Human Papillomavirus Integration and its Role in Cervical Malignant Progression

Rachel Raybould*, Alison Fiander* and Sam Hibbitts*

Obstetrics and Gynaecology, School of Medicine, Cardiff University, CF14 4XN, UK

Abstract: High risk Human Papilloma Virus (HR-HPV) DNA integration into the human genome is one of the key stages in the progression of cervical neoplasia. This mini-review highlights the importance of HPV integration as a potential biomarker for cervical screening and briefly describes the main methods used to detect HPV integration: Amplification of Papillomavirus Oncogene Transcripts (APOT), Restriction Site PCR (RS-PCR), Southern blot and Detection of Integrated Papillomavirus Sequences (DIPS). The potential mechanisms of HPV integration are discussed with a focus on DNA instability; site of integration; and transcriptional regulation of integrants. This article provides an overview of the role HPV integration plays in malignant progression.

Keywords: Human Papillomavirus, integration, cervical cancer, E6 and E7.

INTRODUCTION

Cervical cancer is the second most common cancer among women worldwide. Globally there are around 500,000 new cases and 274,000 deaths due to cervical cancer annually [1]. For centuries, cervical carcinoma has been recognised to behave as a sexually transmitted disease and in the mid 1970’s it was proposed that there was an aetiological link with Human Papillomavirus (HPV) [2-4]. Infection with HPV is now recognised as an essential, but not sufficient, factor for the development of cervical cancer with many HPV lesions regressing without a need for treatment. Current cervical screening serves to detect and treat precancerous cells (Cervical Intraepithelial Neoplasia) before invasive disease develops. A biomarker able to predict malignant progression in women at risk would enhance screening algorithms. High risk HPV (HR-HPV) DNA integration into the human genome is one of the key stages in malignant progression and is therefore a potential biomarker that precedes invasive disease.

HPV GENOME

HPV is strictly epitheliotropic and infects epithelial cells either of the skin or mucous membranes, particularly of the anogenital tract and oropharynx. Viral transcription is tightly regulated and linked spatially and temporally with epithelial differentiation. The HPV genome consists of double stranded DNA (dsDNA) approximately 8kb long contained within a capsid (Fig. 1). Only one coding strand of the dsDNA acts as a template for transcription. The HPV genome is divided into three regions: an Early region containing genes encoding non-structural proteins (E), a Late region containing genes encoding capsid proteins (L), and an Upstream Regulatory Region (URR) (or Long Control Region), which contains a DNA replication origin, transcription regulatory sequences and one or more promoters which control expression of the viral oncoproteins E6 and E7.

HPV PROTEINS

The late proteins, L1 and L2, are expressed late in the HPV lifecycle in the upper, granular layer of the epidermis. The HPV late proteins make up the virion shell and play an important role in mediating virus infectivity. To achieve a successful lifecycle HPV must be able to bind to cell surface receptors. Heparan sulphate proteoglycans on the cell surface are considered the primary binding sites for L1 and L2 from certain HPV evolutionary groups however different secondary receptors are involved for other HPV types.

The Early proteins are predominantly expressed in the basal and suprabasal layers of the epidermis. E5, E6 and E7 are oncogenic proteins (Table 1). E5 expression enhances oncogenic potential [5, 6] but the exact function of E5 remains poorly understood. E6 and E7 expression is essential for maintenance of the transformed state and malignant progression [7, 8] and the roles E6 and E7 play in carcinogenesis are well documented. HPV integration increases the production of E6 and E7 proteins and influences cancer progression through interaction with hTERT, p53 and Retinoblastoma protein (pRB). hTERT is a catalytic subunit of Telomerase that acts to synthesise telomere ends of linear chromosomes during DNA replication. p53 is a transcription factor that regulates cell cycle arrest, apoptosis, senescence, DNA repair and cell metabolism; p53 activity is inhibited by ubiquitin ligase which also ubiquitinates p53 to initiate p53 degradation. pRB is a tumour suppressor protein and interacts with transcription factor E2F to repress the transcription of genes required for the S phase of the cell cycle. In the normal HPV life-cycle expression of E5, E6 and E7 is tightly regulated within cells that are destined to be lost from the
surface epithelial layers, such that they do not pose a carcinogenic threat (Fig. 2).

**HPV INTEGRATION**

There are more than 100 types of HPV [1], which can be sub-divided by their oncogenic risk. HPV16, HPV18, HPV31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV56 and HPV58 are the predominant HR types detected in cervical cancer cases and HR-HPV DNA is often integrated into the human genome. As disease progresses, the risk of integration is higher [9, 10]. Many studies of HPV16 have demonstrated that upon viral integration, variable parts of the HPV genome are disrupted; fragments containing E2, E4 and E5 ORFs are missing whereas the entire E1, E6 and E7 ORFs are integrated and retained [11, 12]. Integration is not part of the normal HPV lifecycle, unlike retroviruses that encode the protein integrase, which facilitates their integration into the host genome. The mechanisms by which HPV integrates its DNA into the human genome are not fully understood although there are a number of hypotheses:

1. **DNA Instability**

   There is considerable debate over whether integration precedes genetic instability or integration arises due to genetic instability. A number of studies have observed that DNA integration occurs in the presence of Double Strand Breaks (DSBs). DSBs occur in regions of DNA in which the DNA repair process has failed. Regions that harbour DNA instability, known as Chromosome Fragile Sites (CFSs), are distributed throughout the genome. Studies have reported an increased frequency of HR-HPV integration in regions of DNA that contains CFSs [13-15]. However, it remains

---

**Fig. (1).** A schematic representation of HPV16: a) A representation of HPV16 episomal DNA contained within the HPV capsid, b) A schematic representation of HPV16 in episomal form, c) A schematic representation of linear HPV16.
unclear as to whether DNA integration is more common in CFSs because DNA is more prone to breakage, as a result of factors such as increased expression of E6 and E7, or because the host DNA contains sequences that increase the likelihood of integration occurring.

Matzner et al. [16] investigated the integration frequency of exogenous DNA into a breast cancer cell line containing many fragile sites within the genome. This study reported preferential integration of exogenous DNA into fragile sites of cellular DNA, supporting the hypothesis that integration occurs due to chromosome instability.

There is also speculation that HPV oncogenes E6 and E7 induce DNA damage [17] and increase the risk of HPV DNA integration [18]. Kessis et al. [18] assessed the integration frequency of a reporter plasmid in a cell line expressing individual E6 or E7 genes from HR and Low-Risk (LR)
HPV types and proposed that the difference in oncogenic potential observed between HR and LR HPV types may be due to the increased ability of HR-HPV to integrate into host DNA.

2. Site of Integration

Integration is random throughout the genome and much research has been carried out investigating sites of HR-HPV integration in cell lines and clinical samples of HPV related cancers at various sites of the body. A systematic review of HR-HPV integration in epithelial dysplasia and infections of the female lower genital tract assessed over 190 HR-HPV integration loci, all of which were randomly located throughout the genome with a predilection for integration in CFs [14]. Integration of HPV DNA into genomic regions harbouring oncogenes may also contribute to the malignant phenotype. A number of studies have observed disruption and/or over-expression of oncogenes as a consequence of integration at specific sites [15, 19-23]; however this data is not conclusive with contradicting reports from different research groups [13, 14].

3. Transcriptional Regulation of Integrants

As mentioned above, HR-HPV DNA integration occurs randomly throughout the human genome but not all integrants will be involved in malignant progression of the host cell. Transcription of integrated HR-HPV is regulated and both host cell and HPV regulatory elements have been implicated in this event. Disruption of regulatory elements may contribute to the fate of the cell.

HPV gene expression is tightly associated with the epithelial cell life cycle. Differentiated and quiescent epithelial cells are the natural host of HPV. When epithelial cells have undergone differentiation, the cells stop expressing genes involved in cell replication [24] and HPV must reactivate the host cell machinery in order to replicate (Fig. 2). In low-grade cervical lesions, expression of HPV E6 and E7 in undifferentiated basal cells is repressed; this is partly due to host transacting repressors of the HPV URR, and as the cell differentiates E7 protein inactivates host proteins that repress transcription and viral gene expression increases. In contrast, high grade lesions that contain integrated HPV have high levels of viral transcription throughout the epithelium [25]. A study that injected an immortalised cell line into nude mice reported that viral transcription was more evident in basal cells where the HPV DNA was mainly integrated [26]. It is probable that a combination of HPV DNA integration and host cell de-repression of viral gene expression contributes to the malignant phenotype.

The physical state of HPV DNA (integrated or episomal) plays a role in influencing viral gene regulation. Episomal HPV within the host cell contributes to regulation of integrant transcription, with E1 and E2 ORFs repressing immortalisation and over-expression of E6 and E7 genes. The loss of E1 or E2 ORFs following integration may therefore provide a selective growth advantage [27]. Loss of episomes from integrant containing cells enables expression of E6 and E7, highlighting the importance of both episomal loss and integration in malignant progression [28]. HPV infected cells can be treated to promote episome loss and recent studies have observed that this may speed up malignant progression [29].

CERVICAL SCREENING

Cervical screening serves to detect and treat precancerous cells before invasive disease develops. However, cytological testing may lack sensitivity and an increasing number of women develop invasive cancer following a routine normal smear result. HPV testing detects high grade lesions but also detects transient infections that are not associated with the development of high grade lesions; thus HPV testing has a good negative predictive value but a poor positive predictive value. HR-HPV DNA integration into the human genome is one of the key stages in the onset of malignant progression and so it makes a very plausible positive predictive biomarker of invasive disease. Not only would development of assays to detect integration improve cervical screening and improve management of women with cervical HPV infections, the assays would be applicable to HPV related cancers in other parts of the body, in men and women e.g. head and neck cancers, vulval cancer and anogenital cancers. Thus understanding the biology and mechanisms behind HPV integration will aid the prevention and/or detection of many HPV related cancers.

HPV VACCINATION

HPV16 and HPV18 are responsible for approximately 70% of cervical cancer cases globally [30] and most studies have focused on investigating integration for these 2 genotypes. Current vaccination programs serve to protect women against HR HPV16 and HPV18 infection. Vaccination protects against HPV16 and HPV18 infections throughout the body; this in turn will reduce the prevalence of HPV related cancers including cancers of the anogenital tract as well as the head and neck. There are two vaccines currently available: Gardasil®, a quadrivalent HPV6/11/16/18 vaccine (Merck) and Cervarix™, a bivalent HPV16/18 vaccine (GSK). Vaccine trials have been conducted for both Cervarix™ (GSK001/007 [31] and PATRICIA [32]) and Gardasil® (Merck 007 [33], FUTURE I [34] and FUTURE II [35]). The vaccine trials followed women from Asia, Australia, Europe, Latin America and North America for up to four years; see Table 2 for an overview of trial designs. The vaccines reduce persistent HPV infection of the targeted HPV types and the HPV6/11/16/18 vaccine eliminated genital warts in the vaccinated group compared to the placebo [36]. In the PATRICIA trial, Cervarix™ was at least 90% effective against persistent infection with HPV16/18; with a 100% reduction of high grade cervical lesions associated with HPV16/18 in the vaccine recipient group compared to the placebo group. The PATRICIA trial also indicated that the benefits of the vaccine would be greater in women that have not had previous exposure to HPV16/18 infections. Clinical trials of the vaccine have also suggested the possibility of cross-protection against other HR-types and there are some concerns that type-replacement may occur with one or more of the other HR types becoming more predominant in cancer cases in the absence of HPV16 and HPV18 [36]. It will therefore be advantageous to develop novel methods to
investigate HPV integration for less prevalent HR-HPV types.

**DETECTION OF HPV INTEGRATION**

There are a number of methods currently employed to detect HR-HPV integrants in the human genome (summary in Table 3). Amplification of Papillomavirus Oncogene Transcripts (APOT) detects transcripts of integrated HPV, Restriction Site PCR (RS-PCR), Southern blot and Detection of Integrated Papillomavirus Sequences (DIPS) detect integrated HPV DNA regardless of its transcriptional status. The methods used to detect HPV integration are important because a transcriptionally active integrant may contribute more to the malignant phenotype. However, integrants that are not transcribed may also contribute to malignant progression by regulating or disrupting the expression of genes that contribute to cervical malignant progression.

The majority of studies have investigated integration in clinical samples whilst a number of studies have chosen to investigate integration mechanisms in carcinoma cell line cultures such as CaSki, HeLa and SiHa. The W12 cell line has been used as a model to study HPV16 integration. This cell line was derived from a low-grade cervical lesion

### Table 2. A Summary of HPV Vaccine Clinical Trials

<table>
<thead>
<tr>
<th>Clinical Trial</th>
<th>Vaccine</th>
<th>Endpoint</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSK001/007</td>
<td>Cervarix™</td>
<td>N=1,113, AR=15-25 years, SF= 6 months, Mean FU= 48 months. Objective: to examine HPV16/18 infection incidence, infection persistence, CIN1+ and adverse events.</td>
<td>Efficacy was 91.6% against HPV16/18 infection and 100% against persistent infection with HPV16/18.</td>
</tr>
<tr>
<td>PATRICIA</td>
<td>Cervarix™</td>
<td>N=18,644, AR=15-25 years, SF= 12 months, Mean FU= 15 months. Objective: to examine HPV16/18 CIN2+, infection persistence, CIN1+ and adverse events.</td>
<td>There was 100% reduction against HPV16/18 CIN2/. Efficacy against 6 month and 12 month HPV16/18 persistent infection was 93.8% and 91.2%, respectively.</td>
</tr>
<tr>
<td>Merck 007</td>
<td>Gardasil®</td>
<td>N=552, AR=16-23 years, SF= 6 months, Mean FU= 60 months. Objective: to examine HPV6/11/16/18 infection persistence, cervical and external genital disease and adverse events.</td>
<td>Persistent infection with HPV6/11/16/18 was reduced by 89%.</td>
</tr>
<tr>
<td>FUTURE I</td>
<td>Gardasil®</td>
<td>N=5,455, AR=16-24 years, SF= 6 months, Mean FU= 36 months. Objective: to examine HPV6/11/16/18 CIN1+, external genital lesions and adverse events.</td>
<td>HPV associated genital warts was reduced by 100%. Vaccination reduced vulvar, vaginal and perianal lesions by 34% and reduced cervical lesions by 20% regardless of HPV type.</td>
</tr>
<tr>
<td>FUTURE II</td>
<td>Gardasil®</td>
<td>N=12,167, AR=15-26 years, SF= 12 months, Mean FU= 36 months. Objective: to examine HPV16/18 CIN2+ and adverse events.</td>
<td>HPV associated genital warts was reduced by 98%. Estimated vaccine efficacy in cervical lesions regardless of HPV type was 17%.</td>
</tr>
</tbody>
</table>

**Abbreviations:** N= number of women, AR= age range, SF= screening frequency, Mean FU= mean follow-up, CIN=Cervical Intraepithelial Neoplasia, VIN= Vulval Intraepithelial Neoplasia, VAIN= Vaginal Intraepithelial Neoplasia.

---

### Table 3. A Summary of the Most Common DNA and RNA Based HR-HPV Integration Detection Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS-PCR</td>
<td>Host/viral genomic regions are amplified by PCR using HPV specific primers and a primer designed to bind to restriction enzyme sites.</td>
<td>Can obtain DNA sequence of host/viral junctions.</td>
<td>Large concentrations of DNA required and labour intensive.</td>
</tr>
<tr>
<td>APOT</td>
<td>RT-PCR followed by PCR using HPV specific primer discriminates HPV miRNAs derived from integrated and episomal viral genomes.</td>
<td>Can obtain DNA sequence of host/viral junctions.</td>
<td>Labour intensive and expensive.</td>
</tr>
<tr>
<td>DIPS</td>
<td>Single-side-specific ligation-mediated PCR. Involves vectorette PCR and suppression PCR to detect integrated HPV DNA.</td>
<td>Can obtain DNA sequence of host/viral junctions.</td>
<td>Labour intensive.</td>
</tr>
<tr>
<td>Southern blot</td>
<td>Cellular DNA digestion and electrophoresis followed by hybridisation of labelled HPV DNA probes to determine the physical state (integrated or episomal) of HPV.</td>
<td>Can reliably distinguish episomal from integrated HPV DNA.</td>
<td>Uses large concentrations of DNA and labour intensive. The use of radio-labelled probes has health and safety implications.</td>
</tr>
<tr>
<td>Real-time PCR</td>
<td>Physical state of HPV is estimated by calculating HPV E2:E6/E7 ratio by real-time PCR amplification of HPV E2 and E6/E7.</td>
<td>Uses small concentrations of DNA and is less labour intensive.</td>
<td>HPV E2:E6/E7 ratio may not reliably distinguish integrated DNA in a background of episomal DNA. Consumables expensive.</td>
</tr>
</tbody>
</table>

**Abbreviations:** RS-PCR= Restriction Site PCR; APOT= Amplification of Oncogene Transcripts; DIPS= Detection of Integrated Papillomavirus.
infected with HPV16 [37]. Early passages of W12 contain HPV16 in an episomal form (ca 100 episomes per cell) and long-term passage produces cells containing fully integrated HPV16 DNA. The properties of the W12 cell line are rare and although cell lines enable practical approaches to investigate mechanisms of HPV integration they are limited and not available for all HR HPV types. Investigation of HPV integration, in large and well defined clinical cohorts, may give insight into natural integration events and determine whether HPV integration can be used as a reliable biomarker to predict malignant progression.

CONCLUSION
HR-HPV integration is vital, but not the only contributing factor to the development of invasive cervical cancer. Integration induces changes in the viral DNA structure that favours an immortalized host cell, although this event is disruptive to the viral lifecycle. Further studies of HPV integration are required in large clinical samples to improve our understanding of HPV integration and its role in the pathogenesis of cervical carcinoma.

REFERENCES
