The current literature fully supports HPV (human papillomavirus)-associated OPSCC (oropharyngeal squamous cell carcinoma) as a unique clinical entity. It affects an unambiguous patient population with defined risk factors, has a genetic expression pattern more similar to cervical squamous cell carcinoma than non-HPV-associated HNSCC (head and neck squamous cell carcinoma), and may warrant divergent clinical management compared with HNSCC associated with traditional risk factors. However, a detailed understanding of the molecular mechanisms driving these differences and the ability to exploit this knowledge to improve clinical management of OPSCC has not yet come to fruition. The present review summarizes the aetiology of HPV-positive (HPV + ) OPSCC and provides a detailed overview of HPV virology and molecular pathogenesis relevant to infection of oropharyngeal tissues. Methods of detection and differential gene expression analyses are also summarized. Future research into mechanisms that mediate tropism of HPV to oropharyngeal tissues, improved detection strategies and the pathophysiological significance of altered gene and microRNA expression profiles is warranted.

Key words: cancer, head and neck squamous cell carcinoma (HNSCC), human papillomavirus (HPV), oropharyngeal squamous cell carcinoma (OPSCC).

INTRODUCTION

HNSCC (head and neck squamous cell carcinoma) is a major global health problem, ranking as the sixth most prevalent cancer worldwide, with >405000 new cases each year, two-thirds of which are in the developing world [1]. Mortality is thought to be 200000 deaths annually, and U.S. deaths occur at a rate of one death every hour. These tumours occur in the oral cavity, larynx, oropharynx and other locations throughout the head and neck (Figure 1). The major risk factors of HNSCC include the synergistic effects of tobacco and ethanol abuse in addition to physical irritation, malnutrition and immune defects. Owing to public health efforts encouraging smoking cessation, the overall incidence of HNSCC has decreased in recent years. In contrast, the incidence of OPSCC (oropharyngeal squamous cell carcinoma) has increased significantly at a rate of 1.3% and 0.6% yearly, between 1973 and 2004, for base of tongue and tonsillar carcinomas respectively. In this same period, the incidence of oral cavity cancer decreased by 1.9% [2,3]. Recent statistics show a striking 225% increase in population-level incidence of HPV (human papillomavirus)-positive (HPV + ) OPSCC between 1984 and 2004, whereas HPV-negative (HPV − ) OPSCC incidence decreased by 50% in the same time period [4]. Oropharyngeal refers to the anatomical location from the junction of hard and soft palate to the plane of the hyoid bone and includes the base of the tongue, palate and (including tonsillar fossae and pillars), uvula, vallecula epiglottica, and lateral and posterior oropharyngeal walls (Figure 1). Molecular studies of malignancies arising in this area have shown that 40–80% of OPSCC diagnosed in the U.S.A. contain HPV and the IARC (International Agency for Research on Cancer) Multicenter Study estimated that 18% of oral and oropharyngeal cancers worldwide are HPV-associated [3,5]. Various studies support a viral aetiopathology for oral cancer in these sites [2–8]. Thus the decline in oral cavity cancer has been attributed to a reduction in tobacco use, whereas the increasing incidence of oropharynx cancer appears to reflect a rise in HPV + malignancy. Similar observations have been made in Europe [9,10].

HPV + cancers have a different biology from that of HPV − cancers, with distinct phenotypic features distinguishing HPV + OPSCC, including poor differentiation, scant keratinization and basaloid phenotype, compared with the typically keratinizing morphology of HPV − tonsillar SCCs (squamous cell carcinomas) [3,11] (Figure 2). Indeed, analysis of HPV + compared with HPV − lesions reveal that tumours harbouring HPV genomic DNA lack many of the characteristic mutations of disease thought to be attributed to traditional risk factors of tobacco and alcohol abuse. Moreover, viral oncoproteins expression and variable degrees of genomic integration have been observed in numerous case studies [9,12–14]. Furthermore, an inverse relationship between p53 mutation and HPV infection has been reported [15], strengthening the evidence in support of a role for HPV in OPSCC aetiology. Overall, HPV + tumours are more likely to have p53 degradation and Rb (retinoblastoma protein) inactivation with resulting p16 up-regulation [16]. In contrast,
HPV − tumours tend to have p53 mutation and down-regulation of p16 [5,11].

In HPV infection, the life cycle of the virus is linked to the differentiation state of the host cell and requires that the host cell remains active in the cell cycle [17,18]. The highest level of viral replication occurs in the granular layer of stratified epithelia where keratinocytes are terminally differentiated and are in the process of enucleation and death [19] (Figure 3). An infection can lead to unchecked cellular proliferation via overriding the mechanisms that halt the cell cycle in addition to altered cellular differentiation.

This pathological state allows for replication of viral DNA in synchrony with chromosomal DNA. The uncontrolled cellular proliferation in the context of manipulated cell-cycle checkpoints leads to chromosomal instability and accumulation of genetic mutations and subsequently the initiation of pre-malignant and malignant lesions. The cellular factors that are co-opted by HPV proteins are involved in the proliferation of cells during their differentiation, including viral E6-mediated ubiquitylation of p53 and viral E7-mediated modulation of Rb (see below).

**AETIOLOGY AND DETECTION**

Certain high-risk subtypes of HPV are causative agents for a number of human neoplasia, including disease of the cervix, anus, penis and vulva. Approximately 291 million women worldwide are carriers of HPV DNA, of which 32% are infected with the high-risk subtypes HPV-16 or HPV-18, or both. Type 16 (85−95 %) and, to a lesser extent, type 18 are the subtypes most commonly identified among HPV + OPSCC [20,21]. Interestingly, the low-risk subtypes HPV-6 and HPV-11 have been detected in oropharyngeal cancers, suggesting that these types may not in fact be benign at this anatomical location [14,22].

Although detection of HPV genomic DNA was first noted in head and neck cancer over 20 years ago [3], an aetiological role is now undeniably supported. The involvement of HPV in head and neck squamous cell carcinogenesis is supported by a series of observations. First, HPV is a virus with broad and essential tropism for epithelial tissues. *In vitro* studies clearly prove that high-risk viral oncoproteins immortalize human keratinocytes, including oral keratinocytes [23]. The high-risk HPV genotypes can be detected in SCC with PCR techniques and fluorescence *in situ* hybridization [13]. In addition, genotype-concordant viral DNA can be found in the lymph nodes of patients with metastatic OPSCC [24]. Moreover, multiple gene signatures detected with DNA microarrays are able to predict HPV-16 prevalence in primary HNSCC with a false discovery rate of <0.2 [25]. Lastly, the established role for high-risk HPV in cervical SCC is foundational support for a causative role for HPV in OPSCC because these distinct anatomical locations share morphological, embryological and immunological features and can be infected with the same virus [5].

HPV + OPSCC has identifiable histology, biochemical characteristics and surrogate markers that give it singular identity. As a result of its prognostic value, the American Joint Committee on Cancer Staging Criteria recommends testing HPV status for oral and pharyngeal cancers [26]. Efficient and effective detection of HPV-associated tumours is an essential component of disease management as well as research. The
HPV (red) infects proliferating cells in the basal layer that are exposed by wounding. The virus replicates in synchrony with cellular DNA replication. The highest level of viral replication occurs in the granular layer, as expression of HPV E6 and E7 maintain proliferation of these normally terminally differentiated cells. Mature virions are released through the cornified layer.

Figure 3  HPV infection of oral mucosa

HPV (red) infects proliferating cells in the basal layer that are exposed by wounding. The virus replicates in synchrony with cellular DNA replication. The highest level of viral replication occurs in the granular layer, as expression of HPV E6 and E7 maintain proliferation of these normally terminally differentiated cells. Mature virions are released through the cornified layer.

optimal method to detect HPV-associated tumours, however, is still unresolved. The histopathological characteristics of HPV + OPSCC are instructive, but not definitive in the detection of HPV. More commonly, PCR, immunohistochemistry and in situ hybridization are used as part of a multi-step detection protocol. Histopathological characteristics of HPV + OPSCC include poor differentiation, scant keratinization and basaloid phenotype (Figure 2). The usefulness of these findings for diagnostic and prognostic purposes is debated. The non-keratinizing OPSCC phenotype correlates positively with immunohistochemical staining for p16 and favourably associates with prognostic data [27]. Others report a positive association between the non-keratinizing phenotype and HPV + disease, but a portion of non-keratinizing OPSCC cases are HPV − [28,29]. Furthermore, histological findings are subject to inter-observer variability. Although there is value in completing a histological profile, other methods have largely replaced histological findings for definitive detection of HPV + OPSCC.

PCR of HPV sequences requires specimens of fresh-frozen or formalin-fixed paraffin-embedded tissue. Tissue sections represent a heterogenous mixture of tumour, stroma and inflammatory cells, such that laser-capture microdissection is necessary to localize the HPV signal to the specific cell type of interest. As mentioned above, the expression of E6 and E7 does not necessarily signal integrated HPV, and whether the virus is integrated or episomal remains unanswered with conventional consensus primer-based PCR. However, since the E2 gene is a common point of alteration in integrated HPV, cells with integrated HPV will not amplify a PCR product using primers specific for the E2 gene. The ratio of E2 to E6 is then assessed and, if very low, is taken to reflect integrated HPV [14]. Although this technique is intriguing, it falls short in terms of sensitivity, because multiple studies have shown that genes besides E2 can be disrupted secondary to viral integration [30–32] and that viral oncoproteins are expressed regardless of integration status. Quantitative real-time PCR has provided data that are important clinically in establishing prognosis and providing additional evidence that OPSCC is a unique clinicopathological entity. The addition of quantification to detection methods has allowed for studies comparing the number of viral genomes present with the disease site and also with patient outcome. The viral load varies widely among OPSCC cases. Recent data show that it correlates with patient outcome, such that a high viral load is associated with improved prognosis. It is thought to occur because most HPV + tumours with high viral load have high p16 and reduced or low Rb, p53 and cyclin D1, whereas low-viral-load tumours show reduced p16, increased cyclin D1 and p53, and normal Rb [8].

As described above, HPV oncoprotein E7 inactivates the Rb gene product, leading to sufficient up-regulation of p16 for detection by immunohistochemical methods. As a result, p16 serves as a surrogate marker for HPV + OPSCC, suggesting HPV viral transcription. p16 is a sensitive marker, elevated in cases involving HPV-16 as well as less frequently present viral serotypes [33–35]. Immunohistochemical analysis does not, however, differentiate between HPV serotypes, and false positives might be present in the context of HPV-unrelated disturbances in the Rb gene product. In situ hybridization permits identification of specific HPV serotypes and direct visualization of the distribution of HPV + cells within the tissue sample. Signal-amplification steps significantly improve the sensitivity of this technique to the point of viral detection with as few as one copy per cell [33]. Complete concordance with PCR results was demonstrated in a small sample when unequivocally positive or negative in situ hybridization findings are present [36]. However, this technique does not provide evidence of viral integration as may be possible with PCR.

HPV VIROLOGY

Overview

HPVs are small, 50–55 nm in diameter, non-enveloped double-stranded DNA viruses, which carry out their life cycle in either mucosal or cutaneous epithelia. Infection may result in an asymptomatic carrier state or a variety of both benign and malignant neoplasia. These viridae have icosahedral capsids composed of 72 capsomeres, surrounding a circular DNA genome of ∼7900 bp [18]. Hundreds of different genotypes have been identified, and the viruses are grouped according to distinctions in type-specific antigens on the virion surface. Each type shares less
than 90% DNA sequence homology in the region of their major capsid protein, L1. Specific clinical manifestations are associated with many of the types and, for the sake of convenience, the more common subtypes are grouped according to high- or low-potential symptoms or the ability to effect cellular transformation. Although there are 15 known high-risk types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 and 82) and 12 low-risk types (6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81 and CP6108), types 16, 18 and 31 are the major types associated with mucosal epithelial cancers, whereas HPV-16 accounts for 90–95% of HPV + OPSCC [37].

The genome can be divided into an early (E) region (containing genes E1, E2, E4, E5, E6 and E7), late (L) region (containing genes L1 and L2), and a URR (upstream regulatory region) (Figure 4). Two major viral promoters induce transcription of polycistronic mRNAs. The DNA itself associates with cellular histones and is compacted into aggregates similar to chromatin [38]. The conformation of the chromatin and epigenetic modifications of histones are significantly altered upon cellular differentiation, and this effectively divides gene expression into early and late events. During early stages of the viral life cycle, early transcripts are initiated by a promoter referred to as either p97 or p105 in the HPV-16/31 or HPV-18 subtypes respectively. The p97 promoter is located just upstream of the E6 ORF (open reading frame) [38]. The major promoter of late genes is located further downstream and varies slightly depending on the virus subtype, but it is generally referred to as p742 [38]. The latter transcriptional events are activated in concert with epithelial differentiation via segregated temporal usage of these promoters. The early genes E1–E7 play a role in regulating, promoting and supporting viral DNA transcription and replication. The late genes, L1 and L2, are transcribed only in productively infected cells and encode the major and minor capsid proteins required for assembly of progeny virions and eventual accumulation and release into the environment.

**Mode of viral entry**

The basal cell is fundamental to papillomavirus infection and may be the only cell within epithelia capable of establishing infection. This dependence on terminally differentiated keratinocytes has made the study of viral entry challenging, because propagation of virions in cell culture does not occur [39]. It is thought that infection occurs at sites of injury in the proliferating basal layer of epithelial surfaces. This proliferation due to microtrauma induces basal cell migration and enhanced cell division, therefore increasing the probability of a productive infection. A remarkably complicated and still poorly understood process is required for binding to the host cell surface, subsequent internalization and intracellular trafficking, but the basic evolutionary strategy is dependent upon processes for excluding access to the differentiated cells and fostering access to the basal cell layers. This is achieved by preferential binding to basement membrane proteins specific to the basal layers [39].

Most papillomavirus types utilize HSPGs (heparan sulfate proteoglycans) located within the basement membrane and extracellular matrix, specifically syndecan-1 that is the HSPG isotype expressed predominantly in epithelia, not just for initial attachment, but also to promote conformational changes in the viral capsid [40]. These interactions are thought to initiate attachment and be essential for infectious internalization, but not directly mediate virion internalization [39]. HPV capsids also bind laminin-5, but the significance of this interaction is not entirely understood [40]. A number of candidate integrins have been studied in vitro and it appears that at least three integrins, α6, β4 and β1, may be involved in secondary HPV binding, although α6 is likely to be paramount [41–43]. Interestingly, there has not been definitive evidence of an amino-acid-binding motif present on the surface of HPV representative of a known recognition motif for any of these host cell receptors [40]. A particular conformational cluster of basic acids is thought to be critical and involved in inducing a conformational change of the capsid upon contact with the host cell surface [40]. Virus internalization may require additional factors such as syndecans and glypicans, two HSPGs also thought to be up-regulated in response to injury. There is compelling evidence for three modes of viral entry: clathrin-mediated endocytosis, caveolar endocytosis, and a clathrin- and caveolar-independent pathway involving tetraspanin-enriched microdomains [40].
HPV-16 is proposed to bind cells and subsequently co-localize with transmembrane tetraspanins CD63 and CD151, although the role of these tetraspanins has not been fully elucidated [44]. After binding and endocytosis, HPV is thought to gain access to the microfilament network via an interaction with L2 and the motor protein complex dynein for transport around the cytoplasm and to the nucleus [45].

Host immune response

Clearly, HPV infections that lead to cancer require alterations in the host cellular immune response. Before viral replication, host immune intervention via TNFα (tumour necrosis factor α) and IL-1 (interleukin 1) production may influence the fate of HPV-infected cells, as a robust response results in antiviral activity and specifically down-regulates HPV-16 E6/E7 mRNA transcription [46]. Alternatively, some authors have suggested that HPV may mediate resistance to TNFα early in its life cycle [47,48]. From studies in women with cervical HPV infection, it has been demonstrated that the majority of women infected mount an effective immune response and clear the infection, ∼10% develop persistent infection, and only 5–10 per 100 000 progress from pre-malignant to malignant cervical disease in developed countries [19]. Similar comparisons are not yet available for head and neck infections. However, evasion of the innate immune response is a biological hallmark of HPV infection in both pre-malignant and malignant disease and results from viral stimulation and/or repression of host signalling pathways [19]. Although the molecular pathways responsible for this evasion are postulated to lie in viral down-regulation of cytokine expression, an alternative or perhaps parallel explanation is the evolutionary niche created by infection of keratinocytes, which have a naturally short lifespan and are generally sheltered from large populations of antigen-presenting cells [19]. Because of the short keratinocyte lifespan, cell lysis is not necessary for viral escape (Figure 3); rather virions are released as cells propagate through the mucosal layers. Nevertheless, the hope of vaccination is to bolster adaptive immunity by introducing highly immunogenic repeat structures based on L1 termed L1 VLPs (virus-like particles). These VLPs are brought into lymph nodes by antigen-presenting cells and are presented to T-cells that will drive the production of a B-cell response and capsid-specific neutralizing antibodies [49].

Type I IFNs (interferons) are potently produced in response to viral nucleic acids. Under normal conditions, various TLRs (Toll-like receptors) such as TLR9 recognize unmethylated CpG DNA, leading to IRF (IFN-regulatory factor) activation and expression of IFNα and IFNβ. IFN subsequently acts in a paracrine fashion to both the respiratory and alimentary tracts and therefore allows access to both inspired and ingested antigens. There are four functional immunological structures of the tonsils: the outer endodermally derived reticular crypt epithelium, the mesodermally derived follicular area, the mantle zone and the germinal centre of the lymphoid follicle. Antigen-presenting cells at the outer epithelium transport antigen to helper T-cells and may subsequently stimulate B-cells to produce IgA antibodies which are transported to the surface. It has been speculated that the immunological function of the tonsils may contribute to harbouring the virus and that transformation of the keratinocytes requires some contribution from the microenvironment created by the lymphocytes and antigen-presenting cells populating the tonsils [51,52,56]. Some studies have suggested that HPV-associated tonsillar SCC originates from the deep invaginations of the cryptic epithelium, whereas non-HPV-related SCCs develop from the surface epithelium [57–59]. The numerous branching crypts that line the palatine and lingual tonsils are common sites of inflammation owing to accumulations of dead lymphocytes and desquamated epithelium [60]. The epithelium of the crypts is reticulated, which refers to a discontinuous epithelium where non-epithelial cells such as lymphocytes and antigen-presenting cells pass between squames. This organization of cells facilitates direct transport of antigens, a favourable environment for contact between the effector cells of the immune system, and is thought to help provide a pool of immunoglobulins [61]. However, this reticulated structure is also a compromise in epithelial integrity, leaving the deep layers of the epithelium exposed to deposition of viral particles [62]. The determinants of infectivity and susceptibility are not well established with respect to HPV and oropharyngeal cancers. It has been postulated that the common endodermal origin of the squamous lining of both the cervix and tonsils, their shared locations at the junction between external and internal environments, and human microbiota at mucosal surfaces at these locations play a role in allowing persistent HPV infection and provide a favourable milieu for carcinogenesis initiation and propagation.

E1

E1 contains a nuclear localization signal, and it has been proposed that maintenance of the viral episome in undifferentiated keratinocytes is a function of a conserved region of E1 unique to anogenital subtypes of the virus [63]. E1 is a 68 kDa protein with ATPase and helicase with 3′ → 5′ activities. As a monomer, E1 weakly binds ATP-rich sequences upstream of the start sites of early-transcribed HPV genes [64–69]. E1-recognition

PATHOGENESIS: A CONSEQUENCE OF VIRAL GENE EXPRESSION

Although HPV has been identified in squamous cell cancers at other head and neck sites, the overwhelming anatomical locations for disease burden associated with HPV are the mucosal lining of the tonsils and the base of tongue (Figure 1B). The pharyngeal tonsillar tissue functions as part of the secondary immune system. The location of the tonsils allows for exposure to both the respiratory and alimentary tracts and therefore allows access to both inspired and ingested antigens. There are four functional immunological structures of the tonsils: the outer endodermally derived reticular crypt epithelium, the mesodermally derived follicular area, the mantle zone and the germinal centre of the lymphoid follicle. Antigen-presenting cells at the outer epithelium transport antigen to helper T-cells and may subsequently stimulate B-cells to produce IgA antibodies which are transported to the surface. It has been speculated that the immunological function of the tonsils may contribute to harbouring the virus and that transformation of the keratinocytes requires some contribution from the microenvironment created by the lymphocytes and antigen-presenting cells populating the tonsils [51,52,56]. Some studies have suggested that HPV-associated tonsillar SCC originates from the deep invaginations of the cryptic epithelium, whereas non-HPV-related SCCs develop from the surface epithelium [57–59]. The numerous branching crypts that line the palatine and lingual tonsils are common sites of inflammation owing to accumulations of dead lymphocytes and desquamated epithelium [60]. The epithelium of the crypts is reticulated, which refers to a discontinuous epithelium where non-epithelial cells such as lymphocytes and antigen-presenting cells pass between squames. This organization of cells facilitates direct transport of antigens, a favourable environment for contact between the effector cells of the immune system, and is thought to help provide a pool of immunoglobulins [61]. However, this reticulated structure is also a compromise in epithelial integrity, leaving the deep layers of the epithelium exposed to deposition of viral particles [62]. The determinants of infectivity and susceptibility are not well established with respect to HPV and oropharyngeal cancers. It has been postulated that the common endodermal origin of the squamous lining of both the cervix and tonsils, their shared locations at the junction between external and internal environments, and human microbiota at mucosal surfaces at these locations play a role in allowing persistent HPV infection and provide a favourable milieu for carcinogenesis initiation and propagation.
sequences in origins of HPV replication are adjacent to E2 DNA-binding sites and the latter act to load E1 on to the origin and facilitate higher-affinity binding, thereby initiating origin-specific viral DNA replication [38]. The recruitment of cellular replication complexes including chaperone proteins and chromatin-remodelling complexes leads to E1 binding host DNA polymerase α [70–72]. E1 proteins bound to DNA oligomerize into an hexameric ring and act as a 3′→5′ helicase to unwind supercoiled DNA. The regulation of E1 activity is not fully understood, but it is known that the protein interacts with mitogen-activated protein kinases and the cell-cycle regulators cyclin E and A, both of which are expressed after cells have progressed through the restriction point [71]. Mutagenesis studies have revealed essential cyclin-binding motifs, serine residues that are CDK (cyclin-dependent kinase) sites and a dominant leucine-rich NES (nuclear export sequence) located on E1 which are important for its nuclear localization and the overall ability of the virus to adapt the cellular regulatory mechanisms to support viral replication [73,74].

E2

E2 is a 50 kDa transcription factor which plays an integral role in transcriptional regulation of early genes, DNA replication and genome maintenance functions. E2 binds and interacts with numerous cellular factors. The C-terminus encodes a dimeric β-barrel that binds DNA and interacts with E1, whereas the N-terminus encodes a transactivation α-helix domain rich in glutamine [75–77]. The URR [also known as the LCR (long control region)] of mucosal HPV DNA contains four consensus palindromic-sequence-specific binding sites for E2 (E2BSs), three of which are in close proximity to E1-recognition sequences [38]. These E2BSs are overlapped by recognition sequences for cellular transcription factors such as TAF (TATA-box-binding-protein-associated factor) and Sp1 (specificity protein 1) [78,79]. During early stages of infection, viral gene expression is minimal and is initiated by transcription factor binding to sequences in the LCR [80]. E2 interacts with these transcription factors as well as C/EBP (CCAAT/enhancer-binding protein) transcription factors, activation domain-modulating factor-1, p300/CBP [CREB (cAMP-response-element-binding protein)-binding protein], p/CAF (p300/CBP-associated factor), NAP-1 (nucleosome-assembly protein 1) and Brd4 [81]. E2 represses only genes from the early promoter and helps to keep virus copy number low in undifferentiated cells [38]. The mechanisms of transcriptional repression are poorly understood, but probably involve E2-mediated recruitment and co-ordination of several distinct cellular pathways. In support of this, genome-wide siRNA (small interfering RNA) screening identified up to 96 cellular genes contributing to repression [82]. Some of these genes were subsequently validated to independently and additively mediate E2-induced repression, including the cellular demethylase JARID1C (Junonji AT-rich interactive domain 1C)/SMCX and EP400, a component of the NuA4/TIP60 [Tat (transactivator of transcription)-interactive protein 60 kDa] histone acetyltransferase complex. High-risk HPV E2 has also been demonstrated to transactivate and up-regulate SR (serine/arginine-rich) proteins, highly conserved host splicing regulators for alternative splicing that are co-opted to participate in viral RNA processing and can act as oncoproteins [83,84]. This leads to dysregulation of the tightly controlled process of constitutive and alternative splicing of host genes and elaborates the involvement of E2 in the pathogenesis of HPV-induced malignancies. E2 and p16INK4a show exclusive immunohistochemical staining patterns in precursor stages of cervical carcinoma in paraffin-embedded clinical samples. Staining for cytokeratin K13, a marker of squamous cell differentiation, overlapped with E2, whereas proliferation markers Ki67 and p63 were inversely correlated [85]. Studies in HeLa cells show that expression of E2 results in a suppression of transcription of E6 and E7, reflecting a role of E2 in early stages of HPV infection [86].

E4 and E5

Many of the cellular pathways co-opted by HPV discussed so far are logically consistent with viral replication. The E4 and E5 proteins are expressed early, but function to modulate the late phase of the viral life cycle. As E4 and E5 are the third and fourth transcripts on polycistronic early mRNA respectively, very little protein is translated because the ribosome-scanning mechanisms of HPV are relatively inefficient [87]. After epithelial differentiation, E4 becomes the most highly expressed viral gene due to utilization of the late gene promoter, therefore placing these transcripts first and second on polycistronic late mRNA [38]. The E4 ORF lacks a unique AUG codon, but is translated with the first five amino acids from E1 because of splicing events, allowing for an E1–E4 fusion protein [38]. In vitro functions of E1–E4 include E4-mediated G1/M-phase arrest via a cyclin-binding motif [88], association with and reorganization of keratin intermediate filaments [89], and regulation of viral gene expression perhaps by interacting with E4-DBD (DNA-binding domain) RNA helicase [90]. In vivo effects remain a matter of debate, but it is known that there is minimal sequence homology between low- and high-risk E4 protein and that the high-risk types are more likely to interact with and possibly collapse keratin networks.

E5 is an 83-amino-acid hydrophobic protein and therefore localizes to membrane-bound compartments including the Golgi, endosomes, endoplasmic reticulum and the nuclear membrane. The function of this protein is controversial. Some studies suggest that, similarly to bovine papillomavirus E5 association with PDGFR (platelet-derived growth factor receptor), HPV E5 activates EGFR (epidermal growth factor receptor) [91,92]. However, mutational analyses have shown that EGFR is not a direct target of E5 action [93]. Instead, E5 affects EGFR turnover from the plasma membrane by binding an endosomal vacuolar ATPase, the result of which is impaired early endosome acidification, increased recycling of receptors to the cell surface and therefore amplified receptor signalling [18,19]. In vitro studies have shown increased phosphorylation of Akt and ERK1/2 (extracellular-signal-regulated kinase 1/2) associated with E5-mediated enhancement of EGFR signalling, leading to increased expression of VEGF (vascular endothelial growth factor) [58]. The contribution of E5 to carcinogenesis or malignant potential in head and neck cancer is unclear; however, because EGFR activity is a prognostic indicator and therapeutic target, further investigation of E5 regulation of EGFR in OPSCC is warranted.

E6 and E7

It has long been thought that the viral genome is maintained in a non-integrated episomal form in benign warts and integrates into the host genome in most cancers. Increased oncogene expression has been demonstrated subsequent to E2 disruption/deletion and integration and is proposed as a central event in oncogenesis [94,95]. As with other oncopogenic viruses, the integration into the host genome appears to be a random event. However, the pattern of integration is clonal. In cervical cancer, integration of high-risk
HPV into host genomic material induces genomic instability in the host and also alters gene expression in the virus. When integrated, the ORF of the viral genome appears to be consistently interrupted such that the viral repressor E2 is lost, resulting in overexpression of two oncoproteins, E6 and E7. E6/E7, when overexpressed, disrupt the function of wild-type Rb and p53, leading to the development of a malignant phenotype. As a selective mechanism, this favourable gene expression profile allows cells with integrated virus a significant growth advantage. Indeed, in anogenital cancer, a high proportion of HPV tumours with integrated virus are observed. Studies from head and neck cancers have revealed that viral integration is not necessary for initiation of oncogenesis and these cancers vary greatly with respect to the physical state of the HPV genome (episomal compared with integrated) [57, 96, 97]. The significance of this discordance is not well understood, neither is the mechanism of overexpression of E6/E7 in tumours of the head and neck. However, in both cases, in a subset of infected patients, viral infection leads to transcriptionally active HPV with high E6/E7 RNA levels, and subsequently to cell proliferation and eventually to genomic instability due in part to mitotic spindle defects that result in aneuploidy and damaged chromosome structure [97, 98]. Interestingly, E6/E7 gene products are required for the maintenance of proliferation in HeLa cells [19]. These cells, which have been cultured for 50 years, have been demonstrated to contain integrated sequences of HPV-18 expressing E6 and E7 and repression at the HPV-18 promoter or knockdown with RNAi (RNA interference) of E6/E7 results in senescence in 10–14 days [19]. Repression of E6/E7 in HeLa cells also results in diminished telomerase activity, CDK activity and c-Myc expression [99].

The upper layers of the mucosal surface, predominantly the stratum spinosum and granulosum, are sites of viral assembly and high-level viral gene expression (Figure 3). This is a key point in understanding HPV infection, as the life cycle of the virus is linked to the differentiation state of the host cell. Under normal conditions, cell division and cell-cycle entry are confined to the basal layer. HPV requires that the host cell remains active in the cell cycle as the cell progresses through the strata because the virus is dependent on the enzymatic machinery of the cell [18]. There appears to be temporal and spatial regulation of viral gene expression. In normal epithelia, the cell cycle is halted after the divisions taking place in the basal layer. An important molecular correlate to this is the activity of p63, an essential regulator of keratinocyte differentiation. An infection may initiate carcinogenesis when the regulation of differentiation is disrupted, the cell undergoes resistance to growth inhibition, evades the immune response, subverts apoptosis and is effectively immortalized. This pathological state allows for replication of viral DNA in synchrony with chromosomal DNA. Importantly, as the cell enters the upper, differentiating, compartment of the epithelium, viral gene expression transitions from minimal to very high, such that the virus is able to utilize the enzymatic machinery of the host cell to produce multiple copies per cell. This replication is due to entry of the host cell into S-phase, a phenomena largely attributed to the viral protein E7, since knockdown of this protein fails to initiate cellular proliferation and viral genome amplification [100].

The uncontrolled cellular proliferation in the context of manipulated cell-cycle checkpoints leads to chromosomal instability and accumulation of genetic mutations (Figure 5). The cellular factors that are co-opted by HPV proteins are involved in the proliferation of cells during their differentiation. The cooperative activities of the viral proteins E6 and E7 are fundamental in the development of genomic instability because they synergize to effect myriad cellular factors involved in cell-cycle progression, signallling pathways involved in cell survival, differentiation, proliferation (independent of direct effects on the cell cycle) and inhibition of apoptosis. The centrality of E6 and E7 in HPV pathogenesis is reflected by the observation that expression in organotypic raft cultures results in phenotypic changes identical to those seen in high-grade cervical squamous neoplasia in vivo [101]. When E6 and E7 are present, it has been observed that these cells have a higher percentage of polyploidy in addition to inappropriate centrosomal duplication [102]. The result of disrupted mitotic events is chromosomal instability. Because cells of stratified squamous epithelia eventually desquamate,
this instability of the structures required for co-ordinated gene expression is likely to cause malignancy only when there is long-term viral presence, possibly in the timeframe of years to decades [19].

E6 is an 18 kDa 150-amino-acid protein that contains two zinc-binding domains with two CXCC (Cys-Xaa-Xaa-Cys) motifs (Figure 6) [38,103]. It is expressed predominantly in the nucleus, but can also be found in the cytoplasm. There are 12 binding proteins for E6 all of which have a permissive effect on the actions of E7. The combined effects of these proteins result in cells which do not effectively segregate and co-ordinate the events of the cell cycle, leading to the accumulation of mutations because cell-cycle checkpoints are attenuated at the Achilles heel: p53. E6 is exquisitely evolved to affect cell immortalization via both direct and indirect mechanisms. First, E6 forms a trimeric complex with p53 and a cellular E3 ubiquitin ligase, termed E6AP (E6-associated protein) [104]. This interaction results in ubiquitination of p53 and subsequent degradation by the 26S proteasome, drastically reducing the half-life of p53 [105]. Furthermore, E6 can inhibit p53-mediated gene transcription by direct binding and inhibition of its transcriptional activities [106]. p53 is known to heterodimerize with various isoforms of the p63 and p73 family. Because the steady-state level of p53 is reduced as a consequence of E6 expression, it is thought that the differentiation process is altered because the relative balance between p53 and p63 is disrupted, possibly leading to poor co-ordination of the genes necessary for normal squamous differentiation [19]. Indirectly, E6 associates with and inactivates three histone acetyltransferases: p300, CBP and ADA3. These increase the stability of p53 via acetylation [107–109]. Degradation of ADA3 has also been proposed to underlie the inhibition of p14/ARF (alternative reading frame)-dependent activation of p53 [110,111]. E6 contributes additional mechanisms to allow for immortalization of cells by activating transcription of TERT (telomerase reverse transcriptase) [112,113]. This is the catalytic subunit of telomerase, an enzyme usually not expressed in somatic cells. Activation is dependent on E6AP, as knockdown fails to recapitulate the phenotype [103]. All high-risk E6 proteins contain a C-terminal motif termed XT/SXV (Xaa-Thr/Ser-Xaa-Val) that mediates binding to cellular proteins with PDZ domains [18,114]. The PDZ domain proteins are also implicated in the oncogenic activity of HPV because mutant E6 proteins that lose their ability to bind p53 can still immortalize cells [114–116], whereas transgenic mice expressing E6 mutant XT/SXV domains do not develop tumours [117]. PDZ proteins are often found in epithelial tight junctions and it is thought that E6 binding modulates signal transduction pathways downstream of these cell–cell contact proteins [38,114].

Cell-cycle- and S-phase-dependent amplification of the HPV DNA in the differentiated upper levels of the epithelium is most prominently set into motion by HPV E7. Many coalescing signalling pathways can induce the cellular transition from G1/S-phase, but the regulating factors essential for binding and subsequently inhibiting transcription factors controlling S-phase DNA-replication machinery are the Rb family members p105 (Rb-tumour suppressor), p107 and p130 [19]. Rb appears to be the predominant tumour suppressor, supported by observations obtained using knockout mouse models as well as the general observation that Rb inactivation (but not p107 or p130) is a hallmark of many sporadic human cancers [118]. These small so-called nuclear phospho-pocket proteins’ regulate the activity of the many members of the E2F family of transcription factors [19]. The cell cycle can progress when co-ordinated signalling pathways provide biological feedback to the Rb family members via CDK complexes, resulting in their hyperphosphorylation and release from the transactivation domain of E2F transcription factors and the dissociation of the chromatin-remodelling proteins, the HDACs. E2F family members bind DNA in the promoter region of many genes involved in DNA synthesis, cell-cycle progression (cyclin A and cyclin E) and mitosis. All E7 proteins from HPV virions have evolved to bind cellular Rb family members, but only high-risk E7 can efficiently target them for degradation thereby disrupting Rb–E2F–HDAC complexes [18,113]. The difference between low-risk E7 and high-risk E7 appears to be attributed to an aspartate or glycine residue at positions 21 and 22 for the high risk and low risk respectively [113]. Upon binding, E7 targets Rb family members for ubiquitin-dependent proteasomal degradation [113]. The result of degraded Rb is constitutive expression of E2F-responsive genes and the hallmark event in a cell transitioning from resting, or G1-phase, to S-phase and subsequent preparation for DNA replication and mitosis.

E7 is a 13 kDa protein, ~100 amino acids in length with conserved sequence homologies with adenovirus E1A and SV40 (simian virus 40) large T-antigen [113] (Figure 7). Expression is predominantly nuclear, but some studies report a cytoplasmic presence [18,119]. Interestingly, E7 has no intrinsic enzymatic or DNA-binding activities [18]. Instead, it is known to bind an almost overwhelming multitude of cellular factors, the biological significance of which is not fully understood. However, E7 represents an exquisitely evolved viral protein that adapts many host signalling pathways to the benefit of viral replication. It is well known that E7 has extremely low immunogenicity, and the
solution structure provides some explanation as to why this may be: the highly conserved N-terminal region is intrinsically unstructured [119]. The N-terminus contains strong interaction domains with the Rb family members within an LXCXE motif with the conserved region 2, in addition to binding domains for p300 and p600 [113,119]. The C-terminus contains a zinc-binding domain composed of two CXXC motifs and subsequently creates a more rigid structure [113,119]. There are multiple binding motifs within the C-terminus, including residues that interact with E2F directly, a weaker Rb interaction domain thought to be important for Rb destabilization in high-risk infections, a binding site for class I HDACs and domains that interact directly with and effectively titrate cyclin/CDKs p21 and p27, both activators of Rb via CDK2 [113]. The binding of E7 to E2F6 has fascinating consequences, as the interaction is thought to maintain a conducive environment for viral replication by recruiting Polycomb group complexes to E2F-responsive promoters [18]. This recruitment is thought to inhibit the physiological repression induced by Polycomb complexes at promoter regions and allow for constitutive E2F gene transcription [18]. This function of E7 is similar to the function of binding class I HDACs. E7 facilitates HDAC removal at other non-E2F promoters allowing transcriptional machinery access to promoters [18]. Furthermore, the E7–HDAC interaction is essential for episomal maintenance [18]. Unlike E6, the expression of high-risk HPV E7 proteins alone can immortalize human keratinocytes at a low rate. In a transgenic mouse model, E7 expression, along with low-dose oestrogen exposure, induced high-grade cervical dysplasia [18]. E7 also may inhibit differentiation by binding and inhibiting the activity of p/CAF, the activity of which has been shown to acetylate Rb1 at a co-ordinated time point so that keratinocytes exit the cell cycle and permanently differentiate [19]. Many of the mechanisms promoting anti-apoptotic activities and driving continual cell-cycle progression are indirect biochemical effects in the sense that viral proteins are not interacting directly with some of the targets that promote these driving forces of malignant progression. In fact, p53 levels increased downstream of constitutive E2F gene transcription, but once again the synergistic relationship between E6 and E7 acts to neutralize the cell-cycle stalling and pro-apoptotic effects of p53. Signalling pathways associated with keratinocyte survival, such as cell–cell contact-mediated E-cadherin signalling are also altered to allow for successful viral replication. For example, E7 up-regulates the Akt kinases, which are well-characterized downstream effectors of E-cadherin and normally encourage pro-survival signals as keratinocytes detach from the basement membrane and stratify upwards [19].

Various studies have suggested alterations in host miRNA (microRNA) and other epigenetic markers that are induced by HPV in cervical cancer [120–126]. For example, experiments in normal human keratinocytes have shown that the E7 viral oncoprotein blocks normal up-regulation of miR-203, a factor that negatively regulates proliferative capacity upon epithelial cell differentiation [120]. E5 expression has also been shown to repress miR-203 and result in increased p63 levels via an unknown mechanism [124]. Furthermore, E5 expression led to significant up-regulation of miR-146a and repression of miR-324-5p [127]. miR-146a up-regulation by E5 may play significant roles in promoting cell-cycle proteins and attenuating immune response to HPV, whereas miR-324-5p alteration may lead to events associated with epithelial–mesenchymal transition such as N-cadherin up-regulation and E-cadherin down-regulation [124]. In cervical cancer tissue and in cell lines, the high-risk HPV oncoprotein E6 inhibits tumour-suppressive miR-34a by destabilizing p53 [122,125]. An analysis of miRNA-expression profiles in HPV+ cervical cancer cell lines, lesions and tumour tissues revealed aberrant expression of miR-218 and targeting of the epithelial cell-specific marker LAMB3 (laminin β3) [123]. An additional miRNA expression profile on cervical cancer cell lines provided data important for understanding the mechanisms by which particular miRNAs are involved in carcinogenesis by correlating miRNA expression and tissue differentiation [121]. The relationship between HPV status and miRNA profiles in OPSCC has not been evaluated.

**DIFFERENTIAL GENE EXPRESSION ANALYSIS**

Detected HPV+ OPSCC has increasingly been subjected to molecular study to identify key aberrations present in affected cells that may improve our understanding and inform on discrepancies in response to therapy and prognosis among tumour populations. Researchers have compared HPV+ OPSCC to HPV+ disease arising in different anatomical sites, HPV− disease and normal oral mucosa, as well as HPV+ tumours in smokers and non-smokers. Key areas of genetic aberrations include DNA replication, DNA repair, cell cycling and chemotheraphy/radiotherapy-sensitivity.

When genome-wide chromosomal profiles compare HPV+ cervical SCCs, HPV+ OPSCCs and HPV− OPSCCs,
were common in all sample groups, whereas a loss at 13q and a gain at 20q were representative of HPV+ HNSCC compared with cancer-free oral mucosa (397 genes) and in comparison of HPV− HNSCC and normal oral tissue [130]. A subgroup of 59 genes was present only in HPV+ HNSCC and not in HPV− HNSCC or normal oral mucosa. The key genomic regions altered in HPV+ disease are presented in Table 1. These genes are predominantly involved in mitosis, DNA repair and transcriptional regulation.

As tobacco use is a known risk factor for HNSCC, smoking status was considered in a gene expression profiling study [131]. Comparison of gene expression patterns from never-smokers that were HPV+ with those that were HPV− revealed that genes involved in viral defence and immune response were down-regulated in HPV+ tumours [IFIT1 (IFN-induced protein with tetratricopeptide repeat 1), IFITM1−IFITM3 (IFN-induced transmembrane protein 1−3), IFI6−IFI16 (IFN, α-inducible protein 6−16), IFI44L (IFN-induced protein 44-like) and OAS2 (2′−5′-oligoadenylate synthetase 2)], consistent with the immunomodulating nature of HPV infections, predisposing to development into malignancies [130,131]. Up-regulated genes include those involved in CDK inhibition [CDKN2C (CDK inhibitor 2C)], G1/S-phase transition [CDC7 (cell-division cycle 7)], replication and DNA repair [RFC4 and RFC5 (replication factor C4 and C5)] and dimerization with E2F [TFDP2 (transcription factor E2F dimerization partner 2)]. For HPV+ smokers compared with non-smokers, differential expression was observed in several genes regulated by p53 or E2F transcription factors. RFC4 was consistently up-regulated in HPV− samples from never-smokers. Other differentially expressed genes based on smoking status include CDC7, MCM2 (minichromosome maintenance 2), CYP4V2 (cytochrome P450 4V2), IGF (insulin-like growth factor) and keratin-associated proteins [KRT6B, KRT10 and KRT11 (keratin 6B, 10 and 11)]. Significant up-regulation of a key component of the mammalian synaptonemal complex, a structure regulating the arrangement of homologous chromosomes in meiosis [SYCP2 (synaptonemal complex protein 2)] and normally transiently expressed in gametes, was also observed.

In addition to modified gene expression profiles, chromosomal aberrations and amplifications are significantly less frequent in HPV+ compared with HPV− OPSCC [34,132]. A group of chromosomal aberrations in OPSCC clearly occur more frequently among HPV− cases, including gains at 11q13 and losses of 3p, 9p and 18q. On the basis of these data, it has been proposed that the number of required genetic alterations for progression to malignancy is lower in HPV+ oncogenesis [132]. These findings are interesting considering that HPV+ disease is associated with a favourable prognosis, as these specific chromosomal aberrations also appear to be significant indicators of clinical outcome. A notable exception is 16q loss, which predominates in, but is not exclusive to, HPV+ OPSCC and may be an independent positive prognostic variable [132]. The 11q13 amplification, however, portends a poorer prognosis regardless of HPV status, but tends to be rare in HPV-related disease [132]. Thus these genetic differences may have dramatic implications for differing clinical management for the two disease phenotypes.

**Figure 8** Common and distinct chromosomal profiles of HPV+ compared with HPV− HNSCC relative to HPV+ cervical SCC (CxSCC)

Hierarchical clusters demonstrate two main groups: one mainly HPV+ and the other HPV− (Figure 8). However chromosomal alterations were also identified specific for the two tissue types as well as pan-SCC aberrations [128]. A second study corroborates enhanced similarity between HPV+ OPSCC and HPV+ cervical SCC relative to HPV− OPSCC, in particular a distinct and larger subset of cell-cycle genes and some testis-specific genes normally expressed in meiotic cells are up-regulated in concordance with HPV infection [127]. These common changes in transcriptional profiles identified between HPV+ HNSCC and cervical carcinoma include down-regulation of genes involved in immune response (interleukins and IFN-induced proteins) and up-regulation of genes within the 20q region along with down-regulation of genes on 13q [129]. Up-regulation of 20q appears to be caused by E7 expression and is related to inactivation of Rb [129].

An interesting study compared gene expression profiles of HPV+ and HPV− oropharyngeal cancer and oral cavity cancer [128]. In oral cavity tumours, no significant difference in gene expression was noted when comparing HPV+ and HPV− specimens. However, analysis of oropharyngeal tumours shows significant differences (347 differentially expressed genes) in HPV+ compared with HPV− lesions. Differences were particularly common among genes involved in DNA regulation and repair, cell cycle, and chemotherapy/radiotherapy-sensitivity. These results underscore the observation that HPV+ oropharyngeal disease represents a divergent biological entity from HPV− disease. An additional study compared gene expression profiles of HPV+ HNSCC with normal oral mucosa. A larger number of genes were differentially expressed in HPV+ HNSCC compared with cancer-free oral mucosa (397 genes) and in comparison of HPV− HNSCC and normal oral tissue [130].

**UNANSWERED QUESTIONS AND FUTURE DIRECTIONS**

As with other HPV-associated malignancies, HPV+ oropharyngeal cancers appear to be a sexually transmitted disease; however, there is evidence that non-sexual transmission can occur [5,133]. HPV+ tumours are more likely to affect white young men (<50 years of age) with no history of alcohol or tobacco abuse.
and are associated with distinct risk factors including more than six lifetime oral-sex partners (odds ratio 3.4), >26 lifetime vaginal-sex partners (odds ratio 3.1), a younger age at first sexual intercourse, and seropositivity for HPV-16 L1 capsid protein (odds ratio 32.2) [114,135]. Similarly, an odds ratio of 230 was reported for the development of oropharyngeal cancer with high-risk HPV DNA detected in the oral cavity when adjusted for alcohol and tobacco use [134–136], and a similar landmark study reported an odds ratio of 14.4 for the association of HPV-16 seropositivity with the risk of OPSCC [136]. In individuals with OPSCC, seropositivity for HPV-16 was strongly associated with no history of tobacco or alcohol abuse [134]. The mechanism predisposing white young men to this type of cancer remains obscure, but is currently supported in the literature both in the U.S.A. and in Europe, where high rates of HPV + OPSCC have been reported in Scandinavian countries [9,10,128]. This is in stark contrast with cervical cancer, where trends show that overall HPV prevalence and HPV-16 prevalence is highest in sub-Saharan Africa [137]. However, the same study showed HPV + women in Europe were significantly more likely to be infected with HPV-16 than were those in sub-Saharan Africa (odds ratio 2.64, P = 0.0002). Since HPV-16 is the most commonly identified virus in OPSCC, widespread vaccination with a HPV vaccine in women may lead to a decline in the incidence of male HPV + OPSCC [138].

Patients with HPV + OPSCC more often present with stage III–IV disease and, paradoxically, data show that the disease responds better to chemotherapy and radiation, such that HPV status is a strong positive prognostic factor for survival [139]. A review of the literature suggests that patients with HPV + HNSCC had a lower risk of death [meta HR (hazard ratio) 0.85; 95% CI (confidence interval) 0.7–1.0] and less chance of recurrence (meta HR 0.62; 95% CI 0.5–0.8) [7]. When anatomical site is taken into consideration, patients with HPV + OPSCC have a 28% reduced risk of death (meta HR 0.72; 95% CI 0.5–1.0) compared with HPV − OPSCC [5]. The biological mechanism that underlies this increased responsiveness to treatment and improved overall survival is currently unknown [15,16,139–142]. Further questions remain regarding the details of why HPV has a tropism for oropharyngeal tissues, how virus enters keratinocytes preferentially at this specific anatomical site and how it is able to induce malignant transformation. The cell- and tissue-level events associated with the distinct histology observed in HPV + OPSCC specimens raises the question of whether HPV + OPSCC arise from a cell type distinct from that of HPV − OPSCC, or if the viral life cycle alters keratin expression and deposition by infected cells. Another area in need of further exploration involves determining the prevalence of HPV-16 in non-cancerous tissues. Interestingly, one survey identified that 13 of 206 cases of tonsillitis and one of 174 normal tonsils were positive for HPV-16 DNA by PCR in the absence of anti-L1 antibody production [143]. Indeed, even in HPV + OPSCC, the question of whether the virus is integrated or episomal remains unanswered. Furthermore, it is currently unknown whether a specific viral load threshold is associated with carcinogenesis and whether viral load is a significant prognostic factor.

In the last few years, HPV status has been linked to racial survival discrepancies in the U.S.A. [144–147]. A study using National Cancer Institute’s SEER (Surveillance, Epidemiology, and End Results) data reflected lower overall survival rates for black patients with HNSCC and particularly OPSCC [146,147]. It is important to note that black patients tend to present at more advanced stages of disease, with 38.6% of patients presenting with stage IV disease compared with 34.0% and 30.2% of Hispanic and white patients respectively [147]. Furthermore, of all head and neck cancer sites considered, OPSCC was associated with the greatest difference in survival between blacks and whites. This finding is supported by results of an additional study showing that only 4% of HNSCC were HPV + in black patients, whereas 34% were HPV + in white patients [145]. Related to this, a critical area for examination is the issue of screening for oral and oropharyngeal infection, dysplastic changes and malignancies in a cost-effective and widespread manner. Research into miRNAs, because of their relative stability in serum and possibly saliva, holds particular interest in this regard.

### CONCLUSIONS

Current clinical literature fully supports HPV-associated OPSCC as a unique entity: it is histologically distinct, affects an unambiguous patient population with defined risk factors and may warrant divergent clinical management compared with HNSCC associated with traditional risk factors. Although epidemiological and clinical studies support HPV + OPSCC as a distinct clinical entity, the pathobiology of HPV-associated OPSCC initiation and progression is poorly understood. A more detailed understanding of molecular events associated with HPV + OPSCC progression is therefore a prerequisite for improved staging and diagnostic/prognostic stratification, as well as for the design of novel therapeutic strategies specific for this population.

---

**Table 1** Key genes up-regulated in HPV + tumours

Data taken from [132,137].

<table>
<thead>
<tr>
<th>Gene up-regulated in HPV + tumours</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDK2A (CDK inhibitor 2A)</td>
<td>Cell-cycle inhibitor, product of E2F transcription</td>
</tr>
<tr>
<td>PCNA (proliferating-cell nuclear antigen)</td>
<td>Cell-cycle-dependent auxiliary protein for DNA polymerase (pol) δ (also required for SV40 DNA replication in vitro)</td>
</tr>
<tr>
<td>RFC4 (replication factor C4)</td>
<td>Accessory protein (with PCNA) for DNA pol δ and ε. Part of a large multisubunit protein complex of DNA-damage sensors called BASSC in breast cancer early-onset 1-associated genome surveillance complex</td>
</tr>
<tr>
<td>MCM2 (minichromosome maintenance complex 2)</td>
<td>Early S-phase protein involved in DNA replication, probably regulated by E2F motifs</td>
</tr>
<tr>
<td>MCM3 (minichromosome maintenance complex 3)</td>
<td>Involved in DNA replication; probably serves as helicase that links a histone chaperone to a histone H3–H4 bridge</td>
</tr>
<tr>
<td>CDC7 (cell-division cycle 7)</td>
<td>Protein kinase essential for the G1/S-phase transition and initiation of DNA replication. May regulate DNA replication by modulating MCM functions</td>
</tr>
<tr>
<td>TYMS (thymidylate synthetase)</td>
<td>Maintains the dTMP pool critical for DNA replication and repair</td>
</tr>
<tr>
<td>CCNE2 (cyclin E2)</td>
<td>Activates CDK2 and is involved in DNA synthesis</td>
</tr>
<tr>
<td>USP1 (ubiquitin-specific protease 1)</td>
<td>Negatively regulates PCNA polyubiquitylation, a process normally induced by DNA-damaging agents</td>
</tr>
<tr>
<td>BRB1 (BRM/SWI-related gene 1)</td>
<td>Involved in a switch in ATP-dependent chromatin-remodelling mechanisms that coincide with the final mitotic division. Note that this gene is up-regulated in HPV– OPSCC</td>
</tr>
</tbody>
</table>

Data taken from [132,137].
REFERENCES


© The Authors Journal compilation © 2012 Biochemical Society


Received 23 November 2011/7 January 2012; accepted 9 February 2012
Published on the Internet 27 March 2012, doi:10.1042/BJ20112017