), and 1 µl of 10 µM H24 primer. Amplification was performed using the following cycle conditions: 72ºC for 3 minutes, to allow the protruding ends of the ligated DNA to be filled in, followed by 25 cycles of 95ºC for 1 minute, 66ºC for 1 minute, 72ºC for 2 minutes, and a final extension of 72ºC for 10 minutes. The PCR products were analysed using 2% agarose gel electrophoresis and ethidium bromide staining. A diffuse smear between 200-900 bp indicated a successful *Mse*I digestion and linker ligation. Once the optimal smear pattern was achieved, the remaining 39 µl of ligated DNA were purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol. The purified ligated DNA was then concentrated in a SpeedVac concentrator to a volume of 24 µl.

3.3.9.1.3 Methylation-sensitive restriction

The purified ligated DNA sample was divided into 2 equal aliquots of 12 µl. One aliquot was restricted with the methylation-sensitive restriction enzyme, McrBC (New England Biolabs). This endonuclease cleaves DNA containing methylcytosine on one or both strands, but will leave unmethylated DNA intact. The recognition site for

McrBC consists of 2 half-sites of the form (G/A)mC, which can be separated by up to 3 kb. The digestion mix contained 2 μl McrBC (10 U/μl), 2 μl 10x NE Buffer 2, 2 μl 10X BSA, 2 μl 10X GTP (10 mM) in a total volume of 20 μl. The second aliquot was mock-digested, with nuclease-free water substituting McrBC in the digestion mix. The digestions were allowed to proceed overnight at 37°C, using a Px2 Thermal Cycler.

3.3.9.1.4 Linker PCR

The McrBC digested and the mock-digested samples were each divided into 4 aliquots (5 μl each) for PCR amplification. Five μl of digested (or mock-digested) DNA were amplified in a 100 μl volume using 50 μl of GoTaq Green Master Mix (Promega), and 2.5 μl of 10 μM H24 primer. This was performed in quadruplicate for all 4 aliquots per digested and mock-digested sample. Amplification was performed using the following cycle conditions: 72ºC for 3 minutes, to allow the protruding ends of the ligated DNA to be filled in, followed by 20 cycles of 95ºC for 1 minute, 66ºC for 1 minute, 72ºC for 2 minutes, and a final extension of 72ºC for 10 minutes. PCR amplification was conducted for only 20 cycles to ensure that the PCR was still in the linear range. Five μl of the PCR products were analysed using 2% agarose gel electrophoresis and ethidium bromide staining to check the diffuse smear between 200-900 bp as above was still present. The amplified samples were then purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol, with the exception that only 11 μl of nuclease-free water were used to elute the DNA in the final step. The four replicates per sample were combined for a total volume of 44 μl each for both the digested and the mock-digested samples. The

NanoDrop ND-100 Spectrophotometer was then used to measure DNA concentration, according to manufacturer's instructions.

3.3.9.1.5 Fluorescent dye labelling

The McrBC-digested and mock-digested amplified samples were labelled with the fluorescent dyes Cy5 and Cy3 respectively, using the Mirus Label IT Kit (Mirus Bio). The labelling mix included 5 μg of the amplified samples, 5 μl of 10x labelling buffer A, and 1 μl of the fluorescent dye (Cy5 or Cy3) in a total volume of 50 μl. Both samples were incubated at 37°C for 3 hours, and the labelled samples were then purified together using the Cyscribe GFX purification kit (Amersham).

3.3.9.2 Microarray hybridisation

A hybridisation solution mix was prepared comprising 80μl DIG Easy Hyb solution (Roche), 4 μl calf thymus DNA (10 mg/ml) (Sigma-Aldrich) and 4 μl yeast total RNA (10 mg/ml) (Invitrogen). To denature the calf thymus DNA and the yeast total RNA, the hybridisation solution mix was incubated at 65°C for 3 minutes and allowed to cool to room temperature. The combined purified Cy3 and Cy5 labelled samples were then added to the hybridisation solution mix, and mixed gently by pipetting.

The hybridisation solution mix containing the combined labelled samples was hybridised to a University Health Network Human CpG Island 12K (HCGI12K) array. This array contains 12,192 CpG Island clones printed onto Corning UltraGAPS™ slides; the clones were originally derived from a CpG island library established at the Wellcome Trust Sanger Institute, Cambridge in the 1990s (Cross *et*

al, 1994). The CpG Island clones are single-spotted in 48 grids arranged in a 12 x 4 format (Figure 3.3); each grid comprises 16 columns and 17 rows of spots. PCR products of Arabidopsis, SA1 and SA2 (SpotReport™ Alien™ cDNA Array Validation System, Stratagene) have also been spotted onto the array as controls.
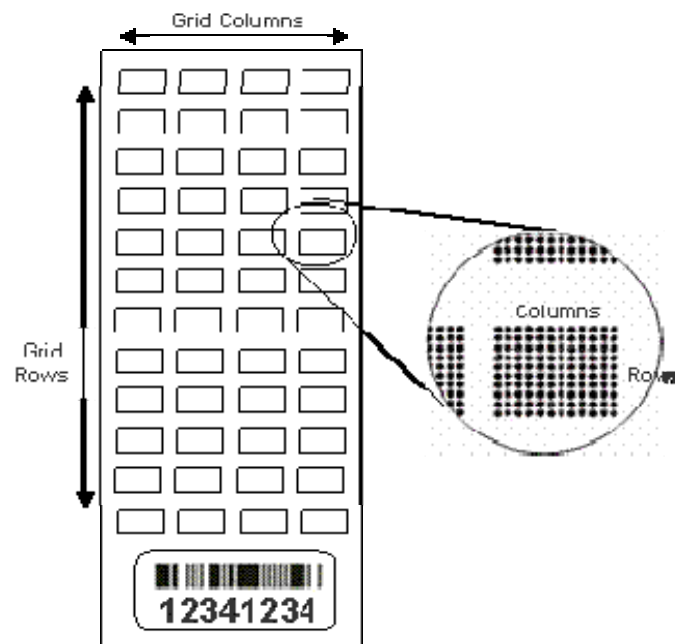
Figure 3.3 Layout of the UHN HCGI12K array.

Hybridisation was performed at 42°C for 16 hours, using the Advalytix SlideBooster hybridisation station (Advalytix), according to the manufacturer's protocol. The SlideBooster uses acoustic wave transducers below the microarray to continuously agitate the solutions on the slide surface, which greatly improves hybridisation and reproducibility of the results. The post-hybridisation wash steps were performed using the Advawash wash station (Advalytix), which provides an automated wash facility for consistent slide washing. The microarray slides were first washed in 2x SSC/0.1%

SDS at 42°C for 5 minutes, followed by 0.1x SSC/0.1% SDS at room temperature for 5 minutes, and finally in 0.1x SCC at room temperature for 5 minutes. The slides were then dried by centrifuging at 1800 rpm for 5 minutes.

3.3.9.3 Data acquisition

After washing and drying, the slides were scanned immediately using a GenePix 4000A scanner (Axon) according to the manufacturer's protocol, and the acquired images were analysed using the GenePix Pro 3.0 software (Axon). The GenePix 4000A scanner contains a dual laser scanner which has been optimised for use with Cy3 and Cy5, and acquires data at these 2 wavelengths (532nm (green) and 635nm (red)) simultaneously. Each spot was defined by the positioning of a grid of circles over the array image. The median pixel intensity within each circle and of the local background (see below) were determined; the net signal for each spot was calculated by subtracting the local background from the corresponding median intensity of that spot. The data were then exported into an Excel spreadsheet for subsequent analyses.

3.3.9.3.1 Computing local background intensities

The local background is calculated using a circular region that is centred on each individual spot ("feature") (Figure 3.4). This region is 3 times the diameter of the corresponding spot. All the pixels in this circular region are used to calculate the median local background pixel intensity unless the pixel is within the spot of interest, or is not completely outside a 2-pixel wide region around each array spot, or unless it lies within a neighbouring array spot.
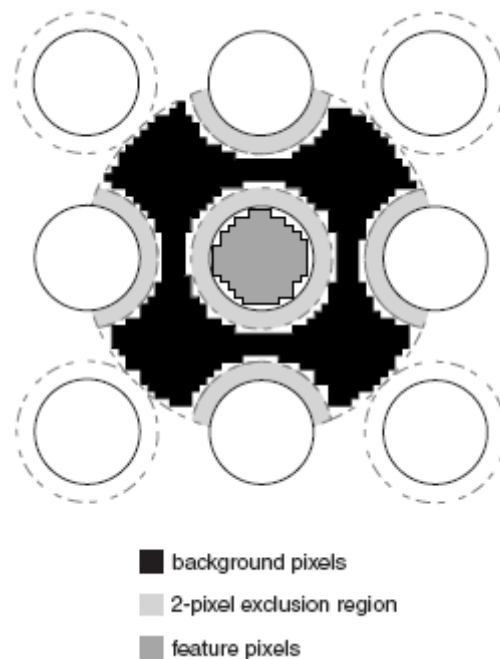
Figure 3.4 Defining background pixels. The black region represents the pixels used for computing background, the dark grey region represents the pixels used to calculate the median intensity of the array spot, and the light grey region represents excluded pixels (from GenePix Pro 3.0 User Manual).

3.3.9.4 Data processing

Prior to data analysis, array spots of low quality were first filtered: Array spots are marked "Not Found" (flag value -50) if the array spot contains fewer than 6 pixels, or if the diameter of the array spot is not within 80-120% of the theoretical spot size, or if the array spot is found in a location than would overlap with an adjacent spot. In addition, for the mock-digested sample, array spots with fluorescent signals close to the background signal ($\leq$25), reflecting PCR or printing failures were also excluded from subsequent data analysis.

3.3.9.5 Data normalisation

Because the Cy3 and Cy5 fluorescent labels are known to have different labelling efficiencies, the array data were normalised using a set of mitochondrial clones. Mitochondrial DNA is known to be completely unmethylated so the cut/mock-cut ratio of the signal intensities for the mitochondrial clones should equal one. There are 120 mitochondrial clones spotted on the UHN HCGI12K array. The median Cy3 and Cy5 signal intensities of all the mitochondrial clones on each array were first calculated. To enable comparison of data between different arrays, the Cy3 and Cy5 mitochondrial signal intensities were normalised to a value of 1000 on each array, ie the Cy3 signal intensity of each mitochondrial clone was multiplied by the median mitochondrial Cy3 signal intensity on that array, and then divided by 1000. This was performed for both the Cy3 and the Cy5 signal intensities. These normalisation factors were then applied to all the probes on the array.

3.3.9.6 Array annotation

Each CpG island clone on the UHN HCGI12K array may be associated with three possible genes located either within the same strand (W), upstream (U), or downstream (D) by a variable distance. To simplify the data analysis Dr Wenbin Wei (Head of Bioinformatics, our Institute) has re-annotated the UHN HCGI12K array - using the chromosome start and end position for each CpG island clone; each clone was allocated the nearest gene using the criteria, "mid-point of the CpG island clone is within 7.5kb upstream of 5' transcriptional start site of that gene through to 100bp downstream of the transcriptional end site." Following this re-annotation the UHN

HCGI12K array now covers 3379 unique genes only. This re-annotation has been used for all the data analyses described in this thesis.

3.3.9.7 Data analysis

A fluorescent ratio was calculated for all the array spots on each array with the mock-cut signal intensity (Cy3) as denominator and the cut signal intensity as numerator (Cy5). The resulting ratio reflects the degree of methylation of each CpG island clone; a ratio approaching zero indicates a methylated CpG island, and a ratio approaching 1 indicates an unmethylated CpG island.

Differentially methylated genes

A t-test was used to compare the individual cut/mock-cut ratios of all the replicates performed between 2 different time-points. A CpG island was considered to be differentially methylated if the p value was <0.05.

3.4 RNA ANALYSIS

3.4.1 RNA Extraction

Total RNA from cell lines was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. The extracted RNA was then stored at -80°C until use. The quality and concentration of the RNA were measured using the NanoDrop ND-100 Spectrophotometer, according to the manufacturer's protocol. The ratio of the absorbance at 260 and 280 nm ($A_{260/280}$) of all the RNA samples was required to be between 1.9 and 2.1.

3.4.2 Reverse transcription-PCR

Reverse transcription-PCR (RT-PCR) was used to analyse the expression of genes. One hundred ng of total RNA from each cell line were reverse transcribed to cDNA using the Quantitect kit (Qiagen), according to the manufacturer's protocol. Fifty ng of cDNA from each cell line were then amplified in a 25 μl volume using 12.5 μl of GoTaq Green Master Mix (Promega), and 12.5 pmol of each primer. Amplification was performed using a Px2 Thermal Cycler and the following cycle conditions: 95ºC for 10 minutes, followed by 40 cycles of 95ºC for 45 seconds, annealing temperature for 1 minute, 72ºC for 2 minutes, and a final extension of 72ºC for 10 minutes. PCR products were analysed using 2% agarose gel electrophoresis and ethidium bromide staining.

3.4.2.1 Designing primers for RT-PCR

The Primer 3 primer design program was used to design primers for RT-PCR (http://www.frodo.wi.mit.edu/primer3/). The pubmed website (http://pubmed.com) was used to obtain the mRNA sequence of the gene of interest, and the sequence was then copied into the Primer 3 program. The options 'pick left primer' and 'pick right primer' were selected and the following parameters for primer design were used: product size 100-300 bp, primer annealing temperature 55-60°C, primer size between 15-25 bp and a GC content between 40-80%. Primers were ordered from Alta Bioscience, UK.

3.4.3 Quantitative-RT-PCR

Quantitative-PCR (Q-PCR) measures PCR amplification as it occurs; data are collected during the exponential phase of the PCR when the quantity of PCR product is directly proportional to the amount of template present, and is therefore more precise. In comparison, standard PCR measures the amount of accumulated PCR product at the end of the PCR cycles, during the plateau phase.

Q-PCR using TaqMan chemistry requires specific PCR primers and a fluorogenic probe for the PCR reaction. The probe contains a reporter dye on the 5' end (FAM or VIC) and a quencher dye on the 3' end (TAMRA or MGB), and anneals downstream to one of the primer sites. When the probe is intact, the proximity of the quencher dye to the reporter dye reduces the amount of fluorescence emitted by the reporter dye. During PCR amplification, the probe is cleaved by the 5' nuclease activity of Taq DNA polymerase, which separates the reporter dye from the quencher dye, and increases the reporter fluorescence signal. Hence the increase in fluorescence intensity is directly proportional to the number of PCR amplicons produced.

Fifty ng of cDNA from each cell line were amplified in a 25 μl volume using 12.5 μl of TaqMan Universal PCR Master Mix (Applied Biosystems), 1.25 μl of 20x TaqMan Gene Expression Assay primer and probe mix for the target gene of interest (Applied Biosystems), and 1.25 μl of 20x TaqMan Gene Expression Assay primer and probe mix for the house-keeping gene (Applied Biosystems). Amplifications were performed in triplicate in a 96 well plate using the ABI PRISM™ 7700 Sequence Detection System (Applied Biosystems) and the following cycle conditions: 50°C for

2 minutes, followed by 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute.

Data were analysed using the ABI PRSIM^{TM} 7700 SDS Software 1.9.1 (Applied Biosystems); fold changes in gene expression between 2 samples were measured using the comparative Ct method for relative quantitation (ddCt). Here the difference between the Ct value of the target gene of interest and the house-keeping gene is designated dCt. One sample is then selected as the reference sample. The difference between the dCt values between different samples compared to the reference sample is then designated ddCt. The absolute fold change in gene expression between each sample compared to the reference sample can then be calculated using the arithmetic formula, $2^{-(ddCt)}$.

3.4.3.1 Designing primers and probes for Q-RT-PCR

Applied Biosystems have already pre-designed primers and probe sequences for many target genes. The target gene of interest and the house-keeping gene were selected and ordered from Applied Biosystems, UK.

3.4.4 Gene expression profiling

Gene expression profiling was performed using the Affymetrix GeneChip® Human Genome U133 Plus 2.0 array (Affymetrix). This array comprises 54,675 probe sets which analyses the expression level of approximately 47,400 transcripts.

3.4.4.1 Synthesis of biotin-labelled, fragmented cRNA

Prior to array hybridisation, biotin-labelled fragmented cRNA was first generated:

3.4.4.1.1 Preparation of Poly-A RNA controls

The Affymetrix GeneChip® Human Genome U133 Plus 2.0 array contains probe sets for several *Bacillus subtilis* genes which are not found in eukaryotic samples (*lys, phe, thr,* and *dap*). These pre-mixed, polyadenylated (poly-A) prokaryotic controls are processed together with the RNA sample, and undergo *in vitro* transcription (IVT) followed by amplification and labelling. Following hybridisation, the signal intensities of these controls then serve as sensitive indicators of target preparation and the labelling reaction efficiency, independent from the quality of the starting RNA sample.

The Poly-A Control Stock (Affymetrix) was serially diluted with the Poly-A Control Dilution Buffer (Affymetrix) prior to use: Firstly, 2μl of the Poly-A Control Stock was added to 38 μl of the Poly-A Control Dilution Buffer (1:20 dilution); 2μl of this first dilution was then added to a further 98 μl of the Poly-A Control Dilution Buffer (1:50 dilution); 2 μl of this second dilution was then added to 18μl of the Poly-A Control Dilution Buffer (1:10 dilution) and 2 μl of this final dilution was added to the cDNA reaction below.

3.4.4.1.2 cDNA synthesis

Total RNA from each cell line was reverse transcribed to generate first-strand complementary DNA (cDNA) followed by converson to double-stranded cDNA,

using the Affymetrix One-Cycle cDNA Synthesis Kit (Affymetrix). All incubation steps were performed in a Px2 Thermal Cycler.

(i) First-strand cDNA synthesis

Firstly, 1 μg of total RNA, 2 μl of diluted poly-A RNA controls (from 3.4.4.1.1) and 1 μl of T7-Oligo(dT) primer (100 pmol/μl), in a total volume of 12 μl, were incubated at 65°C for 10 minutes, followed by 4°C for 2 minutes. Four μl of First Strand Buffer (5x), 2 μl of DTT (0.1M) and 1 μl of dNTPs (10 mM) were then added and the reaction was incubated at 42°C for 2 minutes. One μl of Superscript II Reverse Transcriptase (200 U/μl) was then added and the whole reaction was incubated at 42°C for 1 hour.

(ii) Second-strand cDNA synthesis

130μl of the second-strand master mix (Table 3.5) were then added to each first-strand synthesis sample from (i) for a total volume of 150 μl, and the reaction incubated at 16°C for 2 hours. Two μl of T4 DNA Polymerase (5 U/μl) were added and the reaction was incubated at 16°C for a further 5 minutes. Ten μl of 0.5 M EDTA were then added, and the newly synthesised double-stranded cDNA was immediately purified using the Affymetrix Sample Cleanup Module (Affymetrix) according to the manufacturer's protocol. The purified eluted double-stranded cDNA (12 μl) was used immediately in the *in vitro* transcription (IVT) reaction.

Reagent	Volume
RNase-free Water	91 μl
5X Second- Strand Reaction Mix	30 μl
dNTPs (10 mM)	3 μl
E. coli DNA ligase (10 U/μl)	1 μl
E. coli DNA Polymerase I (10 U/μl)	4 μl
RNase H (2 U/μl)	1 μl
Total Volume	**130 μl**

Table 3.5 Second-strand master mix.

3.4.4.1.3 *In vitro* transcription

Double-stranded cDNA was *in vitro* transcribed to cRNA and labelled with biotin using the Affymetrix IVT Labelling Kit (Affymetrix). Twenty eight μl of IVT mix (Table 3.6) was added to each of the 12 μl of eluted double-stranded cDNA samples, and the reaction was incubated at 37°C overnight.

Reagent	Volume
RNase-free Water	8 μl
10X IVT Labelling Buffer	4 μl
IVT Labelling NTP Mix	12 μl
IVT Labeling Enzyme Mix	4 μl
Total Volume	**40 μl**

Table 3.6 IVT master mix.

The biotin-labelled cRNA was then purified using the Affymetrix Sample Cleanup Module (Affymetrix) according to the manufacturer's protocol. One μl of the eluted purified cRNA was used to quantify the cRNA using the NanoDrop ND-100 Spectrophotometer. The absorbance at 260 nm and 280 nm was measured and the sample was considered of sufficient quality only if the $A_{260/280}$ ratio was between 1.9 and 2.1.

3.4.4.1.4 Fragmentation of cRNA

Prior to hybridisation onto the GeneChip arrays, fragmentation of cRNA has been shown to be critical in obtaining optimal assay sensitivity. Thus, 25 μg of each purified biotin-labelled cRNA sample were incubated with 10 μl of 5x Fragmentation Buffer in a total volume of 50 μl, at 94°C for 35 minutes. The Fragmentation Buffer has been optimized to break down full length cRNA to 35-200 base fragments by metal-induced hydrolysis. The undiluted fragmented biotin-labelled cRNA was stored at -80°C until array hybridisation.

The quality of the labelled cRNA and the fragmented cRNA was assessed using 1.2% agarose gel electrophoresis and ethidium bromide staining; 500ng to 1μg of cRNA should show a smear between 100 bp to 2 kb with most fragments between 500 bp to 1kb, and 1-2μg of fragmented cRNA should show a smear between 35 to 200 bp.

3.4.4.2 Microarray hybridisation

Twenty μg of each fragmented biotin-labelled cRNA sample were hybridised to an Affymetrix GeneChip® Human Genome U133 Plus 2.0 array according to standard Affymetrix protocols at our Institute. Post-hybridisation staining and washing were performed according to the manufacturer's protocol (Affymetrix).

3.4.4.3 Data analysis

Scanned images of the microarray chips were analysed using the Affymetrix GeneChip Operating Software (GCOS) with the default settings except that the target signal was set to 100 for all arrays. Prior to data analysis, the quality reports were first

assessed. An array was considered acceptable if the following criteria were fulfilled: average background <100, percentage of Present calls for all the probe sets on the array >40%, and degradation of GAPDH as measured by the 3'/5' probe set ratio <2. Pairwise comparisons were then performed between different samples and differentially expressed genes were identified using the GCOS change algorithm.

3.4.4.4 Array annotation

One of the challenges associated with the analysis of microarray data is that the annotation files for the arrays are updated every 3 months. To ensure that different datasets are comparable it is clearly important that the datasets are compared using genes annotated with the same annotation version. Re-annotation of the larger datasets have been undertaken by Dr Wenbin Wei (Head of Bioinformatics, our Institute) and Prof Ciaran Woodman (Professor of Cancer Epidemiology, our Institute).

3.5 PROTEIN ANALYSIS

3.5.1 Protein extraction

Protein extraction from cell lines was performed using freshly prepared protein lysis solution (Table 3.7). One protease inhibitor cocktail tablet (Roche) was added to every 10 ml of protein lysis solution. Five hundred µl of protein lysis solution (with protease inhibitor) was added to each cell pellet, pipette-mixed and then incubated on ice for 45 minutes. The sample was then centrifuged at 13,000 rpm at 4°C for 30 minutes. The supernatant containing extracted protein was transferred to a new tube and stored at -80°C until use.

Reagent	Volume
1M Tris (pH 8)	2.5 ml
2M Sodium chloride	3.75 ml
0.5M EDTA (pH 8)	100 µl
1% NP40	500 µl
Nuclease-free water	43.15 ml
Total volume	**50ml**

Table 3.7 Protein lysis solution

3.5.2 Western blotting

Western blot analysis was used to assess protein expression. In this technique, sample proteins are first separated according to their molecular weight using SDS-polyacrylamide gel electrophoresis, followed by transfer of the proteins to nitrocellulose membranes, where they are then probed using antibodies specific to the target protein.

3.5.2.1 Measuring protein concentration

Following protein extraction (section 3.5.1), proteins were first quantified using the Bradford protein assay. The Bradford assay is a colorimetric assay, based on the differential colour change of the Coomassie dye following binding of various concentrations of proteins; without protein the solution is red-brown, when protein binds, the pKa of the dye shifts causing the dye to become blue.

A series of bovine serum albumin (BSA) standards was first prepared (Table 3.8).

BSA standard	Volume of water	Volume of BSA stock (1 mg/ml)	Final concentration (mg/ml)
0	100 μl	0 μl	0
100	90 μl	10 μl	0.1
200	80 μl	20 μl	0.2
300	70 μl	30 μl	0.3
400	60 μl	40 μl	0.4
500	50 μl	50 μl	0.5

Table 3.8 BSA standards used for the Bradford protein assay.

Ten μl of each standard was added to duplicate wells in a 96 well plate (flat-bottom). An aliquot of each protein sample was then diluted 1:10 and 10 μl was also added in duplicate wells to the plate. The Bradford reagent (Bio-Rad Protein Assay Solution) was diluted 1:5 and then 200 μl was carefully added to each standard and sample well. The reaction was incubated at room temperature for 5 minutes before the absorbance at 595nm was read using an ELISA plate reader. A standard curve was then plotted and the protein concentration of the samples was calculated.

3.5.2.2 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

A polyacrylamide gel was first prepared. For the DNMT1 protein (mwt=183kDa) an 8% polyacrylamide gel was prepared, comprising a lower 8% resolving gel and an

upper 5% stacking gel (Table 3.9). The polyacrylamide gel was allowed to polymerise for at least one hour prior to use.

Reagent	8% Resolving Gel	5% Stacking Gel
Water	9.3 ml	4.1 ml
30% ProtoGel*	5.3 ml	1.0 ml
1.5M Tris HCl	5.0 ml (pH 8.8)	750 μl (pH 6.8)
10% SDS	0.2 ml	60 μl
10% ammonium persulphate	0.2 ml	60 μl
TEMED	12 μl	6 μl
Total volume	**10 ml**	**5 ml**

*ProtoGel is a 37.5:1 ratio of acrylamide/methylene bisacrylamide solution (National Diagnostics)

Table 3.9 Reagents required for preparing an 8% polyacrylamide gel.

The polyacrylamide gel was placed into the Bio-Rad Mini-Protean 3 Electrophoresis Cell and the tank was filled with running buffer (0.25M Tris base, 1.92M Glycine, 1% SDS). 2x Laemmli loading buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125M Tris HCl pH 6.8) was added in a 1:1 ratio to 15 μg of protein from each sample. The samples were then heated at 100°C for 5 minutes to denature the proteins and then allowed to cool to room temperature. Ten μl of the Precision Plus Protein Kaleidoscope molecular weight marker (Bio-Rad) were loaded into the first well of the polyacrylamide gel, and the samples were then loaded into the remaining wells. An electrical voltage of 135V was applied for 90 minutes to allow the proteins to separate, stopping just before the bromophenol blue front reached the bottom of the resolving gel. The proteins were then transferred to a nitrocellulose membrane (Invitrogen) in transfer buffer (0.25M Tris base, 1.92M Glycine and 20% methanol), at 90V for 90 minutes.

3.5.2.3 Immunoblotting

Following protein transfer, the nitrocellulose membrane was incubated in 5% blocking solution (5% non-fat dry milk powder in PBS, 0.01% Tween 20) for one hour to prevent non-specific binding of the primary and secondary antibodies to the membrane. The membrane was rinsed with Tris Buffer Saline Tween 20 (TBS-Tween; 1 ml of Tween 20 in 1 l of TBS), and then incubated at 4°C overnight with the primary antibody, diluted to the appropriate concentration with the blocking buffer; specifically DNMT1 (clone 60B1220, mouse monoclonal diluted 1:250; Abcam) or beta-actin (clone-2, mouse monoclonal diluted 1:4000; Santa Cruz). The membrane was then washed in TBS-Tween for 5 minutes with agitation to remove residual primary antibody, followed by incubation with the secondary peroxidase-labelled antibody, again diluted to the appropriate concentration with the blocking buffer, at room temperature for 1 hour. The membrane was then washed with TBS-Tween for 5 minutes with agitation 5 times. Proteins were visualised using the enhanced chemiluminescence system (ECL, Amersham), according to the manufacturer's protocol.

3.5.3 Chromatin Immunoprecipitation

Chromatin proteins have important roles in DNA repair, replication and transcription. Chromatin immunoprecipitation (ChIP) is a technique used to assess the interaction of specific proteins with defined genomic regions.

3.5.3.1 Preparation of cross-linked chromatin

For chromatin immunoprecipitation, untransfected PHFK (passage 3) and PHFK transfected with HPV18 (passage 4/3 and passage 4/8) were grown on J2 3T3 feeder cells in E-medium as described in section 3.1.3. At sub-confluence, the J2 3T3 feeder layer was first removed by EDTA treatment, and the keratinocytes were then isolated by scraping the petri dish and transferred to a universal container. Three volumes of ice-cold PBS/5mM sodium butyrate were added, and the cells were centrifuged at 400 g for 5 minutes at 4°C. The supernatant was discarded and the keratinocytes were re-suspended in 500 μl of cold PBS/5mM sodium butyrate for every $5x10^5$ cells. Native protein-DNA complexes were then cross-linked by treatment with 1% formaldehyde for 8 minutes, followed by the addition of 1.25M glycine to stop the cross-linking. The cells were then centrifuged at 470 g for 5 minutes at 4°C and the supernatant discarded.

3.5.3.2 Sonication

The cells were washed twice with cold PBS/5mM sodium butyrate and then re-suspended in 130 μl of lysis buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% SDS, PMSF, 5 mM sodium butyrate) for every $5x10^5$ cells. The keratinocytes were then sonicated in 500 μl aliquots using the Biorupter Sonicator (Diagenode) set to

'HIGH' for 15 minutes, with 30 seconds on/off. Following sonication the samples were centrifuged at 10,000 g for 10 minutes at 4°C. The supernatant (which contains the chromatin) was removed and stored at -20°C until use. The NanoDrop ND-100 Spectrophotometer was used to measure the chromatin concentration.

To check the adequacy of sonication, a sample of chromatin was de-cross-linked; 2 μg of chromatin was made up to 500 μl with lysis buffer, and 50 μg/ml of proteinase K was added. This mixture was incubated on a thermo-mixer at 68°C overnight at 1300 rpm. The de-cross-linked chromatin was then extracted by phenol/chloroform and precipitated by ethanol (section 3.3.1.2). Following clean up (section 3.3.1.3), the pellet was re-suspended in 30 μl of nuclease-free water. One μg of DNA was run on a 2% agarose gel; a smear between 200-1000 bp was expected, with most fragments around 400 bp. Immunoprecipitation was only performed if sonication was adequate.

2.5.3.3 Immunoprecipitation

One hundred μl of Dynabeads Protein G (Invitrogen) were transferred into a 1.5 ml tube. Using the Dynal Magnetic Particle Concentrator magnet (Invitrogen), the beads were captured and the supernatant was removed. The beads were re-suspended and washed with 500 μl of RIPA buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.5 mM EGTA, 1% Triton, 0.1% SDS, 0.1% sodium deoxycholate, 150 mM sodium chloride), and this wash step was repeated twice. The beads were then re-suspended in 90 μl of RIPA buffer, 2.4 μg of antibody (DNMT1, H3K27me3, or control IgG - Abcam) were added to each tube and this bead-antibody mixture was incubated on a rotator at 4°C for 2 hours.

The antibody-bound beads were then captured using the magnet and the supernatant discarded. Ten µg of chromatin made up to 100 µl with RIPA buffer were added to each tube and the chromatin and antibody-bound beads were incubated on a rotator at 4°C for 2 hours. Following immunoprecipitation, the beads were again captured and washed 3 times with RIPA buffer. The beads were then re-suspended in 100 µl of TE (pH 8).

3.5.3.4 Reversal of cross-links

To remove cross-links, the chromatin-antibody-bound beads were captured, the supernatant discarded, and 150 µl of elution buffer with SDS (20 mM Tris-HCl pH 7.5, 5 mM EDTA, 5 mM sodium butyrate, 50 mM sodium chloride, 1% SDS) and 50 µg/ml of proteinase K were added. This was incubated on a thermo-mixer at 68°C overnight at 1300 rpm. The immunoprecipitated DNA was then extracted by phenol/chloroform and precipitated by ethanol (section 3.3.1.2). Following clean up (section 3.3.1.3), the pellet was re-suspended in 30 µl of nuclease-free water and stored at -20°C until use.

3.5.3.5 Detection

The immunoprecipitated DNA was then analysed by PCR or Q-PCR. For PCR analysis, 50 ng of each immunoprecipiated DNA sample were amplified in a 25 µl volume using 12.5 µl of GoTaq Green Master Mix (Promega) and 0.3 µmol/L primer mix. Amplification was performed as described in section 3.3.3 but for only 30 cycles. The specificity of the binding was assessed by performing PCR of the region of

interest, and also a downstream region. Table 3.10 lists the primer sequences and the corresponding annealing temperatures used.

Name	Primer sequence 5'-3'	Product size (bp)	Annealing temperature (°C)
RARB F	TCCTGGGAGTTGGTGATGTCAG	230	62
RARB R	AAACCCTGCTCGGATCGCTC		
RARB down F	TCAAAGTCAGGATTAGGGGACTAC	229	60
RARB down R	GTCACACTCGTTATTTCACCAAAG		

Table 3.10 Primer sequences used for ChIP-PCR analysis

For Q-PCR analysis, 30 ng of each immunoprecipitated DNA sample were amplified in a 25 µl volume using 12.5 µl SYBR^{TM} Green PCR Master Mix and 0.3 µmol/L primer mix. Amplifications were performed in duplicate in a 96 well plate using the ABI PRISM^{TM} 7700 Sequence Detection System (Applied Biosystems) and the following cycle conditions: 95°C for 15 minutes, followed by 40 cycles of 95°C for 30 seconds, annealing temperature for 1 minute, and 72°C for 2 minutes. The incorporation of SYBR green into double-stranded DNA was measured during each elongation step. The Ct value for the control IgG antibody immunoprecipitated DNA was selected as the reference sample. The absolute fold change between each antibody compared to the control IgG antibody was then calculated using the arithmetic formula, $2^{-(Ct[sample]-Ct[IgG])}$.

Chapter 4

RESULTS 1:

ASSOCIATION BETWEEN SMOKING

AND CDKN2A METHYLATION

4.1 Restatement of Objectives

To investigate the association between smoking status and *CDKN2A* methylation in cervical smears from disease-free women:

- To determine the prevalence and incidence of *CDKN2A* methylation in smokers and never smokers.

- To determine if *CDKN2A* methylation becomes undetectable following smoking cessation.

4.2 Selection of candidate genes for inclusion in the study of smoking-induced epigenetic changes in cervical material

This has to be informed by an assessment of:

a) which genes are methylated as a consequence of tobacco exposure and,

b) how often these genes are methylated in women with cervical intraepithelial neoplasia (CIN).

Table 4.1 describes the prevalence with which methylated forms of a number of TSG are found in cancer-free samples taken from current or former smokers and non-smokers. This analysis was performed by Professor Ciaran Woodman (Professor of Cancer Epidemiology, our Institute).

TSG	Non-smokers	Smokers	Significance
APC	-	42% (68; 1)	-
CDH1	0% (5; 1)	27% (22; 1)	NS
CDH13	4% (47; 2)	15% (301; 3)	NS
CDKN2A	3% (110; 5)	14% (1049; 13)	p=0.0007
CDKN2B	8% (37; 1)	68% (22; 1)	p=0.0000
DAPK1	0% (25; 2)	21% (249; 4)	p=0.0121
FHIT	0% (37; 2)	30% (182; 2)	p=0.0000
MGMT	0% (25; 2)	15% (371; 5)	p=0.0401
PAX5	-	12% (118; 1)	-
RARB	11% (47; 2)	32% (295; 3)	p=0.0032
RASSF1A	0% (47; 2)	4% (479; 5)	NS
TMS1	-	2% (85; 1)	-

Table 4.1 Smoking-associated cancers: prevalence of methylated forms of TSG in cancer-free subjects by smoking status. Pooled analysis reporting as a percentage the detection of methylated forms in cell samples taken from cancer-free smokers and non-smokers. The table includes the number of cases contributing to each estimate and the number of studies from which they are derived in brackets.

The studies from which these estimates were derived were identified using the search terms "methylation", "smokers", "smoking", "non-smokers" and "tobacco". Sixteen studies were identified: 12 studies reported the prevalence of methylated forms in bronchial epithelial cells which were obtained from induced or un-induced sputum samples, bronchial lavages or bronchial aspirates; in 3 studies samples were taken from the oral mucosa using either buccal smears or mouth and throat rinsings; and in one study cervical smears were taken (Kersting *et al*, 2000; Palmisano *et al*, 2000; Belinsky *et al*, 2002; Gilliland *et al*, 2002; Zöchbauer-Müller *et al*, 2003; Chang *et al*, 2004; Grote *et al*, 2004; Kim *et al*, 2004; Kulkarni *et al*, 2004; Lea *et al*, 2004; von Zeidler *et al*, 2004; Belinsky *et al*, 2005; de Fraipont *et al*, 2005; Russo *et al*, 2005; Cirincione *et al*, 2006; Machida *et al*, 2006).

A pooled analysis revealed that methylated forms of *CDKN2A*, *CDKN2B*, *DAPK1*, *FHIT*, *MGMT* and *RARB* are more commonly detected in smokers than in non-smokers (Table 4.1). However, there are a number of caveats:

- The number of non-smokers is small, and in only 2 of these studies was there a significant difference in the prevalence of methylated forms (Chang *et al*, 2004; Kim *et al*, 2004). The small number of healthy non-smokers reflects the difficulty in obtaining a sputum sample in the absence of lung disease, and the reluctance of healthy non-smokers to undergo more invasive procedures.

- None of these studies reported an association between the detection of methylated forms and measures of smoking intensity or duration. Some studies comment on

the absence of an association, others do not appear to have explored this possibility. However, none of the studies was designed to explore such a relationship. Furthermore, almost invariably, the study population includes older individuals who have smoked for many years. Therefore, the failure to reveal a disease-exposure relationship may merely reflect the lack of heterogeneity in the distribution of the exposure variables.

- Given that many of the "healthy" smokers included in these studies were either participating in secondary prevention programmes, or had some form of benign lung disease, it is possible that some already had a cancer which was unrecognised at the time the sample was taken. For example, in only one of these studies was the "cancer-free" status of healthy smokers confirmed by a prolonged period of follow-up during which they were regularly reviewed (Palmisano *et al*, 2000).

There has only been one study which has related the prevalence of methylated forms of a TSG in cervical material to smoking status: methylated forms of *CDKN2A* were detected in 12% (4/34) of smokers with normal smears, compared to 5% (2/38) of non-smokers with normal smears (Lea 2004). However in women with CIN, and in those with invasive cancer, the prevalence of methylated forms was significantly higher in cervical smears taken from smokers than from a non-smoker (Lea *et al*, 2004).

Table 4.2 describes the prevalence with which methylated forms of a number of TSG are found in cervical tissue or exfoliated cells taken from disease-free women, women

with CIN and those with invasive disease. This analysis was again undertaken by Professor Ciaran Woodman. These estimates were generated from 52 studies identified using the search terms "methylation", "cervix", and "cervical". A pooled analysis revealed that 11 genes (*APC, CCNA1, CDH1, CDKN2A, DAPK1, HIC1, IGSF4, RARB, ROBO1, SLIT1* and *SLIT2*) are more likely to be methylated in HGCIN compared to disease-free controls.

Thus 3 genes, *CDKN2A*, *DAPK1* and *RARB* are more commonly methylated both in cancer-free smokers compared to cancer-free non-smokers, and in HGCIN compared to disease-free controls. Two of these genes (*CDKN2A* and *DAPK1*) were selected for further study.

No.	Gene	Disease-free controls	LG CIN	HG CIN	Squamous cell carcinoma	Cervical carcinoma	Adeno-carcinoma
1	APC	18% (90; 4)	32% (37; 1)	34% (38; 1)	24% (238; 5)	32%* (88; 1)	54%* (65; 4)
2	CCNA1	0% (25; 1)	0% (13; 1)	36%* (11; 1)		93%* (30; 1)	
3	CDH1	0% (53; 4)	7% (42; 3)	22%* (60; 3)	61%* (170(3)	47%* (135; 4)	33%* (57; 3)
4	CDKN2A	3% (254; 7)	12%* (120; 6)	29%* (237; 9)	32%* (407; 8)	22%* (372; 7)	20%* (110; 7)
5	DAPK1	1% (184; 5)	6% (69; 3)	30%* (88; 3)	64%* (299; 6)	52%* (180; 3)	39%* (89; 5)
6	HIC1	2% (43; 3)	52%* (54; 2)	70%* (91; 3)	20%* (108; 2)	71%* (79; 1)	63%* (27; 2)
7	IGSF4	0% (25; 3)	0% (29; 2)	39%* (31; 2)	58%* (52; 1)	65%* (23; 1)	
8	RARB	0% (47; 4)	5% (83; 3)	15%* (61; 3)	30%* (117; 3)	40%* (121; 3)	15%* (13; 2)
9	ROBO1	0% (51; 1)	7% (62; 1)	8%* (48; 1)		46%* (119; 1)	
10	SLIT1	0% (40; 1)	0% (48; 1)	10%* (39; 1)		53%* (119; 1)	
11	SLIT2	0% (51; 1)	2% (62; 1)	25%* (48; 1)		64%* (119; 1)	
12	FANCF	0% (18; 1)		0% (37; 1)		30%* (91; 1)	
13	FHIT	0% (50; 4)	3% (76; 2)	2% (63; 2)	12%* (77; 1)	24%* (189; 4)	0% (5; 1)
14	MGMT	3% (206; 6)	4% (93; 3)	7% (74; 3)	11%* (217; 4)	14%* (109; 2)	12%* (51; 3)
15	PTEN	0% (11(1)	15% (27; 2)	0% (11(1)	58%* (62; 1)		
16	RASSF1	3% (29; 3)	3% (58; 3)	1% (73; 3)	17%* (299; 7)	5% (110; 3)	26%* (132; 6)
17	SLIT3	0% (40; 1)	4% (48; 1)	2% (42; 1)		49%* (118; 1)	
18	TERT	0% (14; 1)	0% (13; 1)	0% (31; 1)		62%* (76; 2)	0% (9; 1)
19	TIMP3	0% (8; 1)	0% (13; 1)	16% (31; 1)		11% (171; 3)	55%* (38; 2)
20	C15orf48	0% (21; 1)				36%* (22; 1)	
21	MT1G	5% (21; 1)				55%* (22; 1)	
22	POU2F3	0% (7; 1)			41%* (32; 1)		36% (14; 1)
23	SFRP1	5% (21; 1)				58%* (22; 1)	
24	SPARC	5% (21; 1)				91%* (22; 1)	
25	TFPI2	38% (21; 1)				82%* (22; 1)	
26	TNFRSF10C	0% (12; 1)				100%* (50; 1)	
27	HSPA2		0% (13; 1)	3% (31; 1)		73%* (11(1)	
28	SOCS1		0% (13; 1)	7% (31; 1)		50%* (11; 1)	
29	TWIST1		0% (23; 1)	14% (22; 1)		43%* (56; 1)	
30	SOCS2		23% (13; 1)	45% (31; 1)		64%* (11; 1)	
31	CDH13		5% (41; 2)	14% (63; 2)		46%* (89; 1)	

Table 4.2 Epigenetic changes in cervical neoplasia: prevalence of methylated forms in relation to disease severity and histological type. Pooled analysis of 52 studies reporting the detection of methylated forms in cervical material. Listed are those genes for which methylated forms in tissue or exfoliated cells taken from women with invasive disease were significantly more frequent (p<0.05) than in disease-free controls or when these were unavailable, women with LGCIN. Genes numbered 1-11 are also more likely to be methylated in HGCIN, whereas those numbered 12-19 are not. The presence of methylated forms in HGCIN and in disease-free controls tissue has yet to be established for genes numbered 20-26 and 27-31 respectively. *Estimates significantly different from baseline. The table includes the number of cases contributing to each estimate and the number of studies from which they are derived in brackets. "Cervical cancer" refers to cases of invasive disease with unspecified histological type (from Woodman *et al*, 2007).

4.3 Detection of methylated forms in cervical cell lines

Methylation specific PCR (MSP) and nested MSP of *CDKN2A* and *DAPK1* were first optimised using DNA extracted from 4 cervical cancer cell lines (CaSki, SiHa, HeLa and C33A) and normal human foreskin keratinocytes (NHFK). The MSP and nested MSP primers used are published primers which have previously been well validated in several tumour sites (Table 3.3, Materials and Methods) (Herman *et al*, 1996; Palmisano *et al*, 2000; Esteller *et al*, 2001; Belinsky *et al*, 2002).

4.3.1 Using methylation specific PCR

Using MSP, the *CDKN2A* promoter was found to be methylated in CaSki, but not in NHFK, SiHa, HeLa or C33A (Figure 4.1a). Some unmethylated forms were also present in CaSki. Whilst for the *DAPK1* promoter, methylated forms were detected in SiHa, but not in NHFK, CaSki, HeLa or C33A (Figure 4.1b). Some unmethylated forms were also detected in SiHa.
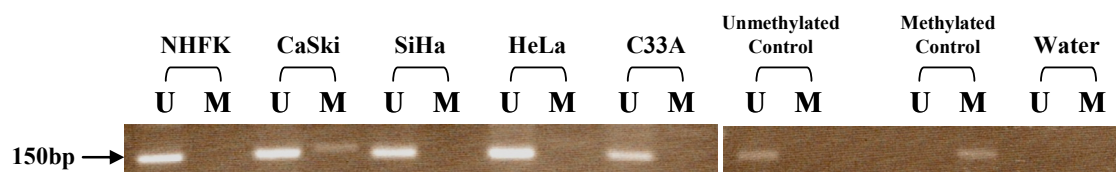
Figure 4.1a MSP for *CDKN2A* in cervical cell lines. The *CDKN2A* promoter was found to be methylated in

CaSki, but not in normal human foreskin keratinocytes (NHFK), SiHa, HeLa or C33A. Unmethylated forms of

the *CDKN2A* promoter were detected in all 5 cell lines. Expected product size: unmethylated primer (U) = 151bp,

methylated primer (M) = 150bp.
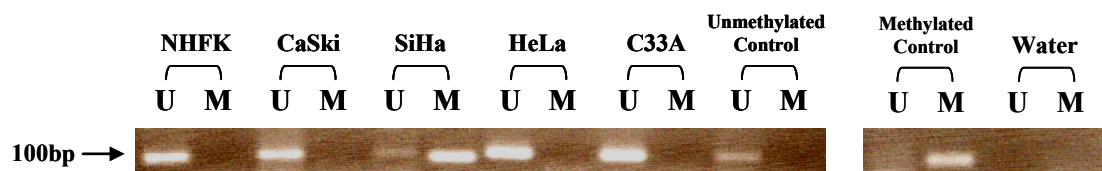
Figure 4.1b MSP for *DAPK1* in cervical cell lines. The *DAPK1* promoter was found to be methylated in SiHa,

but not in NHFK, CaSki, HeLa or C33A. Unmethylated forms of the *DAPK1* promoter were detected in all 5 cell

lines. Expected product size: unmethylated primer (U) = 106bp, methylated primer (M) = 99bp.

4.3.2 Using nested MSP

Nested MSP was performed in all 5 samples. Figure 4.2 shows that a single band of the correct size was obtained for all samples after the stage 1 PCR. Using nested MSP, methylation of the *DAPK1* promoter was detected in all 5 samples (Figure 4.3). Some methylated forms were also detected using the unmethylated control; this could be due to the increased sensitivity of the nested MSP technique or a consequence of false priming secondary to a high number of PCR cycles.

4.3.3. Using lower concentrations of DNA

To determine if lower quantities of DNA could be used for bisulphite modification, different concentrations of HeLa (1 μg, 500 ng, 100 ng, 50 ng and 10 ng) were subject to standard bisulphite modification. The bisulphite modified DNA was then eluted in 50 μl of nuclease-free water in all cases, and 5 μl were used in the MSP reaction. Figure 4.4 shows that following MSP for *DAPK1*, the unmethylated band could be detected even using 10ng of starting DNA for bisulphite modification (ie 1 ng in the MSP), although at this concentration the band was faint compared to higher concentrations of DNA. One hundred ng was thus selected as the starting concentration for bisulphite modification of cervical samples.

4.4 Detection of methylated forms in cervical smears

Next, we wanted to see if we could detect methylation of the promoters of *CDKN2A* and *DAPK1* in cervical smears. A nested MSP method was chosen to investigate methylated forms of TSGs in these cytological samples, as the number of abnormal cells present in the samples was expected to be very low.
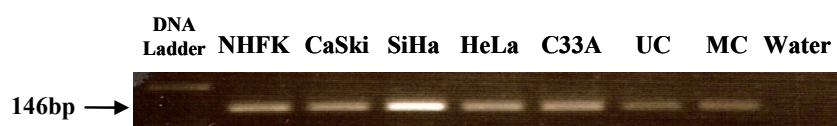
Figure 4.2 Nested MSP for *DAPK1* in cervical cell lines – stage 1 PCR. A single band of the correct size was observed following the stage 1 PCR in all the cervical cell lines and controls. UC, unmethylated control. MC, methylated control.
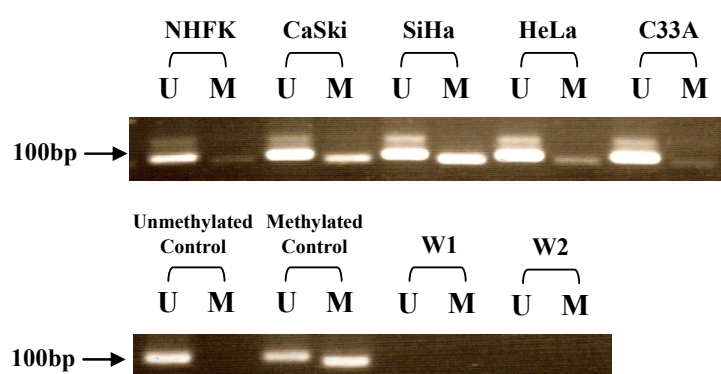
Figure 4.3 Nested MSP for *DAPK1* in cervical cell lines. The *DAPK1* promoter was found to be methylated in all 5 cervical cell lines. Expected product size: unmethylated primer (U) = 106bp, methylated primer (M) = 99bp. W1, template-free control from both stages of PCR. W2, template-free control from stage 2 PCR only.
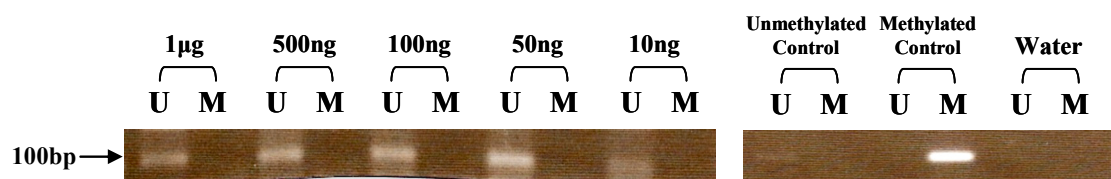
Figure 4.4 MSP for *DAPK1* in HeLa. Bisulphite modification of different starting concentrations of HeLa DNA was performed followed by MSP for *DAPK1*. An unmethylated band could be detected even using 10ng of starting DNA for bisulphite modification. Expected product size: unmethylated primer (U) = 106bp, methylated primer (M) = 99bp.

4.4.1 Detection of methylated forms in cervical smears from women with HGCIN

DNA extraction was first performed on all available cytological samples from 11 women with prevalent HGCIN ie present on first biopsy following study entry. One hundred ng of the DNA from the cervical sample corresponding to the diagnosis of HGCIN (Table 4.3), were then subject to bisulphite modification, followed by nested MSP analysis

Tube No.	Study No.
2	138.2
6	170.2
11	557.2
27	702.2
31	876.2
36	878.3
40	888.2
48	1371.2
56	1448.2
59	1466.2
66	1878.2

Table 4.3 Cervical cytological samples from 11 women who were found to have HGCIN in the cervical biopsy taken at the same visit.

Using nested MSP, methylation of the *DAPK1* promoter was found in 5/11 HGCIN samples (Figure 4.5a). An unmethylated band was detected with all 11 samples. For *CDKN2A*, 5 samples (11, 31, 48, 56 and 66) were non-evaluable, producing no band with either the methylated or the methylated primers. Of the remaining 6 HGCIN samples, methylation of the *CDKN2A* promoter was detected in 1 sample (Figure 4.5b).
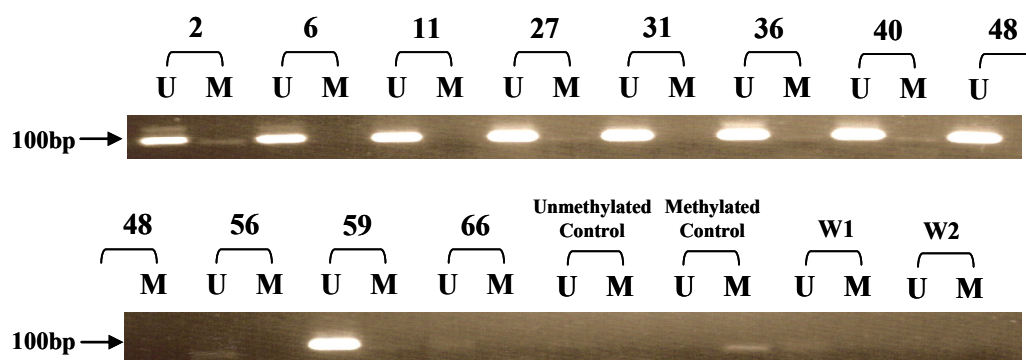
Figure 4.5a Nested MSP for *DAPK1* in 11 HGCIN cervical smear samples. The *DAPK1* promoter was found to be methylated in 5 samples with HGCIN (2, 27, 36, 40, 59). Unmethylated forms of the *DAPK1* promoter was detected in all 11 smear samples. Expected product size: unmethylated primer (U) = 106bp, methylated primer (M) = 99bp. W1, template-free control from both stages of PCR. W2, template-free control from stage 2 PCR only.
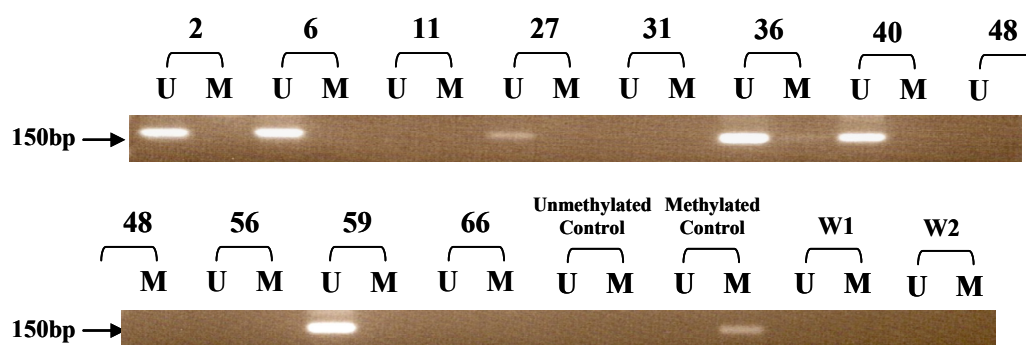
Figure 4.5b. Nested MSP for *CDKN2A* in 11 HGCIN cervical smear samples. 5 HGCIN samples were non-evaluable (11, 31, 48, 56, 66). Of the remaining 6 HGCIN samples, methylation of the *CDKN2A* promoter was found in 1 sample (36). Unmethylated forms of the *CDKN2A* promoter was detected in 6 smear samples (2, 6, 27, 36, 40 and 59). Expected product size: unmethylated primer (U) = 151bp, methylated primer (M) = 150bp. W1, template-free control from both stages of PCR. W2, template-free control from stage 2 PCR only.

4.4.2 Detection of methylated forms in cervical smears from disease-free women

Ten women with serial smear samples, who remained cytologically normal and tested negative for HPV DNA throughout follow-up were selected. To ensure the "disease-free" status of the sample, 1 cervical smear from the middle of the range of all the available cytological samples for each woman was selected for DNA extraction, and bisulphite modification.

Using nested MSP, methylation of the CDKN2A promoter was not detected in any of the 5 evaluable normal smear samples (Figure 4.6a). For DAPK1, of 9 evaluable samples, methylation was detected in 1 sample (Figure 4.6b).

4.4.3 Validation of bisulphite modification using the EZ DNA Methylation-Gold Kit^{TM}

One of the limitations of standard bisulphite modification is that it is a laborious technique, which is impractical if multiple samples require modification. Thus the EZ DNA Methylation-Gold Kit^{TM} was tested to see if this could be used for all further cervical smear samples (section 3.3.2.2, Materials and Methods).

Previously, using MSP, I have shown methylation of the *TFPI2* promoter in CaSki, SiHa and HeLa cervical cell lines, but not in C33A. Methylation was complete in SiHa but some unmethylated forms were also present in CaSki and HeLa (Figure 4.7). The results for NHFK, SiHa, HeLa and the methylated control have also been validated using bisulphite genomic sequencing (Figures 4.8a, 4.8b and 4.8c).
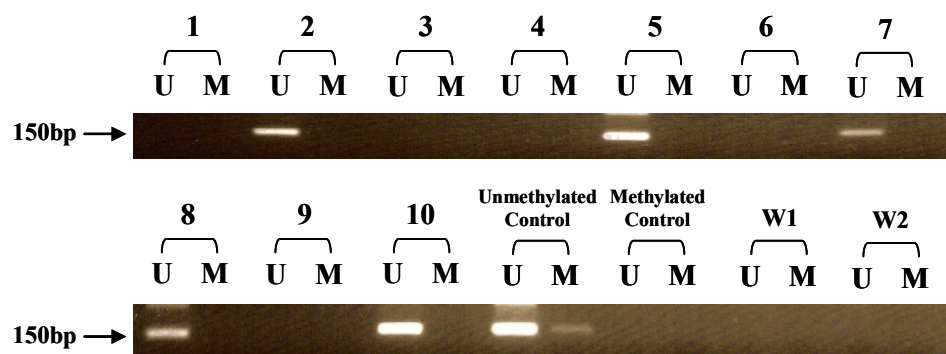
Figure 4.6a Nested MSP for *CDKN2A* in cervical smear samples from 10 disease-free women. 5 samples were non-evaluable (1, 3, 4, 6, 9). Of the remaining 5 normal smear samples, methylation of the *CDKN2A* promoter was not detected in any of the samples. Unmethylated forms of the *CDKN2A* promoter were present in 5 normal smear samples (2, 5, 7, 8, 10). Expected product size: unmethylated primer (U) = 151bp, methylated primer (M) = 150bp. W1, template-free control from both stages of PCR. W2, template-free control from stage 2 PCR only.
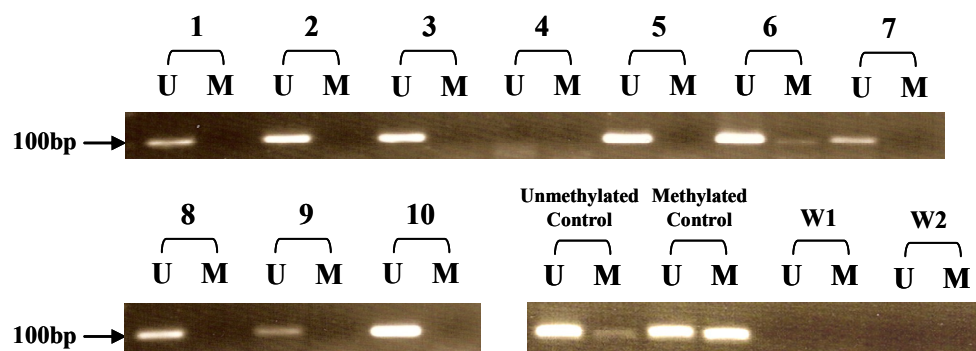
Figure 4.6b Nested MSP for *DAPK1* in cervical smear samples from 10 disease-free women. 1 sample was non-evaluable (4). Of the remaining 9 normal smear samples, methylation of the *DAPK1* promoter was found 1 sample (6). Unmethylated forms of the *DAPK1* promoter were detected in 9 smear samples (1, 2, 3, 5, 6, 7, 8, 9, 10). Expected product size: unmethylated primer (U) = 106bp, methylated primer (M) = 99bp. W1, template-free control from both stages of PCR. W2, template-free control from stage 2 PCR only.
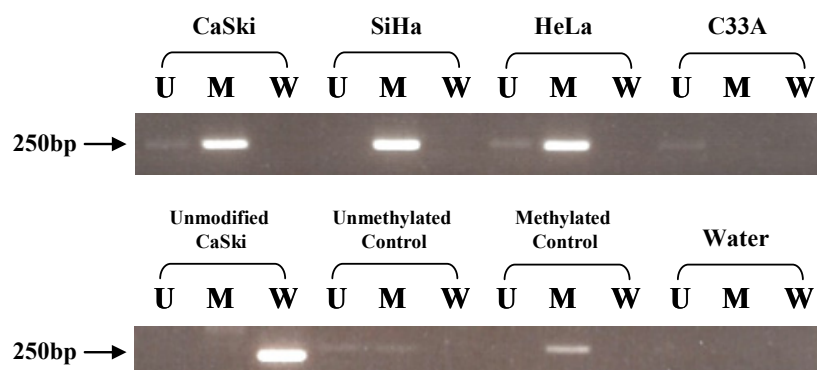
Figure 4.7 MSP for _TFPI2_ in cervical cell lines. The _TFPI2_ promoter was found to be methylated in CaSki, SiHa and HeLa, but not in C33A. Methylation was complete in SiHa, but some unmethylated forms were also present in CaSki and HeLa. The absence of a product with the wild-type primers confirms the completeness of bisulphite modification. Expected product size: unmethylated primer (U) = 261bp, methylated primer (M) = 256bp, wild-type primer (W) = 222bp.
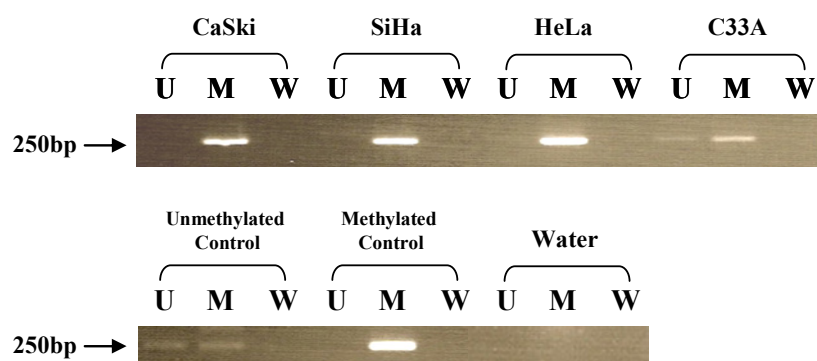
Figure 4.9 MSP for _TFPI2_ in cervical cell lines (using EZ DNA Methylation Gold kit). MSP for _TFPI2_ using EZ modified DNA produced similar results to conventional bisulphite modification. The _TFPI2_ promoter was found to be methylated in CaSki, SiHa, HeLa and C33A. Methylation was complete in CaSki and SiHa, but some unmethylated forms were also present in HeLa and C33A. The absence of a product with the wild-type primers confirms the completeness of bisulphite modification. Expected product size: unmethylated primer (U) = 261bp, methylated primer (M) = 256bp, wild-type primer (W) = 222bp.

Figure 4.8a BGS of the *TFPI2* promoter in NHFK. The methylation status of all 106 CpGs located in the *TFPI2* promoter in NHFK was assessed using bisulphite genomic sequencing (BGS). ○ = unmethylated CpG, ● = methylated CpG. The region covered by the MSP primers is indicated in red.
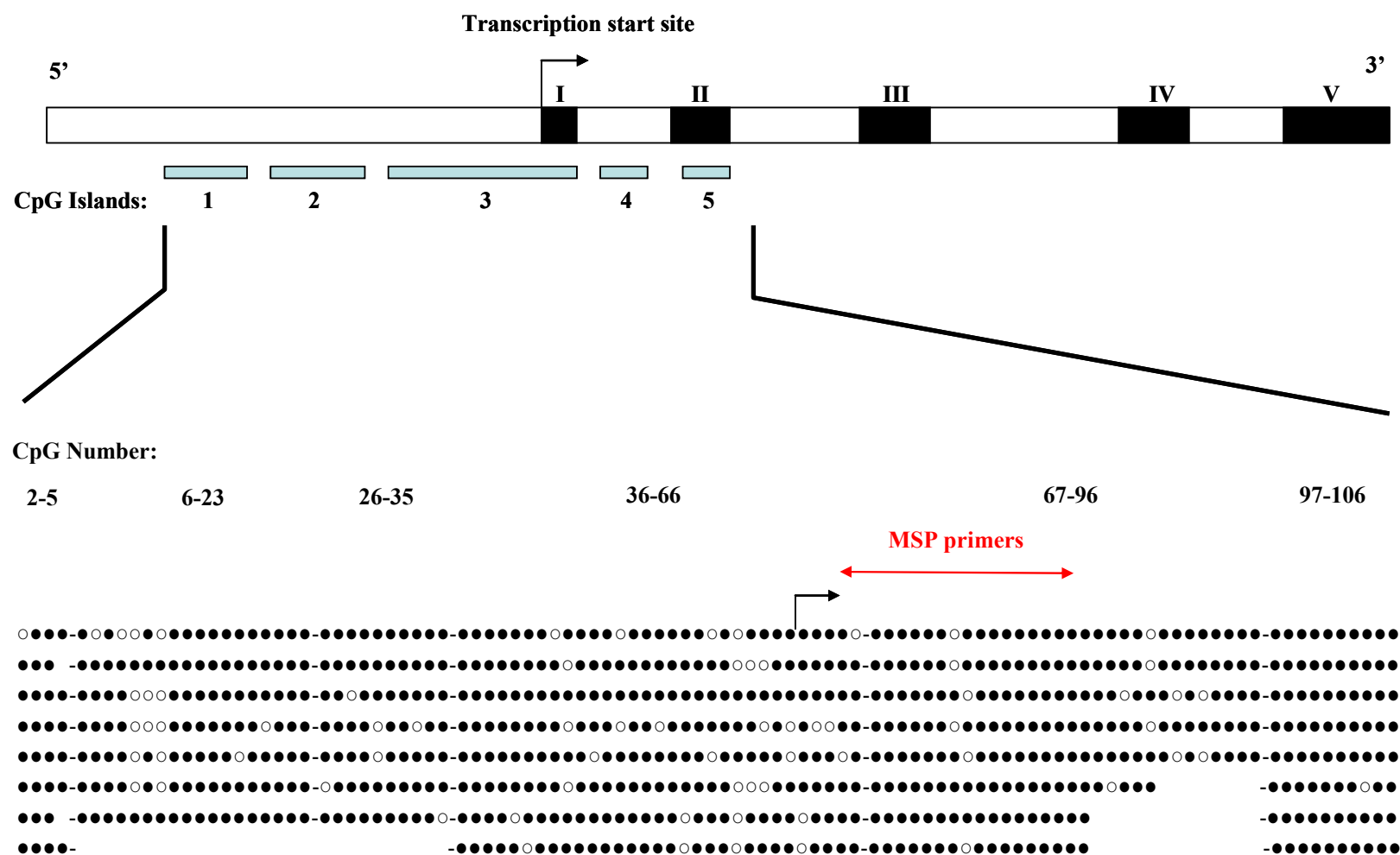
Figure 4.8b. BGS of the *TFPI2* promoter in SiHa. The methylation status of all 106 CpGs located in the TFPI2 promoter in SiHa cells was assessed using bisulphite genomic sequencing (BGS). ○ = unmethylated CpG, ● = methylated CpG. The region covered by the MSP primers is indicated in red.
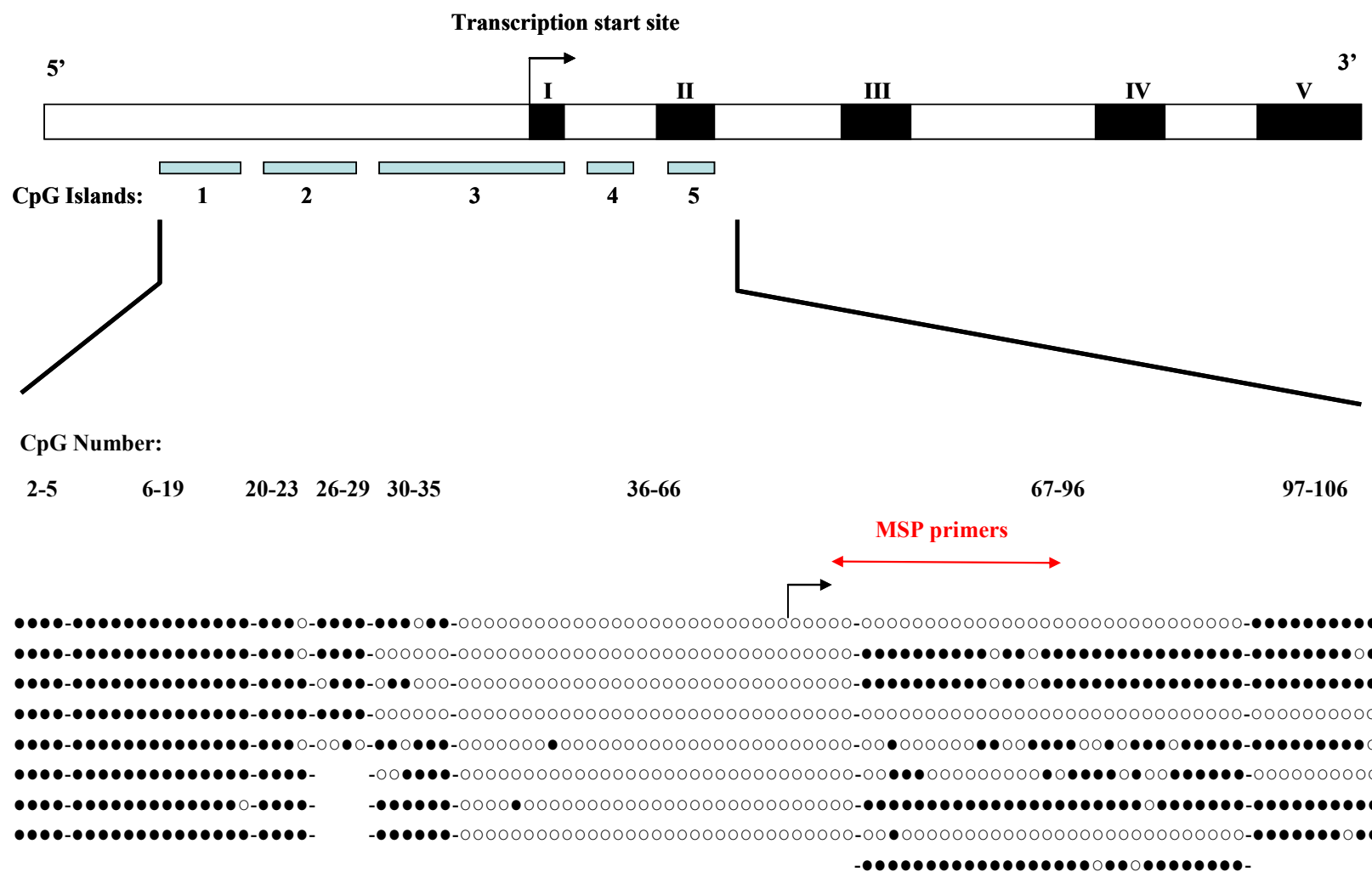
Figure 4.8c BGS of the *TFPI2* promoter in HeLa. The methylation status of all 106 CpGs located in the *TFPI2* promoter in HeLa cells was assessed using bisulphite genomic sequencing (BGS). ○ = unmethylated CpG, ● = methylated CpG. The region covered by the MSP primers is indicated in red.

To show that these results could be replicated using the EZ DNA Methylation-Gold Kit[TM] 500 ng of DNA from CaSki, SiHa, HeLa and C33A were bisulphite modified using the kit. The bisulphite modified DNA was eluted in 20 μl, and 2 μl was used in the MSP reaction. Using MSP, methylation of the *TFPI2* promoter was found in CaSki, SiHa, HeLa and C33A (the latter had previously only been unmethylated) (Figure 4.9). No product was seen with the wild-type primers confirming the completeness of bisulphite modification using the kit. As the MSP results using DNA bisulphite modified using the kit were comparable to those obtained following standard bisulphite modification, all subsequent samples were bisulphite modified using the kit.

4.4.4 Optimising nested MSP conditions in cervical smears

Nested MSP of *CDKN2A* and *DAPK1* in cervical smear samples was optimised using the 11 HGCIN and the 10 normal cervical samples; nested MSP was repeated using a series of different PCR conditions, using DNA bisulphite modified using the kit (see section 3.3.4.3 Materials and Methods):

- Increasing the number of PCR cycles. To increase the sensitivity of the technique.

- Increasing the annealing temperature (Tm) of the methylated primer in the stage 2 PCR. To ensure the highest specificity of amplification of only methylated alleles in the sample (mispriming can occur more readily at lower annealing temperatures).

PCR cycles stage 1:		20	20	30	40	40	40	40	40	40	40	40
PCR cycles stage 2:		25	30	25	30	30	30	30	30	30	30	30
M primer Tm (°C)		60	60	60	60	65	66	68	70	65	66	70
Diluting stage 1 PCR product		No	No	No	No	No	No	No	No	1:50	1:50	1:50
No. with methylation	**HGCIN (n=11)**	0 (0%)	0 (0%)	2 (18%)	9 (82%)	4 (36%)	2 (18%)	1 (9%)	0 (0%)	4 (36%)	4 (36%)	0 (0%)
	Normal (n=10)	1 (10%)	3 (30%)	5 (50%)	7 (70%)	3 (30%)	3 (30%)	0 (0%)	0 (0%)	3 (30%)	3 (30%)	0 (0%)

Table 4.4 Summary of nested MSP for *DAPK1* in HGCIN and normal smears using different PCR conditions.

PCR cycles stage 1:		40	40	40	40
PCR cycles stage 2:		30	40	40	40
Stage 2 PCR cycling times		30s	15s	15s	15s
M primer Tm (°C)		68	68	70	70
Diluting stage 1 PCR product		No	No	No	1:50
No. with methylation	**HGCIN (n=11)**	3 (27%)	3 (27%)	4 (36%)	4 (36%)
	Normal (n=10)	2 (20%)	2 (20%)	3 (30%)	1 (10%)

Table 4.5 Summary of nested MSP for *CDKN2A* in HGCIN and normal smears using different PCR conditions.

- Diluting the stage 1 PCR product prior to amplification in the stage 2 PCR. False positives can occur more readily if too much template is taken forward to the stage 2 PCR.

Tables 4.4 and 4.5 summarise the results of nested MSP for *CDKN2A* and *DAPK1* using different PCR conditions. The optimal PCR conditions for nested MSP in cervical smear samples for *CDKN2A* and *DAPK1* are shown in Table 4.6:

Conditions	*CDKN2A*	*DAPK1*
No of PCR cycles: Stage 1	40	40
No of PCR cycles: Stage 2	40	30
Annealing temp of methylated primer (°C)	70	66
Dilution of stage 1 PCR product	1:50	1:50

Table 4.6 Optimal nested MSP conditions for *CDKN2A* and *DAPK1* in cytological samples.

As no case-control difference was observed between the cytological samples taken from women with HGCIN compared to those from disease-free women for *DAPK1* (Table 4.4), only *CDKN2A* was taken forward for the study of smoking-induced methylation changes in cervical material.

4.5 Association between smoking status and *CDKN2A* methylation in cervical smears from disease-free women

To demonstrate an association between smoking and methylation in disease-free women, nested MSP for *CDKN2A* was performed using cervical smear samples from a longitudinal cohort of 2,011 women (see 3.2.1 Materials and Methods).

4.5.1 Study population

A cohort of 2,011 women aged 15 to 19 years were recruited from a single family planning clinic (a Brook Advisory Centre) between 1988 and 1992. These women were seen every 6 months when a cervical smear was taken and the smoking history was updated. The study populations for all analyses described below comprise subsets of the 1,075 women who were cytologically normal and tested negative for HPV DNA in the cervical smear taken at study entry (incident cohort) (see section 3.2.4 Materials and Methods).

4.5.2 Sensitivity of nested MSP for *CDKN2A*

Many attempts were made to determine the sensitivity of the nested MSP assay using serial dilutions of the methylated control mixed with a fixed amount of the CpGenome Universal unmethylated control (vial B). However, the reproducibility of the result was poor and we believed that this was probably because the sensitivity of the assay was functioning at the limits of detection. The minimal reproducible sensitivity was 1:1000; however at this level the methylated PCR product was always a very strong band when visualised using agarose gel electrophoresis and ethidium bromide staining, while positive methylation PCR products from cytological samples were frequently faint bands, suggesting that the sensitivity of my assay was greater then 1:1000. A pragmatic decision was therefore made that a sample would be considered positive only if a methylated band was obtained on 2 replicates.

4.5.3 Prevalence of *CDKN2A* methylation in smokers and never smokers

I first wanted to conduct a cross-sectional analysis to determine if there was a difference in the prevalence of methylation of the *CDKN2A* promoter in smokers compared to never smokers. This analysis was restricted to women who had normal cervical smears and who tested negative for HPV DNA in cervical samples throughout follow-up (disease-free women).

Within the entire cohort of 2,011 young women, 1,022 had smoked at some point during follow-up (Figure 4.10):
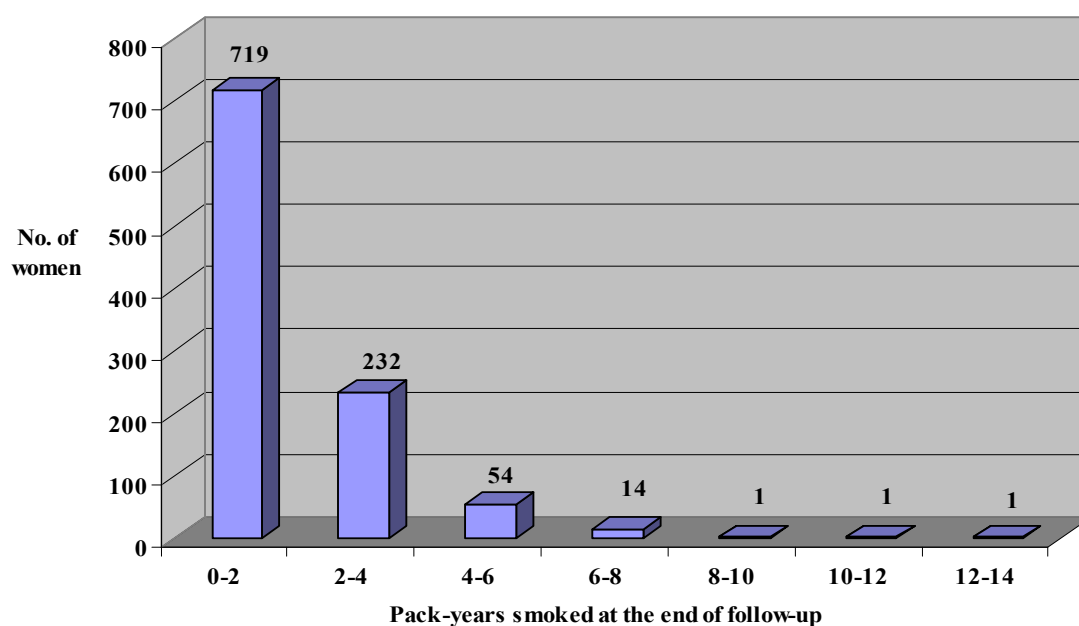
Figure 4.10 Chart showing the distribution of pack-years smoked for all smokers within the entire cohort of 2,011 young women.

Women were categorised as light, moderate or heavy smokers according to whether their cumulative number of pack-years smoked at the end of follow-up fell in the lower (0-0.76 pack-years), middle (0.76-1.83 pack-years) or upper third (1.83-12.11 pack-years) of the distribution of pack-years smoked in the entire cohort of 2011 young women.

Never smokers

25 women who had never smoked before study entry and who remained non-smokers throughout follow-up (never smokers), were randomly selected from the pool of disease-free never smokers (Table 4.7, n=534). To ensure the "disease-free" status of the sample, 1 cervical smear from the middle of the range of all the available cytological samples for each woman was selected for DNA extraction, and bisulphite modification.

Ever smokers before study entry	Smoking status at first interview	Smoking status after first interview	Gave up at some point during follow up	No. of women		Incident smokers	Ex-smokers
N	N	N	-	534			
N	N	Y	N	66		66	
N	N	Y	Y	31		31	31
Y	N	N	-	46			
Y	N	Y	N	34			
Y	N	Y	Y	11			11
Y	Y	N	-	22			22
Y	Y	Y	N	253			
Y	Y	Y	Y	78			78
			Total	**1075**		**97**	**142**

Table 4.7 Smoking status of the 1075 young women who were cytologically normal and HPV-negative at study entry.

Smokers

25 disease-free women who smoked throughout follow-up were randomly selected from the category of heavy smokers. The last cervical smear for each of the smokers was selected for DNA extraction and bisulphite modification.

Nested MSP for *CDKN2A* was performed on the selected cervical smear of these 50 women. The stage 2 PCR was performed in duplicate. A sample was considered: positive for *CDKN2A* methylation when a methylated band was seen in both stage 2 PCR replicates using the methylated primers; negative when only an unmethylated band was detected; and non-evaluable when no band was seen using either the methylated or the unmethylated primers.

Methylation of the *CDKN2A* promoter was detected in 2/21 (9%) evaluable disease-free never smokers (Figure 4.11) and in 7/21 (33%) evaluable disease-free heavy smokers (Figure 4.12) (p=0.06).

Figure 4.11 Nested MSP for *CDKN2A* in never smokers. Using nested MSP, methylation of the *CDKN2A* promoter was detected in 2 replicates in 2/21 (9%) evaluable smears taken from disease-free never smokers (NNN1 AND NNN15). Expected product size: unmethylated primer (U) = 151bp, methylated primer (M) = 150bp. UCA (CpGenome Universal unmethylated DNA (vial A)) and UCB (CpGenome Universal unmethylated DNA (vial B)) were used as positive controls for unmethylated DNA, and MC (methylated control) was used as a positive control for methylated DNA. W1, template-free control from both stages of PCR. W2, template-free control from stage 2 PCR only

.

139

Figure 4.12 Nested MSP in 25 heavy smokers. Using nested MSP, methylation of the *CDKN2A* promoter was detected in 2 replicates in 7/20 (33%) evaluable smears taken from disease-free heavy smokers (S1, S3, S6, S7, S8, S12, S15). Expected product size: unmethylated primer (U) = 151bp, methylated primer (M) = 150bp. UCA (CpGenome Universal unmethylated DNA (vial A)) and UCB (CpGenome Universal unmethylated DNA (vial B)) were used as positive controls for unmethylated DNA, and MC (methylated control) was used as a positive control for methylated DNA. W1, template-free control from both stages of PCR. W2, template-free control from stage 2 PCR only

We next wanted to see if we could detect a dose-response relationship. A further 25 disease-free women were randomly selected from each of the categories of light and moderate smokers. Again the last cervical smear was selected for DNA extraction and bisulphite modification. Using nested MSP, methylation of the *CDKN2A* promoter was detected in 7/21 (33%) evaluable disease-free light smokers (Figure 4.14), and in 9/20 (45%) evaluable disease-free moderate smokers (Figure 4.13). Thus the prevalence of *CDKN2A* methylation did not vary with smoking category; however the range of values of pack-years smoked was not extensive in this population of young women.

Overall, methylation of the *CDKN2A* promoter was detected more frequently in disease-free smokers than in disease-free never smokers; 23/62 (37.1%) evaluable samples taken from smokers, compared to only 2/21 (9.3%) evaluable samples taken from never smokers (Fisher's exact p-value=0.03).

Although this cross-sectional survey did suggest an association between smoking status and the detection of methylated forms of *CDKN2A*, inferences based on such surveys may be confounded by extraneous factors which are associated with both smoking and the detection of *CDKN2A* methylation. Therefore, I next set out to determine if smoking preceded the first detection of TSG methylation, using a subset of women from the incident cohort who first began to smoke after they were enrolled on study (incident smokers).
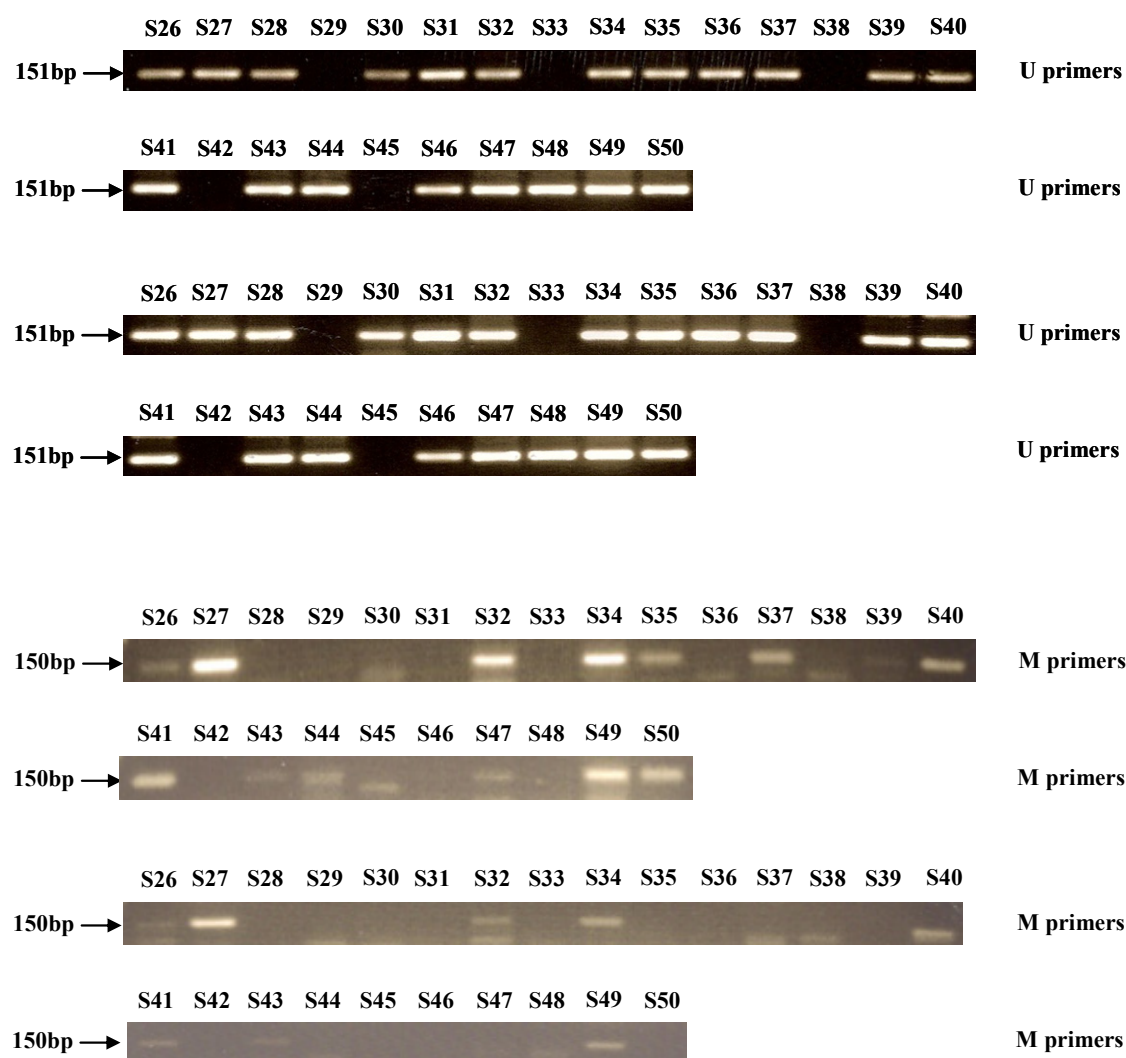
Figure 4.13 Nested MSP in 25 moderate smokers. Using nested MSP, methylation of the *CDKN2A* promoter was detected in 2 replicates in 9/20 (45%) evaluable smears taken from disease-free moderate smokers (S26, S27, S32, S34, S40, S41, S43, S49). Expected product size: unmethylated primer (U) = 151bp, methylated primer (M) = 150bp.
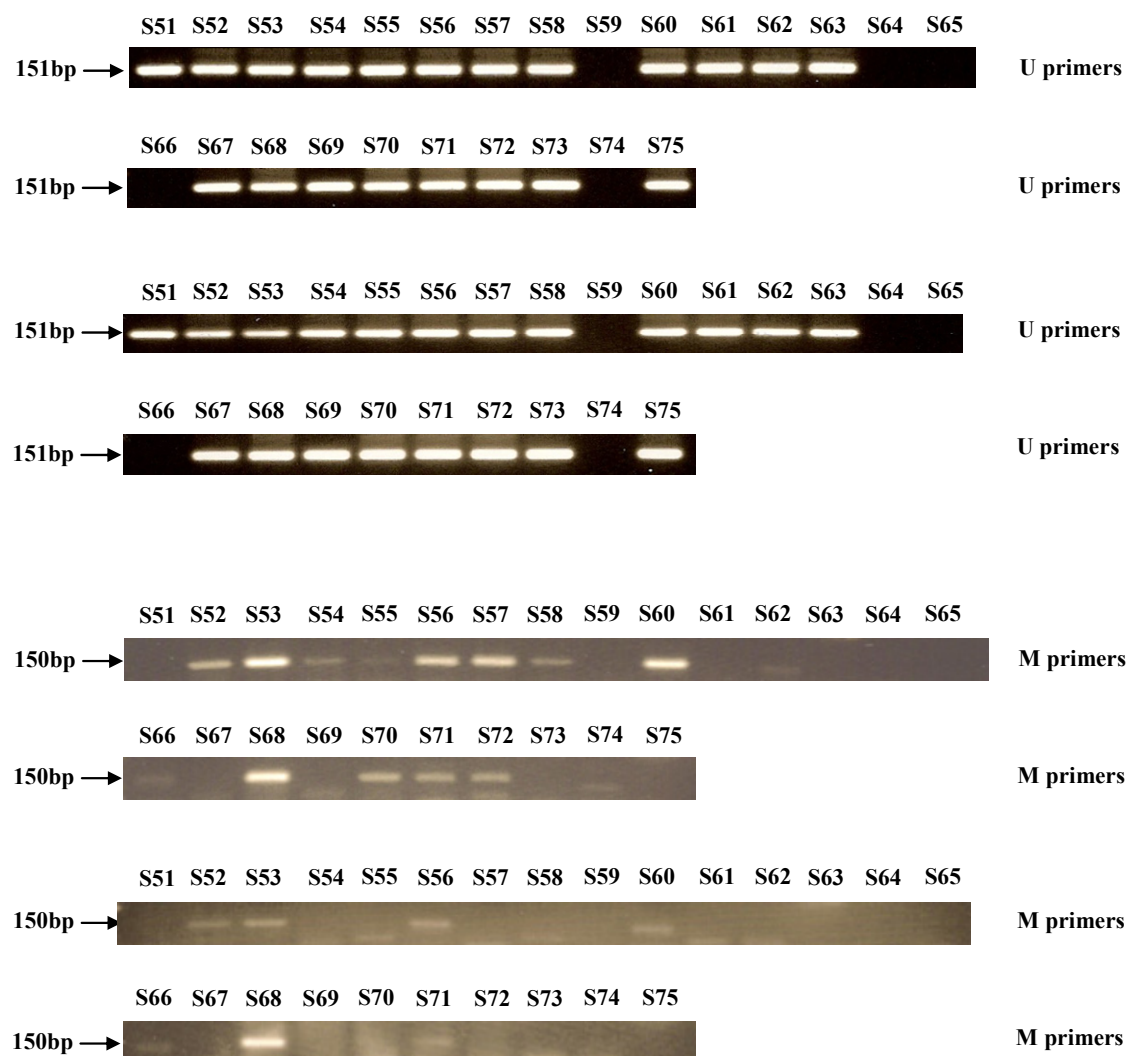
Figure 4.14 Nested MSP in 25 light smokers. Using nested MSP, methylation of the *CDKN2A* promoter was detected in 2 replicates in 7/21 (33%) evaluable smears taken from disease-free light smokers (S52, S53, S56, S60, S66, S68, S71). Expected product size: unmethylated primer (U) = 151bp, methylated primer (M) = 150bp.

4.5.4 Incidence of *CDKN2A* methylation in smokers and never smokers

Next, we compared the incidence of methylation of the *CDKN2A* promoter in disease-free incident smokers with that in disease-free never smokers. From the incident cohort of 1,075 women who were cytologically normal and who tested negative for HPV DNA at study entry, 97 women began smoking during follow-up (Table 4.7) and 60 of these remained cytologically normal and HPV-negative throughout follow-up. All the cervical samples from these 60 women were selected for DNA extraction and bisulphite modification followed by nested MSP for *CDKN2A*. Women were excluded from subsequent analyses when the last sample taken before smoking initiation was unavailable or not evaluable (n=10), when it tested positive for methylated forms of *CDKN2A* (n=3), or if no smoking sample was available or evaluable (n=9) (Table 4.8).

Incident smokers from 1075	**97**
Subsequent test positive for HPV or CAB	37
Pre-smoking smear missing	8
Pre-smoking smear non-evaluable	2
Pre-smoking smear methylated	3
Smoking smears missing	3
Smoking smears non-evaluable	6
Disease-free incident smokers evaluable	**38**

Table 4.8 Breakdown of all incident smokers excluded from subsequent analyses.

Of the remaining 38 women who tested negative for *CDKN2A* methylation in the sample taken immediately before they began to smoke, 8 subsequently tested positive for methylated forms of *CDKN2A* after smoking initiation (Table 4.9); the median time to first detection of methylation of the *CDKN2A* promoter was 266 days (range 153-824).

Each of the 38 women contributing to this incidence analysis was matched, on length of follow-up (within 30 days), with two randomly selected disease-free never smokers. For the 8 women who first tested positive for methylated forms after smoking initiation, the follow-up period was defined as the interval between the last unmethylated sample taken before smoking initiation and the first cervical sample to test positive for methylated forms during that smoking episode. For the remaining 30 women who did not test positive for methylated forms after they started to smoke, the follow-up period was defined as the interval between the last unmethylated sample taken before smoking initiation and the last cervical sample taken during that smoking episode. Nested MSP for *CDKN2A* was performed on the first and last matched sample from each never smoker control. Three controls were excluded from subsequent analyses because of the presence of methylated forms in the first of their two samples (n=2) or when one of the two samples was not evaluable (n=1). Of the remaining 73 controls who tested negative for methylated forms of *CDKN2A* in their first sample, 5 tested positive for these forms in their second sample. Compared with disease-free never smokers, disease-free women who first started to smoke during follow-up had an increased risk of acquiring methylation of the *CDKN2A* promoter (odds ratio=3.67; 95% CI 1.09 to 12.33; χ^2=4.42; 1 df: p=0.04).

	Study no.	Pre-smoking smear	Smoking smears						
			1	2	3	4	5	6	7
1	235	U	M						
2	322	U	M						
3	731	U	M						
4	814	U	M						
5	1500	U	M						
6	1565	U	M						
7	571	U	M	U	M				
8	804	U	U	-	-	M	-		
9	30	U	U						
10	131	U	U						
11	147	U	U						
12	184	U	U						
13	338	U	U						
14	634	U	U						
15	811	U	U						
16	882	U	U						
17	949	U	U						
18	1352	U	U						
19	1362	U	U						
20	1391	U	U						
21	1580	U	U						
22	1843	U	U						
23	1845	U	U						
24	1856	U	U						
25	1922	U	U						
26	2080	U	U						
27	1912	U	-	U					
28	1915	U	-	U					
29	4	U	U	U					
30	756	U	U	U					
31	1020	U	U	U					
32	1581	U	U	U					
33	1759	U	U	U					
34	1679	U	U	U	U				
35	1330	U	U	U	U	-			
36	231	U	U	-	-	U			
37	1151	U	U	U	-	U	U		
38	277	U	U	U	U	U	-	-	U

Table 4.9 Nested MSP results for the 38 incident smokers who tested negative for *CDKN2A* methylation in their last pre-smoking smear. Nested MSP for *CDKN2A* was performed on all the smears taken following smoking initiation, whilst the women remained cytologically normal and HPV-negative. U; unmethylated for *CDKN2A* in 2 replicates. M; methylated for *CDKN2A* in 2 replicates. - = non evaluable sample.

4.5.5 Loss of *CDKN2A* methylation following smoking cessation

Finally, I measured how often *CDKN2A* methylation was reversible following smoking cessation. This analysis was again restricted to disease-free women. From the incident cohort of 1,075 women who were cytologically normal and who tested negative for HPV DNA at study entry, 142 women stopped smoking during follow-up (Table 4.7) and 75 of these remained cytologically normal and HPV-negative throughout follow-up. For these 75 women, the final cervical sample whilst they were still smoking, and all the subsequent cervical samples following smoking cessation were selected for DNA extraction and bisulphite modification. Women were excluded from subsequent analyses when the last sample taken before smoking cessation was unavailable or not evaluable (n=9), when it tested negative for methylated forms of *CDKN2A* (n=40), or if no ex-smoking sample was available or evaluable (n=7) (Table 4.10).

Ex-smokers from 1075	142
Tested positive for HPV or CAB during follow-up	67
Smoking smear missing	3
Smoking smear non-evaluable	6
Not methylated on last smoking smear	40
Ex-smoking smear missing	5
Ex-smoking smear non-evaluable	2
Disease-free ex-smokers evaluable	**19**

Table 4.10 Breakdown of all ex-smokers excluded from subsequent analyses

Of the remaining 19 women who tested positive for *CDKN2A* methylation in their last cervical sample taken immediately before they stopped smoking, 12 subsequently tested negative for *CDKN2A* methylation in one or more cervical samples after smoking cessation; the median time to the first negative sample was 432 days (range 161-1107) (Table 4.11). Of the 7 women who continued to test positive for

methylated forms of *CDKN2A* following smoking cessation, 6 women provided one further sample and 1 woman provided two further samples. However, as the range of follow-up times for these women was only 190 to 893 days (median 266), it would be unwise to conclude that their epigenetic marks were in any meaningful sense "fixed".

	Study no.	Final smoking smear	Ex-smoking smears			
			1	2	3	4
1	250	M	U			
2	706	M	U			
3	751	M	U			
4	828	M	U			
5	1363	M	U			
6	1433	M	U			
7	1458	M	U	U	-	
8	1500	M	U	U		
9	733	M	M	U		
10	949	M	M	U		
11	439	M	-	M	U	U
12	977	M	-	M	-	U
13	1053	M	M	M		
14	329	M	M			
15	411	M	M			
16	606	M	M			
17	973	M	M			
18	1353	M	M			
19	1727	M	M			

Table 4.11 Nested MSP results for the 19 ex-smokers who tested positive for *CDKN2A* methylation in their last smoking smear. Nested MSP for *CDKN2A* was performed on all the smears taken following smoking cessation, whilst the women remained cytologically normal and HPV-negative. U; unmethylated for *CDKN2A* in 2 replicates. M; methylated for *CDKN2A* in 2 replicates. - = non evaluable sample.

4.5.6 Specificity of nested MSP assay

To exclude false priming, ie amplification of unmethylated alleles by the methylated primers, a subset of methylated products was analysed by direct sequencing (n=21): methylation of all CpGs within the *CDKN2A* primer sites was observed in every case (Figure 4.12).

F primer R primer	Sample tube no:
●●●●●●●●●●●●●●●●●●●	Methylated control
●●●○○○○○○○○○○○○●●●●	C2
●●●○○○○○○○○○○○○●●●●	C6
●●●○○○○○○○○○○○○●●●●	C10
○○○○○○○○○○○●●●●	E4
●●●●●●●●●●●●●●●●	E12
○○○○○○○○●●●●	E14
●●●○○○○○○○○○○○○●●●●	M7
●●●○○○○○○○○○○○○●●●●	M171
●●●○○○○○○○○○○○○●●●●	M219
●●●○○○○○○○○○○○○●●●●	M237
●●○○○○○○○○○○○○	P28
●●●○○○○○○○○○○○	P38
○○○○○○○○○○○●●●●	P39
○○○○○○○○○○○●●●●	S3
○○○○○○○○○○○●●●●	S23
○○●●●●○●●●●●●●●	S26
●●●●●○○●●●●○●○●●●●●	S53
●●●○●○●●●●●○●●●●●●	S58
○○○○○○○○○○○●●●●	S78
○○○○○○○○○○○●●●●	S86
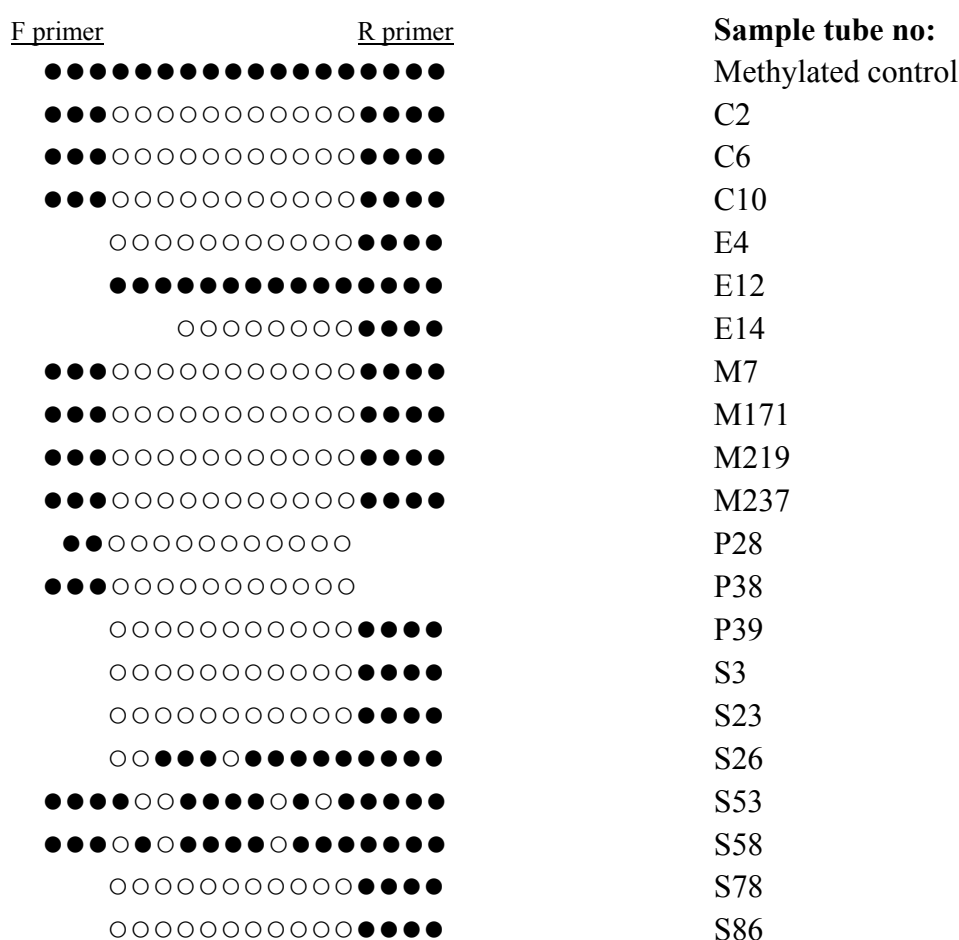
Figure 4.12 Specificity of nested MSP. Summary of direct sequencing results for the extracted methylated band from 21 samples. Methylation of all CpGs within the *CDKN2A* primer sites was observed in every case. The *CDKN2A* forward (F) primer sequence contains 3 CpGs. The *CDKN2A* reverse (R) primer sequence contains 4 CpGs. ○ = unmethylated CpG, ● = methylated CpG.

Chapter 5

RESULTS 2:

ASSOCIATION BETWEEN HPV

INFECTION AND METHYLATION OF

CELLULAR GENES

5.1 Restatement of Objectives

To investigate the spectrum of methylation changes associated with human papillomavirus (HPV) infection:

- To describe and validate the changes in DNA methylation which follow the transfection of PHFK (isolated from a single donor) with episomal HPV18, and which follow long term cultivation of the W12 disease progression model.

- To determine how often changes in DNA methylation observed using these models are recapitulated in cervical neoplasia.

- To investigate possible determinants of virus-associated methylation changes.

- To describe and validate the transcriptional changes which follow the transfection of PHFK (isolated from a single donor) with episomal HPV18 and which follow long-term cultivation of the W12 disease progression model.

- To compare these transcriptional changes with those reported in published gene expression arrays using similar *in vitro* models, and to determine how often these changes are recapitulated in cervical neoplasia.

- To examine the relationship between methylation and transcriptional changes in these models.

- To investigate possible HPV18-induced 'epigenetic switching', using a candidate gene approach.

In this section, I report the results of a pilot study designed to evaluate the use of the UHN HCGI12K arrays in the cervical cancer cell line, HeLa. I first describe the steps involved in data processing, normalisation and analysis before reporting for a panel of candidate genes, how often the methylation status predicted by the array could be confirmed using bisulphite genomic sequencing. Finally, I determine how often genes which were predicted to be methylated on the array have previously been reported to be methylated in cervical neoplasia.

5.2 Assessment of the methylation status of CpG islands in HeLa cells

At the time of this study, the UHN HCGI12K array was the only array-based platform commercially available for the measurement of methylation changes. The technique was first optimised in HeLa cells.

Differential methylation hybridisation (DMH) was performed in HeLa cells with 4 technical replicates as described in section 3.3.9 (Materials and Methods). The HeLa samples restricted with McrBC were labelled with Cy5, and the mock-digested samples were labelled with Cy3.

5.2.1 Data processing

The Cy3 and Cy5 fluorescent intensities were obtained for each hybridised spot (see 3.3.9.3). Prior to data analysis, array spots of low quality were first filtered as described in section 3.3.9.4. After this filtering step, the number of remaining array

spots contributing to the data analysis for all 4 replicates of HeLa is shown in Table 5.1.

Array	No. of spots "Not Found"	No. with signal intensities near background ($\leq$25)	Remaining spots
HeLa 1	1601	805	9786
HeLa 2	768	12	11412
HeLa 3	1710	7	10475
HeLa 4	294	3	11895

Table 5.1 Filtering - spots of low quality were first removed prior to data analysis.

5.2.2 Normalisation

The median Cy3 and Cy5 signal intensities of all the mitochondrial clones on each array were used to calculate a normalisation factor as described in section 3.3.9.5, and this was then applied to all the probes on each corresponding array.

5.2.3 Confirmation of the adequacy of the normalisation step

The log ratios of the mitochondrial clones were plotted for each array. Following normalisation the log ratio should equal zero. Figure 5.1 shows that the log ratios for the mitochondrial clones for each replicate were close to zero.

HeLa 1: Normalised log ratios for mitochondrial clones

HeLa 2: Normalised log ratios for mitochondrial clones

HeLa 3: Normalised log ratios for mitochondrial clones

HeLa 4: Normalised log ratios for mitochondrial clones
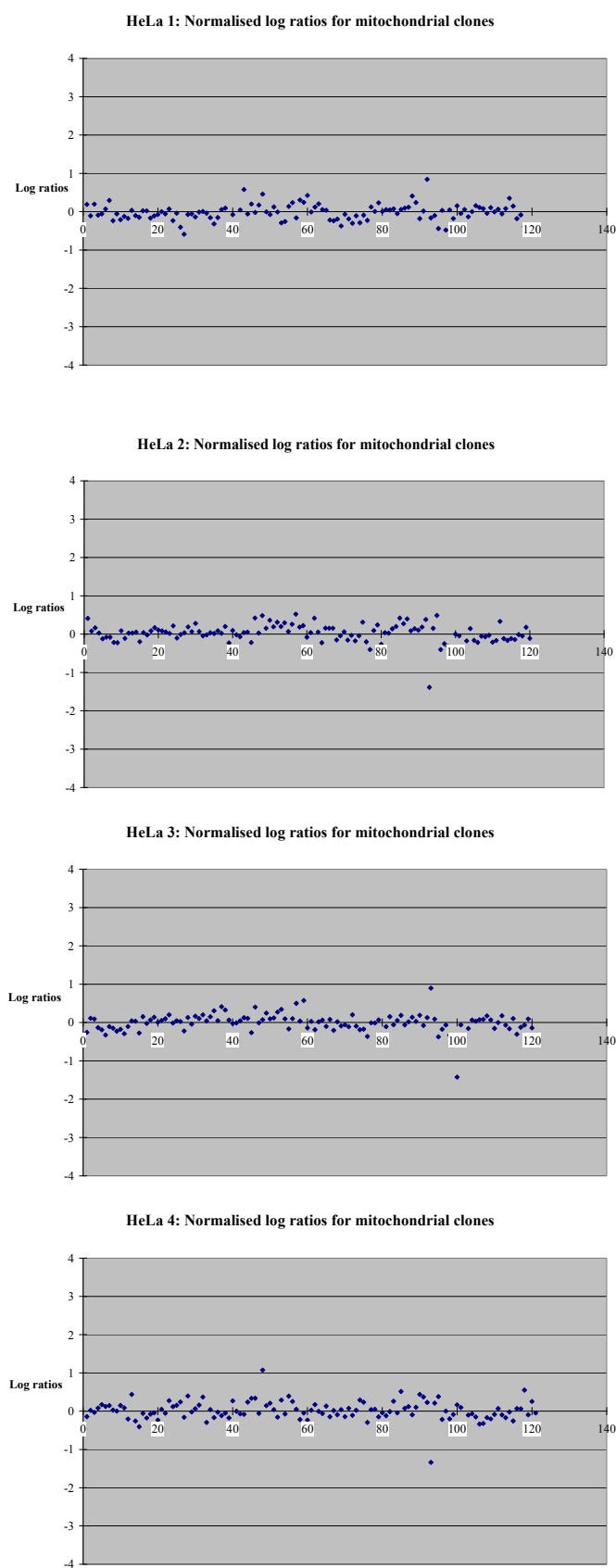
Figure 5.1 Normalised log ratios of the mitochondrial clones for all 4 replicates of HeLa.

The log (mock-cut) vs the log (cut) was plotted to check that the mitochondrial clones cover the full range of the signal intensities. Figure 5.2 shows that the mitochondrial clones on each array cover most of the range of the signal intensities. Graphs of log ratio (M) vs mean log intensity (A) (MA graphs) were then plotted for each array individually. Based on the assumption that most genes will be unchanged, the majority of points on the y axis (log ratio) will be located at zero. Figure 5.3 shows that the log ratio for all probes on each array was located around zero following normalisation.

5.2.4 Data analysis

Having shown that the technical quality of each individual array was adequate, the 4 replicates were then analysed together. Only clones that were informative in all 4 replicates were selected. Any clone with a low quality result in any of the 4 replicates was removed from further analysis: 2932 clones were filtered leaving 9260 common informative clones on all 4 arrays, corresponding to 2804 unique genes.

A fluorescent ratio was calculated for all the remaining spots with the mock-cut signal intensity (Cy3) as denominator and the cut signal intensity (Cy5) as numerator. The resulting ratio reflects the degree of methylation of each CpG island clone; a ratio approaching 0 indicates a methylated CpG island, and a ratio approaching 1 indicates an unmethylated CpG island. Average fluorescent ratios were then calculated using the normalised ratios from each of the 4 technical replicates.
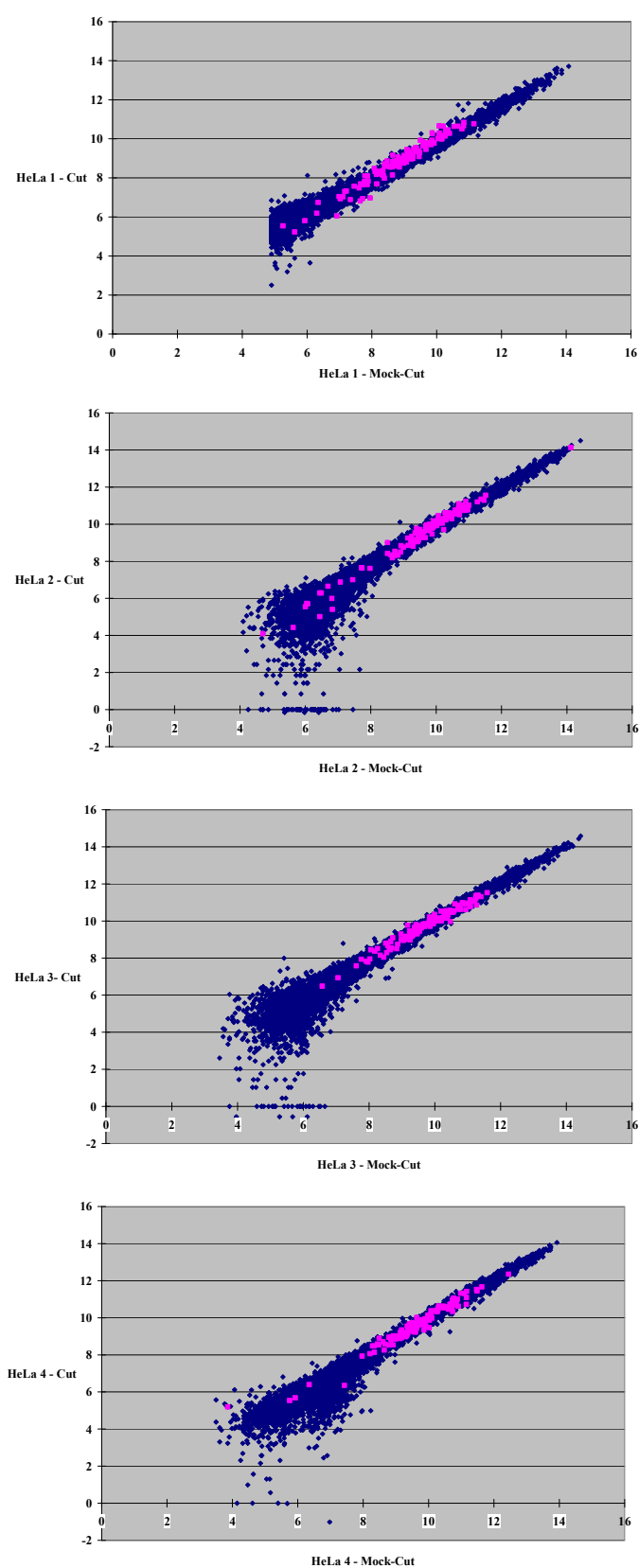
Figure 5.2 Log (mock-cut) vs log (cut) plots for all 4 replicates of HeLa showing that the mitochondrial clones (in pink) cover most of the range of the signal intensities on each array.

a) HeLa 1   **Pre**   **Post**

b) HeLa 2

c) HeLa 3

d) HeLa 4
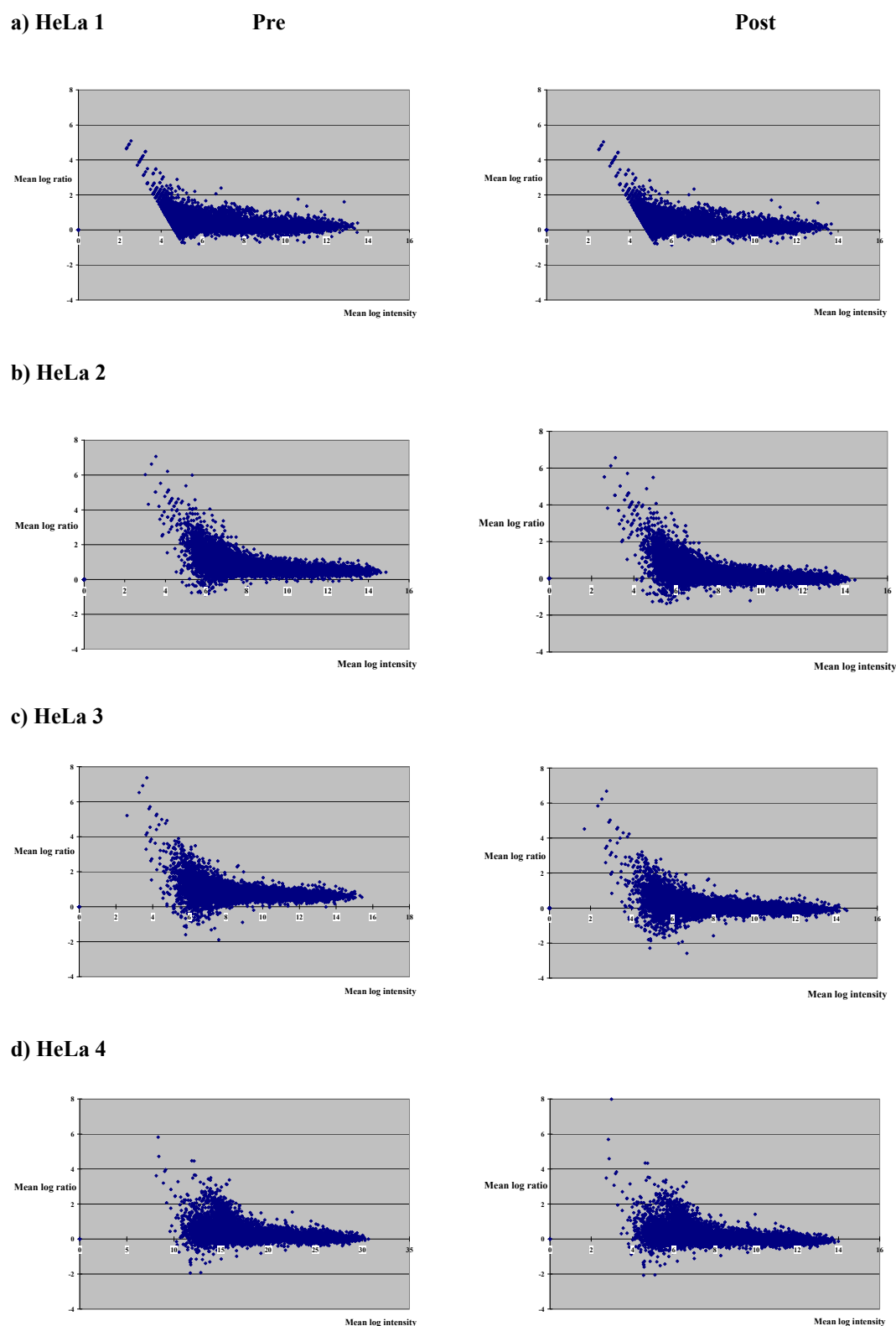
Figure 5.3 Graphs of log ratio (M) vs mean log intensity (A) (MA plots) pre and post-normalisation for all 4 replicates of HeLa.

To further categorise the data, I have used the method published by Nouzova *et al* (2004), in which the methylation status is classified using the cut/mock-cut ratios:

<0.5	= heavily methylated (HM)
0.5-0.67	= methylated (M)
0.68-0.87	= partially methylated (PM)
>0.87	= not methylated (NM)

Table 5.2 and Figure 5.4 show the number of informative clones and the corresponding genes in each of these methylation categories. Genes that were represented more than once on the array were re-allocated to the category with the highest methylation status.

Category	No. of clones	No. of genes
Heavily methylated (<0.5)	163 (2%)	85
Methylated (0.5-0.67)	652 (7%)	298
Partially methylated (0.68-0.87)	3897 (42%)	1366
Not methylated (≥0.88)	4548 (49%)	1055

Table 5.2 Distribution of the methylation status of the common informative clones in HeLa cells and their corresponding genes.
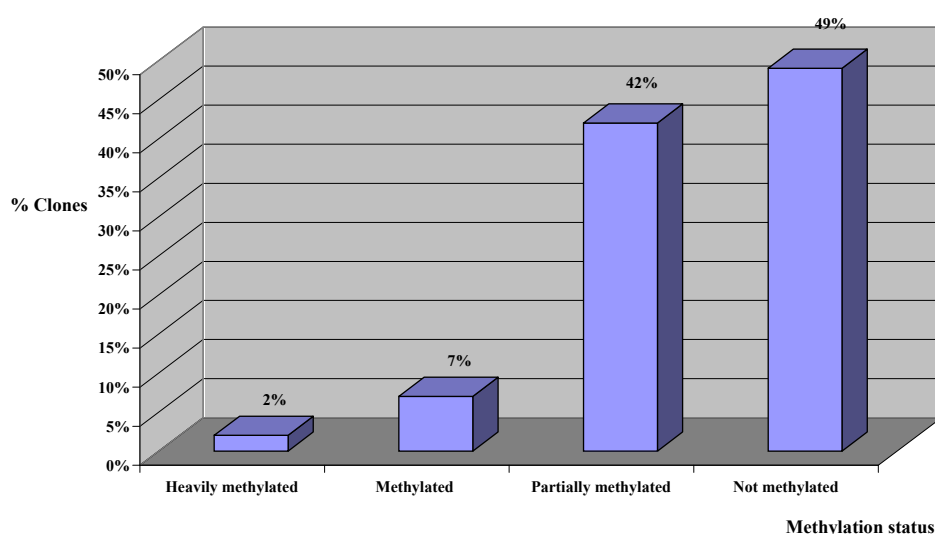
Figure 5.4 Graph showing the distribution of the methylation status of the common informative clones in HeLa cells.

5.2.5 Validation using bisulphite genomic sequencing

To determine if these arbitrary cut-off values accurately predicted methylation status, the CpG islands corresponding to 2 genes from each of the four categories were selected for validation by bisulphite genomic sequencing (BGS) (Table 5.3). The primer sequences used are listed in Table 3.4 (Materials and Methods). The primers were designed to cover all the CpGs present in each CpG island clone selected for analysis by BGS

Category	UHN ID	Gene symbol	Cut/mock-cut ratio
Heavily methylated (<0.5)	UHNhscpg0000969	*MARK1*	0.46
	UHNhscpg0000007	*SOCS3*	0.39
Methylated (0.5-0.67)	UHNhscpg0002561	*RARB*	0.56
	UHNhscpg0011071	*WT1*	0.54
Partially methylated (0.68-0.87)	UHNhscpg0006964	*ATM*	0.84
	UHNhscpg0010927	*CDKN2A*	0.81
Not methylated (≥0.88)	UHNhscpg0010736	*CDH13*	0.88
	UHNhscpg0001366	*ROBO3*	0.95

Table 5.3 Selection of genes for confirmation by bisulphite genomic sequencing.

Bisulphite genomic sequencing was only successful for five genes (*SOCS3, RARB, ATM, CDH13* and *ROBO3*) (Figure 5.5). For *SOCS3* and *RARB*, there was good concordance between the methylation status predicted by the array and that observed by BGS. However, the array results and the BGS data were discordant for 3 genes. Although *ATM* was predicted to be "partially methylated" using the array (the cut/mock-cut ratio was 0.84), it was unmethylated by BGS. Similarly, *CDH13* and *ROBO3* were both predicted to be "not methylated" using the array, but using BGS I have shown them both to be partially methylated. Thus the cut-off values do not reliably predict methylation status determined by BGS when a gene is not methylated or is partially methylated, but do when the gene is methylated or heavily methylated.

a) *SOCS3* ("heavily methylated")

●●●●●●●●●●●
○●●●○●●●●●●
●●●●●●●●●●●
●●●●●●●●●●●
●●●●●●●●●●●
●●●●●●●●●●●
●●●●●●●●●●
○●●●○●●●●●●

4503bp downstream
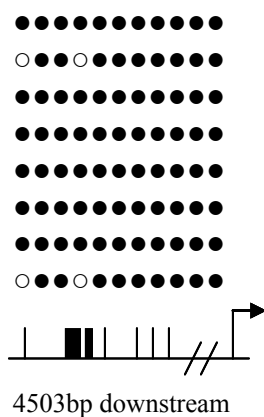
b) *RARB* ("methylated")

●●●●●●●●●●●●●●●●○
●●●●●●●●●●●●●●●●○
●●○○●●●●●●●●●●●●●
●●●●●●●●●●●●●●●●
●●●●●●●●●●●●●○○●○
●●●●●●●●●●●●●●●○
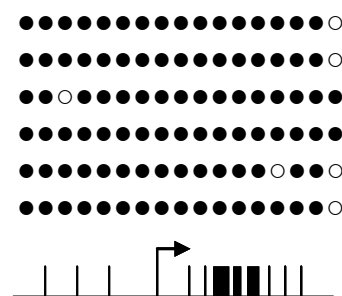
c) *ATM* ("partially methylated")

○○○○○○○○
○○○○○○○○
○○○○○○○○
○○○○○○○○
○○○○○○○○
○○○○○○○○
○○○○○○○○
○○○○○○○○

247bp upstream
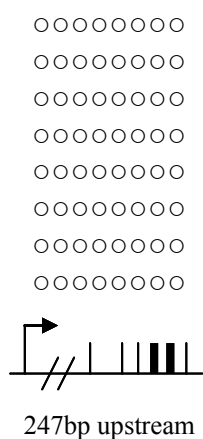
d) *CDH13* ("not methylated")

●●○○●
○●○○
●●○●
●●○●
○●○●
●●○○
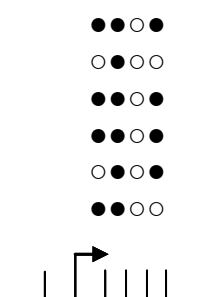
e) *ROBO3* ("not methylated")

●●○●●●●
●●●○●●●
○●○○●●●
○○●○●●●
○○○●●●●
●●○●●○●
●○○●○○○
●●○●●●○

1005bp upstream
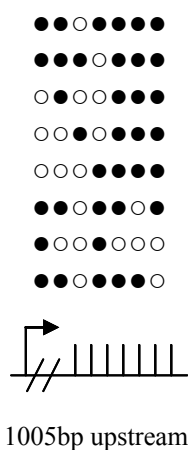
Figure 5.5 Bisulphite genomic sequencing results for HeLa. Each circle represents a CpG site. ●=methylated CpG. ○=unmethylated CpG. Below each set of BGS result is a diagrammatic representation of the location of the CpG island in relation to the transcriptional start site (arrow) of the gene; the vertical lines represent CpG sites. The methylation status predicted by the array is shown in inverted commas.

 5.2.6 External validation

A review of the literature has previously identified 31 genes that are methylated significantly more frequently in women with invasive cervical cancer compared with disease-free controls or women with low-grade CIN (see section 4.2 and Table 4.2). This analysis has recently been updated by Professor Ciaran Woodman, and the number of genes reported to be hypermethylated in cervical neoplasia has now increased to 58. Of these 58 genes, 15 are represented on the UHN HCGI12K arrays in HeLa cells post-filtering (see section 5.2.4). Table 5.4 shows the number of genes predicted in HeLa cells to be "heavily methylated" or "methylated" using the UHN HCGI12K arrays, which have previously been reported to be methylated in cervical neoplasia.

Methylation status in HeLa cells	Methylated in cervical neoplasia		Odds ratio (95% CI)
	Yes	No	
"Not methylated" (n=1055)	2	1053	1 (reference)
"Heavily methylated" or "methylated" (n=383)	8	375	11.3 (2.4-53.3) p=0.0001

Table 5.4 Frequency with which genes classified as "heavily methylated" or "methylated" in HeLa cells are known to be methylated in cervical neoplasia.

This comparison revealed that compared to genes predicted to be "not methylated" in HeLa cells, those genes which were classified as "heavily methylated" or "methylated" were found to be enriched significantly (11-fold) for genes previously reported to be methylated in cervical neoplasia; these 8 genes are listed in Table 5.5.

Genes	Reference
CDH13	Widschwendter 2004; Feng 2005; Henken 2007
HOXA11	Apostolidou 2009
NKX6-1	Lai 2008
ONECUT1	Lai 2008
RARB	Virami 2001; Narayan 2003; Gustafson 2004; Feng 2005; Wisman 2006; Henken 2007
ROBO1	Narayan 2006
SLIT2	Narayan 2006
SOCS3	Shivapurkar 2007

Table 5.5 List of 8 genes predicted to be "heavily methylated" or "methylated" in HeLa cells and also known to be methylated in cervical neoplasia.

In summary, in this section I have shown that using the UHN HCGI12K array in HeLa cells, 2 candidate genes predicted to be "heavily methylated" or "methylated" were confirmed as such using BGS. Furthermore, compared with genes predicted to be "not methylated" in HeLa cells, genes predicted to be "heavily methylated" or "methylated" were also significantly more likely to be known to be methylated in cervical neoplasia. The UHN HCGI12K array was therefore taken forward for use in our 2 *in vitro* models, primary human foreskin keratinocytes following transfection with HPV18 and the W12 disease progression model.

In this section, I describe how the UHN HCGI12K arrays were used to assess the methylation status of cellular genes in untransfected PHFK, and how this changed following transfection of these cells with HPV18. I first describe and compare using the methods outlined in the preceding section, the methylation changes seen at an early (P4) and later passage (P9) before investigating how often changes in methylation status predicted on the array could be confirmed by pyrosequencing. Finally, I determine whether genes which were predicted to have increased methylation following HPV18 transfection were enriched for those previously reported to be methylated in cervical neoplasia.

5.3 Methylation changes following transfection of PHFK with HPV18

The UHN HCGI12K arrays were used to assess the methylation status of cellular genes in untransfected PHFK from a single neonatal donor (#1), and the methylation changes following transfection with HPV18 at an early (P4) and later passage (P9).

Prior to data analysis, arrays spots of low quality were first filtered as described in section 3.3.9.4 (Table A1, Appendix). The data were then normalised using the subset of mitochondrial clones on each array as described in section 3.3.9.5.

For the data analysis all 9 arrays were analysed together. Only clones that were informative in all 9 arrays were selected: 3534 clones were filtered leaving 8658 common informative clones on all 9 arrays, corresponding to 2612 unique genes.

In untransfected PHFK (#1), <1% (35) of the clones were classified as "heavily methylated", 4% "methylated" (352), 20% "partially methylated" (1758) and 75% "not methylated" (6513) (Figure 5.6a).
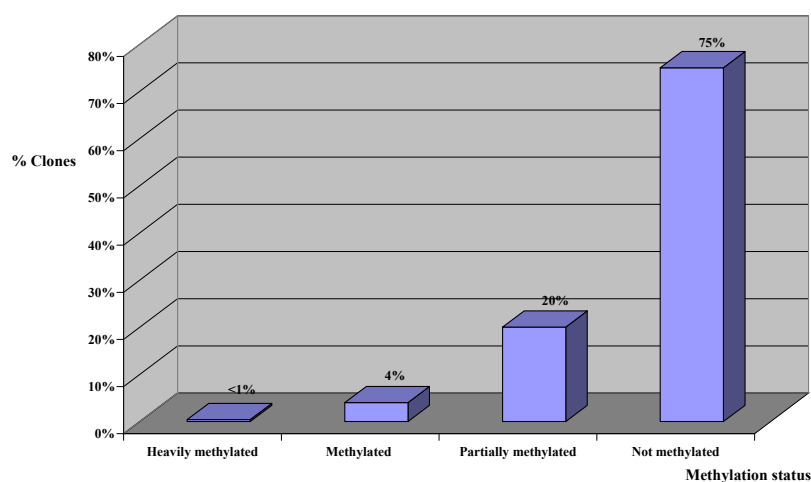
Figure 5.6a Graph showing the distribution of the methylation status of the common informative clones in untransfected PHFK (#1).

At passage 4, in PHFK (#1) transfected with HPV18, 2% (160) of the clones were classified as "heavily methylated", 15% "methylated" (1296), 43% "partially methylated" (3765) and 40% "not methylated" (3437) (Figure 5.6b).
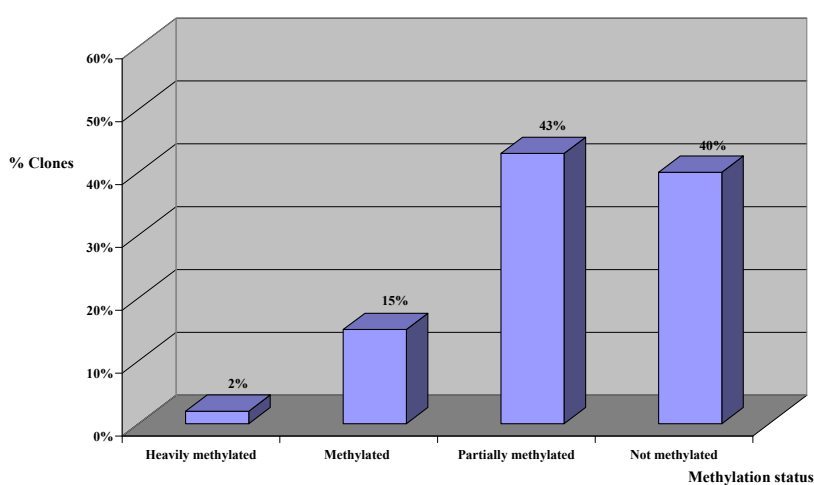
Figure 5.6b Graph showing the distribution of the methylation status of the common informative clones in PHFK (#1) transfected with HPV18 – passage 4.

At passage 9, in PHFK (#1) transfected with HPV18, 2% (181) of the clones were classified as "heavily methylated", 10% "methylated" (843), 48% "partially methylated" (4141) and 40% "not methylated" (3493) (Figure 5.6c).
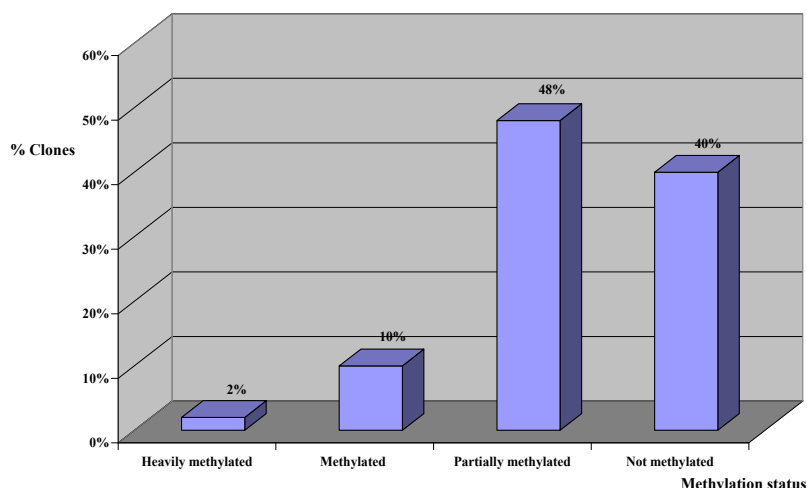
Figure 5.6c Graph showing the distribution of the methylation status of the common informative clones in PHFK (#1) transfected with HPV18 – passage 9.

These graphs show that at passage 4 of PHFK (#1) transfected with HPV18, the distribution of the methylation status of the common informative clones showed an increase in the number of clones classified as "methylated" and "partially methylated", and a decrease in the number of clones classified as "not methylated". However there were no substantial changes in these proportions between passage 4 and passage 9.

5.3.1 Differentially methylated genes

A gene was considered to be differentially methylated if the cut/mock-cut ratios of the associated clone between untransfected PHFK (#1) and PHFK (#1) transfected with HPV18 at passages 4 or 9, were found to be significantly different (p<0.05) using the

t-test (see section 3.3.9.6). Table 5.6 shows the number of clones and their corresponding genes found to be differentially methylated at an early and later passage of PHFK (#1) following transfection with HPV18.

Comparison	Methylation change	No. of clones	No. of genes
PHFK(#1)/HPV18 P4 vs untransfected PHFK	hypermethylated	1546	667
	hypomethylated	83	30
	both*	-	6
PHFK(#1)/HPV18 P9 vs untransfected PHFK	hypermethylated	1698	556
	hypomethylated	103	56
	both*	-	7
PHFK(#1)/HPV18 P9 vs PHFK(#1)/HPV18 P4	hypermethylated	564	172
	hypomethylated	574	291
	both*	-	9

Table 5.6 Number of clones and corresponding genes differentially methylated in PHFK (#1) following transfection with HPV18.

*Both = genes identified as both hypermethylated and hypomethylated in non-overlapping regions

At both an early and late passage of PHFK (#1) following transfection with HPV18, it can be seen that more genes became hypermethylated than hypomethylated compared to untransfected PHFK (#1). In total 986 unique genes became significantly more methylated at early or late passages, or both, while 51 unique genes became significantly less methylated.

5.3.2 *De novo* methylated genes

I also undertook an analysis to identify genes which became methylated *de novo* following transfection of PHFK (#1) with HPV18. A CpG island clone was classified as methylated *de novo* if it was "not methylated" in untransfected PHFK (#1) (>0.87) and then became "methylated" (0.5-0.67) or "heavily methylated" (<0.5) following transfection with HPV18 (early or late passage). Table 5.7 shows the number of

clones and their corresponding genes which became *de novo* methylated following transfection of PHFK (#1) with HPV18 at passages 4 and 9.

De novo methylation	No. of clones	No. of genes
Heavily methylated Early	4	4
Methylated Early	290	167
Both*	-	0
Heavily methylated Late	28	18
Methylated Late	287	136
Both*	-	2

Table 5.7 Number of clones and corresponding genes which became *de novo* methylated following transfection of PHFK (#1) with HPV18.

*Both = genes identified as both "heavily methylated" and "methylated" in non-overlapping regions in cells from the same passage

In total 566 unique CpG island clones became methylated *de novo* following transfection of PHFK with HPV18 at early or late passage, corresponding to 294 unique genes.

Interestingly, most of the genes which became methylated *de novo* at early and late passages did not overlap: only 29 genes were common to both groups (Figure 5.7).
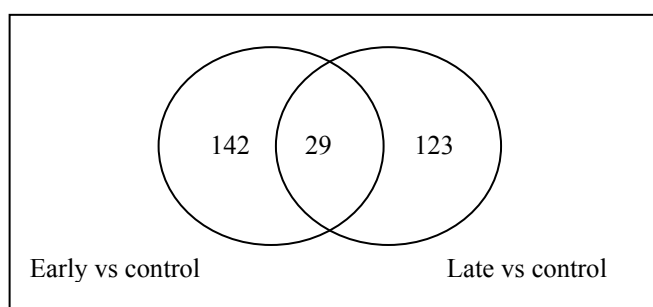
Figure 5.7 Number of genes which became *de novo* methylated in PHFK transfected with HPV 18 compared to untransfected keratinocytes.

142 unique genes which became "methylated" or "heavily methylated" for the first time at early passage (P4) were found to be less methylated at passage 9. Of the 29

genes which were found to be *de novo* methylated at both early and late passages, only 3 of these showed an increase in methylation between early and late passage – from "methylated" to "heavily methylated". The remaining 26 genes became "methylated" at early passage and their methylation status remained unchanged at passage 9. 123 unique genes became *de novo* methylated for the first time at late passage.

5.3.3 Validation of methylation changes using pyrosequencing

Pyrosequencing was used to validate some of the methylation changes observed in PHFK (#1) following transfection with HPV18, using DNA harvested from the same passages used to probe the UHN HCGI12K arrays. All the pyrosequencing analyses were performed twice.

<u>5.3.3.1 Criteria used to select genes for validating methylation changes in PHFK following transfection with HPV18:</u>

Twelve genes (*CDH13, NKX6-1, NOL4, RARB, SLIT2, SOX1, COL5A3, GAP43, KIT, NR2F1, PLXDC2,* and *UNC5B*) were selected which fulfilled both of the following criteria (Table A3, Appendix):

- Not methylated in untransfected PHFK (#1)

- Became "methylated" or "heavily methylated" following transfection of PHFK (#1) with HPV18 at early or late passages (ie *de novo* methylated)

Six of these genes have also previously been reported to be methylated in cervical neoplasia (*CDH13, NKX6-1, NOL4, RARB, SLIT2, SOX1*).

5.3.3.2 Pyrosequencing results

The sequence provided in the annotation file was used to design pyrosequencing primers: one or more primers were designed to cover all (or nearly all) the CpGs present in each CpG island probe. The primer sequences used are listed in Table A5 (Appendix). Pyrosequencing analysis was performed in untransfected PHFK (#1), and in PHFK (#1) following transfection with HPV18 at passages 4 and 9.

Of the 12 genes, 1 (*SLIT2*) could not be analysed further because no CpG sites were present in the sequence provided in the annotation file; this must be considered a false positive result, the reasons for which will be discussed later. Two genes (*SOX1* and *KIT*) repeatedly failed the pyrosequencing run despite producing a single band of the correct size following PCR amplification. Pyrosequencing analysis was successful for the remaining 9 genes. *De novo* methylation was confirmed in only 3 genes (*NKX6-1*, *PLXDC2* and *RARB*) (Table 5.8). For *NKX6-1*, *de novo* methylation was observed at CpG 9 and for *PLXDC2*, this was observed at CpG 1. For *RARB*, *de novo* methylation of CpGs 1-3 was found at passage 4 following transfection with HPV18. In addition, in untransfected keratinocytes, 3 CpGs (numbers 13, 18 and 19) in *RARB* were already methylated, and these showed a further increase in methylation at passage 4 following transfection with HPV18. For all 3 genes (*NKX6-1, PLXDC2* and *RARB*) pyrosequencing also demonstrated progressive methylation at all the CpG sites described between passages 4 and 9, whilst the UHN HCG12K array predicted no further change or a decrease in methylation between these 2 stages.

The pyrosequencing results for the remaining 6 genes (*CDH13, COL5A3, GAP43, NOL4, NR2F1, UNC5B*) are shown in Table 5.9; de novo methylation was not demonstrated at any CpG site for these 6 genes. Furthermore, although all 6 genes were predicted by the array to be not methylated in untransfected keratinocytes, pyrosequencing revealed significant methylation at all the CpG sites for 3 genes (*CDH13, NR2F1* and *UNC5B*).

Gene	Sample	Cut/mock-cut ratio	% Methylation																		
			CpG 1	CpG 2	CpG 3	CpG 4	CpG 5	CpG 6	CpG 7	CpG 8	CpG 9	CpG 10	CpG 11	CpG 12	CpG 13	CpG 15	CpG 16	CpG 17	CpG 18	CpG 19
NKX6-1	Untransfected PHFK#1	0.94	0.0	4.2	16.6	12.2	6.5	0.0	3.6	3.2	**0.0**	5.9	6.8	10.8	0.0	4.8				
	PHFK#1/HPV18 P4	0.67	0.0	1.8	11.3	9.7	5.9	2.3	5.9	3.7	**5.0**	4.2	5.1	9.9	0.0	0.0				
	PHFK#1/HPV18 P9	0.65	2.7	1.8	10.7	5.3	5.5	3.4	4.6	6.1	**10.9**	4.9	9.5	10.3	0.0	6.4				
	MC	-	-	-	-	-	-	91.3	90.9	89.1	95.0	81.1	96.3	96.8	76.3	98.0				
PLXDC2	Untransfected PHFK#1	0.99	**7.7**	12.0	3.9	2.7	10.8	15.8												
	PHFK#1/HPV18 P4	0.58	**13.3**	12.3	4.6	3.5	10.6	19.5												
	PHFK#1/HPV18 P9	0.75	**18.9**	14.8	6.0	3.4	13.8	24.7												
	MC	-	98.8	93.6	97.6	94.6	95.9	88.5												
RARB	Untransfected PHFK#1	0.93	**3.7**	**5.1**	**3.8**	3.5	1.9	2.3	3.7	2.4	4.7	2.9	7.0	5.8	**9.9**	7.0	4.1	4.9	**13.0**	**17.1**
	PHFK#1/HPV18 P4	0.56	**15.1**	**16.8**	**13.7**	4.9	4.0	3.5	5.0	2.0	4.7	4.6	14.2	7.0	**14.6**	7.6	4.3	5.5	**24.7**	**30.3**
	PHFK#1/HPV18 P9	0.95	**19.6**	**22.7**	**17.9**	5.1	3.8	4.7	5.1	4.1	9.1	5.2	14.0	10.5	**19.6**	4.5	3.4	7.0	**22.2**	**34.0**
	MC	-	94.8	97.2	95.7	93.1	96.4	60.1	90.4	77.6	92.6	95.2	100.0	91.0	97.0	100.0	100.0	100.0	95.9	91.4

Table 5.8a Pyrosequencing results for selected candidate genes in the PHFK-HPV18 cell line. This table shows the number of CpGs analysed in the selected CpG island for each gene, and the methylation percentage at each CpG site using cells from untransfected PHFK (from donor #1), and PHFK from the same donor transfected with HPV18 at passage (P) 4 and P9. The corresponding cut/mock-cut ratios using the UHN HCGI12K array are shown. MC, methylated control**.**

No.	Gene	Sample	Cut/mock-cut ratio	% Methylation										
				CpG 1	CpG 2	CpG 3	CpG 4	CpG 5	CpG 6	CpG 7	CpG 8	CpG 9	CpG 10	CpG 11
1	COL5A3	Untransfected PHFK#1	0.88	2.3	3.0	0.0	3.6	0.0	3.8	0.0				
		PHFK#1/HPV18 P4	0.67	3.2	3.9	2.5	3.8	0.0	5.5	4.9				
		PHFK#1/HPV18 P9	0.67	3.2	3.5	2.9	2.6	0.0	4.6	4.1				
		MC	-	96.2	95.4	95.5	93.7	74.6	100.0	98.9				
2	NOL4	Untransfected PHFK#1	1.16	9.7	15.3	16.6	0.0	11.3	2.7	5.1	0.0	6.2		
		PHFK#1/HPV18 P4	1.00	10.1	16.3	12.2	0.0	8.9	7.5	5.8	9.6	4.9		
		PHFK#1/HPV18 P9	0.63	6.7	14.4	11.5	0.0	9.8	5.1	9.3	9.7	7.4		
		MC	-	-	99.1	91.2	88.4	56.7	95.4	97.4	100.0	89.4		
3	NR2F1	Untransfected PHFK#1	0.88	50.9	56.9	65.0	36.1	64.2	13.5	54.2	59.6	62.6	36.0	56.7
		PHFK#1/HPV18 P4	0.50	49.6	54.3	61.7	18.8	61.4	11.3	51.6	59.0	52.5	41.2	52.6
		PHFK#1/HPV18 P9	0.88	56.7	62.0	67.2	12.6	62.4	10.2	62.5	76.3	77.4	66.0	78.9
		MC	-	60.3	83.3	100.0	99.5	93.6	44.3	100.0	100.0	95.5	97.4	95.5
				CpG 12	CpG 13	CpG 14	CpG 15	CpG 16	CpG 17	CpG 18	CpG 19	CpG 20	CpG 21	
		Untransfected PHFK#1	0.88	53.2	47.2	54.4	39.5	55.0	32.6	57.9	57.2	60.2	46.8	
		PHFK#1/HPV18 P4	0.50	53.6	46.0	46.7	35.0	49.3	30.8	52.7	51.3	54.1	47.9	
		PHFK#1/HPV18 P9	0.88	76.9	67.0	57.0	65.1	64.8	42.5	61.9	69.2	74.6	63.3	
		MC	-	100.0	91.8	83.9	100.0	78.2	72.7	73.3	87.4	100.0	96.2	

Table 5.8b Pyrosequencing results for selected candidate genes in the PHFK-HPV18 cell line II. This table shows the number of CpGs analysed in the selected CpG island for the remaining 6 genes, and the methylation percentage at each CpG site using cells from untransfected PHFK (from donor #1), and PHFK from the same donor transfected with HPV18 at passage (P) 4 and P9. The corresponding cut/mock-cut ratios using the UHN HCGI12K array are shown. MC, methylated control**.**

No.	Gene	Sample	Cut/mock-cut ratio	% Methylation				
				CpG 1	CpG 2	CpG 3	CpG 4	CpG 5
4	CDH13	Untransfected PHFK#1	0.88	86.2	88.4	96.7	70.1	100.0
		PHFK#1/HPV18 P4	0.68	82.3	90.3	94.7	70.5	100.0
		PHFK#1/HPV18 P9	0.77	87.3	87.9	89.7	73.4	94.5
		MC	-	96.5	91.7	97.8	74.0	100.0
5	GAP43	Untransfected PHFK#1	0.88	3.6	5.7	7.8	0.0	
		PHFK#1/HPV18 P4	0.56	4.4	4.0	11.0	6.4	
		PHFK#1/HPV18 P9	0.61	4.7	5.2	10.9	4.2	
		MC	-	92.0	91.8	91.8	66.9	
6	UNC5B	Untransfected PHFK#1	0.92	15.7	16.9	15.6	19.3	
		PHFK#1/HPV18 P4	0.48	16.1	16.9	13.1	11.3	
		PHFK#1/HPV18 P9	0.76	15.2	20.6	13.2	6.5	
		MC	-	89.1	89.9	76.7	85.3	

Table 5.8b (cont)

5.3.4 External validation: Comparison with genes previously reported to be methylated in cervical neoplasia

Next, the results of the UHN HCGI12K arrays performed in PHFK (#1) following transfection with HPV18, were compared with genes previously reported to be methylated in cervical neoplasia. Of the 58 genes, 18 were represented on the UHN HCGI12K arrays in the PHFK-HPV18 cell line (post-filtering) (see section 5.2.6).

Change in methylation status		Methylated in cervical neoplasia		Odds ratio (95% CI)
		Yes	No	
Significantly more methylated E or L	No (n=1626)	10	1616	1 (reference)
	Yes (n=986)	8	978	OR = 1.3 95% CI: 0.5-3.4 p = 0.5565
De novo methylated E or L	No (n=2318)	13	2305	1 (reference)
	Yes (n=294)	5	289	OR = 3.0 95% CI: 1.1-8.7 p = 0.0260

Table 5.9 Frequency with which genes differentially more methylated or *de novo* methylated following transfection of PHFK (#1) with HPV18, are known to be methylated in cervical neoplasia.

Of these 18 genes, 8 (*CDH13, HSPA2, ONECUT1, OPCML, RARB, RB1, ROBO1, SLIT2*) were found to be significantly more methylated at early or late passages following transfection of PHFK (#1) with HPV18 (Table 5.9, top panel). However, compared to genes which did not become significantly more methylated, those which did were not significantly enriched for genes previously reported to be methylated in cervical neoplasia.

When the analysis was restricted to genes that became *de novo* methylated (Table 5.9, bottom panel), of the 18 genes known to be methylated in cervical neoplasia, 5 (*CDH13, NKX6-1, RARB, SLIT2, SOX1*) were found to be *de novo* methylated at early or late passages following transfection of PHFK (#1) with HPV18. This analysis revealed that compared to genes that did not become *de novo* methylated, genes that did were significantly enriched (3-fold) for genes previously reported to be methylated in cervical neoplasia. The 5 genes identified were *CDH13, NKX6-1, RARB, SLIT2,* and *SOX1*.

In summary, in this section I have shown using the UHN HCGI12K array platform that 986 genes were significantly more methylated, and 51 genes significantly less methylated in PHFK (#1) following transfection with HPV18. *De novo* methylation was also observed in 294 unique genes, and in this group there was a significant enrichment for genes previously reported to be methylated in cervical neoplasia. Pyrosequencing analysis was successful in 9 genes, but *de novo* methylation was confirmed in only 3 genes.

In this section, I describe how the UHN HCGI12K arrays were used to assess the methylation status of cellular genes in early (P11) and late (P56) passages of the W12 cell line, a disease progression model. I first describe and compare using the methods outlined in preceding sections, the methylation changes seen at early and late passages, before investigating how often changes predicted on the array could be confirmed by pyrosequencing. Finally, I determine whether genes which were predicted to have increased methylation following progression of the W12 cell line were enriched for those previously reported to be methylated in cervical neoplasia.

5.4 Methylation changes in the W12 disease-progression model

The UHN HCGI12K arrays were also used to assess the methylation status of W12 cells at passages 11 and 56, as described in section 5.3.

For the data analysis all 6 arrays were analysed together. Only clones that were informative in all 6 arrays were selected: 1500 clones were filtered leaving 10692 common informative clones on all 6 arrays, corresponding to 3014 unique genes.

In W12 P11 cells, 2% (248) of the clones were classified as "heavily methylated", 5% "methylated" (568), 21% "partially methylated" (2215) and 72% "not methylated" (7661) (Figure 5.8a).
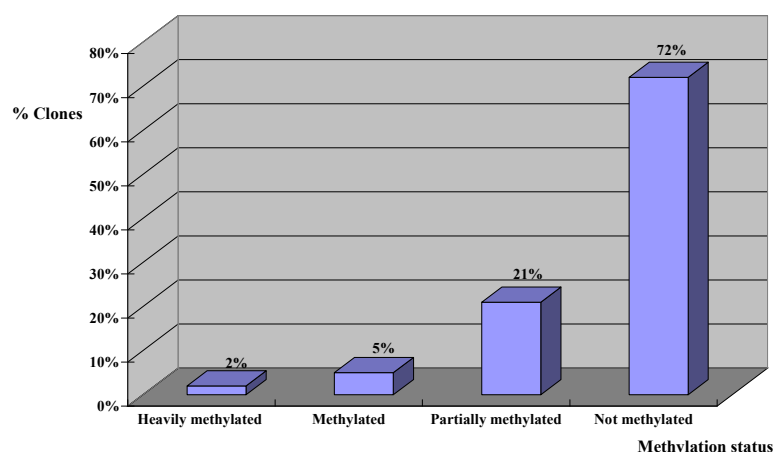
Figure 5.8a Graph showing the distribution of the methylation status of the common informative clones in W12 passage 11 cells.

In W12 P56 cells, 3% (340) of the clones were classified as "heavily methylated", 7% "methylated" (771), 21% "partially methylated" (2277) and 68% "not methylated" (7304) (Figure 5.8b).
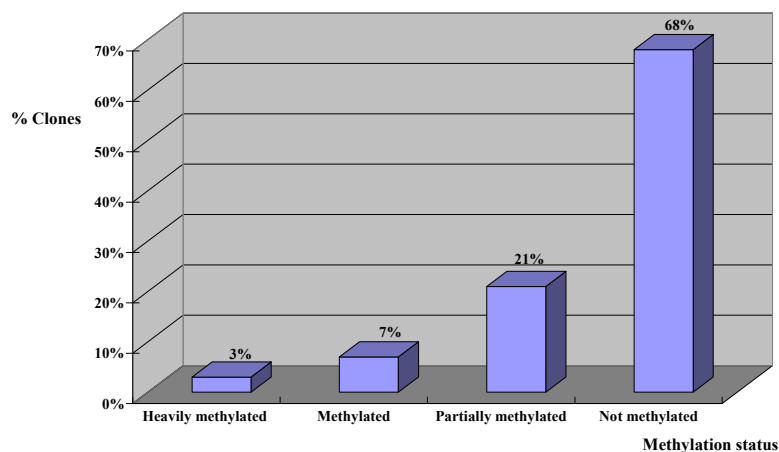
Figure 5.8b Graph showing the distribution of the methylation status of the common informative clones in W12 passage 56 cells.

These graphs show that there was no substantial change in the distribution of the methylation status of the common informative clones between W12 passage 11 and W12 passage 56.

5.4.1 Differentially methylated genes

A gene was considered to be differentially methylated if the cut/mock-cut ratios of the associated clone between W12 passage 11 and 56, were found to be significantly different ($p<0.05$) using the t-test (see section 3.3.9.6). Table 5.10 shows that 108 and 81 genes were significantly more and less methylated respectively, between W12 passages 11 and 56.

Comparison	Change in methylation	No. of clones	No. of genes
	hypermethylated	193	108
W12 P56 vs W12 P11	hypomethylated	145	81
	Both*	-	3

Table 5.10 Number of clones differentially methylated between W12 passages 11 and 56, and their corresponding genes.

*Both = genes identified as both hypermethylated and hypomethylated in non-overlapping regions

5.4.2 Validation of methylation changes using pyrosequencing

Pyrosequencing was used to validate some of the methylation changes observed between W12 passages 11 and 56, using DNA harvested from the same passages used to probe the UHN HCGI12K arrays. All the pyrosequencing analyses were performed twice.

5.4.2.1 Criteria used to select genes for validating methylation changes in the W12 disease progression model:

Of the 108 genes which became hypermethylated between W12 passages 11 and 56, 8 genes were selected for validation (Table A4, Appendix); 4 of these (*BARD1, FANCL, HTRA1, PHOX2B*) were known to have tumour suppressor activity and 4 genes

(*CDH12, DOCK5, PAX2, PPP2CA*) demonstrated the largest change in methylation status between passage 11 and 56.

5.3.2.2 Pyrosequencing results

The sequence provided in the annotation file was used to design pyrosequencing primers (see section 5.3.3.2). The primer sequences used are listed in Table A5 (Appendix).

Of the 8 genes selected for validation, 2 *(BARD1* and *PHOX2B)* were very CpG rich and it was not possible to design any suitable primers. Two more (*CDH12* and *DOCK5*) could not be analysed because no CpG sites were present in the sequence provided in the annotation file. These must be considered false positive results. For 1 gene (*PPP2CA*) pyrosequencing analysis failed despite multiple attempts. Pyrosequencing analysis was successful for the remaining 3 genes. Increased methylation was confirmed in only 1 gene, *PAX2*, at CpG 4 (Table 5.11). For the 2 other genes (*FANCL* and *HTRA1*) an increase in methylation was not demonstrated at any CpG site (Table 5.11). Furthermore, although *FANCL* and *HTRA1* were both predicted by the array to be "not methylated" in W12 P11 cells, pyrosequencing revealed significant methylation at all the CpG sites for both genes.

Gene	Sample	Cut/mock-cut ratio	% Methylation									
			CpG 1	**CpG 2**	**CpG 3**	**CpG 4**	**CpG 5**	**CpG 6**	**CpG 7**	**CpG 8**	**CpG 9**	**CpG 10**
FANCL	W12 P11	1.00	93.0	39.8	62.2	39.5	63.3	63.6	68.2	72.4	66.0	40.3
	W12 P56	0.73	77.9	35.1	36.0	47.2	68.7	67.5	73.7	63.6	62.4	29.3
	MC		97.9	59.1	73.6	74.3	71.4	70.6	77.6	79.9	49.7	30.9
HTRA1	W12 P11	0.89	39.8	33.3	98.1	82.2	86.4					
	W12 P56	0.63	42.8	35.6	97.4	82.3	87.7					
	MC		67.6	66.2	99.3	68.5	95.4					
PAX2	W12 P11	0.96	3.8	44.0	36.4	**7.0**	5.9	5.4	3.4	5.9	3.6	0.0
	W12 P12	-	5.3	41.9	36.4	**6.7**	0.0	5.3	0.0	0.0	0.0	0.0
	W12 P13	-	4.8	44.5	34.4	**7.2**	6.0	5.9	3.8	5.1	4.1	6.1
	W12 P14	-	4.8	44.3	36.2	**9.5**	0.0	4.8	0.0	4.9	5.1	0.0
	W12 P15	-	6.6	45.8	36.2	**9.3**	5.9	6.8	0.0	4.5	0.0	5.2
	W12 P56	0.6	7.8	45.0	33.0	**17.7**	8.7	10.4	7.9	6.3	5.7	0.0
	MC		100.0	87.7	86.0	63.0	87.3	83.1	69.7	83.9	69.7	86.1
			CpG 11	**CpG 12**	**CpG 13**	**CpG 14**	**CpG 15**	**CpG 16**	**CpG 17**	**CpG 18**	**CpG 19**	**CpG 20**	**CpG 21**
	W12 P11	0.96	5.0	3.1	6.0	6.0	4.2	6.4	5.1	3.8	6.2	4.4	4.2
	W12 P56	0.6	4.2	5.9	8.6	12.1	7.6	8.3	14.3	5.9	7.4	4.8	4.9
	MC		100.0	94.6	100.0	94.3	85.7	91.0	93.4	63.0	94.7	82.9	64.6

Table 5.11 Pyrosequencing results for selected candidate genes in the W12 cell line. This table shows the number of CpGs analysed in the selected CpG island for each gene, and the methylation percentage at each CpG site using cells from W12 passage (P) 11 and P56. For the PAX2 gene, methylation of CpGs1-10 (primer set 1) was also analysed in cells taken from W12 P12-15. The corresponding cut/mock-cut ratios using the UHN HCGI12K array are shown. MC, methylated control.

To assess if the observed *de novo* methylation of *PAX2* is related to the loss of episomal forms of HPV16 in the W12 cell line, pyrosequencing of the PAX2 CpG island was also performed in W12 cells from passages 11 to 15 (Table 5.11); I have previously shown that complete loss of viral episomes, with the emergence of integrated forms alone, occurs at passage 12 in this W12 cell line (Figure 5.9).

	P11	P12	P13	P14	P15		P56
PAX2 - CpG 4	7.0%	6.7%	7.2%	9.5%	9.3%		17.7%
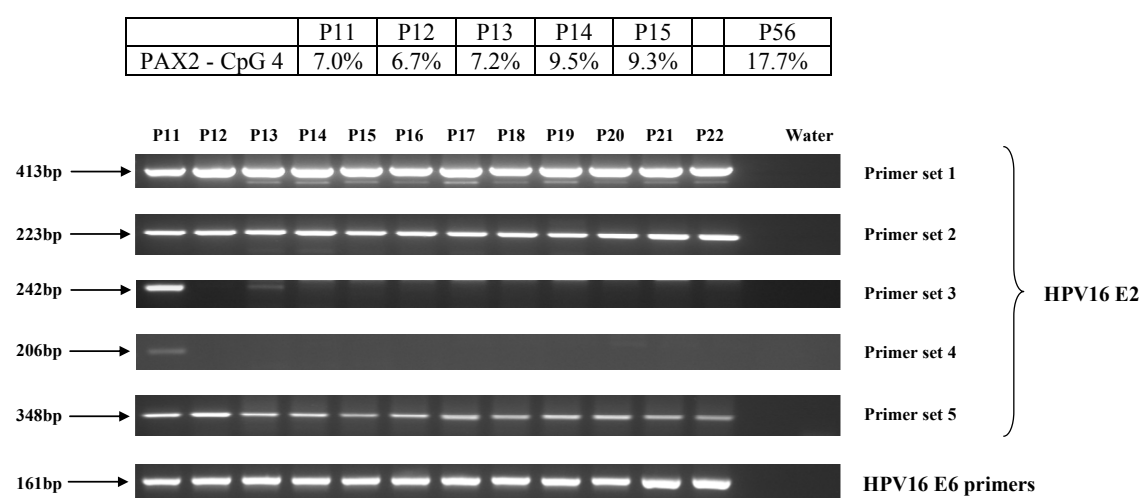
Figure 5.9 Viral integration in the W12 cell line. Five overlapping primers had been designed (by Dr Constandinou-Williams) to cover the HPV16 E2 gene (primer sets 1-5). PCR using DNA from serial passages of the W12 cell line shows that complete loss of HPV16 episomes occurs at passage 12 in this W12 cell line, in the presence of intact HPV16 E6 gene (ie presence of integrated forms only). Water was used as a template-free control for the PCR. The top panel shows for the *PAX2* gene, the percentage methylation of CpG 4 between W12 P11-P15.

Using pyrosequencing, methylation of CpG 4 of the *PAX2* gene did not change substantially between P11 and P15, but was increased at P56. This analysis showed that the increased methylation of *PAX2* in the W12 cell line is a late event, and one that cannot be directly linked to viral integration.

5.4.3 External validation: Comparison with genes known to be methylated in cervical neoplasia

The results of the UHN HCGI12K arrays performed in the W12 cell line were compared with genes previously reported to be methylated in cervical neoplasia. Of the 58 genes, 21 were represented on the UHN HCGI12K arrays in the W12 cell line (post-filtering) (see section 5.2.4).

Change in methylation status		Methylated in cervical neoplasia		Odds ratio (95% CI)
		Yes	No	
Hypermethylated in W12 P56 vs P11	No (n=2909)	20	2889	1 (reference)
	Yes (n=105)	1	104	OR = 1.4 95% CI: 0.2-10.4 p = 0.7485

Table 5.12 Frequency with which genes differentially more methylated in W12 P56 compared to P11 are known to be methylated in cervical neoplasia.

Of these 21 genes, one gene (*CDH13*) was found to be significantly more methylated at P56 compared to P11 (Table 5.12). However, compared with genes that did not become significantly more methylated during long term cultivation of the W12 cell line, genes that did were not significantly enriched for genes previously reported to be methylated in cervical neoplasia.

My observation that only a small number of genes became differentially methylated between W12 passages 11 and 56 could be explained if most of the methylation changes have already taken place by W12 passage 11. To test this hypothesis, I examined how often genes previously reported to be methylated in cervical neoplasia were already methylated at W12 passage 11 (Table 5.13).

Methylation status at p11	Methylated in cervical neoplasia		Odds ratio (95% CI)
	Yes	No	
Not methylated (n=1713)	9	1704	1 (reference)
Methylated (n=1301)	12	1289	OR = 2.0 95% CI: 0.8-4.9 p = 0.1275

Table 5.13 Frequency with which genes known to be methylated in cervical neoplasia were detected in W12 P11 cells, stratified by methylation status.

Table 5.13 shows that of the 21 genes known to be methylated in cervical neoplasia, 12 were already methylated ("partially methylated", "methylated" or "heavily methylated") at W12 passage 11. This analysis revealed that compared to genes that were not methylated at W12 P11, genes that were methylated were twice as likely to be known to be methylated in cervical neoplasia, although this was not statistically significant.

In summary, in this section I have shown that genes which became more methylated following long term cultivation of the W12 cell line were not significantly enriched for genes known to be methylated in cervical neoplasia. This may in part be explained by my observation that many of these genes were already methylated at W12 passage 11. Pyrosequencing analysis was successful for 3 genes, but increased methylation was confirmed in only 1 of these.

5.5 Determinants of virus-associated methylation changes

In this section, I investigate why the methylation status of some cellular genes is changed following transfection with HPV18, and others are not. I consider two hypotheses discussed in the introduction to this thesis. First, that the likelihood of a change in methylation status can be predicted from a gene's CpG content and second, that genes which are reported to be marked by both H3K27me3 and H3K4me3 in human embryonic stem cells (HES) are more likely to be hypermethylated following transfection with this oncogenic virus. The analysis was restricted to genes found to be more methylated in the PHFK-HPV18 cell line only; the W12 cell line was not explored further because only a small number of differentially methylated genes had been identified.

5.5.1 Datasets used in the analyses

Three datasets were used in these analyses:

- The UHN HCGI12K arrays performed in the PHFK-HPV18 cell line (n=2612).

- A CpG content array. Weber *et al* (2007) classified genes as having low, intermediate or high CpG content. The total number of genes on this array was 15,609. Following re-annotation of this dataset 14,899 named genes remained.

- An array describing the distribution of histone marks. Zhao *et al* (2007) used chromatin immunoprecipitation coupled with the paired-end ditags sequencing strategy to profile 2 histone marks, H3K4me3 and H3K27me3, in HES. The total number of genes on this array was 17,468. Following re-annotation of this dataset, 16,950 genes remained.

The final number of genes contributing to the analyses comprised 1885 genes common to all 3 datasets.

5.5.2 Determinants of HPV18-induced methylation changes

A univariate analysis was undertaken to examine if the susceptibility to methylation change in cellular genes was associated with CpG content and also to determine if genes marked by H3K4me3 or H3K27me3 in HES were more likely to be methylated in PHFK following transfection with HPV18 (Table 5.14).

Determinant	Methylation increased (n=728)	Methylation not increased (n=1157)	Odds ratio (95% CI)	Methylation decreased (n=32)	Methylation not decreased (n=1853)	Odds ratio (95% CI)
Low CpG content (n=250)	94	156	1 (reference)	2	248	1 (reference)
Intermediate CpG content (n=231)	81	150	0.9 (0.6-1.3) p = 0.5636	4	227	2.1 (0.4-12.0) p = 0.3577
High CpG content (n=1404)	553	851	1.1 (0.8-1.4) p = 0.5936	26	1378	2.3 (0.6-9.9) p = 0.2349
Not marked by H3K4me3/ H3K27me3 (n=247)	91	156	1 (reference)	3	244	1 (reference)
Marked by H3K4me3 alone (n=1320)	506	814	1.1 (0.8-1.4) p = 0.6578	21	1299	1.3 (0.4-4.4) p = 0.6584
Marked by H3K4me3 & H3K27me3 (n=318)	131	187	1.2 (0.9-1.7) p = 0.2933	8	310	2.2 (0.5-8.0) p = 0.2668

Table 5.14 Distribution of the determinants of the gain and loss of methylation following transfection of PHFK with HPV18 – a univariate analysis.

In the top panel of Table 5.14, the 1885 genes included in the analysis were stratified by CpG content; the CpG content was low in 250 genes, intermediate in 231 genes and high in 1404 genes. The proportion of genes in each category with increased

methylation following transfection with HPV18 was first examined, taking the low CpG content category as the reference (Table 5.14, top left panel). Of the 250 genes with low CpG content, methylation increased in 94/250 (38%). Of the 231 genes with intermediate CpG content, methylation increased in 81/231 (35%). Of the 1404 genes with high CpG content, methylation increased in 553/1404 (39%). Thus there was no association between CpG content and increased methylation. This assessment was repeated for the genes with decreased methylation (Table 5.14, top right panel). These analyses show that CpG content does not significantly increase the likelihood of a gene becoming more or less methylated following transfection of PHFK with HPV18.

In the bottom panel of Table 5.14, the 1885 genes included in the analysis were stratified by whether the genes were known to be marked by neither H3K4me3 nor H3K27me3, by H3K4me3 alone, or by both H3K4me3 and H3K27me3 in human embryonic stem cells; 247 genes were not marked by either H3K4me3 or H3K27me3 in HES, 1320 genes were marked by H3K4me3 alone, and 318 genes were marked by both H3K4me3 and H3K27me3. Next, the proportion of genes in each category with increased methylation following transfection with HPV18 was examined, taking the category of genes not marked by either H3K4me3 or H3K27me3 as the reference (Table 5.14, bottom left panel). Of the 247 genes which were not marked by either H3K4me3 or H3K27me3 in HES, methylation increased in 91/247 (37%). Of the 1320 genes which were marked by H3K4me3 alone in HES, methylation increased in 506/1320 (38%). Of the 318 genes which were marked by both H3K4me4 and H3K27me3 in HES, methylation increased in 131/318 (41%). Thus, being marked by H3K4me3 or H3K27me3 in HES does not increase the likelihood of a gene becoming

more methylated. This analysis was repeated for the genes with decreased methylation (Table 5.14, bottom right panel). These analyses show that being marked by H3K4me3 or H3K27me3 in HES cells does not significantly increase the likelihood of a gene becoming more or less methylated following transfection of PHFK with HPV18.

In summary, in this section I have shown that neither CpG content, nor the presence of H3K4me3 or H3K27me3 marks in HES are predictive of changes in methylation following transfection of PHFK with HPV18.

In this section, I summarise the changes found when gene expression arrays were used to profile the transcriptional changes observed in PHFK following transfection with HPV18, and those found following long term-cultivation of the W12 cell line. I compare these changes with the results of one other study which has examined the transcriptional consequences of HPV18 infection and with a number of published arrays which compare the gene expression profile of tissue taken from women with cervical neoplasia with that taken from disease-free controls. Finally, I explore the extent to which the changes observed on my methylation arrays predicted the transcriptional changes seen on my gene expression arrays.

5.6.1 Transcriptional changes following transfection of PHFK with HPV18

The Affymetrix GeneChip® Human Genome U133 Plus 2.0 arrays were used to assess the transcriptional changes between untransfected PHFK (#1) and PHFK (#1) transfected with HPV18 at passages 4 and 9, as described in section 3.4.4 (Materials and Methods).

Table 5.15 shows the number of probe sets that were differentially expressed or unchanged following transfection of PHFK (#1) with HPV18.

Comparison	No. of probe sets		
	Increased	Decreased	No change
PHFK/HPV18 P4 vs untransfected PHFK	6580	4751	43344
PHFK/HPV18 P9 vs untransfected PHFK	7130	4752	42793
PHFK/HPV18 P9 vs PHFK/HPV18 P4	2187	1917	50571

Table 5.15 Number of probe sets differentially expressed following transfection of PHFK (#1) with HPV18.

Thus it can be seen that most of the probe sets were unchanged following transfection of PHFK with HPV18.

Table 5.16 shows the number of unique genes that were differentially expressed following transfection with HPV18. For genes with more than one probe set, the probe set with the highest baseline expression level (in untransfected PHFK) was selected.

Comparison	No. of genes		
	Increased	Decreased	Both
PHFK/HPV18 P4 vs untransfected PHFK	3158	3171	246
PHFK/HPV18 P9 vs untransfected PHFK	3397	3127	298
PHFK/HPV18 P9 vs PHFK/HPV18 P4	1342	1369	63

Table 5.16 No of unique genes differentially expressed following transfection of PHFK (#1) with HPV18.

There was an additional layer of complexity in that a subset of genes was identified with probe sets that were both up and down-regulated following transfection. For this subset of genes, the probe set for each unique gene with the highest expression value in untransfected keratinocytes, was used to re-allocate the final expression status of these genes following transfection (ie irrespective of the subsequent direction of change). Table 5.17 summarises the final number of genes used for all subsequent analyses.

Comparison	No. of genes	
	Increased	Decreased
PHFK/HPV18 P4 vs untransfected PHFK	3267	3308
PHFK/HPV18 P9 vs untransfectedPHFK	3521	3301
PHFK/HPV18 P9 vs PHFK/HPV18 P4	1370	1404

Table 5.17 No of unique genes differentially expressed following transfection of PHFK (#1) with HPV18 – no overlap.

Most of the genes (>70%) that were up-regulated in PHFK (#1) following transfection with HPV18 were up-regulated at both early and late passages (Figure 5.10).
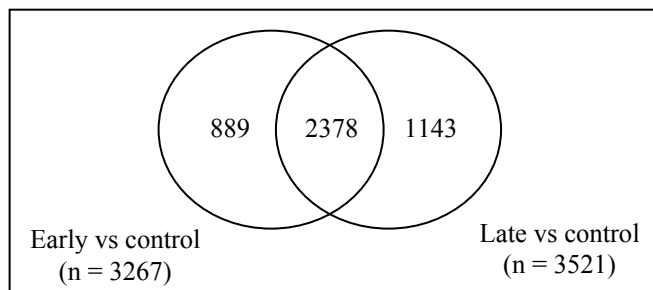
Figure 5.10 Genes up-regulated following transfection of PHFK (#1) with HPV18 at early and late passages.

Similarly, most of the genes (>80%) that were down-regulated in PHFK (#1) following transfection with HPV18, were down-regulated at both early and late passages (Figure 5.11).
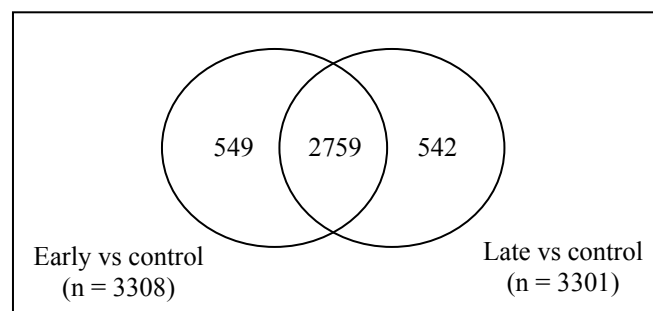
Figure 5.11 Genes down-regulated following transfection of PHFK (#1) with HPV18 at early and late passages.

5.6.1.1 External validation: Comparison with published HPV-PHFK arrays

There has only been one other published study which has compared the transcriptional profile of PHFK transfected with HPV18. Karstensen *et al* (2006) transfected the genome of wild type HPV18 into PHFK from a single donor, following which both untransfected and transfected keratinocytes were cultured on J2 fibroblast feeder cells. The HPV18 remained episomal following transfection. Affymetrix GeneChip® Human Genome U133A arrays were used to compare the gene expression profiles

using RNA harvested from untransfected PHFK (passage 3-5) and PHFK transfected with HPV18 (passage 5-9). Four independent biological replicates were processed and hybridised. Cross-comparisons between all 4 replicates in untransfected and transfected keratinocytes were performed: genes which were called 'Present' and 'Increased' in at least 13 out of 16 comparisons, and which had an average signal log ratio (SLR) of greater than 1 were classified as 'up-regulated', and genes which were called 'Decreased' in at least 13 out of 16 comparisons and an average SLR of -1 were classified as 'down-regulated'. In this analysis, 147 and 82 genes were reported to be up-regulated and down-regulated respectively.

Following re-annotation of this dataset, 145 named genes were found to be up-regulated and 81 remained down-regulated: Of 145 genes up-regulated on the Karstensen array, 60 genes (41%) were also found to be up-regulated on our PHFK-HPV18 array at early passage, 48 (33%) at late passage and 64 (44%) at either early or late passages (Figure 5.12). Of 81 genes down-regulated on the Karstensen array, 50 genes (62%) were also found to be down-regulated on our PHFK-HPV18 array at early passage, 47 (58%) at late passage and 53 (65%) at either early or late passages (Figure 5.13).
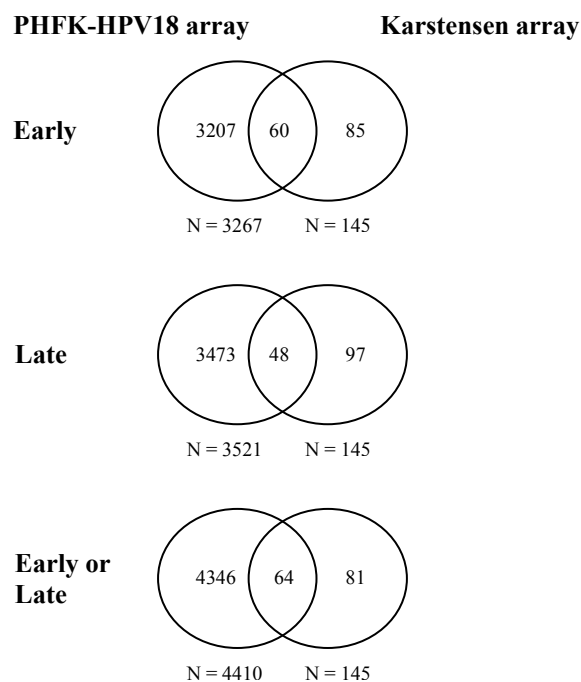
Figure 5.12 Number of genes found to be up-regulated on my PHFK-HPV18 array, the Karstensen array, or both.
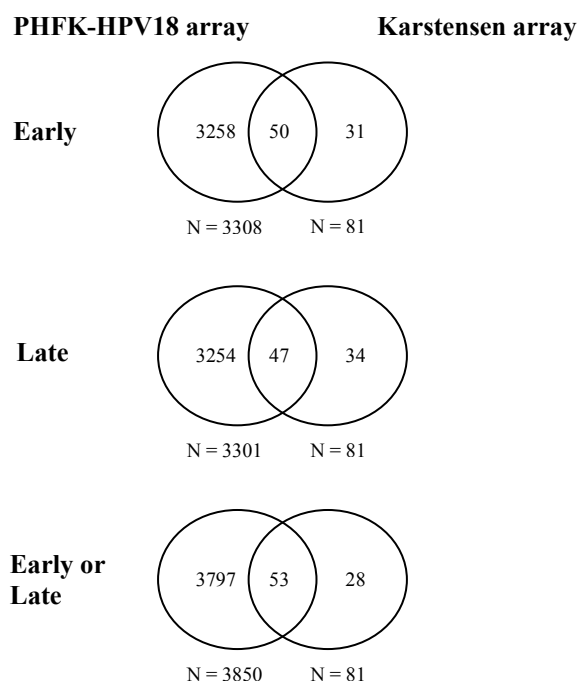
Figure 5.13 Number of genes found to be down-regulated on my PHFK-HPV18 array, the Karstensen array, or both.

Next, I measured the number of genes which were concordantly or discordantly regulated on both arrays (Table 5.18). I defined concordantly regulated genes as those that were either up-regulated in both arrays or down-regulated in both arrays; and discordantly regulated genes as those that were either up-regulated on my PHFK-HPV18 array but down-regulated on the Karstensen array or vice versa. This analysis was repeated for the PHFK-HPV18 array results at early passage, late passage and either early or late passage.

	PHFK/HPV18 P4	PHFK/HPV18 P9	PHFK/HPV18 P4 or P9
Concordant	110 (49%)	95 (42%)	117 (52%)
Discordant	17 (8%)	26 (11%)	28 (12%)

Table 5.18 Comparison between the Karstensen array and my PHFK-HPV18 array: number of concordant and discordantly regulated genes.

This analysis revealed that 42-52% of the differentially regulated genes identified on the Karstensen array were concordantly expressed on my PHFK-HPV18 array, whilst only 8-12% of these were discordantly expressed.

5.6.2 Transcriptional changes in the W12 disease progression model

The Affymetrix GeneChip® Human Genome U133 Plus 2.0 arrays were also used to measure the transcriptional changes that occur during long term cultivation of the W12 cell line, as described in section 3.4.4. This analysis used RNA from passages 11 and 56. Table 5.19 shows the number of probe sets that were differentially expressed or unchanged between these 2 passages.

Comparison	No. of probe sets		
	Increased	Decreased	No change
W12 P56 vs W12 P11	5196	7408	42071

Table 5.19 Number of probe sets differentially expressed between W12 P11 and W12 P56.

Table 5.20 shows the number of unique genes that were differentially expressed during long term cultivation of the W12 cell line. For genes with more than one probe set, the probe set with the highest expression level in W12 P11 cells was selected.

Comparison	No. of genes		
	Increased	Decreased	Both
W12 P56 vs W12 P11	3671	3939	215

Table 5.20 Number of unique genes differentially expressed between W12 P11 and W12 P56.

Again a subset of genes was identified with probe sets that were both up and down-regulated on disease progression. For this subset of genes, the probe set with the highest expression value in P11 cells for each unique gene was used to reallocate the final expression status of these genes following transfection (ie irrespective of the subsequent direction of change). Table 5.21 summarises the final number of genes used for all subsequent analyses.

Comparison	No. of genes	
	Increased	Decreased
W12 P56 vs W12 P11	3560	3835

Table 5.21 Number of unique genes differentially expressed between W12 P11 and W12 P56 – no overlap.

5.6.3 External validation: Comparison with published cervical cancer arrays

The results from the Affymetrix GeneChip® Human Genome U133 Plus 2.0 arrays performed in our 2 *in vitro* models were compared with the results from published arrays that have compared the gene expression profile of cervical cancer or CIN with that of tissue taken from disease-free controls, and which used the same Affymetrix platform as that employed in our experiments: 5 studies were identified (Table A6, Appendix).

Table 5.22 lists for each of these arrays, the numbers of cases and controls, and the criteria used by the authors to identify differentially expressed genes. All the genes from these arrays were re-annotated using the Nov 09 Affymetrix annotation.

First Author	Platform	Samples	Criteria	No. of genes	
				Increased	Decreased
Pyeon	U133 Plus 2.0	8 normal cervices and 20 cervical cancers (Re-analysed in-house by Dr Wei using GEO dataset)	Rank Product with cut-off set at 10% percentage of false positives	1301	1842
Rosty	U133A	5 normal cervices, 5 cell lines, 35 primary tumours including 5 replicates	T-test with 5% false discovery rate	999	N/A
Santin	U133A	11 primary cervical cancer cell lines derived from stage IB-IIA cervical cancers and 4 normal cervical keratinocytes cell lines	FC > 2; Wilcoxon rank-sum p value = 0.009	206	214
Wong	U133A	18 normal cervices and 29 cervical SCC	FC > 2 in at least two-thirds cervical SCC compared to normal cervices	20	81
Zhai	U133A	10 normal cervices and 21 cervical SCC	FC > 1.5; p < 0.001	952	746
Zhai	U133A	7 HSIL and 21 cervical SCC	FC > 1.5; p < 0.001	300	441

Table 5.22 Published cervical cancer arrays: number of cases and controls, criteria used to identify differentially expressed genes, and the numbers of differentially expressed genes.

(FC = fold change)

First, the number of differentially expressed genes in these published cervical cancer arrays that were concordantly regulated in the W12 array, the PHFK-HPV18 array, or both was assessed (Table 5.23)

First Author	Samples	No. of genes differentially expressed on array	No. of genes concordantly expressed in:		
			W12 (P56 vs P11)	PHFK (P9 vs untransfected)	W12 and/or PHFK
Pyeon	8 normal cervices and 20 cervical cancers	3143	764 (24%)	460 (15%)	1106 (35%)
Rosty	5 normal cervices, 5 cell lines, 35 primary tumours including 5 replicates	999	551 (55%)	179 (18%)	652 (65%)
Santin	11 primary cervical cancer cell lines derived from stage IB-IIA cervical cancers and 4 normal cervical keratinocytes cell lines	420	236 (56%)	143 (34%)	321(76%)
Wong	18 normal cervices and 29 cervical SCC	101	41 (41%)	47 (47%)	66 (65%)
Zhai	10 normal cervices and 21 cervical SCC	1698	660 (39%)	544 (32%)	1034 (61%)
Zhai	7 HSIL and 21 cervical SCC	741	246 (33%)	173 (23%)	364 (49%)

Table 5.23 Frequency with which genes that were differentially expressed in the published cervical cancer arrays were concordantly regulated on the W12 array, the PHFK-HPV18 array, or both.

The number of differentially expressed genes on the published cervical cancer arrays ranged from 101 to 3143; this was not surprising given the different criteria used to identify differentially expressed genes (see Table 5.22). It can be seen that 24-56% of the differentially expressed genes on the cervical cancer arrays were concordantly expressed on the W12 array, and 15-47% were concordantly expressed on the PHFK-HPV18 array. Remarkably, 76% of the genes found to be differentially expressed in primary cervical cultures derived from early stage disease when compared with cervical keratinocyte cell lines taken from disease-free controls, were concordantly expressed in the W12 array, the PHFK-HPV18 array or both.

Next, in order to determine if the overlap with the cervical cancer arrays increases with further passaging, the number of differentially expressed genes in the published

cervical cancer arrays that were concordantly expressed at early and late passages of

PHFK (#1) transfected with HPV18 was compared (Table 5.24).

First Author	Samples	No. of genes differentially expressed on array	No. of genes concordantly expressed in:	
			PHFK(#1)/HPV18 P4	PHFK(#1)/HPV18 P9
Pyeon	8 normal cervices and 20 cervical cancers	3143	432 (15%)	460 (15%)
Rosty	5 normal cervices, 5 cell lines, 35 primary tumours including 5 replicates	999	182 (18%)	179 (18%)
Santin	11 primary cervical cancer cell lines derived from stage IB-IIA cervical cancers and 4 normal cervical keratinocytes cell lines	420	159 (38%)	143 (34%)
Wong	18 normal cervices and 29 cervical SCC	101	48 (49%)	47 (47%)
Zhai	10 normal cervices and 21 cervical SCC	1698	571 (34%)	544 (32%)
Zhai	7 HSIL and 21 cervical SCC	741	171 (23%)	173 (23%)

Table 5.24 Frequency with which genes that were differentially expressed in the published cervical cancer arrays were concordantly regulated on the PHFK-HPV18 array: comparison of early and late passages.

The proportion of differentially expressed genes on the cervical cancer arrays that were concordantly regulated at P4 ranged from 15-49%, and at P9 ranged from 15-47%. This analysis shows that the extent to which the transcriptional profile of PHFK transfected with HPV18 mimics the cervical transcriptome does not increase with increasing passages.

5.6.4 Comparison of CpG island and gene expression arrays

Next, I wanted to examine the relationship between the methylation changes and the gene expression changes described above for both the PHFK-HPV18 cell line and the W12 disease progression model.

5.6.4.1 Relationship between methylation and transcriptional changes in PHFK following transfection with HPV18

The subset of genes used for these analyses comprised 2484 genes common to the UHN HCGI12K arrays (post-filtering) and the Affymetrix GeneChip® Human Genome U133 Plus 2.0 arrays performed in the PHFK (#1) following transfection with HPV18.

Tables 5.25a and 5.25b show the methylation and gene expression status of all 2484 genes in PHFK (#1) following transfection with HPV18 at passages 4 and 9, respectively. No change in methylation or gene expression was detected in 45% of the genes at either passage 4 or 9. Of the remaining genes, it can be seen that substantially more genes became hypermethylated than hypomethylated, whilst a similar proportion of genes became up- or down-regulated at these time points.

No. of genes		Gene Expression		
		Increased (n=490)	Decreased (n=490)	Unchanged (n=1504)
Methylation	Increased (n=635)	126	138	371
	Decreased (n=24)	3	5	16
	Unchanged (n=1825)	361	347	1117

Table 5.25a Number of genes differentially methylated or differentially expressed following transfection of PHFK (#1) with HPV18 – passage 4.

No. of genes		Gene Expression		
		Increased (n=515)	Decreased (n=495)	Unchanged (n=1474)
Methylation	Increased (n=526)	103	117	306
	Decreased (n=48)	8	8	32
	Unchanged (n=1910)	404	370	1136

Table 5.25b Number of genes differentially methylated or differentially expressed following transfection of PHFK (#1) with HPV18 – passage 9.

A univariate analysis was then undertaken to examine if there was any correlation between methylation and transcriptional changes in our PHFK-HPV18 cell line (Tables 5.26a and 5.26b).

Methylation change	Expression decreased	Expression not decreased	Odds ratio (95% CI)	Expression increased	Expression not increased	Odds ratio (95% CI)
Methylation not increased (n=1849)	352	1497	1 (reference)	364	1485	1 (reference)
Methylation increased (n=635)	138	497	1.2 (0.9-1.5) p = 0.1409	126	509	1.0 (0.8-1.3) p = 0.9319
Methylation not decreased (n=2460)	485	1975	1 (reference)	487	1973	1 (reference)
Methylation decreased (n=24)	5	19	1.1 (0.4-2.9) p = 0.8910	3	21	0.6 (0.2-1.9) p = 0.3713

Table 5.26a Relationship between methylation and transcriptional changes in PHFK (#1) following transfection with HPV18 - passage 4.

For example, at passage 4 of PHFK (#1) transfected with HPV18, 635 genes were found to have increased methylation and 1849 genes had no increase in methylation (Table 5.26a, top left panel). Of 635 genes with increased methylation, expression was decreased in 138/635 (22%), and of 1849 genes with no increase in methylation,

expression was decreased in 352/1849 (19%). Thus no global association between increased methylation and decreased gene expression was found.

These analyses were repeated for the common subset of 2484 genes at passage 9 of PHFK (#1) transfected with HPV18 (Table 5.26b):

Methylation change	Expression decreased	Expression not decreased	Odds ratio (95% CI)	Expression increased	Expression not increased	Odds ratio (95% CI)
Methylation not increased (n=1958)	378	1580	1 (reference)	412	1546	1 (reference)
Methylation increased (n=526)	117	409	1.2 (0.9-1.5) p =0.1342	103	423	0.9 (0.7-1.2) p =0.4633
Methylation not decreased (n=2436)	487	1949	1 (reference)	507	1929	1 (reference)
Methylation decreased (n=48)	8	40	0.8 (0.4-1.7) p = 0.5679	8	40	0.8 (0.4-1.6) p = 0.4828

Table 5.26b Relationship between methylation and transcriptional changes in PHFK (#1) following transfection with HPV18 - passage 9.

These analyses show that there was no significant association between genes which gain or lose methylation and those with increased or decreased expression, following transfection of PHFK (#1) with HPV18, at early or late passage.

5.6.4.2 Relationship between methylation and transcriptional changes in the W12 cell line

The subset of genes used for these analyses comprised 2877 genes common to the UHN HCGI12K arrays (post-filtering) and the Affymetrix GeneChip® Human Genome U133 Plus 2.0 array performed in the W12 cell line.

Table 5.27 shows the number of differentially methylated and differentially expressed genes in the common subset of 2877 genes. No change in methylation or gene expression was detected in 1494 (52%) genes.

No. of genes		Gene Expression		
		Increased (n=599)	Decreased (n=686)	Unchanged (n=1592)
Methylation	Increased (n=100)	21	22	57
	Decreased (n=76)	19	16	41
	Unchanged (n=2701)	559	648	1494

Table 5.27 Number of genes differentially methylated or differentially expressed between W12 P11 and W12 P56

A univariate analysis was then undertaken to examine if there was any correlation between methylation changes and gene expression changes in the W12 cell line (Table 5.28).

Methylation change	Expression decreased	Expression not decreased	Odds ratio (95% CI)	Expression increased	Expression not increased	Odds ratio (95% CI)
Methylation not increased (n=2777)	664	2113	1 (reference)	578	2199	1 (reference)
Methylation increased (100)	22	78	0.9 (0.6-1.6) p = 0.6595	21	79	1.0 (0.6-1.6) p = 0.9640
Methylation not decreased (n=2801)	670	2131	1 (reference)	580	2221	1 (reference)
Methylation decreased (n=76)	16	60	0.8 (0.5-1.5) p = 0.5627	19	57	1.3 (0.8-2.2) p = 0.3630

Table 5.28 Relationship between methylation and transcriptional changes in the W12 disease progression model.

For example, at W12 P56, 100 genes were found to have increased methylation and 2777 genes had no increase in methylation (Table 5.28, top left panel). Of 100 genes

with increased methylation, expression was decreased in 22/100 (22%), and of 2777 genes with no increase in methylation, expression was decreased in 664/2777 (24%). Thus no association between increased methylation and decreased gene expression was found. These analyses show that there was no significant association between genes which gain or lose methylation and those with increased or decreased expression, following long term cultivation of the W12 cell line.

In summary, I have shown that the transcriptional changes observed in PHFK following transfection with HPV18 are consistent with those previously reported. In addition I have shown that many of the genes found to be differentially expressed in 5 published cervical cancer arrays were concordantly regulated in PHFK following transfection with HPV18, the W12 cell line (the disease progression model), or both. This suggests that our 2 *in vitro* models can provide useful observations relevant to cervical neoplasia. However, no significant relationship between the global methylation and transcriptional changes identified in PHFK (#1) following transfection with HPV18 was observed. Similarly, in the W12 disease progression model, no significant relationship between the global methylation and transcriptional changes between P11 and P56 was identified.

In this section, I investigate using pyrosequencing whether the increased methylation which I observed in the retinoic acid receptor beta gene, *RARB*, in PHFK following transfection with HPV18 was also seen following the transfection of keratinocytes isolated from a different donor with HPV18 and with HPV16. I also explore the methylation status of the *RARB* promoter in two HPV positive and in one HPV negative cervical cancer cell lines. I next explore changes in the expression of *RARB* and of the DNA methyltransferases in PHFK following transfection with HPV18. Finally, I investigate using chromatin immunoprecipitation how the binding of DNMT1 and of H3K27me3 to the *RARB* promoter changes following transfection, having first optimised this assay in HeLa cells.

5.7 Investigation of possible 'epigenetic switching' using a candidate gene approach

In a previous section (5.3.3.2), I confirmed the *de novo* methylation of 3 genes (*NKX6-1, PLXDC2* and *RARB*) in PHFK (#1) following transfection with HPV18. These genes were of interest because they were not expressed and not methylated in untransfected keratinocytes, but they then became methylated for the first time following transfection with HPV18. These genes were therefore potential 'epigenetic switching' genes (see section 1.1.5.3). For 2 genes (*NKX6-1* and *PLXDC2*), *de novo* methylation was observed at only one CpG site, whilst *de novo* methylation was observed at 3 CpG sites in the *RARB* promoter. *RARB* was thus selected for further analysis.

5.7.1 *RARB* is hypermethylated following transfection of a different PHFK donor with HPV18

I first set out to determine if the methylation changes in the *RARB* promoter were specific to the PHFK isolated from donor #1. Pyrosequencing analysis of the *RARB* promoter was performed in untransfected PHFK from a second donor (#2), and in these PHFK (#2) following transfection with HPV18 at passages 4 and 9. These transfection experiments had been performed by Dr Christothea Constandinou-Williams (post-doctoral fellow, our Institute).

Using pyrosequencing, an increase in methylation at CpGs 1-3 was also detected in the *RARB* promoter in PHFK (#2) following transfection with HPV18 at passage 4 (Table 5.29, Figure 5.14b). Methylation of these 3 CpG sites also increased between early and late passage.

5.7.2 *RARB* is hypermethylated following transfection of PHFK with 2 variants of HPV16

I next investigated if these methylation changes were specific to HPV18. Pyrosequencing analysis was performed in PHFK (#2) following transfection with 2 variants of HPV16 – HPV16B and HPV16K. These 2 variants were originally isolated from condylomata acuminate lesions and they differ by one amino acid residue in the L1 major capsid protein (Kirnbauer *et al*, 1993). These transfection experiments had also been performed by Dr Christothea Constandinou-Williams.

Gene	Sample	% Methylation																		
		CpG 1	CpG 2	CpG 3	CpG 4	CpG 5	CpG 6	CpG 7	CpG 8	CpG 9	CpG 10	CpG 11	CpG 12	CpG 13	CpG 15	CpG 16	CpG 17	CpG 18	CpG 19
RARB	Untransfected PHFK #1	**3.7**	**5.1**	**3.8**	3.5	1.9	2.3	3.7	2.4	4.7	2.9	7.0	5.8	**9.9**	7.0	4.1	4.9	**13.0**	**17.1**
	PHFK#1/HPV18 P4	**15.1**	**16.8**	**13.7**	4.9	4.0	3.5	5.0	2.0	4.7	4.6	14.2	7.0	**14.6**	7.6	4.3	5.5	**24.7**	**30.3**
	PHFK#1/HPV18 P9	**19.6**	**22.7**	**17.9**	5.1	3.8	4.7	5.1	4.1	9.1	5.2	14.0	10.5	**19.6**	4.5	3.4	7.0	**22.2**	**34.0**
	MC	94.8	97.2	95.7	93.1	96.4	60.1	90.4	77.6	92.6	95.2	100.0	91.0	97.0	100.0	100.0	100.0	95.9	91.4
	Untransfected PHFK #2	**10.2**	**12.9**	**9.6**	6.5	3.1	5.6	4.8	4.4	6.3	0.0	5.0	6.6	9.9	8.6	6.0	4.9	13.7	18.4
	PHFK#2/HPV18 P4	**14.6**	**18.0**	**14.2**	5.8	5.1	8.9	3.4	5.0	5.5	2.9	7.0	5.9	11.7	7.9	5.5	6.0	16.0	20.5
	PHFK#2/HPV18 P9	**20.9**	**23.0**	**19.0**	7.9	3.3	9.1	5.3	5.7	6.4	3.7	7.7	4.1	10.9	6.8	5.7	4.5	15.2	17.0
	Untransfected PHFK #2	**10.2**	**12.9**	**9.6**	6.5	3.1	5.6	4.8	4.4	6.3	0.0	5.0	6.6	**9.9**	8.6	6.0	4.9	13.7	**18.4**
	PHFK#2/HPV16B P4	**19.9**	**25.5**	**19.3**	6.3	4.0	7.4	3.5	5.7	6.2	3.5	8.6	5.9	**17.6**	5.6	4.6	4.7	9.0	**30.6**
	PHFK#2/HPV16B P9	12.1	12.4	10.9	5.4	0.0	6.6	3.3	4.1	7.0	5.0	13.2	7.4	**25.8**	6.0	4.5	3.2	10.8	**32.7**
	Untransfected PHFK #2	**10.2**	**12.9**	**9.6**	6.5	3.1	5.6	4.8	4.4	6.3	0.0	**5.0**	**6.6**	**9.9**	8.6	6.0	4.9	**13.7**	**18.4**
	PHFK#2/HPV16K P4	**17.3**	**19.1**	**14.7**	11.4	4.4	6.3	5.5	3.9	3.2	0.0	**7.5**	**6.1**	**13.2**	6.1	5.0	4.8	**12.9**	**13.3**
	PHFK#2/HPV16K P9	12.4	13.3	11.7	5.3	4.1	8.6	8.2	5.5	7.4	5.1	**30.9**	**14.4**	**22.6**	11.6	5.7	5.6	**25.7**	**29.0**
	CaSki	**93.9**	**96.9**	**43.9**	**67.2**	**80.1**	**79.4**	**94.1**	**83.5**	**94.8**	**92.4**	**98.4**	**91.2**	**95.4**	-	-	-	**83.0**	**67.1**
	HeLa	**87.1**	**88.6**	**31.2**	**17.9**	**19.5**	1.9	**25.6**	7.2	9.6	**39.6**	**93.1**	**22.4**	**72.4**	-	-	-	**61.9**	**77.0**
	C33A	8.1	4.0	4.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	5.8

Table 5.29 Pyrosequencing analysis of *RARB*. This table shows the methylation percentage at each CpG site in the *RARB* promoter. Pyrosequencing was performed in PHFK from 2 different donors, transfected with HPV18 and with 2 variants of HPV16 (B and K), and also in 3 cervical cancer cell lines; CaSki (HPV16 positive), HeLa (HPV18 positive) and C33A (HPV negative). MC, methylated control**.**

Using pyrosequencing, *de novo* methylation of CpGs 1-3 was again detected at early passage, but unlike the 2 PHFK isolates transfected with HPV18, methylation of CpGs 1-3 decreased between passages 4 and 9 (Table 5.29, Figures 5.14c and 5.14d). These changes were common to both the PHFK-HPV16B and the PHFK-HPV16K cell lines.

In addition, an increase in methylation of a further 2 CpGs (13 and 19) was seen in both HPV16 cell lines. This was also detected with PHFK (#1) following transfection with HPV18 but not with the second donor. Additional increases in methylation were also detected in CpGs 11, 12 and 18 for the PHFK-HPV16K cell line alone.

These results show that methylation of CpGs 1-3 of the *RARB* promoter is commonly observed following transfection with HPV, irrespective of the PHFK isolate or the infecting high-risk HPV type.

5.7.3 *RARB* is hypermethylated in HPV positive cervical cell lines

Pyrosequencing analysis of the *RARB* promoter was also performed using DNA from 3 cervical cancer cell lines: CaSki, a HPV16-positive cell line; HeLa, a HPV18-positive cell line; and C33A, a cervical cancer cell line that does not contain HPV. Using pyrosequencing, no methylation was detected at any CpG sites in the *RARB* promoter in the C33A cell line (Table 5.29). Conversely, in the 2 HPV-positive cell lines (CaSki and HeLa), most of the CpG sites in the *RARB* promoter were found to be heavily methylated (Table 5.29). These results are consistent with a role for HPV-associated methylation changes in the *RARB* promoter.
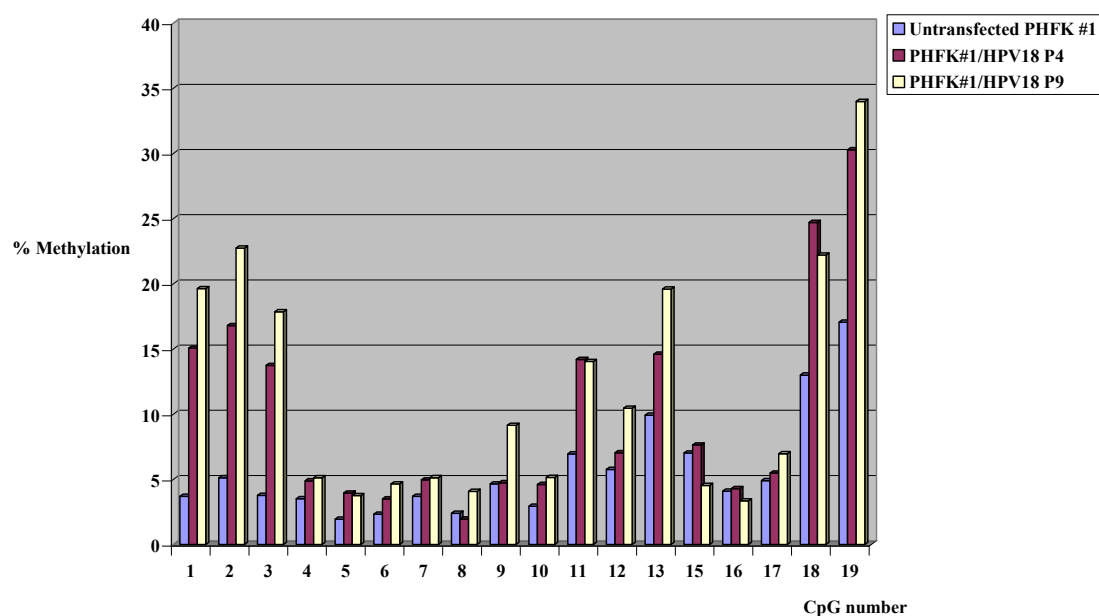
Figure 5.14a *RARB* **pyrosequencing analysis in untransfected PHFK (#1), and in PHFK (#1) following transfection with HPV18 at early and late passages.**
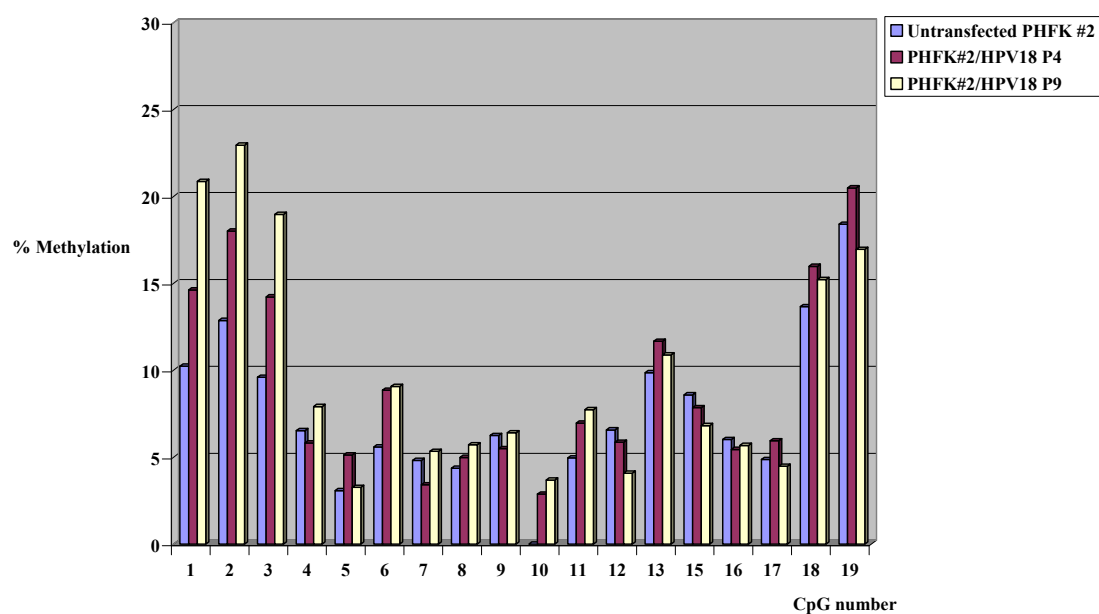
Figure 5.14b *RARB* **pyrosequencing analysis in untransfected PHFK (#2), and in PHFK (#2) following transfection with HPV18 at early and late passages.**
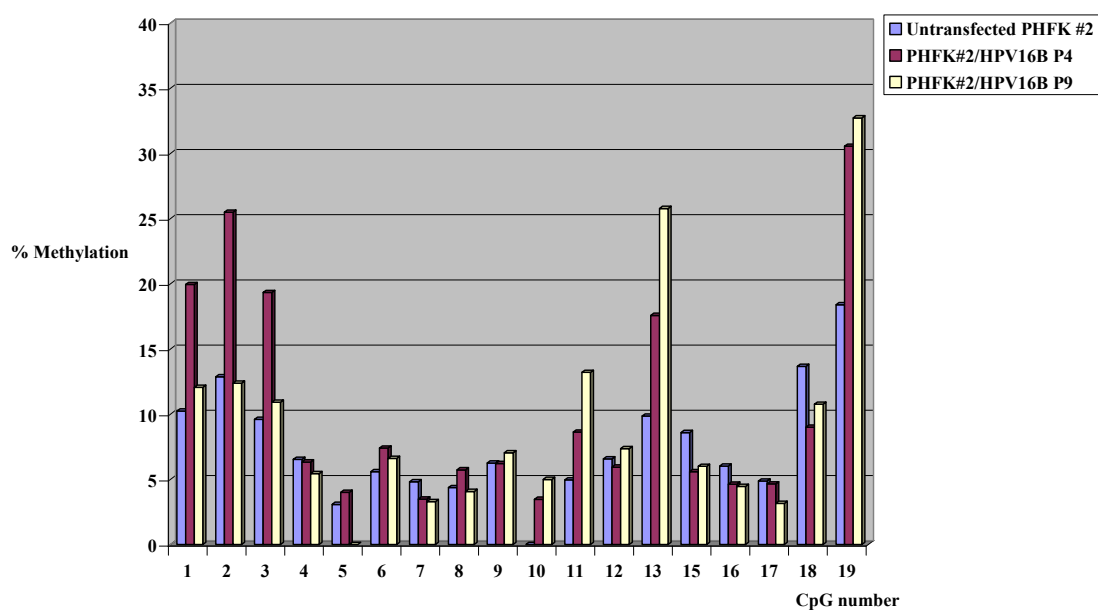
Figure 5.14c *RARB* **pyrosequencing analysis in untransfected PHFK (#2), and in PHFK (#2) following transfection with HPV16B at early and late passages.**
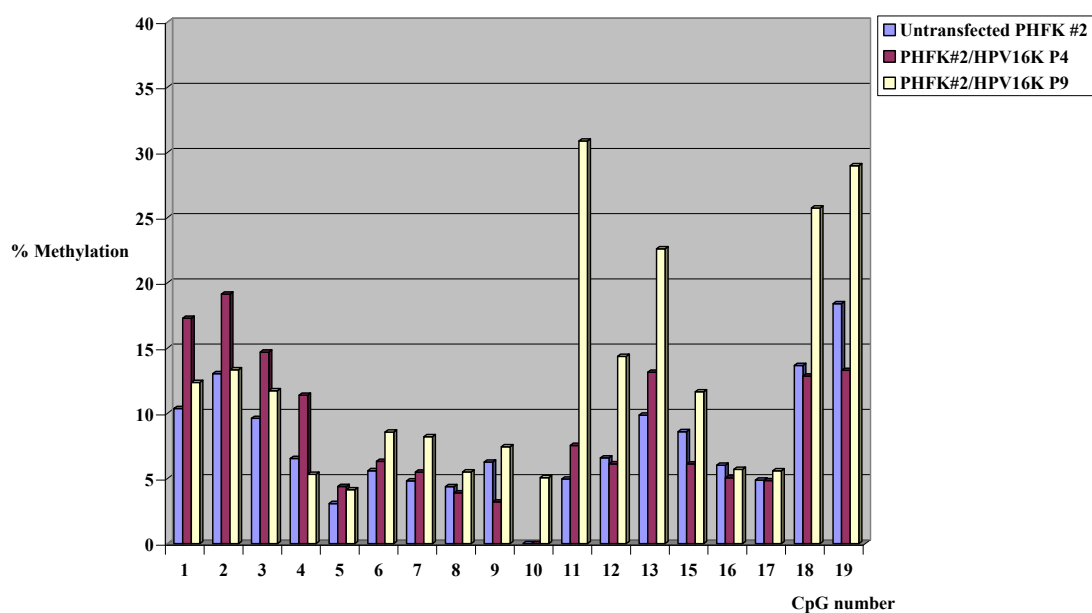
Figure 5.14d *RARB* **pyrosequencing analysis in untransfected PHFK (#2), and in PHFK (#2) following transfection with HPV16K at early and late passages.**

5.7.4 *RARB* is not expressed in PHFK (#1) before and after transfection with HPV18

The Affymetrix GeneChip® Human Genome U133 Plus 2.0 arrays performed in PHFK (#1) transfected with HPV18 showed that the expression of *RARB* was very low in untransfected cells, and this expression remained very low following transfection with HPV18 at early and late passage (Figure 5.15).
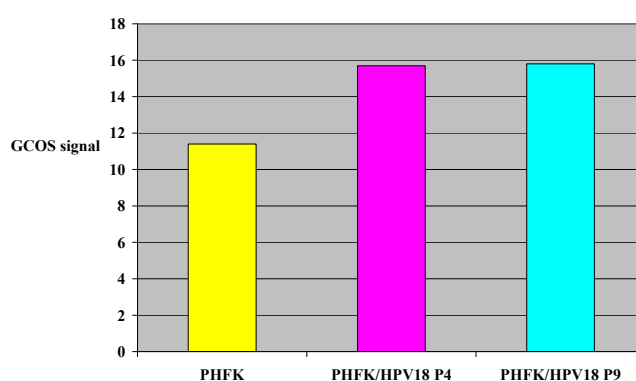
Figure 5.15 Expression of *RARB* in PHFK. The GCOS signal for *RARB* in untransfected PHFK (donor #1) is very low, and remains very low following transfection of PHFK (from the same donor) with HPV18 at early (P4) and late (P9) passage.

Analysis of *RARB* expression was performed using RNA harvested from the same passages used to probe the expression arrays. Using RT-PCR, the expression of *RARB* was found to be undetectable in untransfected PHFK, and remained undetectable following transfection with HPV18 at passages 4 and 9 (Figure 5.16), confirming the expression array findings.
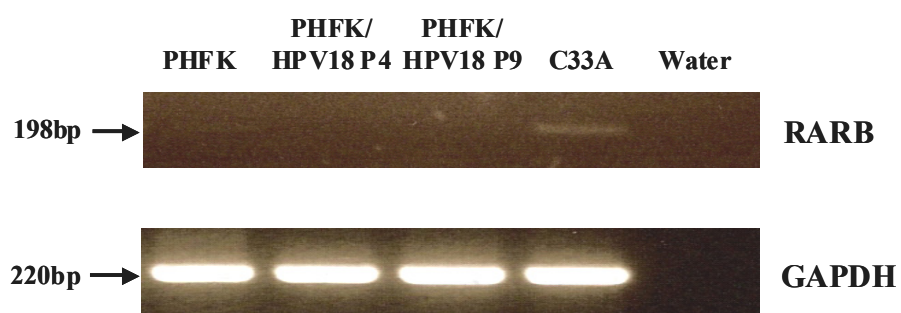
Figure 5.16 RT-PCR for *RARB* in PHFK. The expression of *RARB* was found to be undetectable in PHFK (from donor #1), and remained unchanged following transfection of PHFK (from the same donor) with HPV18, at early (P4) and late (P9) passage. C33A cervical cell line DNA was used as a positive control. GAPDH was used as a house-keeping gene.

5.7.5 Expression of DNA methyltransferases in PHFK following transfection with HPV18

The Affymetrix GeneChip® Human Genome U133 Plus 2.0 arrays performed in PHFK (#1) transfected with HPV18 showed that DNA methyltransferase 1 (DNMT1) was expressed in untransfected cells, that this expression increased significantly following transfection with HPV18 at early passage, but that there was no further significant change in expression between early and late passage (Figure 5.17a). The array also showed that the expression of DNMT3A and 3B was low (GCOS signals <100) in untransfected PHFK (#1), and that this did not change following transfection with HPV18 at early or late passages (Figures 5.17b and 5.17c).

Q-RT-PCR analysis of DNA methyltransferase expression was performed using RNA harvested from the same passages used to probe the expression arrays. By this analysis, DNMT1 was shown to be up-regulated (5-fold) in PHFK (#1) following transfection with HPV18 at early passage, but down-regulated between early and late passages (Figure 5.18a). For DNMT3A and 3B, no significant change in the expression of both genes was detected following transfection with HPV18 at early or late passages (Figures 5.18b and 5.18c), using Q-RT-PCR.

Western blotting for DNMT1 was undertaken to determine if the differential gene expression of DNMT1 observed in PHFK (#1) following transfection with HPV18 was translated to the protein level. This analysis was performed 3 times.
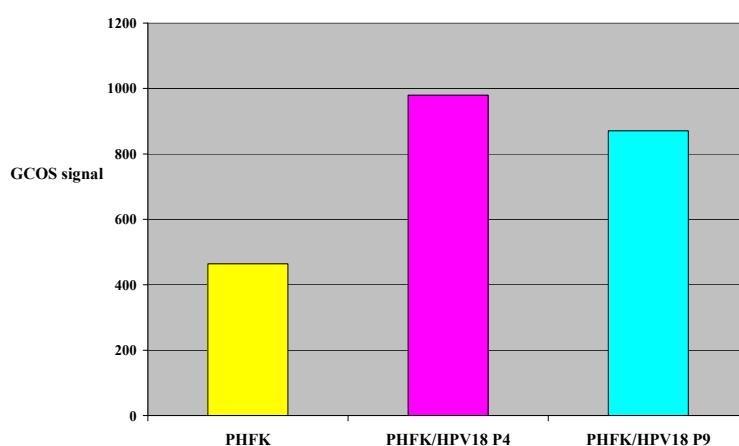
Figure 5.17a Expression of DNMT1 in PHFK. GCOS signals for DNMT1 in untransfected PHFK (from donor #1), and following transfection of PHFK (from the same donor) with HPV18 at early (P4) and late (P9) passage.
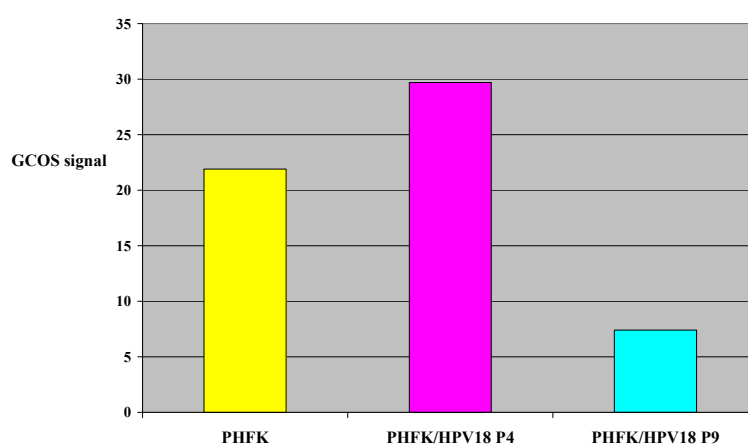
Figure 5.17b Expression of DNMT3A in PHFK. GCOS signals for DNMT3A in untransfected PHFK (from donor #1), and following transfection of PHFK (from the same donor) with HPV18 at early (P4) and late (P9) passage.
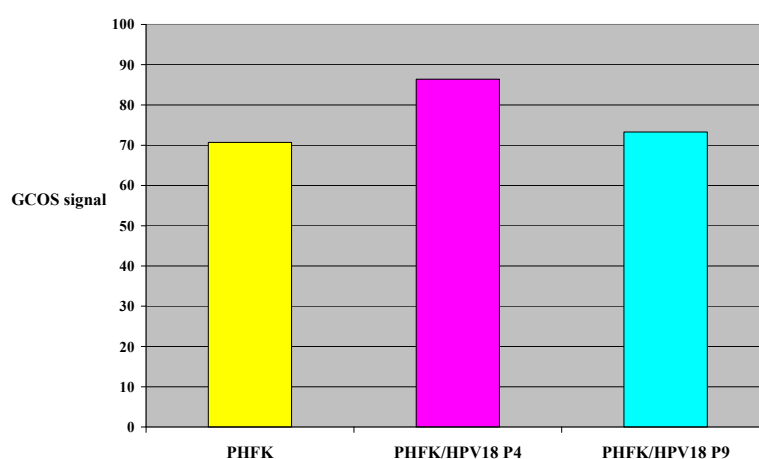
Figure 5.17c Expression of DNMT3B in PHFK. GCOS signals for DNMT3B in untransfected PHFK (from donor #1), and following transfection of PHFK (from the same donor) with HPV18 at early (P4) and late (P9) passage.

211
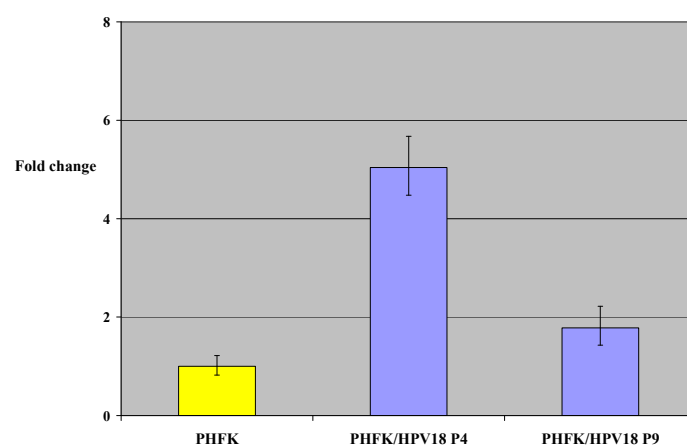
Figure 5.18a DNMT1 Q-RT-PCR in PHFK transfected with HPV18. Relative expression of DNMT1 in PHFK (from donor #1) transfected with HPV18 at early (P4) and late (P9) passage, compared to in untransfected PHFK (from the same donor).
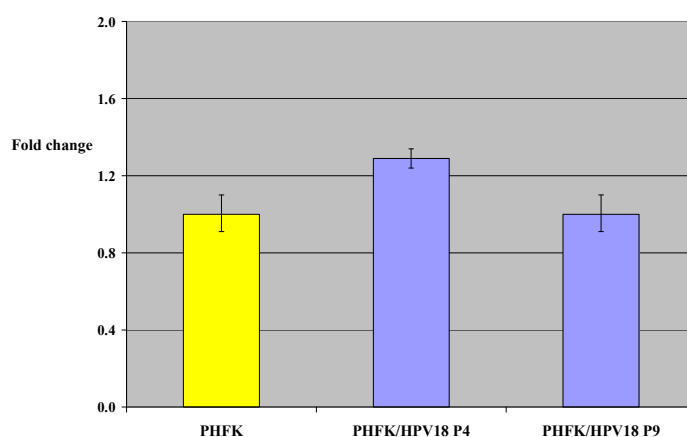
Figure 5.18b DNMT3A Q-RT-PCR in PHFK transfected with HPV18. Relative expression of DNMT3A in PHFK (from donor #1) transfected with HPV18 at early (P4) and late (P9) passage, compared to in untransfected PHFK (from the same donor).
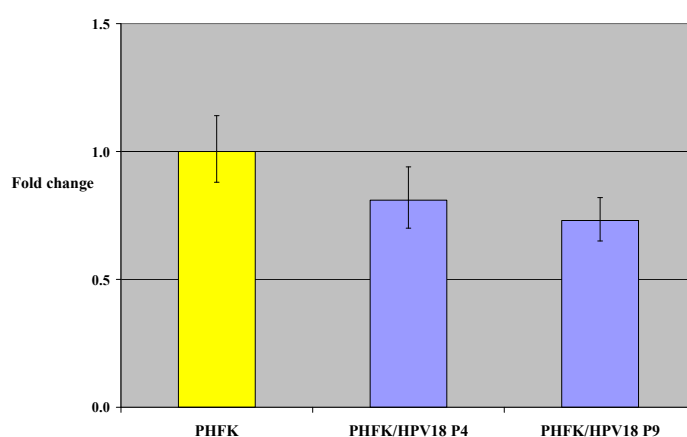
Figure 5.18c DNMT3B Q-RT-PCR in PHFK transfected with HPV18. Relative expression of DNMT3B in PHFK (from donor #1) transfected with HPV18 at early (P4) and late (P9) passage, compared to in untransfected PHFK (from the same donor).

PHFK   PHFK/HPV18   PHFK/HPV18
Passage 4   Passage 9

183kDa ⟶   **DNMT1**

42kDa ⟶   **β-Actin**
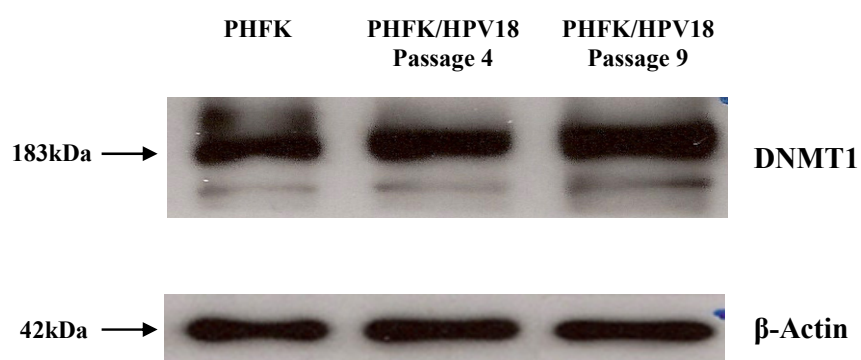
Figure 5.19 Western blotting for DNMT1 in PHFK. The DNMT1 protein increases

following transfection of PHFK (from donor #1) with HPV18 at early passage (P4), with

further increased expression at late passage (P9). β-actin was used as a loading control.

Using western blotting, an increase in the expression of DNMT1 protein was seen following transfection of PHFK with HPV18 at early passage, with a further increase in expression observed between early and late passages (Figure 5.19).

5.7.6 Chromatin immunoprecipitation

Using RT-PCR, I have shown that the expression of *RARB* is undetectable in untransfected PHFK (#1). However, I have also shown using pyrosequencing, that the *RARB* promoter is not methylated in untransfected PHFK (#1). Following transfection of PHFK (#1) with HPV18, the expression of RARB remained undetectable. However pyrosequencing demonstrated *de novo* methylation of the *RARB* promoter, which increased with increasing passages. As I had also shown, using Q-RT-PCR and western blotting, that the expression of DNMT1 increased following transfection of PHFK (#1) with HPV18, I therefore set out to determine if these changes were a consequence of epigenetic switching (ie if *RARB* was initially silenced by H3K27me3, but then acquires DNA methylation silencing with loss of the H3K27me3 mark).

To determine if DNMT1 or H3K27me3 were bound to the *RARB* promoter in untransfected PHFK (#1) and if there was an increase in DNMT1 binding in these PHFK following transfection with HPV18, chromatin immunoprecipitation (ChIP) assays using antibodies specific to DNMT1 and H3K27me3 were performed. To preclude non-specific enrichment by ChIP, the IgG antibody was used as a negative control, and a no antibody control was also included in each experiment.

The ChIP assay was first optimised in HeLa cells. Wu *et al* (2008) have previously demonstrated binding of both DNMT1 and H3K27me3 proteins to the HOXA7 upstream regulatory region in HeLa cells. This enrichment was specific to the promoter region (in Region E). No similar enrichment was seen in Region I, which is located between exon 1 and exon 2 (Wu *et al*, 2008). Using antibodies specific to DNMT1 and H3K27me3, I performed the ChIP assay twice in HeLa cells. Figure 5.20 shows enrichment of both DNMT1 and H3K27me3 proteins in Region E of the HOXA7 gene, with no binding observed in Region I, thus reproducing the results from the paper published by Wu *et al*. The enrichment of DNMT1 and H3K27me3 was confirmed using quantitative PCR (Q-PCR) (Figure 5.21). This ChIP assay was thus taken forward for use in PHFK cells.
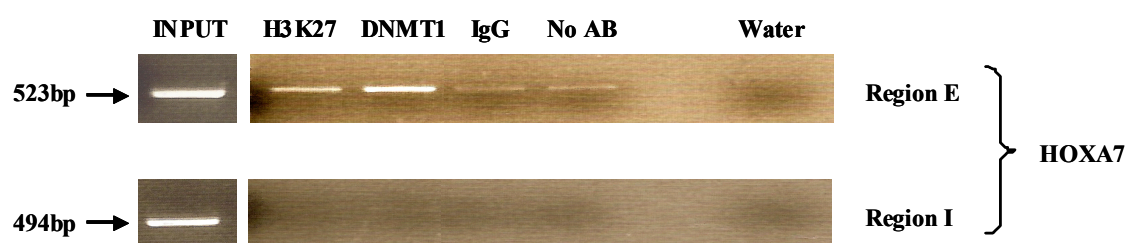
Figure 5.20 ChIP analysis in HeLa cells. PCR for 30 cycles reveals enrichment of the H3K27me3 and DNMT1 proteins in region E of the HOXA7 promoter, relative to IgG (non-specific antibody control), and the no antibody control (*No AB*). Conversely, in region I (located between exon 1 and exon 2 of the HOXA7 gene), no enrichment was observed. Water was used as a template-free control for the PCR.
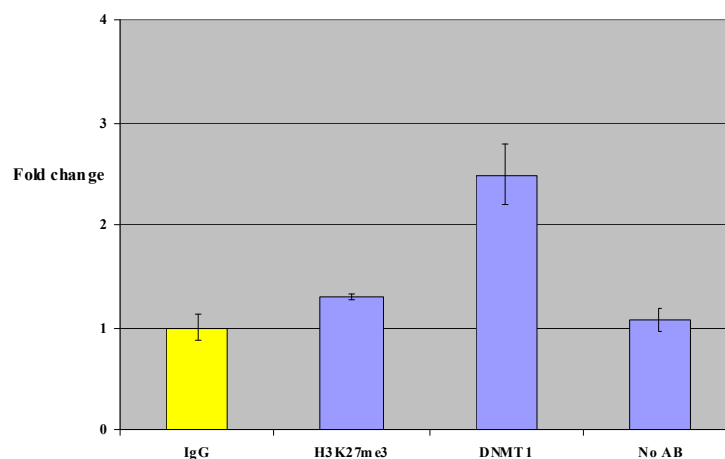
Figure 5.21 ChIP analysis in HeLa cells – HOXA7. Q-PCR confirms enrichment of the H3K27me3 and DNMT1 proteins in region E of the HOXA7 promoter, relative to IgG (non-specific antibody control). No AB, no antibody control.

The ChIP assay was next performed in untransfected PHFK (#1) and in PHFK (#1) transfected with HPV18 at early passage (P4) using antibodies specific to DNMT1 and H3K27me3. In untransfected PHFK (#1), a 2.5 fold increase of the DNMT1 protein at the RARB promoter relative to the IgG antibody negative control was detected, whilst there was no enrichment of the H3K27me3 protein (Figure 5.22a).
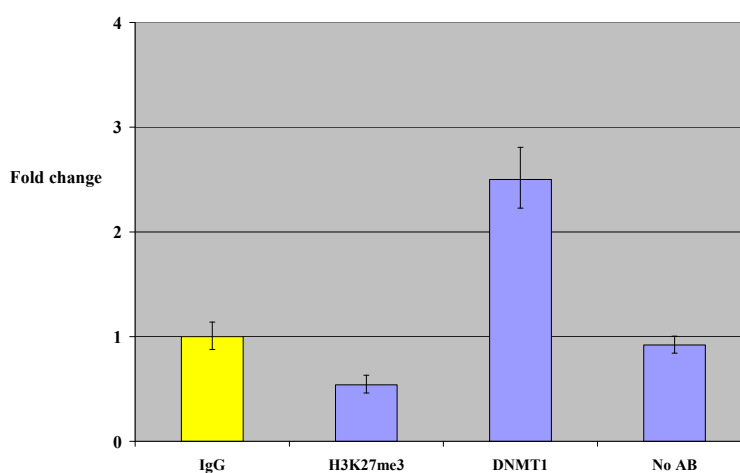
Figure 5.22a ChIP analysis in untransfected PHFK (#1) – RARB. Q-PCR reveals a 2.5 fold increase of the DNMT1 protein at the *RARB* promoter relative to IgG (non-specific antibody control), but no enrichment of the H3K27me3 protein. No AB, no antibody control.

Following transfection of PHFK (#1) with HPV18, there was a marked enrichment of both the DNMT1 (7.5 fold) and H3K27me3 (8-fold) proteins relative to the IgG antibody negative control (Figure 22b).
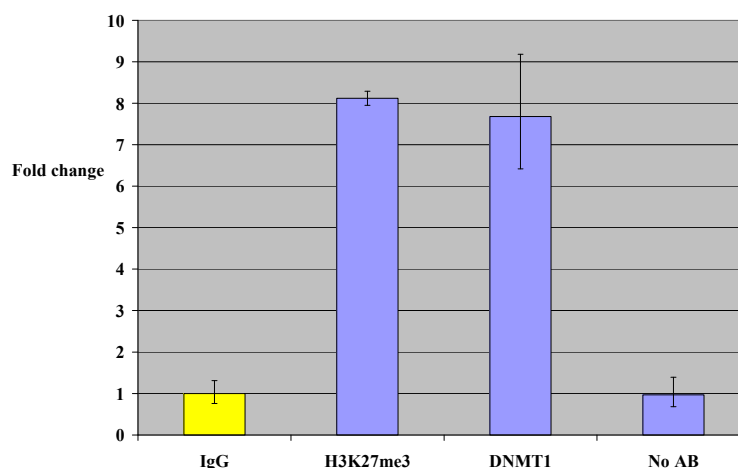
Figure 5.22b ChIP analysis in PHFK (#1) transfected with HPV18 at passage 4 – RARB. Q-PCR reveals enrichment of the H3K27me3 (8 fold) and DNMT1 (7.5 fold) proteins at the RARB promoter relative to IgG (non-specific antibody control). No AB, no antibody control.

The enrichment observed in each case was specific to the *RARB* promoter: no enrichment was observed in the downstream region of *RARB* (region D, located 2.5kb downstream from the end of the first exon) for either protein in untransfected PHFK (#1) or in keratinocytes transfected with HPV18 (Figure 5.23). This suggests that the increased binding of DNMT1 following transfection of PHFK (#1) with HPV18 may play a role in the observed *de novo* methylation in these cells.
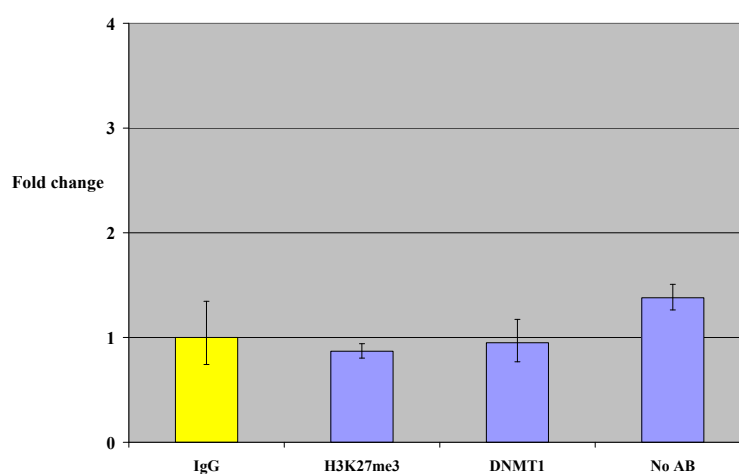
Figure 5.23a ChIP analysis in untransfected PHFK (#1) – RARB region D. Q-PCR reveals no enrichment of the H3K27me3 or DNMT1 proteins in *RARB* region D (located 2.5kb downstream from the end of the first exon), relative to IgG (non-specific antibody control). No AB, no antibody control.
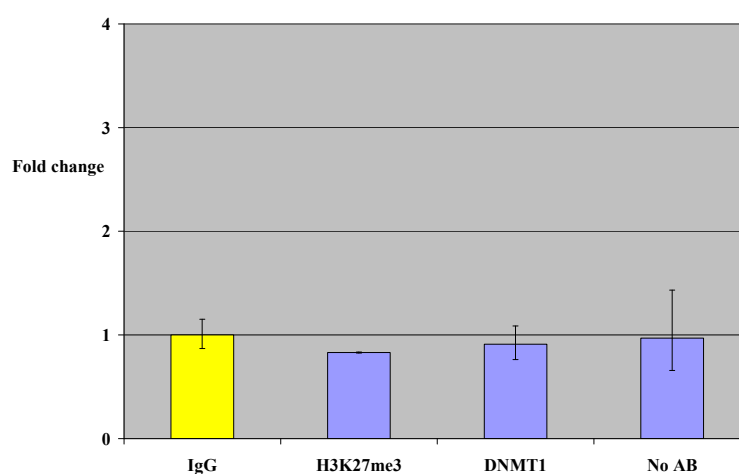
Figure 5.23b ChIP analysis in PHFK (#1) transfected with HPV18 at passage 4 – RARB region D. Q-PCR reveals no enrichment of the H3K27me3 or DNMT1 proteins in *RARB* region D (located 2.5kb downstream from the end of the first exon), relative to IgG (non-specific antibody control). No AB, no antibody control.

In summary, in this section I have shown that *de novo* methylation of the *RARB* promoter occurs in PHFK following transfection with both HPV16 and HPV18, irrespective of the PHFK isolate. I have also shown that methylation of the *RARB* promoter is high in 2 HPV-positive cell lines, but is absent in C33A, a HPV-negative cervical cancer cell line. In addition, I have demonstrated up-regulation of DNMT1 in PHFK transfected with HPV18. Finally, I have observed increased binding of DNMT1 and H3K27me3 at the *RARB* promoter following transfection with HPV18.

Chapter 6

DISCUSSION

In this chapter I begin by discussing my smoking and *CDKN2A* methylation results, before moving on to discuss the results of my work looking at the contribution of HPV infection to the methylation of cellular genes.

6.1 Association between smoking and CDKN2A methylation

Using the cervix uteri as a model, I have shown that smoking increases the risk of acquiring methylated forms of the tumour suppressor gene, *CDKN2A*, the epigenetic inactivation of which is frequently associated with the pathogenesis of neoplasia at many sites (Herman *et al*, 1995; Merlo *et al*, 1995). The *CDKN2A* gene product binds to and inhibits the cyclin dependent kinases 4 and 6, which normally phosphorylate the serine/threonine residues of the retinoblastoma (Rb) protein (Serrano *et al*, 1993; Weinberg, 1995). Thus, inactivation of *CDKN2A* enables initial phosphorylation of the Rb protein and subsequent progression of the cell cycle through G1 into S phase (Serrano *et al*, 1993; Weinberg, 1995). Our choice of study design and our unique study population allowed me to reveal, for the first time, the relationship between an incident exposure (smoking initiation) and the subsequent appearance of an epigenetic change (methylation of *CDKN2A*).

These findings are consistent with those of a cross-sectional study reporting a non-significant two-fold excess of methylated forms of *CDKN2A* in normal cervical smears taken from smokers, compared with those taken from non-smokers (Lea *et al*, 2004); the higher prevalence of methylated forms of *CDKN2A* observed in my study may be because I used nested MSP, a more sensitive assay (Palmisano *et al*, 2000). Compared with never-smokers, current smokers who are cancer-free, but not

necessarily disease-free, are also reported to have a higher prevalence of methylated forms of the tumour suppressor genes fragile histidine triad (*FHIT*) and O-6-methylguanine-DNA methyltransferase (*MGMT*) in bronchial lavages and plasma, respectively (Kim *et al*, 2004; Belinsky *et al*, 2005).

I have also shown that loss of *CDKN2A* methylation can occur following smoking cessation. Consistent with this finding, *in vitro* and animal studies have shown that cigarette smoke and nicotine-induced methylation changes in TSG can be reversed following withdrawal of the exposure (Watson *et al*, 2004; Bachman *et al*, 2006; Soma *et al*, 2006; Liu *et al*, 2007). For example, in human oesophageal squamous epithelial cells, short term exposure to nicotine was sufficient to induce methylation of *FHIT*, with attenuation of *FHIT* protein expression; methylation of *FHIT* persisted throughout moderate to long-term nicotine exposure (Soma *et al*, 2006). Following cessation of nicotine exposure, irrespective of the duration of exposure, demethylation of *FHIT* was observed in association with re-expression of the FHIT protein (Soma *et al*, 2006).

6.2 Mechanism of smoking-induced methylation changes

The mechanism of smoking-induced methylation changes remains incompletely understood. In 1996, Belinsky *et al* published the first report demonstrating an increase in DNA methyltransferase activity and gene expression in A/J mice exposed to the tobacco specific carcinogen NNK, in association with a marked increase in overall methylation. Nicotine-induced methylation of *FHIT* in association with increased DNMT3A expression, has also been demonstrated in human oesophageal

squamous epithelial cells (Soma *et al*, 2006). In normal human liver tissue samples, expression of DNA methyltransferase was found to be significantly higher in smokers compared with non-smokers (Hammons *et al*, 1999). In patients with lung cancer, DNMT1 expression was also found to be significantly higher in smokers compared with never smokers, and in those with adenocarcinoma histology, patients who have smoked more than 65 pack-years had a 4.17 increased risk of increased DNMT1 expression compared to those with a less than 45 pack-year smoking history (Kwon *et al*, 2007). More recently, it was found that the DNMT1, 3A and 3B proteins were highly expressed in a coordinate manner in 100 lung cancer samples (independent of cell proliferation), particularly in smokers (Lin *et al*, 2007). Furthermore, the increased expression of DNMT1 and DNMT3B was significantly associated with hypermethylation of the tumour suppressor genes, *FHIT*, *CDKN2A* and *RARB,* and using tissue chromatin-immunoprecipitation in a subset of samples, strong binding of DNMT1 and DNMT3B to the methylated promoters of these 3 TSG was observed (Lin *et al*, 2007). Thus exposure to tobacco carcinogens induces increased expression of the DNMTs which results in increased DNA methylation.

However, rather than a general increase in promoter methylation across the genome, several lines of evidence demonstrate that smoking-induced DNA methylation is both gene- and tissue-specific. Using murine models, Belinsky *et al* (1996) examined DNA methyltransferase activity in different histological cell types of murine lungs before and after exposure to NNK; increased DNA methyltransferase activity was observed only in alveolar type II cells, which are thought to be the target cells leading to lung tumour formation, in association with an increase in DNA methylation level, whilst

no increase in DNA methyltransferase activity or DNA methylation was observed in the Clara (non-target) cells. In patients with lung cancer, DNMT1 expression was found to be significantly higher in adenocarcinomas compared with squamous cell carcinomas, while the average number of pack-years smoked was higher in squamous cell carcinomas than in adenocarcinomas, thus suggesting that the relationship between smoking and DNMT1 expression is tissue-specific (Kwon *et al*, 2007). Now, the p16 and p14ARF proteins are both encoded from the same locus on chromosome 9p21 by alternative splicing. In 185 primary lung cancers, frequent methylation of *p16* but no methylation of *p14ARF* was detected, and the prevalence of *p16* methylation was significantly higher in squamous cell carcinomas compared with adenocarcinomas (Kim *et al*, 2001). Similarly, *RASSF1A* and *BLU* are adjacent genes located on chromosome 3p21.3. Methylation of both genes was detected at a similar frequency in a series of primary lung cancers (47% and 43% respectively), however the prevalence of RASSF1A methylation was significantly higher in smokers, particularly in adenocarcinomas, whilst BLU methylation was more common in never-smokers (Marsit *et al*, 2005).

Furthermore such gene-specific DNA methylation appears to be modulated by exposure to different carcinogens. Issa *et al* (1996) demonstrated that methylation of the oestrogen receptor gene (*ER*) in lung tumours in rats depended on the carcinogen exposure; the incidence of *ER* methylation in NNK-induced lung tumours was low (16.7%), compared to 38.1% in x-ray-induced lung tumours and 81.8% in plutonium-induced lung tumours. Similarly, Vuillemenot *et al* (2004) observed a high incidence of *RARB* methylation in lung tumours induced in mice following exposure to cigarette

smoke or NNK (90% and 85% respectively), compared to 56% in methylene chloride-induced lung tumours and 60% in vinyl carbamate-induced lung tumours.

It remains unknown what causes some loci to be more susceptible to smoking-induced methylation. The dynamic control of *cis*-acting elements may play a role. CpG islands are commonly flanked by repetitive sequences (eg Alu repeats) which are frequently methylated in normal tissues, and these may act as "*de novo* methylation centres" from which methylation can spread into adjacent sequences (Graff *et al*, 1997; Turkor & Bestor, 1997). Simultaneously, many CpG islands contain multiple Sp-1 binding sites located both upstream and downstream of the transcriptional start site, which are thought to protect CpG islands from the spread of methylation from methylation centres (Graff *et al*, 1997). It has been hypothesised that such *cis*-acting elements may become disrupted following exposure to tobacco smoke rendering some genes more vulnerable to methylation (Kim *et al*, 2003). The direct genotoxic effect of carcinogens in tobacco smoke may also play a role. Carcinogens in tobacco smoke can induce single- and double-strand breaks in DNA. Recently a high methylation index (defined as the methylation of 3 or more gene-specific promoters) in the sputum of cancer-free smokers, was reported to be associated with a reduced ability to repair double strand breaks (Leng *et al*, 2008). Indeed it has previously been demonstrated that DNMT1 is rapidly recruited to sites of DNA damage (Mortusewicz *et al*, 2005; Cuozzo *et al*, 2007).

6.3 Smoking-induced methylation and cervical neoplasia

Smoking is an independent risk factor for cervical neoplasia (see 1.4, introduction), however smoking does not increase the risk of acquiring HPV infection, and it does not prolong the duration of HPV infection, thus other mechanisms must be involved to account for the association of smoking with cervical neoplasia (Collins *et al*, 2010). In order to determine if smoking-induced methylation changes are causally associated with cervical neoplasia, it will now be important to assess if the observed smoking-induced methylation of *CDKN2A* in disease-free women, increases either the risk of women acquiring cytological abnormality, or the risk of progression to high-grade CIN in women with low-grade disease, or both.

6.4 Implications for smoking cessation

Smoking at an early age has been reported to increase the risk of both cervical and lung cancer, independent of smoking duration or intensity (Hegmann *et al*, 1993; Nordlund *et al*, 1997; Knoke *et al*, 2004; International Collaboration of Epidemiological Studies of Cervical Cancer, 2006b). Meta-analyses suggest that smoking cessation interventions have little effect on the behaviour of adolescents and young adults (Grimshaw & Stanton, 2006; Myung *et al*, 2009). Qualitative studies exploring smoking cessation in adolescent smokers have revealed that young smokers generally discount distant health risks such as cancer, because they believe that they don't smoke enough to suffer any negative consequences or that they will have quit smoking before the cigarettes can harm them, and perceive themselves to be less vulnerable to the well known long term adverse health effects of smoking (Balch, 1998; Goldman *et al*, 1998; McVea *et al*, 2009). It has been proposed that providing

more information on the immediate adverse heath consequences might help to challenge such disengagement beliefs (justifications to continue smoking) in young smokers (Kleinjan *et al*, 2009). My work provides such information. Perhaps the most salient motivational finding is the reversibility of epigenetic changes following smoking cessation.

6.5 Reservations and limitations of smoking data

Even allowing for possible inaccuracy in self-reporting, smoking behaviour is unlikely to account for all cases of *CDKN2A* methylation observed in my study. In women who have never smoked, epigenetic changes may be explained by other, as yet undefined, risk factors. For example, exposure to low-dose airborne benzene has been associated with the methylation of *p15* (*CDKN1B*) in healthy individuals (Bollati *et al*, 2007).

One limitation of this study may be the presence of sampling bias in the matched analysis. Disease-free incident smokers were first identified who tested negative for *CDKN2A* methylation in the sample taken immediately before they began to smoke. In this group serial samples were then tested until methylated forms of *CDKN2A* were detected or the end of follow-up was reached. For the controls – never smokers – only 2 samples were tested for the presence of *CDKN2A* methylation. However this was not a major source of bias because the median number of samples tested in the incident smokers was 2.5. Another limitation of this study was the absence of a control group when evaluating the impact of smoking cessation on *CDKN2A*

methylation. Thus it remains possible that loss of *CDKN2A* methylation may not be directly associated with smoking cessation.

6.6 Technical considerations

I have used nested MSP to detect methylated forms of *CDKN2A* in cervical cytological samples taken from disease-free young women. This technique has previously been reported to have a sensitivity of 1:50,000 (Palmisano *et al*, 2000). However, I have struggled to determine reproducibly the sensitivity of my nested MSP assay, possibly because the assay may be functioning at the limits of detection. The minimum consistent sensitivity of my nested MSP assay was 1:1000, however the PCR band was very bright at this dilution, while the methylated bands observed in some cytological samples were faint, suggesting that the sensitivity of my nested MSP assay was greater than 1:1000. Nevertheless, a lower nested MSP assay sensitivity of 1:1000 compared to the initial reports of 1:50,000 has also been reported by other authors (Cirincione *et al*, 2006).

The high sensitivity of nested MSP is mainly achieved through the use of high number of PCR cycles. However, this in turn, increases the risk of false positive results as a consequence of mispriming (see section 3.3.4.3). I have attempted to minimise this problem by increasing the annealing temperature of the methylated primer, to ensure the highest specificity of amplification of only methylated alleles in the sample, and by diluting the stage 1 PCR product prior to use in the stage 2 PCR (section 3.3.4.3). Furthermore, the use of unbiased stage 1 PCR primers that recognise the bisulphite-modified template but do not discriminate between methylated and

unmethylated alleles, should also eliminate incompletely bisulphite-modified template, which can also be a source of false positives; although since the introduction of kits for bisulphite modification, this is now rarely seen. To definitively exclude the presence of false priming, I have sequenced a subset of methylated products, and in every case, methylation of all the CpGs within the *CDKN2A* primer sites was observed, suggesting that the optimisation steps have helped to minimise the occurrence of false priming. The relative absence of methylation of CpGs located in the region between the 2 *CDKN2A* primers is different to the heavy methylation pattern previously described for *CDKN2A* in this region, in invasive samples (Herman *et al*, 1996). However, this may not be unexpected given that the cytological samples used in this thesis have all been taken from disease-free young women. Furthermore, the consistent methylation of CpGs within the *CDKN2A* primer sites may suggest that these primer sites are acting as "*de novo* methylation centres" as discussed above. This could be confirmed by performing sequencing analysis on methylated samples taken from different stages of cervical carcinogenesis.

In a few samples a double PCR band was observed following amplification with the methylated primers – with a second band smaller than the expected PCR product (eg Figure 4.11). Mispriming in a different region of DNA with some areas of similarity to the primer sequence may account for this: some of these smaller bands were analysed by direct sequencing, and in every case no recognisable sequence was detected (data not shown).

The primers I have used for nested MSP are published primers which have previously been well validated in several tumour sites (see section 4.3). Nevertheless, it is important to note that the expected product size following *CDKN2A* stage 1 PCR amplification is quite large - 273bp. This could be a potential source of bias, since bisulphite modification typically fragments DNA into small fragments. Despite this reservation, my direct sequencing data clearly confirms that PCR amplication of bisulphite modifified DNA at this size is possible, as has been demonstrated in other tumour sites (Herman *et al*, 1996; Palmisano *et al*, 2000; Esteller *et al*, 2001; Belinsky *et al*, 2002).

In some of the initial nested MSP experiments a PCR product was sometimes not observed following amplification of the unmethylated control DNA or the methylated control DNA with the unmethylated or the methylated primers, respectively (eg Figures 4.5 and 4.6). Degradation of bisulphite-modified unmethylated and methylated control DNA following repeat freeze-thaw cycles is the most likely explanation for this. In all subsequent experiments, the unmethylated and the methylated control DNA were bisulphite-modified just prior to use, and this problem was not observed again.

6.7 Identification of HPV-induced methylation and transcriptional changes

In an attempt to reveal epigenetic changes relevant to the pathogenesis of cervical neoplasia, I have performed gene expression and CpG island arrays on 2 *in vitro* models, a HPV replication model and a disease progression model.

I have shown that the transcriptional changes observed in PHFK following transfection with HPV18 are consistent with those previously reported. There has only been one published study which has also compared the transcriptional profile of PHFK transfected with HPV18, and approximately 50% of the differentially regulated genes in this published array were concordantly regulated in my PHFK-HPV18 array (Karstensen *et al*, 2006). I have also compared the transcriptional changes observed in my 2 *in vitro* models with those reported in 5 published studies that have compared the gene expression profile of cervical cancer or CIN with that of tissue taken from disease-free controls, and which used the same Affymetrix platform that I have used (Rosty *et al*, 2005; Santin *et al*, 2005; Wong *et al*, 2006; Pyeon *et al*, 2007; Zhai *et al*, 2007); many of the differentially regulated genes in these published arrays were also concordantly regulated on either the PHFK-18 array, the W12 array, or both. These comparisons suggest that our 2 *in vitro* models can provide useful observations relevant to cervical neoplasia.

Using the UHN HCGI12K array platform, I have identified differentially methylated genes in our 2 *in vitro* models. However the external validation has been poor. Genes found to be differentially methylated in both the PHFK-HPV18 model and the W12 cell line were not significantly enriched for genes previously reported to be

methylated in cervical neoplasia. Although genes found to be *de novo* methylated in the PHFK-HPV18 model did show an increased frequency of genes known to be methylated in cervical neoplasia. Furthermore, I have observed only a small number of differentially methylated genes between passages 11 and 56 of the W12 cell line. If most of the methylation changes have already taken place by passage 11, then this may explain the small number of differentially methylated genes observed during further cultivation of the W12 cell line; however an analysis of methylated genes at passage 11 alone revealed only a non-significant two-fold increased frequency of genes previously reported to be methylated in cervical neoplasia, thus this cannot fully explain the small number of changes observed. Validation of the predicted methylation changes has also been poor; pyrosequencing analysis was successful in 9 genes in the PHFK-HPV18 model, but the methylation change was confirmed in only 3 genes (*NKX6-1*, *PLXDC2* and *RARB*), similarly, in the W12 cell line pyrosequencing analysis was successful for 3 genes, but increased methylation was confirmed in only 1 gene (*PAX2*).

I have also attempted to explore possible determinants of HPV-induced methylation changes. However, I have found that neither promoter CpG content, nor the reported presence of H3K4me3 or H3K27me3 marks in human embryonic stem cells, were predictive of changes in methylation following transfection of PHFK with HPV18. Intermediate promoter CpG content (weak CpG islands) and the presence of a stem cell-like chromatin pattern (consisting of the simultaneous presence of the active mark H3K4me3 and the repressive mark H3K27me3), have both been reported to predispose genes to becoming methylated (Ohm *et al*, 2007; Weber *et al*, 2007).

Furthermore, I have found no significant relationship between the methylation and transcriptional changes observed in PHFK transfected with HPV18 compared to untransfected keratinocytes. Similarly, in the W12 disease progression model, no significant relationship between the methylation and transcriptional changes between passages 11 and 56 was observed.

These disappointing results may either be due to the fact that the *in vitro* models assessed are not good models of changes occurring during cervical neoplasia, or that the CpG island array platform is a poor predictor of methylation changes. Now, as I have already shown that the external validation of the transcriptional changes observed in our 2 *in vitro* models was good, this suggests that the CpG island array platform was the main reason for the observed disappointing methylation results. Differential methylation hybridisation using CpG island microarrays has actually been successfully used by many investigators to reveal differentially methylated genes in many tumour sites (Huang *et al*, 1999; Yan *et al*, 2001; Wei *et al*, 2002; van Doorn *et al*, 2005; Wang *et al*, 2005; Lu *et al*, 2009; Lu *et al*, 2010). However, it is interesting to note that virtually all the published studies have utilised custom CpG island arrays. In particular, there is a significant paucity of publications reporting differentially methylated genes identified using the same UHN HCGI12K array platform as that used for my work - while this array has been commercially available for over 8 years now. This suggests that the limitation of the CpG island arrays observed in my thesis may be particularly limited to the UHN HCGI12K array platform.

One reason the UHN HCGI12K array may have a limited ability to assess changes in promoter methylation, may relate to its design. Each array contains 12,192 CpG island clones which were originally derived from a CpG island library established by Sally Cross at the Wellcome Trust Sanger Institute, Cambridge in the 1990s; CpG islands were isolated from male human genomic DNA using a methyl-CpG binding domain column (Cross *et al*, 1994). The CpG islands clones were only sequenced subsequently hence the selection of CpG island clones for use on this commercial array was not informed by a pre-determined knowledge of the location of these CpG islands in the genome (Heisler *et al*, 2005). Now, not all CpG islands are located in the proximal promoter regions of human genes, thus a proportion of the clones on this CpG island array cannot reveal aberrant methylation of promoter regions. The fact that DNA sequence information only became available later also helps to explain the false positive findings on my arrays; some clones which were identified as becoming differentially methylated did not contain any CpG sites in their sequence. Indeed, initial characterisation of the arrays suggested that approximately 80% of the clones corresponded to unique CpG islands but the remainder comprised ribosomal DNA, repetitive elements and mitochondrial DNA (Cross *et al*, 1994).

Another limitation of the UHN HCGI12K array platform is that each CpG island clone is associated with three possible genes located either within the same strand (W), upstream (U), or downstream (D) of the transcriptional start site by a variable distance. Indeed many of the CpG islands are actually located a considerable distance from the transcriptional start site of the allocated gene. In an attempt to simplify the data analysis, Dr Wenbin Wei, head of Bioinformatics in our Institute re-annotated the

UHN HCGI12K array - using the chromosome start and end position for each CpG island clone; each clone was allocated the nearest gene using the criterion, "mid-point of the CpG island clone is within 7.5kb upstream of 5' transcriptional start site of that gene through to 100bp downstream of the transcriptional end site". However, following this re-annotation the UHN HCGI12K array was found to cover only 3379 unique genes.

There are also a number of reasons as to why I may have failed to reveal an underlying relationship between promoter CpG content, stem cell-like chromatin pattern and DNA methylation. Firstly, this may simply reflect the limited ability of the UHN HCGI12K array platform to detect methylation changes. Secondly, this analysis was inevitably restricted to a small subset of genes; only 1885 genes were included in the analyses of promoter CpG content and stem cell chromatin pattern. Thirdly, there may be no relationship between promoter CpG content, stem cell chromatin pattern and DNA methylation in PHFK transfected with HPV18. Indeed recent analyses of mouse embryonic stem cells have suggested that DNA methylation and histone modifications target different genes for repression, and an analysis of normal and malignant prostate cancer cells has demonstrated that gene promoters enriched with H3K27me3 are not more likely to become methylated in cancer (Fouse *et al*, 2008; Kondo *et al*, 2008).

There are also a number of reasons as to why I have not found an association between global methylation and transcriptional changes. Firstly, this may again simply reflect the limited ability of the UHN HCGHI12K array to predict methylation changes.

Secondly, after filtering, the subset of genes available for these analyses was small. Thirdly, there may be no significant relationship between DNA methylation and gene expression. Although promoter methylation can directly repress gene expression (section 1.1.4.1, Introduction), several recent reports have revealed that DNA methylation frequently occurs in genes that are already repressed by Polycomb-mediated histone modifications in embryonic stem cells (Ohm *et al*, 2007; Schlesinger *et al*, 2007; Widschwendter *et al*, 2007). Furthermore methylation- associated gene silencing may be more a function of local histone modifications rather than the DNA methylation *per se*. The recent ability to integrate information from both methylation and transcriptional arrays is now beginning to provide us with a more complete understanding of the relationship between global methylation and transcriptional changes. Indeed, several recent studies have revealed a relatively low negative correlation between DNA methylation and gene expression (Acevedo *et al*, 2008; O'Riain *et al*, 2009; Andrews *et al*, 2010).

For these reasons, it will be important to repeat the array experiments performed in our 2 *in vitro* models using the high resolution promoter tiling arrays now commercially available.

6.8 HPV18-induced up-regulation of DNMT1

Using the PHFK-HPV18 model, I have shown that transfection with HPV18 induced a 5-fold increase in the expression of DNMT1, with no change in the transcription of DNMT3A or DNMT3B. In addition, the up-regulation of DNMT1 was associated with increased DNMT1 protein expression. Up-regulation of DNMT1 has also been

demonstrated using Q-RT-PCR in PHFK isolated from different donors following transfection with both HPV16 and HPV18 (Dr Afaf Diyaf, personal communication). Furthermore, HPV-induced up-regulation of the DNA methyltransferases has previously been reported in studies of both HPV-immortalised keratinocyte cell lines and cell lines derived from cervical cancer (Garner-Hamrick *et al*, 2004; Santin *et al*, 2005). Thus, up-regulation of DNMT1 appears to be responsible for the methylation changes observed in our PHFK-HPV18 model.

Interestingly, while up-regulation of DNMT1 was observed in PHFK transfected with HPV18 at early passage, this was followed by a decrease in expression between early and late passages, although at late passage DNMT1 expression was still higher (almost double) compared to that in untransfected cells. However, while DNMT1 mRNA expression decreased between early and late passages of the PHFK-HPV18 cell line, I have also shown that the expression of the DNMT1 protein increased progressively following transfection with HPV18. The reason for the observed discrepancy between the mRNA and the protein expression of DNMT1 is currently unclear. Previous attempts to correlate mRNA levels with protein expression have met with varying success. While for some genes, such as *ERBB2* (*HER2/neu*), the expression levels measured using RT-PCR, immunohistochemistry and fluorescence *in situ* hybridisation have shown a highly significant correlation (Schlemmer *et al*, 2004), for many genes the correlation between mRNA and protein has been poor (Chen *et al*, 2002; Lichtinghagen *et al*, 2002; Greenbaum *et al*, 2003; Pascal *et al*, 2008). The fact that proteins can have very different half-lives as a consequence of varied protein synthesis and degradation, has been proposed as one possible

contributing factor (Greenbaum *et al*, 2003). Post-transcriptional and post-translational modifications may also contribute to the poor correlation. For example, one study has revealed that genes with a high degree of variable expression during the cell cycle, indicative of the cell controlling mRNA levels at different points of the cell cycle to achieve desired protein levels, showed a high correlation between mRNA and protein levels (r=0.89) (Greenbaum *et al*, 2003). In contrast, genes with minimal variation in expression throughout the cell cycle showed minimal correlation between mRNA and protein expression (r=0.2), suggesting that the cell may be controlling such genes at the post-translational level, and thus the mRNA level is somewhat independent of the final protein concentration (Greenbaum *et al*, 2003). Indeed, it has recently been demonstrated in porcine cells that although the abundance of DNMT1 mRNA plays an important role in protein regulation, DNMT1 is mainly regulated post-transcriptionally (Giraldo *et al*, 2009).

In addition, although I have shown that the expression of the DNMT1 protein increased progressively following transfection with HPV18, I have also observed a decrease in methylation in some genes between early and late passages of the PHFK-HPV18 model. To date no demethylase enzyme has been identified. However, one possibility for the observed hypomethylation may be that there is a change in the abundance of proteins which regulate the binding of DNMT1 to the gene promoters between early and late passages of the PHFK-HPV18 model; as cells divide during cultivation, the resulting lower bound DNMT1 levels will fail to fully re-establish the methylation patterns in all the daughter cells (passive demethylation). For example, in mice it has been found that *Lsh*, a member of the SNF2 chromatin remodelling family,

is necessary for recruitment of DNMT3B to its gene target, and depletion of *Lsh* results in DNA hypomethylation (Fan *et al*, 2008). However *Lsh* does not associate with DNMT1 (Zhu *et al*, 2006). Another example is poly (ADP-ribose) polymerase (PARP). High intracellular levels of PARP-1 and ADP-ribose polymers can bind to the DNMT1 promoter, and inhibit its enzymatic activity, which will again lead to passive hypomethylation during subsequent rounds of cell division (Guastafierro *et al*, 2008; Caiafa *et al*, 2009; Zampieri *et al*, 2009). It will clearly be interesting to explore these hypotheses further.

6.9 HPV-induced methylation of *RARB*

In addition to HPV18-induced up-regulation of DNMT1, I have also demonstrated *de novo* methylation of the *RARB* promoter in PHFK following transfection with HPV18, and this was associated with increased binding of DNMT1 at the *RARB* promoter. To my knowledge, this is the first time that HPV-induced up-regulation of one of the DNMTs has been associated with subsequent methylation of a TSG. Pyrosequencing analysis also revealed increased methylation of *RARB* in PHFK following transfection with HPV16, suggesting that methylation of *RARB* may commonly follow infection with high-risk HPV types.

RARB was also found to be methylated in the only other study which has analysed methylation changes in PHFK following transfection with high-risk HPV types; using methylation specific-multiplex ligation dependent probe amplification, methylation of *RARB* was detected in both HPV16 and HPV18 immortalised cells but not in

untransfected keratinocytes (Henken *et al*, 2007). However, in this study, no attempt was made to demonstrate binding of the DNMTs to the *RARB* promoter.

Down-regulation of *RARB* has been reported in many tumours including cervical carcinoma, suggesting that loss of *RARB* expression may be an important event in carcinogenesis (Xu *et al*, 1994; Geisen *et al*, 1997; Widschwendter *et al*, 1997; Xu *et al*, 1997b; Qiu *et al*, 1999; Lotan *et al*, 2000). However, while *RARB* is expressed in normal cervical epithelial cells it is not normally expressed in the skin or in keratinocytes derived from the foreskin (Zelent *et al*, 1989; Geisen *et al*, 1997; Reichrath *et al*, 1997). In these tissues, the important role of the retinoic acid receptors (RAR) in cell growth and differentiation may then be mediated primarily by the RAR alpha and gamma subtypes, which are highly expressed here; indeed *RAR-gamma* expression is predominantly confined to keratinising squamous epithelia, and the expression of *RARB* and *RAR-gamma* appears to be mutually exclusive (Dollé *et al*, 1990; Viallet & Dhouailly, 1994; Reichrath *et al*, 1997). I have also shown that *RARB* was not expressed in untransfected PHFK and remained so following transfection with HPV18. This raises the question as to why HPV18 causes methylation of *RARB* in a setting where *RARB* is already silenced.

Several *in vitro* studies have revealed that *RARB* inhibits proliferation through inhibition of the transcription factor activator protein-1 (AP-1) via both ligand-dependent and ligand-independent mechanisms (van der Burg *et al*, 1995; Lee *et al*, 1996; Lin *et al*, 2000; De-Castro Arce *et al*, 2004). AP-1 is a heterodimeric protein containing proteins from 2 major families, Jun (c-Jun, JunB, JunD) and Fos (c-Fos,

FosB, Fra-1, Fra-2) (Angel & Karin, 1991). Constitutive activation of AP-1 is known to be a common mechanism for tumour promotion in epithelial tissues irrespective of the tumour site (Angel & Karin, 1991; Prusty & Das, 2005; Hussain *et al*, 2009). In addition, AP-1 is also essential for transcriptional regulation of the HPV viral oncogenes; site-directed mutagenesis of the AP-1 binding site in the upstream regulatory region (URR) of HPV16 or HPV18 almost completely abolishes transcription of the E6 and E7 genes (Offord & Beard, 1990; Mack & Laimins, 1991; Thierry *et al*, 1992; Soto *et al*, 1999). Recently, it has been shown that ectopic expression of *RARB* selectively abrogates binding of AP-1 to the HPV18 URR independent of ligand, by selective degradation of the c-Jun protein. The inhibition of AP-1 binding induced *de novo* methylation and nuclear condensation of the URR and resulted in strong down-regulation of the transcription of the HPV18 viral oncogenes (De-Castro Arce *et al*, 2007). Thus, in HPV-associated cancers, methylation of *RARB* may be a mechanism for maintaining transcription of the HPV oncogenes E6 and E7.

Currently, it remains unknown which of the HPV viral oncogenes may be responsible for the HPV-induced methylation of *RARB*, although E7 appears to be the most likely candidate. Previously, HPV16 E7 has been found to associate with DNMT1 (Burgers *et al*, 2006); 293T cells were co-transfected with pcDNA3-E7-F (which expresses Flag-tagged HPV16 E7) and Myc-DNMT1 (which expresses Myc-tagged full length DNMT1), followed by immunoprecipitation of the cell extracts with an anti-Flag antibody. DNMT1 was detected in the immunoprecipitated fraction by Western blot analysis. Furthermore, HPV16 E7 was found to bind directly to DNMT1; GST-tagged HPV16 E7 bound strongly to baculovirus-expressed DNMT1, and deletion analyses

demonstrated that this binding was mediated by the CR3 zinc finger domain of HPV16 E7 (Burgers *et al*, 2006). In addition, the binding of HPV16 E7 to DNMT1 resulted in a significant up-regulation of DNMT1 enzymatic activity (Burgers *et al*, 2006). However, in this study, the increased methyltransferase activity was not linked with the subsequent methylation of a TSG. The high-risk E7 protein is also known to bind to and degrade the Rb tumour suppressor protein, which releases the E2F transcription factor from Rb inhibition (Dyson *et al*, 1989). Now it has previously been demonstrated that DNMT1 contains E2F binding sites in its promoter region, and that binding of E2F regulates transcription of DNMT1, thus this may be an additional mechanism by which the high-risk E7 protein can cause up-regulation of DNMT1 (McCabe *et al*, 2005).

It is interesting to note that in untransfected PHFK, DNMT1 was already bound at a low level to the *RARB* promoter, in the absence of detectable DNA methylation. It is possible that a low level of methylation is present at the *RARB* promoter, which may be contributing to the silenced *RARB* expression, but that the sensitivity of pyrosequencing precluded its detection. Alternatively it is known that the N-terminal catalytic region of DNMT1 can directly repress transcription, by recruiting histone deacetylases and other co-repressor proteins such as DMAP1, independently of its methyltransferase activity, and this may account for silenced *RARB* expression in the absence of DNA methylation (Rountree *et al*, 2000).

The pattern of *de novo* methylation of the *RARB* promoter following transfection with HPV18 or HPV16, appears to target certain CpG loci, rather than consecutive

adjacent CpGs, suggesting the existence of several "*de novo* methylation centres". This pattern may simply reflect the methylation pattern seen at the earliest stages of HPV-induced carcinogenesis - just following HPV infection. Consistent with this, pyrosequencing analysis of the *RARB* promoter in the HPV-positive cervical cancer cell lines, CaSki and SiHa, revealed that most of the CpG sites in the *RARB* promoter were heavily methylated. However, it remains possible that the observed non-consecutive pattern of methylation is artefactual; bisulphite genomic sequencing of the *RARB* promoter would confirm the genuine existence of such a pattern if this pattern was also observed in individual clones.

6.10 Relationship between DNA methylation and H3K27me3

Following transfection with HPV18, I have also demonstrated increased binding of H3K27me3 at the *RARB* promoter. Consistent with this result, up-regulation of EZH2 (which catalyses tri-methylation of lysine 27 at histone H3) was observed on my gene expression arrays following transfection with HPV18. Up-regulation of EZH2 has also been observed following infection of primary human keratinocytes with HPV18 E6 and E7 genes grown in an organotypic raft culture system (Garner-Hamrick *et al*, 2004), and down-regulation of EZH2 has been identified following inhibition of HPV18 E6 and E7 expression by RNA interference in HeLa cells (Kuner *et al*, 2007). Furthermore, in MCF-7 cells, the ectopic expression of high risk (HPV16 and HPV18) and low risk (HPV6b and HPV11) E7 has been reported to activate the EZH2 promoter, via E7-mediated release of E2F from pRb and the other pocket proteins (Holland *et al*, 2008). However, when these experiments were repeated in primary human cervical or foreskin keratinocytes, HPV18 E7, in contrast to the other E7

proteins examined, did not seem to activate the EZH2 promoter (Holland *et al*, 2008). This latter finding is in contrast with my findings, where EZH2 was up-regulated following transfection of PHFK with HPV18. The reason for this discrepancy is not clear.

Both H3K27me3 and EZH2 have been reported to be enriched at the promoters of many genes that are hypermethylated and silenced in cancer (McGarvey *et al*, 2006). However, the functional relationship between these 2 repressive mechanisms remains unclear. Vire *et al* (2005) demonstrated a direct interaction between EZH2 and the DNMTs both *in vitro* and *in vivo*, but several subsequent studies have provided compelling evidence that H3K27me3 and DNA methylation function independently (McGarvey *et al*, 2007; Kondo *et al*, 2008). The presence of both silencing marks (DNA methylation and H3K27me3) at the *RARB* promoter has been reported previously in the colon cancer cell line, SW48, where marked elevation of H3K27me3 and high DNA methylation was observed at the *RARB* promoter, in association with gene silencing. However, in the prostate cancer cell line, PC3, marked elevation of H3K27me3 was also observed at the *RARB* promoter but with relatively low levels of promoter DNA methylation (Kondo *et al*, 2008). This clearly demonstrates that *RARB* can be targeted by both DNA methylation and H3K27me3, and suggests that the epigenetic silencing of *RARB* may be tissue and cell-line specific.

Our viral replication model which allows for sequential observations to be made at successive passages does provide the opportunity to resolve the temporal ambiguity about the order in which the H3K27me3 and methylation marks are imposed. I have

observed increased binding of both DNMT1 and H3K27me3 in the 4th passage following transfection with HPV18, and it will be interesting to perform ChIP experiments in the preceding passages.

6.11 Limitations of ChIP experiments

The main limitation of this section of my work is that after troubleshooting my experiments, sufficient material was only available for the chromatin immunoprecipitation experiments to be performed once. It will therefore be important to repeat the ChIP experiments using PHFK from other donors that have been transfected with HPV16 and HPV18 which are available in our laboratory, to confirm that the observed increased binding of DNMT1 and H3K27me3 at the *RARB* promoter is reproducible and also not unique to the PHFK donor used for my work.

6.12 Conclusions

In conclusion, using the cervix uteri as a model, I have shown that smoking increases the risk of acquiring methylated forms of *CDKN2A,* in disease-free young women, which may become undetectable following smoking cessation. In addition, using a HPV replication model, I have also shown that HPV18 infection is followed by the up-regulation of the DNA methyltransferase DNMT1 which binds to the TSG, *RARB*, resulting in its *de novo* methylation.

APPENDIX

Array	No. of spots "flagged"	No. with signal intensities near background ($\leq$25)	Remaining spots
Untransfected PHFK 1	444	1	11747
Untransfected PHFK 2	733	11	11448
Untransfected PHFK 3	474	5	11713
PHFK/HPV18 Early 1	684	2	11506
PHFK/HPV18 Early 2	417	10	11765
PHFK/HPV18 Early 3	492	13	11687
PHFK/HPV18 Late 1	3222	15	8955
PHFK/HPV18 Late 2	1064	1	11127
PHFK/HPV18 Late 3	924	3	11265

Table A1. Filtering - number of spots filtered from each PHFK array prior to data analysis.

Array	No. of spots "flagged"	No. with signal intensities near background ($\leq$25)	Remaining spots
W12 P11 a	769	36	11387
W12 P11 b	766	53	11373
W12 P11 c	736	41	11415
W12 P56 a	1050	43	11099
W12 P56 b	593	33	11566
W12 P56 c	805	21	11366

Table A2. Filtering - number of spots filtered from each W12 array prior to data analysis.

Gene	UHN ID	Normalised cut/mock-cut ratios			No. of CpGs in probe	No. of CpGs analysed
		Untransfected PHFK#1	PHFK#1/HPV18 Early	PHFK#1/HPV18 Late		
CDH13	UHNhscpg0002264	0.88	0.68	0.77	5	5
COL5A3	UHNhscpg0004024	0.88	0.69	0.67	42	7
GAP43	UHNhscpg0001120	0.88	0.56	0.61	42	4
KIT	UHNhscpg0007511	0.89	0.54	0.65	9	-
NKX6-1	UHNhscpg0005115	0.94	0.67	0.65	15	14
NOL4	UHNhscpg0001134	1.16	1.00	0.63	9	9
NR2F1	UHNhscpg0002537	0.88	0.50	0.88	29	21
PLXDC2	UHNhscpg0001333	0.99	0.58	0.75	6	6
RARB	UHNhscpg0002561	0.93	0.56	0.95	19	18
SLIT2	UHNhscpg0010610	1.05	0.71	0.55	0	-
SOX1	UHNhscpg0009290	0.91	0.66	0.79	46	-
UNC5B	UHNhscpg0005197	0.92	0.48	0.76	5	4

Table A3. Genes selected for pyrosequencing analysis in the PHFK-HPV18 cell line.

Gene	UHN ID	Normalised cut/mock-cut ratios		No. of CpGs in probe	No. of CpGs analysed
		W12 P11	W12 P56		
BARD1	UHNhscpg0002602	0.99	0.78	19	-
CDH12	UHNhscpg0008894	0.83	0.46	0	-
DOCK5	UHNhscpg0003319	0.77	0.36	0	-
FANCL	UHNhscpg0006579	1.00	0.73	10	10
HTRA1	UHNhscpg0000708	0.89	0.63	5	5
PAX2	UHNhscpg0010216	0.96	0.60	21	21
PHOX2B	UHNhscpg0000388	0.68	0.39	24	-
PPP2CA	UHNhscpg0011351	0.87	0.43	8	-

Table A4. Genes selected for pyrosequencing analysis in the W12 cell line.

Name	Primer sequence 5'-3'	Product size (bp)	Annealing temperature (°C)
FANCL PYRO F1	ATTGTTTATAGAGGGGGATATGTT		
FANCL PYRO R1*	CTCCATTCCCTCCTCCTATAACTA	105	50
FANCL SEQ1	ATAGAGGGGGATATGTTAT		
FANCL PYRO F2*	TTTTGAAGAGGTTGAGGAAGTTT		
FANCL PYRO R2	AAAATCCTCCACAAAATCCTCAT	181	50
FANCL SEQ2	CTCCACAAAATCCTCAT		
HTRA1 PYRO F1*	ATAAAAATTGTGAAGGATATTGGT		
HTRA1 PYRO R1	CCACATATAAACACCAACTAAACTCT	217	50
HTRA1 SEQ1	ACCAACTAAACTCTAAACTT		
HTRA1 PYRO F2	TTTGTGTTTTATGTTTAGTTGGTTA		
HTRA1 PYRO R2*	TCCATTACATATAAAAAATTAACTTCTC	185	50
HTRA1 SEQ2	TTATGTTTAGTTGGTTAGAG		
PAX2 PYRO F	GAGGAAGGGAAATAAGTTAAAGAAAT		
PAX2 PYRO R*	CCCTAAAATTCAACTTAAAATCCAA	287	52
PAX2 SEQ1	GATTAAGAAGGAAGGGATAG		
PAX2 SEQ2	AAGGGGGTTAGGGTT		
Key: * = biotinylated primer			

Table A5. Primers for pyrosequencing (W12)

Name	Primer sequence 5'-3'	Product size (bp)	Annealing temperature (°C)
CDH13 PYRO F1	TTGGATTTAATTATTTTTTTGTTGAAA		
CDH13 PYRO R1*	TCTAATTCATACCAAACATCTACC	269	50
CDH13 SEQ1	TTTTTGTTGAAATTATTGAT		
CDH13 PYRO F2	AGATGTGTTTAGTTGGGATATTGT		
CDH13 PYRO R2*	TCTAATTCATACCAAACATCTACC	216	52
CDH13 SEQ2	TTGTTTTTGTATGTTAAAAT		
NKX6-1 PYRO F1*	AATTGGAAGGTAGGAAAGTGAATT		
NKX6-1 PYRO R1	TCCCCCTAAATAACTAACACCAA	165	52
NKX6-1 SEQ1	AACACCAACAAACTAACTAC		
NKX6-1 PYRO F2	GTTGGTGTTAGTTATTTAGGGGGAATT		
NKX6-1 PYRO R2*	TCAAATCCAAACCATACTACCACTC	151	54
NKX6-1 SEQ2	GTTAGTTATTTAGGGGGAA		
NOL4 PYRO F1	TTGGGAAGTTATTTATGTTGAGTAAT		
NOL4 PYRO R1*	CCATTAAAATCCCACTAACAAATACCTT	380	50
NOL4 SEQ1	TTGTTTTTTGTTTTTAGGA		
NOL4 PYRO F2	TAGGGTTGATGTTTAGTTTTTTAAGTGTTT		
NOL4 PYRO R2*	CCATTAAAATCCCACTAACAAATACCTT	212	54
NOL4 SEQ2	GATAGTTGAGTAAATTAGTATTAGATG		
RARB PYRO F1*	TAGTTGGGTTATTTGAAGGTTAGTA		
RARB PYRO R1	CACTTCCTACTACTTCTATCACACAA	110	50
RARB SEQ1	CACACAAAATAAAAAATTAA		
RARB PYRO F2	GAGTTGTTTGAGGATTGGGATGT		
RARB PYRO R2*	CCAAAAAAATCCCAAATTCTCC	141	54
RARB SEQ2	AGTTGTTTGAGGATTGG		
RARB PYRO F3/4	TGGGATTTTTTTGGGAATTT		
RARB PYRO R3/4*	CCATCAAACTCTACCCCTTTTTTA	178	52
RARB SEQ3	GGATTTTTTTGGGAATTT		
RARB SEQ4	TGTTTGTTATAATTTATGAT		
SOX1 PYRO F	TGTGGGGTTTGGATTGTGAAG		
SOX1 PYRO R*	CTCCCAAAACCCCTACAAATTC	143	54
SOX1 SEQ	TTGGATTGTGAAGGTTT		
COL5A3 PYRO F	GGGAAATGTAGTTTTAGTGTTGG		
COL5A3 PYRO R*	CCCAACAACTCCCCATAAAA	157	52
COL5A3 SEQ	GGTGATAAATAGGAGTAAGG		
GAP43 PYRO F*	GGGGAGAAATAAAGAGATGAGATT		
GAP43 PYRO R	TACCTAATTTCCCTTCTCCAAATCAA	168	52
GAP43 SEQ	TTTCCCTTCTCCAAAT		
KIT PYRO F	AGGAGATAGGTATAGAGGGGTGAA		
KIT PYRO R*	CCTCCCAAATTTATCCTCTTCC	189	52
KIT SEQ	GGGGTGAAGTAATTTGA		
NR2F1 PYRO F1	AAGTGTAGAGGGTTAGGTAGGTTG		
NR2F1 PYRO R1*	CACCAAACAAACACAAAATTCTT	186	50
NR2F1 SEQ1	GGTTAGGTAGGTTGGGTG		
NR2F1 PYRO F2	TGTTTGGGAGTAGTTTTAATTGGT		
NR2F1 PYRO R2*	AAAAAATCTCTTTTTCCCATACTT	249	50
NR2F1 SEQ2	TTTAGTGTTTTTAGATTTTG		
NR2F1 PYRO F3	GGATGGATATGTAAGGTTAGTTGAAGTT		
NR2F1 PYRO R3*	AAACCAATACCCCAACCAACC	191	62
NR2F1 SEQ3	TTGAAGTTGTTTTTAGATAG		
PLXDC2 PYRO F1	TGGAGGGAAAAGAAAAGAAAATG		
PLXDC2 PYRO R1*	CACCTTACCTAAAATCCCCAAATA	217	52
PLXDC2 SEQ1	TGTGTAATTTAAATGATTGT		
PLXDC2 PYRO F2	TGGGGATTTTAGGTAAGGTGAATT		
PLXDC2 PYRO R2*	ACACACCCCTTCAACCATAAATA	174	52
PLXDC2 SEQ2	TTTTTGAGTGAAATTTATTA		
UNC5B PYRO F	TGGAGTAGGAAAATTGGAGTAGAA		
UNC5B PYRO R*	TAAAACCAAACCCTCACTCCTAA	163	50
UNC5B SEQ	GGAAAATTGGAGTAGAAAG		
Key: * = biotinylated primer			

Table A5. Primer sequences for pyrosequencing (PHFK-HPV18 cell line)

First Author	Platform	Samples	Methods
Pyeon	U133 Plus 2.0	8 normal cervices and 20 cervical cancers	Laser capture microdissected (LCM) samples; 2 rounds of amplification
Rosty	U133A	5 normal cervices, 5 cell lines, 35 primary tumours including 5 replicates	Tumour samples contained >50% malignant cells
Santin	U133A	11 primary cervical cancer cell lines derived from stage IB-IIA cervical cancers and 4 normal cervical keratinocytes cell lines	All cell lines harvested for RNA extraction at a confluence of 60–80%; 2-15 passages; no significant differences in the no. of passages between normal and cancer cell lines
Wong	U133A	18 normal cervices and 29 cervical cancers (SCC)	Tumour samples contained >80% malignant cells
Zhai	U133A	10 normal cervices, 7 HSIL and 21 cervical cancers (SCC)	LCM samples; 2 rounds of amplification

Table A6. Published cervical cancer arrays: gene expression profiling in patients with cervical neoplasia and cancer-free controls using the U133 Affymetrix platform.

REFERENCES

Abele R, Clavel M, Dodion P, et al (1987). The EORTC Early Clinical Trials Cooperative Group experience with 5-aza-2'-deoxycytidine (NSC 127716) in patients with colo-rectal, head and neck, renal carcinomas and malignant melanomas. *Eur J Cancer Clin Oncol* **23**: 1921-1924.

Acevedo L, Bieda M, Green R, Farnham PJ (2008). Analysis of the mechanisms mediating tumor-specific changes in gene expression in human liver tumors. *Cancer Res* **68**: 2641-2651.

Alazawi W, Pett M, Arch B, *et al* (2002). Changes in cervical keratinocytes gene expression associated with integration of human papillomavirus 16. *Cancer Res* **62**: 6959-6965.

Ali S, Astley SB, Sheldon TA, Peel KR, Wells M (1994). Detection and measurement of DNA adducts in the cervix of smokers and non-smokers. *Int J Gynecol Cancer* **4**:188-193.

Andrae B, Kemetli L, Sparen P, *et al* (2008). Screening-preventable cervical cancer risks: evidence from a nationwide audit in Sweden. *J Natl Cancer Inst* **100**: 622-629.

Andrews J, Kennette W, Pilon J, *et al* (2010). Multi-platform whole-genome microarray analyses refine the epigenetic signature of breast cancer metastasis with gene expression and copy number. *PLoS One* **5**: e8665.

Angel P, Karin M (1991). The role of Jun, Fos, and the AP-1 complex in cell-proliferation and transformation. *Biochim Biophys Acta* **1072**: 129-157.

Antequera F, Bird A. Number of CpG islands and genes in human and mouse (1993). *Proc Natl Acad Sci USA* **90**: 11995-11999.

Antilla A, Pukkala E, Soderman B, *et al* (1999). Effect of organised screening on cervical incidence and mortality in Finland, 1963-1995: recent increase in cervical cancer incidence. *Int J Cancer* **83**: 59-65.

Apostolidou S, Hadwin R, Burnell M, *et al* (2009). DNA methylation analysis in liquid-based cytology for cervical cancer screening. *Int J Cancer* **125**: 2995-3002.

Arends MJ, Buckley CH, Wells M (1998). Aetiology, pathogenesis, and pathology of cervical neoplasia. *J Clin Pathol* **51**: 96-193.

Arora P, Kim EO, Jung JK, Jang KL (2008). Hepatitis C virus core protein downregulates E-cadherin expression via activation of DNA methyltransferase 1 and 3b. *Cancer Lett* **261**: 244-252.

Avner P and Heard E (2002). X-chromosome inactivation: counting, choice and initiation. *Nature Rev Genet* **2**: 59-67.

Bachman AN, Curtin GM, Doolittle DJ, Goodman JI (2006). Altered methylation in gene-specific and GC-rich regions of DNA is progressive and non-random during promotion of skin tumorigenesis. *Toxicol Sci* **91**: 406-418.

Bachman KE, Rountree MR, Baylin SB (2001). Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J Biol Chem* **276**: 32282-32287.

Bachmann KE, Park BH, Rhee I, *et al* (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 IE, *et al* (2004). The human papillomavirus-18 genome is efficiently targeted by cellular DNA methylation. *Virology* **324**: 483-492.

Badal V, Chuang LS, Tan EH, *et al* (2003). CpG methylation of human papillomavirus type 16 DNA in cervical cancer cell lines and in clinical specimens: genomic hypomethylation correlates with carcinogenic progression. *J Virol* **77**: 6227-6234.

Balch GI (1998). Exploring perceptions of smoking cessation among high school smokers: input and feedback from focus groups. *Prev Med* **27**: A55–A63.

Bastien N, McBride AA (2000). Interaction of the papillomavirus E2 protein with mitotic chromosomes. *Virology* **270**: 124-134.

Baylin S (2005). DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol* **2 Suppl 1**:S4-S11.

Belinsky SA, Nikula KJ, Baylin SB, Issa JP (1996). Increased cytosine DNA-methyltransferase activity is target-cell-specific and an early event in lung cancer. *Proc Natl Acad Sci USA* **93**: 4045-4050.

Belinsky SA, Nikula KJ, Palmisano WA, *et al* (1998). Aberrant methylation of p16^{INK4a} is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc Natl Acad Sci USA* **95**: 11891-11896.

Belinsky SA, Palmisano WA, Gilliland FD, *et al* (2002). Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res* **62**: 2370-2377.

Belinsky SA, Klinge DM, Dekker JD, *et al* (2005). Gene promoter methylation in plasma and sputum increases with lung cancer risk. *Clin Cancer Res* **11**: 6505-6511.

Bernstein BE, Mikkelsen TS, Xie X, *et al* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315-326.

Bestor TH (1988). Cloning of a mammalian DNA methyltransferase. *Gene* **74**: 9-12.

Bestor TH (2005). Transposons reanimated in mice. *Cell* **122**: 322-325.

Bhattacharjee B, Sengupta S (2006). CpG methylation of HPV 16 LCR at E2 binding site proximal to P97 is associated with cervical cancer in presence of intact E2. *Virology* **354**: 280-285.

Bird AP (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* **8**: 1499-1504.

Bird AP (1986). CpG-rich islands and the function of DNA methylation. *Nature* **321**: 209-213.

Bird AP, Wolffe AP (1999). Methylation-induced repression--belts, braces, and chromatin. *Cell* **99:** 451-454.

Bird A (2002). DNA methylation patterns and epigenetic memory. *Genes Dev* **16**: 6-21.

Bollati V, Baccareli A, Hou L, *et al* (2007). Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer Res* **67**: 876-879.

Bosch XM, de Sanjose S (2003). Chapter 2: human papillomavirus and cervical cancer – burden and assessment of causality. *J Natl Cancer Inst Monogr* **31**: 3-13.

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**: 2536-2539.

Bray F, Carstensen B, Moller H, et al (2005). Incidence trends of adenocarcinoma of the cervix in 13 European Countries. *Cancer Epidemiol Biomarkers Prev* **14**: 2191-2199.

Bryan JT, Brown DR (2000). Association of the human papillomavirus type 11 E1()E4 protein with cornified cell envelopes derived from infected genital epithelium. *Virology* **277**: 262-269.

Bulk S, Visser O, Rozendal L, Verheijen RH, Meijer CJ (2005). Cervical cancer in the Netherlands 1989-1998: decrease of squamous cell carcinoma in older women, increase of adenocarcinoma in younger women. *Int J Cancer* **113**: 1005-1009.

Burgers WA, Blanchon L, Pradhan S, de Launoit Y, Kouzarides T, Fuks F (2007). Viral oncoproteins target the DNA methyltransferases. *Oncogene* **26**: 1650-1655.

Burnett TS, Sleeman JP (1984). Uneven distribution of methylation sites within the human papillomavirus 1a genome: possible relevance to gene expression. *Nucleic Acids Res* **12**: 8847-8860.

Caiafa P, Guastafierro T, Zampieri M (2009). Epigenetics: poly(ADP-ribosyl)ation of PARP-1 regulates genomic methylation patterns. *FASEB J* **23**: 672-678.

Cameron EE, Bachman KE, Myöhänen S, Herman JG, Baylin SB (1999). Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat Genet* **21**: 103-107.

Chang HW, Ling GS, Wei WI, Yuen AP (2004). Smoking and drinking can induce p15 methylation in the upper aerodigestive tract of healthy individuals and patients with head and neck squamous cell carcinoma. *Cancer* **101**: 125-132.

Chen G, Gharib TG, Huang CC, *et al* (2002). Discordant protein and mRNA expression in lung adeoncarcinomas. *Mol Cell Proteomics* **1**: 304-313.

Chen T and Li E (2004). Structure and function of eukaryotic DNA methyltransferases. *Curr Top Dev Biol* **60**: 55-89.

Chen XS, Garcea RL, Goldberg I, Casini G, Harrison SC (2000). Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Mol Cell* **5**: 557-567.

Cheng S, Schmidt-Grimminger DC, Murant T, Broker TR, Chow LT (1995). Differentiation-dependent up-regulation of the human papillomavirus E7 gene reactivates cellular DNA replication in suprabasal differentiated keratinocytes. *Genes Dev* **9**: 2335-2349.

Cirincione R, Lintas C, Conte D, *et al* (2006). Methylation profile in tumor and sputum samples of lung cancer patients detected by spiral computed tomography: a nested case-control study. *Int J Cancer* **118**: 1248-1253.

Clark SJ, Harrison J, Molloy PL (1997). Sp1 binding is inhibited by (m)Cp(m)CpG methylation. *Gene* **195**: 67-71.

Colella S, Shen L, Baggerly KA, Issa JPJ, Krahe R (2003). Sensitive and quantitative universal pyrosequencing methylation analysis of CpG sites. *Biotechniques* **35**: 146-150.

Collins S, Mazloomzadeh S, Winter H, *et al* (2002). High incidence of cervical human papillomavirus infection in women during their first sexual relationship. *BJOG* **109**: 96-98.

Collins SI, Constandinou-Williams C, Wen K, *et al* (2009). Disruption of the E2 gene is a common and early event in the natural history of cervical human papillomavirus infection: a longitudinal cohort study. *Cancer Res* **69**: 3828-3832.

Collins S, Rollason TP, Young LS, Woodman CB (2010). Cigarette smoking is an independent risk factor for cervical intraepithelial neoplasia in young women: a longitudinal study. *Eur J Cancer* **46**: 405-411.

Comb M, Goodman HM (1990). CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic Acids Res* **18**: 3975-3982.

Conger KL, Liu JS, Kuo SR, Chow LT, Wang TS (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.

Constantinides PG, Jones PA, Gevers W (1977). Functional striated muscle cells from non-myoblast precursors following 5-azacytidine treatment. *Nature* **267**: 364-366.
Conway MJ, Meyers C (2009). Replication and assembly of human papillomaviruses. *J Dent Res* **88**: 307-317.

Costello JF, Frühwald MC, Smiraglia DJ, *et al* (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* **24**: 132-138.

Cripe TP, Haugen TH, Turk JP, *et al* (1987). Transcriptional regulation of the human papillomavirus-16 E6-E7 promoter by a keratinocyte-dependent enhancer, and by viral E2 trans-activator and repressor gene products: implications for cervical carcinogenesis. *EMBO J* **6**: 3745-3753.

Critchlow CW, Wolner-Hanssen P, Eschenbach DA, *et al* (1995). Determinants of cervical ectopia and of cervicitis: age, oral contraception, specific cervical infection, smoking, and douching. *Am J Obstet Gynecol* **173**: 534-543.

Cross SH, Charlton JA, Nan X, Bird AP (1994). Purification of CpG islands using a methylated DNA binding column. *Nat Genet* **6**: 236-244.

Csankovski G, Nagy A, Jaenisch R (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J Cell Biol* **153**: 773-784.

Cui H, Cruz-Correa M, Giardiello FM, *et al* (2003). Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science* **299**: 1753-1755.

Cuozzo C, Porcellini A, Angrisano T, *et al* (2007). DNA damage, homology-directed repair, and DNA methylation. *PLoS Genet* **7**: 1144-1162.

Dall KL, Scarpini CG, Roberts I, *et al* (2008). Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions. *Cancer Res* **68**: 8249-8259.

Dammann R, Strunnikova M, Schagdarsurengin U, *et al* (2005). CpG island methylation and expression of tumour-associated genes in lung carcinoma. *Eur J Cancer* **41**: 1223-1236.

Danos O, Katinka M, Yaniv M (1980). Molecular cloning, refined physical map and heterogeneity of methylation sites of papillomavirus type 1a DNA. *Eur J Biochem* **109**: 457-461.

Daskalakis M, Nguyen TT, Nguyen C, *et al* (2002). Demethylation of a hypermethylated P15/INK4B gene in patients with myelodysplastic syndrome by 5-Aza-2'-deoxycytidine (decitabine) treatment. *Blood* **100**: 2957-2964.

de Fraipont F, Moro-Sibilot D, Michelland S, *et al* (2005). Promoter methylation of genes in bronchial lavages: a marker for early diagnosis of primary and relapsing non-small cell lung cancer? *Lung Cancer* **50**: 199-209.

De Geest K, Turyk ME, Hosken MI, *et al* (1993). Growth and differentiation of human papillomavirus type 31b positive human cervical cell lines. *Gynecol Oncol* **49**: 303-310.

De Smet C, De Backer O, Faraoni I, *et al* (1996). The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. *Proc Natl Acad Sci USA* **93**: 7149-7153.

De-Castro Arce J, Soto U, van Riggelen J, *et al* (2004). Ectopic expression of nonliganded retinoic acid receptor beta abrogates AP-1 activity by selective degradation of c-Jun in cervical carcinoma cells. *J Biol Chem* **279**: 45408-45416.

De-Castro Arce J, Göckel-Krzikalla E, Rösl F (2007). Retinoic acid receptor beta silences human papillomavirus-18 oncogene expression by induction of de novo methylation and heterochromatinization of the viral control region. *J Biol Chem* **282**: 28520-28529.

Ding DC, Chiang MH, Lai HC, *et al* (2009). Methylation of the long control region of HPV16 is related to the severity of cervical neoplasia. *Eur J Obstet Gynecol Reprod Biol* **147**: 215-220.

Doerfler W, Hohlweg U, Müller K, *et al* (2001). Foreign DNA integration--perturbations of the genome--oncogenesis. *Ann N Y Acad Sci* **945**: 276-288.

Doll R, Hill AB (1950). Smoking and carcinoma of the lung. Preliminary report. *Br Med J* **2**: 739-748.

Dollé P, Ruberte E, Leroy P, Morriss-Kay G, Chambon P (1990). Retinoic acid receptors and cellular retinoid binding proteins. I. A systematic study of their differential pattern of transcription during mouse organogenesis. *Development* **110**: 1133-1151.

Dong SM, Sun DI, Benoit NE, *et al* (2003). Epigenetic inactivation of RASSF1A in head and neck cancer. *Clin Cancer Res* **9**: 3635-3640.

Doorbar J, Foo C, Coleman N, *et al* (1997). Characterization of events during the late stages of HPV16 infection in vivo using high-affinity synthetic Fabs to E4. *J Virol* **238**: 40-52.

Doorbar J (2005). The papillomavirus life cycle. *J Clin Virol* **32S**: S7-S15.

Doorbar J (2006). Molecular biology of human papillomavirus infection and cervical cancer. *Clinical Sci (Lond)* **110**: 525-541.

Dürst M, Seagon S, Wanschura S, zur Hausen H, Bullerdiek J (1995). Malignant progression of an HPV16-immortalized human keratinocyte cell line (HPKIA) in vitro. *Cancer Genet Cytogenet* **85**: 105-112.

Duffy CL, Phillips SL, Klingelhutz AJ (2003). Microarray analysis identifies differentiation-associated genes regulated by human papillomavirus type 16 E6. *Virology* **314**: 196-205.

Dyson N, Howley PM, Münger K, Harlow E (1989). The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science* **243**: 934-937.

Egawa K (2003). Do human papillomaviruses target epidermal stem cells? *Dermatology* **207**: 251-254.

Egger G, Liang G, Aparicio A, Jone PA (2004). Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**: 457-463.

Eguchi K, Kanai Y, Kobayashi K, Hirohashi S (1997). DNA hypermethylation at the D17S5 locus in non-small cell lung cancers: its association with smoking history. *Cancer Res* **57**: 4913-4915.

Enokida H, Shiina H, Urakami S, *et al* (2006). Smoking influences aberrant CpG hypermethylation of multiple genes in human prostate carcinoma. *Cancer* **106**: 79-86.

Epstein MA, Achong BG, Barr YM (1964). Virus particles in cultured lymphoblasts from Burkitt's lymphoma. *Lancet* **1**: 702-703.

Esteller M (2005). Aberrant DNA methylation as a cancer-inducing mechansim. *Annu Rev Pharmacol Toxicol* **45**: 629-656.

Esteller M, Corn PG, Baylin SB, Herman JG (2001). A gene hypermethylation profile of human cancer. *Cancer Res* **61**: 3225-3229.

Esteller M. Epigenetics in cancer (2008). *New Eng J Med* **358**: 1148-1159.

Fahrner JA, Eguchi S, Herman JG, Baylin SB (2002). Dependence of histone modifcations and gene expression on DNA hypermethylation in cancer. *Cancer Res* **62**: 7213-7218.

Fan T, Schmidtmann A, Xi S, *et al* (2008). DNA hypomethylation caused by Lsh deletion promotes erythroleukemia development. *Epigenetics* **3**: 134-142.

Fehrmann F, Klumpp DJ, Laimins LA (2005). Human papillomavirus type 31 E5 protein supports cell cycle progression and activates late viral functions upon epithelial differentiation. *J Virol* **77**: 2819-2831.

Feinberg AP, Vogelstein B (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**: 89-92.

Feinberg AP and Tycko B (2004). The history of cancer epigenetics. *Nat Rev Cancer* **4**: 143-153.

Feng Q, Balasubramanian A, Hawes SE, *et al* (2005). Detection of hypermethylated genes in women with and without cervical neoplasia. *J Natl Cancer Inst* **97**: 273-282.

Fernandez AF, Esteller M (2010). Viral epigenomes in human tumorigenesis. Oncogene **29**: 1405-1420.

Ferreux E, Lont AP, Horenblas S, *et al* (2003). Evidence for at least three alternative mechanisms targeting the p16INK4A/cyclin D/Rb pathway in penile carcinoma, one of which is mediated by high-risk human papillomavirus. *J Pathol* **201**: 109-118.

Filion GJ, Zhenilo S, Salozhin S, *et al* (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol Cell Biol* **26**: 169-181.

Finnen RL, Erickson KD, Chen XS, Garcea RL (2003). Interactions between papillomavirus L1 and L2 capsid proteins. *J Virol* **77**: 4818-4826.

Flanagan JM (2007). Host epigenetic modifications by oncogenic viruses. Br J Cancer 96: 183-188.

Flores ER, Allen-Hoffmann BL, Lee D, Lambert PF (2000). The human papillomavirus type 16 E7 oncogene is required for the productive stage of the viral life cycle. *J Virol* **74**: 6622-6631.

Fouse SD, Shen Y, Pellegrini M, *et al* (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**: 160-169.

Franceschi S, Herrero R, Clifford GM, *et al* (2006). Variations in the age-specific curves of human papillomavirus prevalence in women worldwide. *Int J Cancer* **119:** 2677-2684.

Frommer M, McDonald LE, Millar DS, *et al* (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci USA* **89**: 1827-1831.

Fujita N, Watanabe S, Ichimura T, *et al* (2003). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *J Biol Chem* **278**: 24132-24138.

Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nature Genet* **24**: 8-91.

Fuks F, Burgers WA, Godin N, Kasai M, Kouzarides T (2001). Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J* **20**: 2536-2544.

Fuks F (2005). DNA methylation and histone modifications: teaming up to silence genes. *Curr Opin Genet Dev* **15**: 490-495.

Gal-Yam EN, Egger G, Iniguez L, *et al* (2008). Frequent switching of polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc Natl Acad Sci USA* **105**: 12979-12984.

Gardiner-Garden M, Frommer M (1987). CpG islands in vertebrate genomes. *J Mol Biol* **196**: 261-282.

Garner-Hamrick PA, Fostel JM, Chien WM, *et al* (2004). Global effects of human papillomavirus type 18 E6/E7 in an organotypic keratinocyte culture system. *J Virol* **78**: 9041-9050.

Gasco M, Bell AK, Heath V, *et al* (2002). Epigenetic inactivation of 14-3-3 sigma in oral carcinoma: association with p16(INK4A) silencing and human papillomavirus negativity. *Cancer Res* **62**: 2072-2076.

Gaston K, Fried M (1995). CpG methylation and the binding of YY1 and ETS proteins to the Surf-1/Surf-2 bidirectional promoter. *Gene* **157**: 257-259.

Geisen C, Denk C, Gremm B, *et al* (1997). High-level expression of the retinoic acid receptor beta gene in normal cells of the uterine cervix is regulated by the retinoic acid receptor alpha and is abnormally down-regulated in cervical carcinoma cells. *Cancer Res* **57**: 1460-1467.

Gien LT, Beauchemin MC, Thomas G (2010). Adenocarcinoma: a unique cervical cancer. *Gynecol Oncol* **116**: 140-146.

Gilliland FD, Harms HJ, Crowell RE, *et al* (2002). Glutathione S-transferase P1 and NADPH quinone oxidoreductase polymorphisms are associated with aberrant promoter methylation of P16(INK4a) and O(6)-methylguanine-DNA methyltransferase in sputum. *Cancer Res* **62**: 2248-2252.

Giraldo AM, Vaught TD, Fu L, *et al* (2009). Gene expression pattern and down-regulation of DNA methyltransferase 1 using siRNA in porcine somatic cells. *Gene Expr* **14**: 251-263.

Gloss B, Bernard HU (1990). The E6/E7 promoter of human papillomavirus type 16 is activated in the absence of E2 proteins by sequence-aberrant Sp1 distal element. *J Virol* **64**: 5577-5584.

Goldman LK, Glantz SA (1998). Evaluation of antismoking advertising campaigns. *JAMA* **279**: 772-777.

Graff JR, Herman JG, Myohanen S, Baylin SB, Vertino PM (1997). Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in *de novo* methylation. *J Biol Chem* **272**: 22322-22329.

Grassman K, Rapp B, Mashek H, Petry KU, 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. *J Virol* **70**: 2339-2349.

Greenbaum D, Colangelo C, Williams K, Gerstein M (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* **4**: 117.

Greger V, Passarge E, Hoppping W, Mesmer E, Horsthemke B (1989). Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum Genet* **83**: 155-158.

Grimshaw GM, Stanton A (2006). Tobacco cessation interventions for young people. *Cochrane Database Syst Rev* **4**: CD003289.

Grote HJ, Schmiemann V, Kiel S, *et al* (2004). Aberrant methylation of the adenomatous polyposis coli promoter 1A in bronchial aspirates from patients with suspected lung cancer. *Int J Cancer* **110**: 751-755.

Guastafierro T, Cecchinelli B, Zampier, M, *et al* (2008). CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J Biol Chem* **283:** 21873–21880.

Gustafson KS, Furth EE, Heitjan DF*, et al* (2004). DNA methylation profiling of cervical squamous intraepithelial lesions using liquid-based cytology specimens: an approach that utilizes receiver-operating characteristic analysis. *Cancer* **102**: 259-268.

Hammons GJ, Yan Y, Lopatina NG, *et al* (1999). Increased expression of hepatic DNA methyltransferase in smokers. *Cell Biol Toxicol* **15**: 389-394.

Harper DM, Franco EL, Wheeler CM, *et al* (2004). Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. *Lancet* **364**: 1757-1765.

Harper DM, Franco EL, Wheeler CM, *et al* (2006). Sustained efficacy up to 4.5 years of a bivalent L1 virus-like particle vaccine against human papillomavirus types 16 and 18: follow-up from a randomised control trial. *Lancet* **367**: 1247-1255.

Hasegawa M, Nelson HH, Peters E, et al (2002). Patterns of gene promoter methylation in squamous cell cancer of the head and neck. *Oncogene* **21**: 4231-4236.

Hata K, Okano M, Lei H, Li E (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129**: 1983-1993.

Hawley-Nelson P, Vousden KH, Hubbert NL, Lowy DR, Schiller JT (1989). HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *EMBO J* **8**: 3905-3910.

Heard E (2005). Recent advances in X-chromosome inactivation. *Curr Opin Cell Biol* **16**: 247-255.

Heard E, Rougeulle C, Arnaud D, *et al* (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* **107**: 727-738.

Hebner CM, Laimins LA (2006). Human papillomaviruses: basic mechanisms of pathogenesis and oncogenicity. *Rev Med Virol* **16**: 83-97.

Hecht SS (2003). Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nature Rev Cancer* **3**: 733-744.

Hegmann KT, Fraser AM, Keaney RP, *et al* (1993). The effect of age at smoking initiation on lung cancer risk. *Epidemiology* **4**: 444-448.

Heisler LE, Torti D, Boutros PC, *et al* (2005). CpG Island microarray probe sequences derived from a physical library are representative of CpG Islands annotated on the human genome. *Nucleic Acids Res* **3**: 2952-2961.

Hemminki K, Li X, Vaittinen P (2002). Time trends in the incidence of cervical and other genital squamous cell carcinomas and adenocarcinomas in Sweden 1958-1996. Eur J Obstet Gynecol Reprod Biol 101: 64-69.

Hendrich B, Bird A (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18**: 6538-6547.

Henken FE, Wilting SM, Overmeer RM, *et al* (2007). Sequential gene promoter methylation during HPV-induced cervical carcinogenesis. *Br J Cancer* **97**; 1457-1464.

Herdman MT, Pett MR, Roberts I, *et al* (2006). Interferon-beta treatment of cervical keratinocytes naturally infected with human papillomavirus 16 episomes promotes

rapid reduction in episome numbers and emergence of latent integrants. *Carcinogenesis* **27**: 2341-2353.

Herman JG, Merlo A, Mao L, *et al* (1995). Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res* **55**: 4525–4530.

Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci USA* **93**: 9821-9826.

Herman JG, Baylin SB (2003). Gene silencing in cancer in association with promoter hypermethylation. *New Engl J Med* **349**: 2042-2054.

Hino R, Uozaki H, Murakami N, *et al*. Activation of DNA methyltransferase 1 by EBV latent membrane protein 2A leads to promoter methylation of *PTEN* gene in gastric carcinoma. *Cancer Res* 2009; 69: 2766-2774.

Hirochika H, Borker TR, Chow LT (1987). Enhancers and trans-acting E2 transcription factors of papillomviruses. *J Virol* **61**: 2599-2606.

Ho GY, Burk RD, Klein S, *et al* (1995). Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. *J Natl Cancer Inst* **87**: 1365-1371.

Ho GYF, Bierman R, Beardsley L, *et al* (1998). Natural history of cervicovaginal papillomavirus infection in young women. *N Engl J Med* **338**: 423-428.

Holland D, Hoppe-Seyler K, Schuller B, *et al* (2008). Activation of the enhancer of zeste homologue 2 gene by the human papillomavirus E7 oncoprotein. *Cancer Res* **68**: 9964-9972.

Holliday R (1987). The inheritance of epigenetic defects. *Science* **238**: 163-170.

Howlett RI, Marrett LD, Innes MK, Rosen BP, McLachlin CM (2007). Decreasing incidence of cervical adenocarcinoma in Ontario: is this related to improved endocervical Pap test sampling? *Int J Cancer* **120**: 362-367.

Hu G, Liu W, Hanania EG, *et al* (1995). Suppression of tumorigenesis by transcription units expressing the antisense E6 and E7 messenger RNA (mRNA) for the transforming proteins of the human papilloma virus and the sense mRNA for the retinoblastoma gene in cervical carcinoma cells. *Cancer Gene Ther* **2**: 19-32.

Huang TH, Perry MR, Laux DE, *et al* (1999). Methylation profiling of CpG islands in human breast cancer cells. *Hum Mol Genet* **8**: 459-470.

Hussain S, Bharti AC, Salam I, *et al* (2009). Transcription factor AP-1 in esophageal squamous cell carcinoma: alterations in activity and expression during human papillomavirus infection. *BMC Cancer* **9**: 329.

Hutt JA, Vuillemenot BR, Barr EB, *et al* (2005). Life-span inhalation exposure to mainstream cigarette smoke induces lung cancer in B6C3F1 mice through genetic and epigenetic pathways. *Carcinogenesis* **26**: 1999-2009.

International Agency for Research on Cancer (1995). Human papillomaviruses. *IARC Monographs on the Evaluation of the Carcinogenic Risks of Chemicals to Humans* **64**: 1-409.

International Agency for Research on Cancer (2004). Tobacco smoke and involuntary smoking. *IARC Monographs on the Evaluation of the Carcinogenic Risks of Chemicals to Humans* **83**: 1-1438.

International Agency for Research on Cancer (2007). Human papillomaviruses. *IARC Monographs on the Evaluation of the Carcinogenic Risks of Chemicals to Humans* **90**: 1-678.

International Collaboration of Epidemiological Studies of Cervical Cancer (2006a). Comparison of risk factors for invasive squamous cell carcinoma and adenocarcinoma of the cervix: collaborative reanalysis of individual data on 8,097 women with squamous cell carcinoma and 1,374 women with adenocarcinoma from 12 epidemiological studies. *Int J Cancer* **120**: 885-891.

International Collaboration of Epidemiological Studies of Cervical Cancer, Appleby P, Beral V, *et al* (2006b). Carcinoma of the cervix and tobacco smoking: collaborative reanalysis of individual data on 13,541 women with carcinoma of the cervix and 23,017 women without carcinoma of the cervix from 23 epidemiological studies. *Int J Cancer* **118**: 1481-1495.

International Collaboration of Epidemiological Studies of Cervical Cancer (2006c). Cervical carcinoma and reproductive factors: collaborative reanalysis of individual data on 16,563 women with cervical carcinoma and 33,542 women without cervical carcinoma from 25 epidemiological studies. *Int J Cancer* **119**: 1108-1124.

International Collaboration of Epidemiological Studies of Cervical Cancer, Appleby P, Beral V, *et al* (2007). Cervical cancer and hormonal contraceptives: collaborative reanalysis of individual data for 16,573 women with cervical cancer and 35,509 women without cervical cancer from 24 epidemiological studies. *Lancet* **370**: 1609-1621.

International Collaboration of Epidemiological Studies of Cervical Cancer (2009). Cervical carcinoma and sexual behaviour: collaborative reanalysis of individual data on 15,461 women with cervical carcinoma and 29,164 women without cervical carcinoma from 21 epidemiological studies. *Cancer Epidemiol Biomarkers Prev* **18**:1060-1069.

Issa JP, Baylin SB, Belinsky SA (1996). Methylation of the estrogen receptor CpG island in lung tumors is related to the specific type of carcinogen exposure. *Cancer Res* **56**: 3655-3658.

Issa JP (2004). CpG island methylator phenotype in cancer. *Nat Rev Cancer* **4**: 988-993.

Issa JP (2007). DNA methylation as a therapeutic target in cancer. *Clin Cancer Res* **13**: 1634-1637.

Jacobson DL, Peralta L, Farmer M, *et al* (1999). Cervical ectopy and the transformation zone measured by computerized planimetry in adolescents. *Int J Gynaecol Obstet* **66**: 7-17.

Jaenisch R (1997). DNA methylation and imprinting: why bother? *Trends Genet* **13**: 323-329.

Jaenisch R, Bird A (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33**: 245-254.

Jarmalaite S, Kannio A, Anttila S, Lazutka JR, Husgafvel-Pursiainen K (2003). Aberrant p16 promoter methylation in smokers and former smokers with nonsmall cell lung cancer. *Int J Cancer* **106**: 913-918.

Jeon S, Allen-Hoffmann BL, Lambert PF (1995). Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J Virol* **69**: 2989-2997.

Jones CJ, Schiffman MH, Kurman R, Jacob P, Benowitz NL (1991). Elevated nicotine levels in cervical lavages from passive smokers. *Am J Public Health* **81**: 378-379.

Jones PA, Taylor SM (1980). Cellular differentiation, cytidine analogs and DNA methylation. *Cell* **20**: 85-93.

Jones PA, Baylin SB (2002). The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3**: 415-428.

Jones PL, Veenstra GJ, Wade PA, *et al* (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**: 187-191.

Kalantari M, Calleja-Macias IE, Tewari D, *et al* (2004). Conserved methylation patterns of human papillomavirus type 16 DNA in asymptomatic infection and cervical neoplasia. *J Virol* **78**: 12762-12772.

Kalantari M, Lee D, Calleja-Macias IE, Lambert PF, Bernard HU (2008). Effects of cellular differentiation, chromosomal integration and 5-aza-2'-deoxycytidine treatment on human papillomavirus-16 DNA methylation in cultured cell lines. *Virology* **374**: 292-303.

Kalantari M, Chase DM, Tewari KS, Bernard HU (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**: 311-320.

Kantarjian H, Oki Y, Garcia-Manero G, *et al* (2007). Results of a randomized study of 3 schedules of low-dose decitabine in higher-risk myelodysplastic syndrome and chronic myelomonocytic leukemia. *Blood* **109**: 52-57.

Karpf AR, Matsui S (2005). Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. *Cancer Res* **65**: 8635-8639.

Karstensen B, Poppelreuther S, Bonin M, *et al* (2006). Gene expression profiles reveal an up-regulation of E2F and down-regulation of interferon targets by HPV18 but no changes between keratinocytes with integrated or episomal viral genomes. *Virology* **353**: 200-209.

Kersting M, Friedl C, Kraus A, *et al* (2000). Differential frequencies of p16(INK4a) promoter hypermethylation, p53 mutation, and K-ras mutation in exfoliative material mark the development of lung cancer in symptomatic chronic smokers. *J Clin Oncol* **18**: 3221-3229.

Keshet I, Schlesinger Y, Farkash S, *et al* (2006). Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat Genet* **38**: 149-153.

Kikuchi S, Yamada D, Fukami T, *et al* (2006). Hypermethylation of the TSLC1/IGSF4 promoter is associated with tobacco smoking and a poor prognosis in primary nonsmall cell lung carcinoma. *Cancer* **106**: 1751-1758.

Kim DH, Nelson HH, Wiencke JK, *et al* (2001). p16(INK4a) and histology-specific methylation of CpG islands by exposure to tobacco smoke in non-small cell lung cancer. *Cancer Res* **61**: 3419-3424.

Kim DH, Kim JS, Ji YI, (2003). Hypermethylation of RASSF1A promoter is associated with the age at starting smoking and a poor prognosis in primary non-small cell lung cancer. *Cancer Res* **63**: 3743-3746.

Kim GD, Ni J, Kelesoglu N, Roberts RJ, Pradhan S (2002). Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *EMBO J* **21**: 4183-195.

Kim H, Kwon YM, Kim JS, *et al* (2004). Tumor-specific methylation in bronchial lavage for the early detection of non-small-cell lung cancer. *J Clin Oncol* **22**: 2363-2370.

Kim K, Garner-Hamrick PA, Fisher C, Lee D, Lambert PF (2003). Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection. *J Virol* **77**: 12450-12459.

Kirnbauer R, Taub J, Greenstone H, *et al* (1993). Efficient self-assembly of human papillomavirus type 16 L1 and L1-2 into virus-like particles. *J Virol* **67**: 6929-6936.

Kleinjan M, van den Eijnden RJ, Engels RC (2009). Adolescents' rationalizations to continue smoking: the role of disengagement beliefs and nicotine dependence in smoking cessation. *Addict Behav* **34**: 440–445.

Klingelhutz AJ, Foster SA, McDougall JK (1996). Telomerase activation by the E6 gene product of human papillomavirus type 16. *Nature* **380**: 79-82.

Klose RJ, Bird AP (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**: 89-97.

Knight GL, Turnell AS, Roberts S (2006). Role for Wee1 in inhibition of G2-to-M transition through the cooperation of distinct human papillomavirus type 1 E4 proteins. *J Virol* **80**: 7416-7426.

Knoke JD, Shanks TG, Vaughn JW, Thun MJ, Burns DM (2004). Lung cancer mortality is related to age in addition to duration and intensity of cigarette smoking. *Cancer Epid Biomarkers Prev* **13**: 949-957.

Kondo Y, Shen L, Cheng AS, *et al* (2008). Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. *Nat Genet* **40**: 741-750.

Koutsky LA, AultK, Wheeler CM, *et al* (2002). A controlled trial of a human papillomavirus type 16 vaccine. *New Engl J Med* **346**: 1645-1651.

Kovacic MB, Castle PE, Herrero R, et al (2006). Relationships of human papillomavirus type, qualitative viral load, and age with cytologic abnormality. *Cancer Res* **66**: 10112-10119.

Kulkarni V, Saranath D (2004). Concurrent hypermethylation of multiple regulatory genes in chewing tobacco associated oral squamous cell carcinomas and adjacent normal tissues. *Oral Oncol* **40**: 145-153.

Kuner R, Vogt M, Sultmann H, *et al* (2007). Identification of cellular targets for the human papillomavirus E6 and E7 oncogenes by RNA interference and transcriptome analyses. *J Mol Med* **85**:1253 – 1262.

Kwon YM, Park JH, Kim H, *et al* (2007). Different susceptibility of increased DNMT1 expression by exposure to tobacco smoke according to histology in primary non-small cell lung cancer. *J Cancer Res Clin Oncol* **133**: 219-226.

Lai HC, Lin YW, Huang TH, *et al* (2008). Identification of novel DNA methylation markers in cervical cancer. *Int J Cancer* **123**: 161-167.

Laimins LA (1996). Human papillomaviruses target differentiating epithelium for virion production and malignant conversion. *Semin Virol* **7**: 305-313.

Lander ES, Linton EM, Birren B *et al* (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.

Laurson J, Khan S, Chung R, Cross K, Raj K (2010). Epigenetic repression of E-cadherin by human papillomavirus 16 E7 protein. *Carcinogenesis* **31**: 918-926.

Lea JS, Coleman R, Kurien A, *et al* (2004). Aberrant p16 methylation is a biomarker for tobacco exposure in cervical squamous cell carcinogenesis. *Am J Obstet Gynecol* **190**: 674-679.

Lee EJ, Lee BB, Kim JW, *et al* (2006). Aberrant methylation of Fragile Histidine Triad gene is associated with poor prognosis in early stage esophageal squamous cell carcinoma. *Eur J Cancer* **42**: 972-980.

Lee HY, Dawson MI, Walsh GL, *et al* (1996). Retinoic acid receptor- and retinoid X receptor-selective retinoids activate signaling pathways that converge on AP-1 and inhibit squamous differentiation in human bronchial epithelial cells. *Cell Growth Differ* **7**: 997-1004.

Lee JO, Kwun HJ, Jung JK, *et al* (2005). Hepatitis B virus X protein represses E-cadherin expression via activation of DNA methyltransferase 1. *Oncogene* **24**: 6617-6625.

Lee TI, Jenner RG, Boyer LA, *et al* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **126**: 301-313.

Lehman CW, Botchan MR (1998). Segregation of viral plasmids depends on tethering to chromosomes and is regulated by phosphorylation. *Proc Natl Acad Sci USA* **95**: 4338-4343.

Lehnertz B, Ueda Y, Derijck AA, *et al* (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* **13**: 1192-1200.

Leng S, Stidley CA, Willink R, *et al* (2008). Double-strand break damage and associated DNA repair genes predispose smokers to gene methylation. *Cancer Res* **68**: 3049-3056.

Li E, Bestor TH, Jaenisch R (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**: 915-926.

Li E, Beard C, Jaenisch R (1993). Role for DNA methylation in genomic imprinting. *Nature* **366**: 362-365.

Li E (2002). Chromatin modification and epigentic reprogramming in mammalian development. *Nature Rev Genet* 3: 662-673.

Li HP, Leu YW, Chang SW (2005). Epigenetic changes in virus-associated human cancers. *Cell Res* **15**: 262-271.

Lichtinghagen R, Musholt PB, Lein M, *et al* (2002). Different mRNA and protein expression of matrix metalloproteinases 2 and 9 and tissue inhibitor of metalloproteinases 1 in benign and malignant prostate tissue. *Eur Urol* **42**: 398-406.

Lin F, Xiao D, Kolluri SK, Zhang XK (2000). Unique anti-activator protein-1 activity of retinoic acid receptor beta. *Cancer Res* **60**: 3271-3280.

Lin RK, Hsu HS, Chang JW, *et al* (2007). Alteration of DNA methyltransferases contribute to 5'CpG methylation and poor prognosis in lung cancer. *Lung Cancer* **55**: 205-213.

Liu H, Zhou H, Boggs SE, *et al* (2007). Cigarette smoke induces demethylation of prometastatic oncogene synclein-gamma in lung cancer cells by down-regulation of DNMT3B. *Oncogene* **26**: 5900-5910.

Liu J, Lian Z, Han S, *et al* (2006). Downregulation of E-cadherin by hepatitis B virus X antigen in hepatocellular carcinoma. *Oncogene* **25**: 1008-1017.

Liu Y, Lan Q, Siegfried JM, Luketich JD, Keohavong P (2006). Aberrant promoter methylation of p16 and MGMT genes in lung tumors from smoking and never-smoking lung cancer patients. *Neoplasia* **8**: 46-51.

Lotan Y, Xu XC, Shalev M, *et al* (2000). Differential expression of nuclear retinoid receptors in normal and malignant prostates. *J Clin Oncol* **18**: 116-121.

Lu CY, Hsieh SY, Lu YJ, *et al* (2009). Aberrant DNA methylation profile and frequent methylation of KLK10 and OXGR1 genes in hepatocellular carcinoma. *Genes Chromosomes Cancer* **48**: 1057-1068.

Lu YJ, Wu CS, Li HP, *et al* (2010). Aberrant methylation impairs low density lipoprotein receptor-related protein 1B tumor suppressor function in gastric cancer. *Genes Chromosomes Cancer* **49**: 412-424.

Machida EO, Brock MV, Hooker CM, *et al* (2006). Hypermethylation of ASC/TMS1 is a sputum marker for late-stage lung cancer. *Cancer Res* **66**: 6210-6218.

Mack DH, Laimins LA (1991). A keratinocytes-specific transcription factor, KRF-1, interacts with AP-1 to activate expression of human papillomavirus type 18 in squamous epithelial cells. *Proc Natl Acad Sci USA* **88**: 9102-9106.

Marsit CJ, Liu M, Nelson HH, *et al* (2004). Inactivation of the Fanconi anemia/BRCA pathway in lung and oral cancers: implications for treatment and survival. *Oncogene* **23**: 1000-1004.

Marsit CJ, Kim DH, Liu M, *et al* (2005). Hypermethylation of RASSF1A and BLU tumor suppressor genes in non-small cell lung cancer: implications for tobacco smoking during adolescence. *Int J Cancer* **114**: 219-223.

Marsit CJ, McClean MD, Furniss CS, Kelsey KT (2006). Epigenetic inactivation of the SFRP genes is associated with drinking, smoking and HPV in head and neck squamous cell carcinoma. *Int J Cancer* **119**: 1761-1766.

Marsit CJ, Karagas MR, Danaee H, *et al* (2007). Carcinogen exposure and gene promoter hypermethylation in bladder cancer. *Carcinogenesis* **27**: 112-116.

Masterson PJ, Stanley MA, Lewis AP, Romanos MA (1998). A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase alpha-primase p68 subunit. *J Virol* **72**: 7407-7419.

Mathew A, George PS (2009). Trends in incidence and mortality rates of squamous cell carcinoma and adenocarcinoma of cervix—worldwide. *Asian Pac J Cancer Prev* **10**: 645-650.

Matsukura T, Koi S, Sugase M (1989). Both episomal and integrated forms of human papillomavirus type 16 are involved in invasive cervical cancers. *Virology* **172**: 63-72.

Matsuzaki K, Deng G, Tanaka H, *et al* (2005). The relationship between global methylation level, loss of heterozygosity, and microsatellite instability in sporadic colorectal cancer. *Cancer Res* **11**: 8564-8569.

Mayer W, Niveleau A, Walter J, Fundele R and Haaf T (2000). Demethylation of the zygotic paternal genome. *Nature* **403**: 501-502.

McCabe MT, Davis JN, Day ML (2005). Regulation of DNA methyltransferase 1 by the pRb/E2F1 pathway. *Cancer Res* **65**: 3624-3632.

McCann MF, Irwin DE, Walton LA, *et al* (1992). Nicotine and cotinine in the cervical mucus of smokers, passive smokers, and nonsmokers. *Cancer Epidem Biomarkers Prev* **1**: 125-129.

McGarvey KM, Fahrner JA, Greene E, *et al* (2006). Silenced tumor suppressor genes reactivated by DNA demethylation do not return to a fully euchromatic chromatin state. *Cancer Res* **66**: 3541-3549.

McGarvey KM, Greene E, Fahrner JA, Jenuwein T, Baylin SB (2007). DNA methylation and complete transcriptional silencing of cancer genes persist after depletion of EZH2. *Cancer Res* **67**: 5097-5102.

McLaughlin-Drubin ME, Munger K (2008). Viruses associated with human cancer. *Biochim Biophys Acta* **1782**: 127-150.

McVea KL, Miller DL, Creswell JW, McEntarrfer R, Coleman MJ (2009). How adolescents experience smoking cessation. *Qual Health Res* **19**: 580–592.

Merlo A, Herman JG, Mao L, *et al* (1995). 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med* **1**: 686–692.

Modis Y, Trus BL, Harrison SC (2002). Atomic model of the papillomavirus capsid. *EMBO J* **21**: 4754-4762.

Mortusewicz O, Schermelleh L, Walter J, Cardoso MC. Leonhardt H (2005). Recruitment of DNA methyltransferase I to DNA repair sites. *Proc Natl Acad Sci USA* **102**: 8905-8909.

Münger K, Phelps WC, Bubb V, Howley PM, Schlegel R (1989). The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *J Virol* **63**: 4417-4421.

Mund C, Hackanson B, Stresemann C, Lübbert M, Lyko F (2005). Characterization of DNA demethylation effects induced by 5-Aza-2'-deoxycytidine in patients with myelodysplastic syndrome. *Cancer Res* **65**: 7086-7090.

Muñoz N, Bosch X, de Sanjosé S, *et al* (2003). Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* **348**: 518-527.

Myung SK, McDonnell DD, Kazinets G, Seo HG, Moskowitz JM (2009). Effects of Web- and computer-based smoking cessation programs: meta-analysis of randomized controlled trials. *Arch Intern Med* **169**: 929-937.

Nakata S, Sugio K, Uramoto H, *et al* (2006). The methylation status and protein expression of CDH1, p16(INK4A), and fragile histidine triad in nonsmall cell lung carcinoma: epigenetic silencing, clinical features, and prognostic significance. *Cancer* **106**: 2190-2199.

Nan HM, Song YJ, Yun HY, Park JS, Kim H (2005). Effects of dietary intake and genetic factors on hypermethylation of the hMLH1 gene promoter in gastric cancer. *World J Gastroenterol* **11**: 3834-3841.

Nan X, Meehan RR, Bird A (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res* **21**: 4886-4892.

Nan X, Campoy FJ, Bird A (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88**: 471-481.

Nan X, Ng HH, Johnson CA, *et al* (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**: 386-389.

Narayan G, Arias-Pulido H, Koul S*, et al* (2003). Frequent promoter methylation of CDH1, DAPK, RARB, and HIC1 genes in carcinoma of cervix uteri: its relationship to clinical outcome. *Mol Cancer* **2**: 24.

Nees M, Geoghegan JM, Hyman T, *et al* (2001). Papillomavirus type 16 oncogenes downregulate expression of interferon-responsive genes and upregulate proliferation-associated and NF-kappaB-responsive genes in cervical keratinocytes. *J Virol* **75**: 4283-4296.

Ng HH, Zhang Y, Hendrich B, *et al* (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* **23**: 58-61.

Ng HH, Jeppensen P, Bird A (2000). Active repression of methylated genes by the chromosomal protein MBD1. *Mol Cell Biol* **20**: 1394-1406.

Nordlund LA, Carstensen JM, Pershagen G (1997). Cancer incidence in female smokers: a 26-year follow-up. *Int J Cancer* **73**: 625-628.

Nouzova M, Holtan N, Oshiro MM, *et al* (2004). Epigenomic changes during leukemia cell differentiation: analysis of histone acetylation and cytosine methylation using CpG island microarrays. *J Pharmacol Exp Ther* **311**: 968-981.

O'Nions J, Brooks LA, Sullivan A, *et al* (2001). P73 is over-expressed in vulval cancer principally as the delta-2 isoform. *Br J Cancer* **85**: 1551-1556.

O'Riain C, O'Shea DM, Yang Y, *et al* (2009). Array-based DNA methylation profiling in follicular lymphoma. *Leukemia* **23**: 1858-1866.

Offord EA, Beard P (1990). A member of the activator protein 1 family found in keratinocytes but not in fibroblasts required for transcription from a human papillomavirus type 18 promoter. *J Virol* **64**: 4792-4798.

Ohm JE, McGarvey KM, Yu X, *et al* (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* **39**: 237-242.

Okano M, Bell DW, Haber DA, Li E (1999). DNA methyltransferases dnmt3a and dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**: 247-257.

Ostör AG (1993). Natural history of cervical intraepithelial neoplasia: a critical review. *Int J Gynecol Pathol* **12**: 186-192.

Oswald J, Engemann S, Lane N, *et al* (2000). Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* **10**: 475-478.

Paavonen J, Jenkins D, Bosch FX, *et al* (2007). Efficacy of a prophylactic adjuvanted bivalent L1 virus-like-particle vaccine against infection with human papillomavirus types 16 and 18 in young women: an interim analysis of a phase III double-blind, randomised controlled trial. *Lancet* **369**: 2161-2170.

Paavonen J, Naud P, Salmerón J, *et al* (2009). Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. *Lancet* **374**: 301-314.

Palmisano WA, Divine KK, Saxxomanno G, *et al* (2000). Predicting lung cancer by detecting aberrant promoter methylation in sputum. *Cancer Res* **60**: 5954-5958.

Parker JN, Zhao W, Askins KJ, Broker TR, Chow LT (1997). Mutational analyses of differentiation-dependent human papillomavirus type 18 enhancer elements in epithelial raft cultures of neonatal foreskin keratinocytes. *Cell Growth Differ* **8**: 751-762.

Parkin DM (2006). The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* **118**: 3034-3044.

Parkin DM, Bray F (2006). Chapter 2: the burden of HPV-related cancers. *Vaccine* **24 (Suppl 3)**: S11-S25.

Pascal LE, True LD, Campbell DS, *et al* (2008). Correlation of mRNA and protein levels: cell type-specific gene expression of cluster designation antigens in prostate. *BMC Genomics* **9**: 246.

Patel D, Huang SM, Baglia LA, McCance DJ (1999). The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. *EMBO J* **18**: 5061-5072.

Pett MR, Alazawi WOF, Roberts I, *et al* (2004). Acquisition of high-level chromosomal instability is associated with integration of human papillomavirus type 16 in cervical keratinocytes. *Cancer Res* **64**: 1359-1368.

Pett MR, Herdman MT, Palmer RD, *et al* (2006). Selection of cervical keratinocytes containing integrated HPV16 associates with episome loss and an endogenous antiviral response. *Proc Natl Acad Sci USA* **103**: 3822-3827.

Phillips DH, Ni Shé M (1993). Smoking-related DNA adducts in human cervical biopsies. *IARC Sci Publ* **124**: 327-330.

Plath K, Fang J, Mlynarczyk-Evans SK, *et al* (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**: 131-135.

Prendergast GC, Ziff EB (1991). Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science* **251**: 186-189.

Prokhortchouk A, Hendrich B, Jørgensen H, *et al* (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev* **15**: 1613-1618.

Prusty BK, Das BC (2005). Constitutive activation of transcription factor AP-1 in cervical cancer and suppression of human papillomavirus (HPV) transcription in HeLa cells by curcumin. *Int J Cancer* **113**: 951-960.

Pulling LC, Vuillemenot BR, Hutt JA, Devereux TR, Belinsky SA (2004). Aberrant promoter hypermethylation of the death-associated protein kinase gene is early and frequent in murine lung tumors induced by cigarette smoke and tobacco carcinogens. *Cancer Res* **64**: 3844-3848.

Pyeon D, Newton MA, Lambert PA, *et al* (2007). Fundamental differences in cell cycle dergulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. *Cancer Res* **67**: 4605-4619.

Qiu H, Zhang W, El-Naggar AK, *et al* (1999). Loss of retinoic acid receptor-beta expression is an early event during esophageal carcinogenesis. *Am J Pathol* **155**: 1519-1523.

Ramsahoye BH, Biniszkiewicz W, Lyko F *et al* (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci USA* **97**: 5237-5242.

Ravenel JD, Broman KW, Perlman EJ, *et al* (2002). Loss of imprinting of insulin-like growth factor-II (IGF2) gene in distinguishing specific biologic subtypes of Wilms tumor. *J Natl Cancer Inst* **93**: 1698-1703.

Reichrath J, Mittmann M, Kamradt J, Muller SM (1997). Expression of retinoid-X receptors (-alpha,-beta,-gamma) and retinoic acid receptors (-alpha,-beta,-gamma) in normal human skin: an immunohistological evaluation. *Histochem J* **29**: 127-133.

Reik W, Walter J (2001). Genomic imprinting: parental influence on the genome. *Nature Rev Genet* **2**: 21-32.

Remmink AJ, Walboomers JM, Helmerhorst TJ, *et al* (1995). The presence of persistent high-risk HPV genotypes in dysplastic cervical lesions is associated with progressive disease: natural history up to 36 months. *Int J Cancer* **61**: 316-311.

Remus R, Kämmer C, Heller H, *et al* (1999). Insertion of foreign DNA into an established mammalian genome can alter the methylation of cellular DNA sequences. *J Virol* **73**: 1010-1022.

Rhee I, Bachman KE, Park BH, *et al* (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416**: 552-556.

Rhee I, Jair KW, Yen RW, *et al* (2000). CpG methylation is maintained in human cancer cells lacking DNMT1. *Nature* **404**: 1003-1007.

Richart RM (1990). A modified terminology for cervical intraepithelial neoplasia. *Obstet Gynecol* **75**: 131-133.

Ringrose L, Ehret H, Paro R (2004). Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. *Mol Cell* **16**: 641-653.

Robertson KD, Ait-Si-Ali S, Yokochi T, et al (2000). DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nature Genet* **25**: 338-342.

Romanczuk H, Thierry F, Howley PM (1990). Mutational analysis of cis elements involed in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters. *J Virol* **64**: 2849-2859.

Romanczuk H, Howley PM (1992). Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. *Proc Natl Acad Sci USA* **89**: 3159-3163.

Ronaghi M, Uhlén M and Nyrén P (1998). DNA sequencing: a sequencing method based on real-time pyrophosphate. *Science* **281**: 363-365.

Rosty C, Sheffer M, Tsafrir D, *et al* (2005). Identification of a proliferation gene cluster associated with HPV E6/E7 expression level and viral DNA load in invasive cervical carcinoma. *Oncogene* **24**: 7094-7104.

Rougier N, Bourchis D, Gomes DM, *et al* (1998). Chromosome methylation patterns during mammalian preimplantation development. *Genes Dev* **12**: 2108-2113.

Rountree M, Bachman KE, Baylin SB (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nature Genet* **25**: 269-277.

Ruesch MN, Laimins LA (1998). Human papillomavirus oncoproteins alter differentiation-dependent cell cycle exit on suspension in semisolid medium. *Virology* **250**: 19-29.

Russo AL, Thiagalingam A, Pan H, *et al* (2005). Differential DNA hypermethylation of critical genes mediates the stage-specific tobacco smoke-induced neoplastic progression of lung cancer. *Clin Cancer Res* **11**: 2466-2470.

Sado T, Fenner MH, Tan SS, *et al* (2000). X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol* **225**: 294-303.

Sakai T, Toguchida J, Ohtani N, *et al* (1991). Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am J Hum Genet* **48**: 880-888.

Santin AD, Zhan F, Bignotti E, *et al* (2005). Gene expression profiles of HPV16- and HPV18-infected early stage cervical cancers and normal cervical epithelium: identification of novel candidate molecular markers for cervical cancer diagnosis and therapy. *Virology* **331**: 269-291.

Sarraf SA, Stancheva I (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol Cell* **15**:595-605.

Sasieni P, Castanon A, Cuzick J (2009). Screening and adenocarcinoma of the cervix. *Int J Cancer* **125**: 525-529.

Sasson IM, Haley NJ, Hoffman D, *et al* (1985). Cigarette smoking and neoplasia of the uterine cervix: smoke constituents in cervical mucus. *N Engl J Med* **312**: 315-316.

Sato N, Anirban M, Fukishama N, *et al* (2003). Frequent hypomethylation of multiple genes overexpressed in pancreatic ductal adenocarcinoma. *Cancer Res* **63**: 4158-4166.

Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM (1990). The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **63**:1129-1136.

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

Schlemmer BO, Sorensen BS, Overgaard J, *et al* (2004). Quantitative PCR--new diagnostic tool for quantifying specific mRNA and DNA molecules: HER2/neu DNA quantification with LightCycler real-time PCR in comparison with immunohistochemistry and fluorescence in situ hybridization. *Scand J Clin Lab Invest* **64**: 511-522.

Schlesinger Y, Straussman R, Keshet I, *et al* (2007). Polycomb-mediated methylation of Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* **39**: 232-236.

Schwartz SM, Weiss NS (1986). Increased incidence of adenocarcinoma of the cervix in young women in the United States. *Am J Eidemiol* **124**: 1045-1047.

Sedman SA, Barbosa MS, Vass WC, et al (1991). The full-length E6 protein of human papillomavirus type 16 has transforming and trans-activating activities and cooperates with E7 to immortalize keratinocytes in culture. *J Virol* **65**: 4860-4866.

Serrano M, Hannon GJ, Beach D (1993). A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* **366**: 704-707.

Shivapurkar N, Sherman ME, Stastny V, *et al* (2007). Evaluation of candidate methylation markers to detect cervical neoplasia. *Gynecol Oncol* **107**: 549-553.

Silverman LR, Demakos EP, Peterson BL, *et al* (2002). Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. *J Clin Oncol* **20**: 2429-2440.

Smith HO, Tiffany MF, Qualls CR, Key CR (2000). The rising incidence of adenocarcinoma relative to squamous cell carcinoma of the uterine cervix in the United States – 24 year population based study. *Gynecol Oncol* **78**: 97-105.

Smith JS, Linsay L, Hoots B, *et al* (2007). Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions. *Int J Cancer* **121**: 621-632.

Soma T, Kagoni J, Kawabe A, *et al* (2006). Nicotine induces the fragile histidine triad methylation in human esophageal squamous epithelial cell. *Int J Cancer* **119**: 1023-1027.

Sorm F, Pískala A, Cihák A, Veselý J (1964). 5-Azacytidine, a new, highly effective cancerostatic. *Experientia* **20**: 202-203.

Soto U, Das BC, Lengert M, *et al* (1999). Conversion of HPV 18 positive non-tumorigenic HeLa-fibroblast hybrids to invasive growth involves loss of TNF-alpha mediated repression of viral transcription and modifications of the AP-1 transcription complex. *Oncogene* **18**: 3187-3198.

Stanley MA, Browne HM, Appleby M, Minson AC (1989). Properties of a non-tumorigenic human cervical keratinocyte cell line. *Int J Cancer* **43**: 672-676.

Stanley M (2006). Immune responses to human papillomavirus. *Vaccine* **24 Suppl 1**: S16-22.

Stanley MA, Pett MR, Coleman N (2007). HPV: from infection to cancer. *Biochem Soc Trans* **35**:1456-1460.

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.

Stevens A, Lowe J (1992). **Histology**. 1st ed. London: Mosby-Year Book Europe Ltd.

Suga Y, Miyajima K, Oikawa T, *et al* (2008). Quantitative p16 and ESR1 methylation in the peripheral blood of patients with non-small cell lung cancer. *Oncol Rep* **20**: 1137-1142.

Sugawara K, Fujinaga K, Yamashita T, Ito Y (1983). Integration and methylation of shope papillomavirus DNA in the transplantable Vx2 and Vx7 rabbit carcinomas. *Virology* **131**: 88-99.

Suzuki H, Gabrielson E, Chen W, et al (2002). A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. *Nat Genet* **31**: 141-149.

Sved J and Bird A (1990). The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc Natl Acad Sci USA* **87**: 4692-4696.

Swafford DS, Middeton SK, Palmisano WA, *et al* (1997). Frequent aberrant methylation of p16^{INK4a} in primary rat lung tumors. *Mol Cell Biol* **17**: 1366-1374.

Takai D and Jones PA (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci USA* **99**: 3740-3745.

Thain A, Jenkins O, Clarke AR, Gaston K (1996). CpG methylation directly inhibits binding of the human papillomavirus type 16 E2 protein to specific DNA sequences. *J Virol* **70**: 7233-7235.

The FUTURE II Study Group (2007). Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. *New Engl J Med* **356**: 1915-1927.

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**: 3740-3748.

Thorland EC, Myers SL, Persing DH, et al (2000). Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. *Cancer Res* **60**: 5916-5921.

Tomizawa Y, Iijima H, Nomoto T, *et al* (2004). Clinicopathological significance of aberrant methylation of RARbeta2 at 3p24, RASSF1A at 3p21.3, and FHIT at 3p14.2 in patients with non-small cell lung cancer. *Lung Cancer* **46**: 305-312.

Tost J, Dunker J, Gut IG (2003). Analysis and quantification of multiple methylation variable positions in CpG islands by pyrosequencing. *Biotechniques* **35**: 152-156.

Toyooka S, Maruyama R, Toyooka KO, *et al* (2003). Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer. *Int J Cancer* **103**: 153-160.

Tsai CN, Tsai CL, Tse KP, Chang HY, Chang YS (2002). The Epstein-Barr virus oncogene product, latent membrane protein 1, induces the downregulation of E-cadherin gene expression via activation of DNA methyltransferases. *Proc Natl Acad Sci USA* **99**: 10084-10089.

Turan T, Kalantari M, Calleja-Macias IE, *et al* (2006). Methylation of the human papillomavirus-18 L1 gene: a biomarker of neoplastic progression? *Virology* **349**: 175-183.

Turan T, Kalantari M, Cuschieri K, *et al* (2007). High-throughput detection of human papillomavirus-18 L1 gene methylation, a candidate biomarker for the progression of cervical neoplasia. *Virology* **361**: 185-193.

Turkor MS, Bestor TH (1997). Formation of methylation patterns in the mammalian genome. *Mutat Res* **386**: 119-130.

Vaissière T, Hung RJ, Zaridze D, *et al* (2009). Quantitative analysis of DNA methylation profiles in lung cancer identifies aberrant DNA methylation of specific genes and its association with gender and cancer risk factors. *Cancer Res* **69**: 243-252.

van der Burg B, Slager-Davidov R, van der Leede BM, de Laat SW, van der Saag PT (1995). Differential regulation of AP1 activity by retinoic acid in hormone-dependent and -independent breast cancer cells. *Mol Cell Endocrinol* **112**: 143-152.

van Doorn R, Zoutman WH, Dijkman R, *et al* (2005). Epigenetic profiling of cutaneous T-cell lymohoma: promoter hypermethylation of multiple tumour suppressor genes including BCL7a, PTPRG, and p73. J Clin Oncol 23: 3886-389.

Van Tine BA, Kappes JC, Banerjee NS, *et al* (2004). Clonal selection for transcriptionally active viral oncogenes during progression to cancer. *J Virol* **78**: 11172-11186.

Venter JC, Adams MD, Myers EW *et al* (2001). The sequence of the human genome. *Science* **291**: 1304-1351.

Viallet JP, Dhouailly D (1994). Retinoic acid and mouse skin morphogenesis. I. Expression pattern of retinoic acid receptor genes during hair vibrissa follicle, plantar, and nasal gland development. *J Invest Dermatol* **103**: 116-121.

Villa LL, Costa RL, Petta CA, *et al* (2005). Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. *Lancet Oncol* **6**: 271-278.

Villa LL, Costa RL, Petta CA, *et al* (2006). High sustained efficacy of a prophylactic quadrivalent human papillomavirus types 6/11/16/18 L1 virus-like particle vaccine through 5 years of follow-up. *Br J Cancer* **95**: 1459-1466.

Viré E, Brenner C, Deplus R, *et al* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**: 871-874.

Virmani AK, Muller C, Rathi A, *et al* (2001). Aberrant methylation during cervical carcinogenesis. *Clin Cancer Res* **7**: 584-589.

von Zeidler SV, Miracca EC, Nagai MA, Birman EG (2004). Hypermethylation of the p16 gene in normal oral mucosa of smokers. *Int J Mol Med* **14**: 807-811.

Vuillemenot BR, Pulling LC, Palmisano WA, Hutt JA, Belinsky SA (2004). Carcinogen exposure differentially modulates RAR-β promoter hypermethylation, an early and frequent event in mouse lung carcinogenesis. *Carcinogenesis* **25**: 623-629.

Vuillemenot BR, Hutt JA, Belinsky SA (2006). Gene promoter hypermethylation in mouse lung tumors. *Mol Cancer Res* **4**: 267-273.

Wade PA, Gegonne A, Jones PL, *et al* (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat Genet* **23**: 62-66.

Walboomers JM, Jacobs MV, Manos MM, *et al* (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* **189**: 12-19.

Walsh CP, Chaillet R, Bestor TH (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genet* **20**: 116-117.

Wang RY, Gehrke CW, Ehrlich M (1980). Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res* **8**:4777-4790.

Wang Y, Yu Q, Cho AH, *et al* (2005). Survey of differentially methylated promoters in prostate cancer cell lines. *Neoplasia* **7**: 748-760.

Watson RE, Curtin GM, Hellmann GM, Doolittle DJ, Goodman JI (2004). Increased DNA methylation in the *HOXA5* promter region correlates with decreased expression of the gene during tumor promotion. *Mol Carcinog* **41**: 54-66.

Weber M, Hellmann I, Stadler MB, *et al* (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* **39**: 457-466.

Wei SH, Chen CM, Strathdee G, *et al* (2002). Methylation microarray analysis of late-stage ovarian carcinomas distinguishes progression-free survival in patients and identifies candidate epigenetic markers. *Clin Cancer Res* **8**: 2246-2252.

Weinberg RA (1995). The retinoblastoma protein and cell cycle control. *Cell* **81**: 323-330.

Wells M (1996). Female genital tract. In: Underwood JCE (ed). **General and systematic pathology**. 2nd ed. London: Churchill Livingstone.

Wentzensen N, Vinokurova S, von Knebel Doeberitz M (2004). Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive carinoma of the female lower genital tract. Cancer Res 64: 3878-3884.

Werness BA, Levine AJ, Howley PM (1990). Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science* **248**: 76-79.

Wettstein FO, Stevens JG (1983). Shope papillomavirus DNA is extensively methylated in non-virus-producing neoplasms. *Virology* **126**: 493-504.

Widschwendter A, Ivarsson L, Blassnig A, *et al* (2004). CDH1 and CDH13 methylation in serum is an independent prognostic marker in cervical cancer patients. *Int J Cancer* **109**: 163-166.

Widschwendter M, Berger J, Daxenbichler G, *et al* (1997). Loss of retinoic acid receptor beta expression in breast cancer and morphologically normal adjacent tissue but not in the normal breast tissue distant from the cancer. *Cancer Res* **57**: 4156-4161.

Widschwendter M, Fiegl H, Egle D, *et al* (2007). Epigenetic stem cell signature in cancer. *Nat Genet* **39**: 157-158.

Wilkins RJ (1988). Genomic imprinting and carcinogenesis. *Lancet* **1**: 329-331.

Wilson R, Ryan GB, Knight GL, Laimins LA, Roberts S (2007). The full-length E1E4 protein of human papillomavirus type 18 modulates differentiation-dependent viral DNA amplification and late gene expression. *Virology* **362**: 453-460.

Wilting SM, Snijders PJF, Meijer GA, *et al* (2006). Increased gene copy numbers at chromosome 20q are frequent in both squamous cell carcinomas and adenocarcinomas of the cervix. *J Pathol* **209**: 220-230.

Wisman GB, Nijhuis ER, Hoque MO, *et al* (2006). Assessment of gene promoter hypermethylation for detection of cervical neoplasia. *Int J Cancer* **119**: 1908-1914.

Wolff EM, Liang G, Cortez CC, *et al* (2008). RUNX3 methylation reveals that bladder tumors are older in patients with a history of smoking. *Cancer Res* **68**: 6208-6214.

Wong YF, Cheung TH, Tsao GSW, *et al* (2006). Genome-wide gene expression profiling of cervical cancer in Hong Kong women by oligonucleotide array. *Int J Cancer* **118**: 2461-2469.

Woodman CBJ, Collins S, Winter H, *et al* (2001). Natural history of cervical human papillomavirus infection in young women: a longitudinal cohort study. *Lancet* **357**: 1831-1836.

Woodman CB, Collins S, Rollason TP, *et al* (2003). Human papillomavirus type 18 and rapidly progressing cervical intraepithelial neoplasia. *Lancet* **361**: 40-43.

Woodman CB, Collins SI, Young LS (2007). The natural history of cervical HPV infection: unresolved issues. Nat Rev Cancer 7: 11-22.

Woodworth CD, Bowden PE, Doniger J, *et al* (1988). Characterization of normal human exocervical epithelial cells immortalized in vitro by papillomavirus types 16 and 18 DNA. *Cancer Res* **48**: 4620-4628.

Wright TC Jr, Massad LS, Dunton CJ, *et al* (2007). 2006 consensus guidelines for the management of women with cervical intraepithelial neoplasia or adenocarcinoma in situ. *J Low Genit Tract Dis* **11**: 223-239.

Wu MF, Cheng YW, Lai JC, *et al* (2005). Frequent p16INK4a promoter hypermethylation in human papillomavirus-infected female lung cancer in Taiwan. *Int J Cancer* **113**: 440-445.

Wu X, Gong Y, Yue J, et al (2008). Cooperation between EZH2, NSPc1-mediated histone H2A ubiquitination and Dnmt1 in HOX gene silencing. *Nucl Acids Res* **36**: 3590-3599.

Wynder EL and Graham EA (1950). Tobacco smoking as a possible etiologic factor in bronchogenic carcinoma. *J Am Med Assoc* **143**: 329-336.

Xu XC, Ro JY, Lee JS, *et al* (1994). Differential expression of nuclear retinoid receptors in normal, premalignant, and malignant head and neck tissues. *Cancer Res* **54**: 3580-3587

Xu XC, Sneige N, Liu X, *et al* (1997). Progressive decrease in nuclear retinoic acid receptor beta messenger RNA level during breast carcinogenesis. *Cancer Res* **57**: 4992-4996.

Xu XC, Sozzi G, Lee JS, *et al* (1997b). Suppression of retinoic acid receptor beta in non-small-cell lung cancer in vivo: implications for lung cancer development. *J Natl Cancer Inst* **89**: 624-629.

Yan PS, Chen CM, Shi H, *et al* (2001). Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res* **61**: 8375-8380.

Yan PS, Efferth T, Chen HL, *et al* (2002). Use of CpG island microarrays to identify colorectal tumors with a high degree of concurrent methylation. *Methods* **27**: 162-169.

Yanagawa N, Osakabe M, Hayashi M, Tamura G, Motoyama T (2008). Detection of HPV-DNA, p53 alterations, and methylation in penile squamous cell carcinoma in Japanese men. *Pathol Int* **58**: 477-482.

Yang AS, Doshi KD, Choi SW, *et al* (2006). DNA methylation changes after 5-aza-2'-deoxycytidine therapy in patients with leukemia. *Cancer Res* **66**: 5495-5503.

Yoder JA, Soman NS, Verdine GL, Bestor TH (1997). DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J Mol Biol* **270**: 385-395.

Yu T, Ferber MJ, Cheung TH (2005). The role of viral integration in the development of cervical cancer. *Cancer Genet Cytogenet* **158**: 27-34.

Zampieri M, Passananti C, Calabrese R, *et al* (2009). Parp1 localizes within the Dnmt1 promoter and protects its unmethylated state by its enzymatic activity**.** *PLoS One* **4**: e4717.

Zelent A, Krust A, Petkovich M, Kastner P, Chambon P (1989). Cloning of murine alpha and beta retinoic acid receptors and a novel receptor gamma predominantly expressed in skin. *Nature* **339**: 714-717.

Zhai Y, Kuick R, Ota I, *et al* (2007). Gene expression analysis of preinvasive and invasive cervical squamous cell carcinomas identifies HOXC10 as a key mediator of invasion. *Cancer Res* **67**: 10163-10172.

Zhang Y, Ng HH, Erdjument-Bromage H, *et al* (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13**:1924-1935.

Zhao XD, Han X, Chew JL, *et al* (2007). Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveal distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**: 286-298.

Zheng ZM, Baker CC (2006). Papillomavirus genome structure, expression, and post-translational regulation. *Front Biosci* **11**: 2286-2302.

Zhu H, Geiman TM, Xi S, *et al* (2006). Lsh is involved in de novo methylation of DNA. *EMBO J* **25**: 335-345.

Zimmermann H, Degenkolbe R, Bernard HU, O'Connor MJ (1999). The human papillomavirus type 16 E6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator CBP/p300. *J Virol* **73**: 6209-6219.

Zöchbauer-Müller S, Lam S, Toyooka S, *et al* (2003). Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. *Int J Cancer* **107**: 612-616.

zur Hausen H (2002). Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* **2**: 342-350.